

О.П. Балановский

ГЕНОФОНД ЕВРОПЫ



МОСКВА 2015

О.П. Балановский

ГЕНОФОНД ЕВРОПЫ

Товарищество научных изданий КМК
Москва 2015

Балановский О.П. Генофонд Европы. М.: Тов-во научн. изданий КМК. 2015. 354 с.

Монография посвящена генофонду народонаселения Европы – исследованию его пространственной изменчивости и его истории.

Исследования генофондов бурно развиваются благодаря использованию митохондриальной ДНК, Y-хромосомы и полногеномных панелей маркеров. Но обобщающие работы по мтДНК и Y-хромосоме в Европе устарели, поскольку опубликованы более десятилетия назад до выхода основной массы частных научных публикаций. Поэтому в книге проанализированы как собственные данные автора, так и литературные: созданные базы данных включили более 130 тысяч образцов по мтДНК и более 140 тысяч – по Y-хромосоме.

Параллельное изучение генофонда Европы по маркерам Y-хромосомы, мтДНК и полногеномным панелям обеспечило синтез результатов этих трех генетических систем. Такой синтез актуален не только для генетиков, но и для смежных наук – антропологии, археологии, лингвистики, истории, все активнее включающих данные генетики в комплексные исследования этногенеза.

В книге рассмотрены и данные по древней ДНК, позволяющие – эпоха за эпохой – проследить историю генофонда Европы. А совместный анализ количественных лингвистических и генетических данных выявляет непрямую связь языков и генофондов. Особую актуальность для российской науки имеют исследования славянских и северокавказских народов, которым посвящены отдельные главы.

Издание осуществлено при финансовой поддержке Российского фонда фундаментальных исследований по проекту № 15-06-07016, не подлежит продаже



ПРЕДИСЛОВИЕ

Нет, если мне нужна Европа, то такая,
Где перед лужицей в вечерний час дитя
Сидит на корточках, кораблик свой пуская,
В пахучем сумраке Бог весть о чем грустя.

Артур Рембо (Перевод Б.К. Лившица)

Генетическое изучение популяций человека уже сотню лет привлекает многих исследователей, среди которых виднейшие представители отечественной и мировой биологии: А.С. Серебровский, Н.К. Кольцов, В.В. Бунак, Г.Ф. Дебец, Я.Я. Рогинский, Ю.П. Алтухов, Ю.Г. Рычков, Р. Виллемс, L.L. Cavalli-Sforza, W. Bodmer, а также представители смежных дисциплин – археологии, лингвистики, прикладной математики, палеогеографии. За свою долгую историю геногеография использовала самые разные маркеры – от групп крови до фамилий, от отдельных ДНК-маркеров до полных геномов – и обнаружила множество частных закономерностей, причем большинство из них проявлялись вновь и вновь при анализе каждого очередного типа маркеров. Кроме множества частных закономерностей, и в зарубежной, и в российской школах геногеографии был разработан и ряд общих концепций, включая теорию демической диффузии, концепцию обобщенного гена, принцип эквидистантности и ряд других. Предлагаемая книга продолжает традицию систематического геногеографического изучения отдельных регионов мира и посвящена генофонду народонаселения Европы (для краткости – генофонду Европы).

При этом генофонд Европы – и так самый подробно изученный среди всех регионов мира. Для европейской науки он стал полигоном для проверки всех главных генетических концепций и методов. Например, вопрос о палеолитическом или более позднем времени формирования основных черт европейского генофонда является, пожалуй, наиболее широко обсуждаемым вопросом в популяционной генетике человека. Об этом свидетельствует и число статей на эту тему (в том числе в журналах *Nature* и *Science*), и авторитет их авторов [Ammerman, Cavalli-Sforza, 1984; Cavalli-Sforza et al., 1994; Richards et al., 1996; Richards et al., 2000; Semino et al., 2000; Barbujani, Bertorelle, 2001; Chikhi et al., 2002; Haak et al., 2005, 2009, 2015; Bramanti et al., 2009; Allentoft et al., 2015 и т.д.].

Да и по другим вопросам генетической структуры народов Европы написано столько работ, что их перечисление потребовало бы отдельной монографии. Так стоит ли снова писать о генофонде Европы?

Стоит. Дело в том, что среди этого множества работ очень мало обобщающих. Например, работы, обобщающие изменчивость мтДНК и Y-хромосомы в Европе, последний раз проводились более десятилетия назад [Richards et al., 2000; Rosser et al., 2000; Semino et al., 2000; Richards et al., 2002].

подавляющее большинство последовавших крупных работ ставило задачей изучение отдельных регионов Европы или по мтДНК [Helgason et al., 2001, 2003; Meinila et al., 2001; Malyarchuk et al., 2002, 2003, 2004, 2006, 2008; Бермишева и др., 2002; Orekhov et al., 1999; Pfeiffer et al., 1999; Pereira et al., 2004; Tambets et al., 2004; Goodacre et al., 2005; Falchi et al., 2006; Grzybowski et al., 2007; Lappalainen et al., 2008; Alvarez-Iglesias et al., 2009; Santos et al., 2003; Garcia et al., 2011; Karachanak et al., 2012], или по Y-хромосоме [Behar et al., 2003; Харьков и др., 2004, 2005б; Cinnioglu et al., 2004; Di Giacomo et al., 2003; Brion et al., 2005; Flores et al., 2003, 2004; Tambets et al., 2004; Alonso et al., 2005; Goncalves et al., 2005; Kayser et al., 2005; Pericic et al., 2005; Capelli et al., 2006, 2007; Lappalainen et al., 2006, 2008; Adams et al., 2008; Balanovsky et al., 2008, 2011; Battaglia et al., 2009; Fechner et al., 2008; Varzari et al., 2009; King et al., 2011; Yunusbaev et al., 2012; многие другие работы]. Ряд работ был посвящен также отдельным гаплогруппам [Di Giacomo et al., 2004; Rootsi et al., 2007; Onofri et al., 2008; Tofanelli et al., 2009; Derenko et al., 2006, 2007, 2010; Cruciani et al., 2007, 2010; Myres et al., 2011; Underhill et al., 2010; Mendez et al., 2011 и другие работы].

Но обобщающий анализ в масштабе всей Европы после работ 2000-2002 годов не проводился. А за это время не только на порядок возрос объем данных о мтДНК и Y-хромосоме, но и появились данные по полногеномным панелям маркеров, стремительно рванулись вперед исследования древней ДНК, не стояли на месте и исследования смежных наук – особенно лингвистов.

Поэтому представляется своевременным данное исследование, в котором сделана попытка собрать воедино, проанализировать, подытожить и по возможности синтезировать наши сегодняшние знания о генофонде народонаселения Европы.

В предисловиях принято говорить о том, кому адресована книга. Но разве менее важно, кому она *не* адресована? Должен же я предупредить не только о том, что есть в этой книге, но и о том, чего в ней нет! Нет в ней погони за каждой новой веточкой Y-хромосомы, за каждым новым методом анализа полногеномных данных, за каждой частностью, какой бы любопытной – или широко обсуждаемой в узких интернет-кругах – она ни была. И читателям, любящим все это, книга *не* адресована. Им она противопоставлена, потому что в фокусе книги – популяция в целом, а не отдельные веточки разросшегося дерева гаплогрупп одной из генетических систем. Книга, ставящая целью учесть наинovelейшие новости в быстро развивающейся области, была бы обречена устареть еще до своего выхода из печати. А самое главное, ее автора и читателей подстерегала бы опасность за веточками не увидеть леса. Ведь развитие технологий столь стремительно, что стало модно лишь бежать за модой, не останавливаясь для того, чтобы оглянуться – а что же действительно новое мы видим, кроме бесконечных новых веточек.

Поэтому я ставил совсем другую цель – оглядеть не веточки, а сам лес, не частности, а общую структуру генофонда Европы, не отдельные факты, а основные закономерности. Для этого широко использовались как старые, так и самые новые достижения популяционной генетики. Например, построены карты практически *всех* гаплогрупп Y-хромосомы, которые полиморфны в Европе и по которым на сегодняшний день уже есть достаточно данных об их частотах в коренном населении. А в списке литературы можно найти четырнадцать статей (в том числе моих собственных), опубликованных в том же 2015 году, что и сама эта книга, причем более десяти из этих работ столь важны, что они составили основу отдельных разделов книги. Есть в этой книге и то, что еще не было опубликовано в статьях. Но новизна данных была не самоцелью, а лишь одним из многих средств, подчиненным настоящей цели – *выяснению структуры генофонда народов Европы и его истории*. Достичь этой цели в полной мере вряд ли возможно, но стоит попробовать. Лишь тем читателям, которым эта цель интересна, и адресована книга.

БЛАГОДАРНОСТИ

Я глубоко признателен своим учителям Е.В. Балановской и Р. Виллемсу, которые ввели меня в область геногеографии и филогеографии; признателен В.В. Запорожченко, Р.С. Сычеву, А.С. Пшеничному, взявшим на себя труд наполнения баз данных по мтДНК и Y-хромосоме; рад случаю упомянуть своих прекрасных помощниц А.Т. Агджоян, М.И. Чухряеву, Р.А. Схаляхо, М.А. Кузнецову, Х.Д. Дибирову, О.А. Балаганскую, И.Э. Теучеж – результаты и их труда вошли в описанное исследование, а на плечи двух первых легла и часть работы непосредственно по книге; благодарю коллег W. Naak, S. Rootsi, A.B. Дыбо, О.А. Мудрака, С.М. Кошеля, О.М. Утевскую, Э.А. Почешхову, Л.А. Атраментову, М.Б. Лавряшину, Л.И. Тегако, М.И. Чурносова, С. Der Sarkisyan, А.С. Касьяна, Е.И. Кушнеревич, В.И. Хартановича, А.П. Бужилову, Л.М. Епископосьяна и многих других, в сотрудничестве с которыми собирались образцы или разрабатывались вопросы их анализа. Мне очень помогло сообщество нашего междисциплинарного сайта генофонд.рф – его бессменный автор Н.В. Маркина и критический ум Л.С. Клейна. И сугубая благодарность – моим родителям П.Н. Яшуку и Е.В. Балановской, оказавшим огромную и разнообразную помощь при подготовке этой книги.

Исследование проводилось более 10 лет, и разные его аспекты поддерживались в разное время более чем двадцатью грантами, включая международный The Genographic Project, грант РФФ 14-04-00827, грант РФФ 06-06-00640, гранты Программ Президиума РАН «Динамика генофондов», «Фундаментальные науки – медицине». Благодарную роль в возникновении и развитии моей лаборатории сыграли директор ИОГен РАН член-корр. Н.К. Янковский и Программа Президиума РАН «Молекулярная и клеточная биология». Благодарю и лабораторию исторической генетики МФТИ, поддерживавшую вместе с РФФ полногеномный анализ Y-хромосомы. А наиболее постоянная финансовая поддержка все эти годы приходила от РФФИ: за это время под моим руководством выполнены 4 инициативных (10-04-01603, 07-04-00340, 04-04-49664-а, 13-04-01711), 8 экспедиционных и 3 стажерских проекта РФФИ, и конечно же, издательский грант на эту книгу 15-06-07016-д.

Без помощи этих фондов и энтузиазма моих коллег эта книга, конечно же, не смогла бы состояться.

ГЛАВА 1.

СОЗДАНИЕ БАЗ ДАННЫХ О ГЕНОФОНДАХ МИРА

Наше исследование генофонда Европы в контексте мирового генофонда основывается на собственных и литературных данных по изменчивости в населении мира Y-хромосомы, митохондриальной ДНК и полногеномных панелей маркеров, объединенных в специально разработанных базах данных. В данной главе описываются созданные базы данных и дается характеристика собственным результатам (которые включены во все базы данных, но также во многих случаях анализируются и отдельно).

Но этим содержание главы не исчерпывается – будет логичным еще до начала изложения результатов рассмотреть и самые необходимые аспекты методических вопросов – особенностей генотипирования Y-хромосомы, мтДНК и древней ДНК, методы статистического и филогеографического анализа, обсудить вопрос о генетических датировках и программное обеспечение для создания картографических атласов генофонда, занимающих важное место во всем исследовании.

1.1. БАЗА ДАННЫХ ПО Y-ХРОМОСОМЕ

КРАТКАЯ ХАРАКТЕРИСТИКА Y-BASE

ПРЕДНАЗНАЧЕНИЕ И ИСТОРИЯ СОЗДАНИЯ

За последние 15 лет изучение изменчивости Y-хромосомы стало одним из основных направлений в русле популяционно-генетических, молекулярно-антропологических и филогеографических исследований популяций человека. Поэтому многие научные коллективы остро нуждаются в быстром и свободном доступе ко всему массиву уже накопленной информации. Достижения молекулярной генетики получили широкое признание в смежных науках, изучающих историю человечества – в антропологии, археологии, лингвистике, истории, этнологии, палеоэкологии. Однако отсутствие удобной информационной системы, обеспечивающей простой доступ к генетическим данным, резко ограничивает комплексные исследования населения мира специалистами в разных областях естественных и гуманитарных наук.

Поэтому мы поставили перед собой задачу объединить все доступные в литературе данные и обширные собственные архивы, несущие разнообразную информацию о полиморфизме Y-хромосомы в популяциях человека. База данных разрабатывалась под руководством автора в течение ряда лет (2006–2012 гг.) и продолжает совершенствоваться.

В 2009–2012 годах создание этой базы данных являлось одним из важных направлений темы НИР «Анализ распространения гаплотипов митохондриальной ДНК и Y-хромосомы у народов мира на основе создания геоинфосистем», выполнявшейся в Медико-генетическом научном центре РАМН

под руководством автора. В 2010–2012 годах создание базы данных было поддержано целевым грантом РФФИ, что позволило резко интенсифицировать работу над базой данных. Хотя эта многолетняя работа по созданию базы велась главным образом на чистом энтузиазме ее разработчиков, однако, одно время важна была также финансовая поддержка международного проекта «Genographic», и сейчас работа интенсивно продолжается все так же под руководством автора в Институте общей генетики РАН.

Программирование выполнено Андреем Войсковским. В разработке алгоритмов важную роль сыграл Андрей Пшеничнов, основную часть работы по наполнению базы выполняли Роман Сычев и Андрей Пшеничнов, а в части подготовки собственных результатов нашего коллектива к вводу в базу данных – Хадижат Дибирова. Большую помощь в получении полнотекстовых версий статей-источников оказали Mark Haber и Wolfgang Haak. Окончательное редактирование таблиц популяций и их подготовка для картографирования (исключение одних и объединение других выборок в соответствии с разнообразными критериями) выполнены автором.

В настоящее время одна из рабочих версий базы данных реализована в рамках онлайн-инфосистемы (отдельные разделы которой представлены на нашем сайте www.genofond.ru), и продолжение работы заключается в онлайн-реализации остальных разделов и наполнении базы данных новой публикуемой информацией. Основные характеристики базы данных, существующей на данный

Таблица 1.1. Характеристика двух ведущих баз данных по Y-хромосоме

ПАРАМЕТР	Y-base (создана под руководством автора)	YHRD (основной зарубежный аналог)
Объем (образцов) по SNP маркерам	144 464	17 502
Объем (образцов) по STR маркерам	32 054	143 044
Число популяций	2474	нет информации
Число источников	238	нет информации
Размещение	www.genofond.ru	www.yhrd.org

момент и использованной для изучения генофонда Европы и других регионов мира, представлены в таблице 1.1.

Важно отметить, что созданная нами база Y-base в соответствии с целями популяционно-генетических исследований посвящена в первую очередь SNP-маркерам, хотя содержит и STR-гаплотипы. Напротив, основной аналог – международная криминалистическая база данных YHRD – нацелена в первую очередь на STR-маркеры, хотя содержит также и SNP-маркеры. Теоретически, обе базы основываются в основном на литературных данных и поэтому могли бы перекрываться почти полностью. Но на деле основная направленность баз формирует и их объем: Y-base содержит данные по 144 тысячам образцов по SNP-маркерам, а YHRD – по 143 тысячам STR гаплотипов. Такая близость показателей хотя и случайна, но довольно выразительна. При этом показательно и на порядок меньшее в каждой базе число образцов с альтернативными маркерами: 32 тысячи образцов с STR маркерами в нашей базе и 17 тысяч образцов с SNP маркерами в YHRD. Это объясняется тем, что обязательным условием для включения данных в нашу базу было наличие SNP маркеров, а STR мы включали только в том случае, если они опубликованы для тех же образцов. А YHRD поступал наоборот: обязательным условием было только наличие STR маркеров. Поэтому можно считать, что эти две базы взаимно дополняют друг друга и в целом охватывают практически все опубликованные данные по населению мира.

Все остальное содержание данного раздела посвящено техническим вопросам создания базы данных – подходам к сбору данных и к программной реализации базы, описанию двух важнейших модулей автоматической обработки данных, а также трех разделов базы, уже размещенных на нашем сайте www.genofond.ru.

ОБИЛИЕ ДАННЫХ ПОТРЕБОВАЛО СОЗДАНИЯ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ

В обширной и бурно развивающейся области генетических исследований популяций мира маркеры Y-хромосомы являются сейчас одними из

наиболее информативных и широко изучаемых генетических систем. Поэтому массив накопленных данных огромен и возрастает ежемесячно по мере появления новых публикаций. Задача сбора этих данных и создания электронной библиотеки теоретически проста – она хотя и трудоемка, но облегчается тем, что подавляющее большинство статей публикуются на английском языке и их аннотации, как правило, доступны в базе данных PubMed на сервере *ncbi*. Однако получение полнотекстовых версий статей, не имеющих бесплатного доступа, требует дополнительных усилий и времени. Тем не менее, в целом эта задача решается рутинно, и многие профильные лаборатории (в том числе и наш коллектив) располагают достаточно обширными электронными библиотеками.

Намного более сложным оказывается шаг от «достаточно обширной» к почти исчерпывающей электронной библиотеке по данной тематике, поскольку этот необходимый этап подразумевает включение неанглоязычных статей (в первую очередь на китайском и русском языках), а также поиск многих статей, не индексируемых в PubMed и других центральных хранилищах, например, статьи в многочисленных сборниках, монографиях, трудах различных конференций. В отслеживании этих публикаций нашему коллективу способствуют налаженные связи с рядом ведущих международных генетических центров. В отношении российских, белорусских, украинских, казахстанских публикаций, часто несущих важные и нигде более не опубликованные сведения по популяциям этих стран, большую помощь оказывают наши коллеги в пределах бывшего СССР. В целом, наш коллектив находится в благоприятных условиях для создания подробной коллекции тематических публикаций по изменчивости Y-хромосомы в популяциях мира и имеет многолетний опыт создания подобных тематических библиотек.

УНИФИЦИРОВАННАЯ ХАРАКТЕРИСТИКА ПОПУЛЯЦИЙ

Однако самым сложным является не сбор информации, а задача ее объединения в единой базе данных. При решении этой задачи на первый план

выходят вопросы разработки наиболее полного «паспорта» данных и унификации их формата.

Нашим коллективом разработан, пожалуй, наиболее детальный и выстраданный «паспорт» для каждой изученной популяции – набор параметров, по которым характеризуется популяция при внесении ее в базу данных. В этот набор полей включается не только условное название популяции из оригинальной публикации (чем зачастую ограничиваются составители большинства сводок), но и обязательное указание ее этнической принадлежности, страны и региона внутри страны, точных географических координат, стратегии формирования выборки и ее размер. Вся эта информация является чрезвычайно важной для последующей интерпретации данных и отбора популяций для каждого конкретного вида анализа.

Важность такого подхода поясним примером – зачем, например, нужен такой, казалось бы третьестепенный, параметр как «стратегия формирования выборки». При изучении словацкой выборки была обнаружена высокая частота генетических вариантов, характерных для коренного населения Индии. Если бы эта выборка была обозначена в нашей базе данных как «Словакия» или тем более как «словаки», эти результаты были бы необъяснимы. Однако внесение в БД сведений, что стратегия формирования данной выборки была нацелена на создание представительной картины генофонда разных географических регионов Словакии, причем принципиально без учета этнической принадлежности обследуемых, позволило правильно интерпретировать этот на первый взгляд парадоксальный результат. Наличие «индийских» вариантов объяснялось тем, что в выборку попали не только словаки, но и цыгане Словакии: в ряде районов этой страны заметную долю населения составляют цыгане, в генофонде которых, по результатам многих исследований, сохраняется память об их индийском происхождении.

УНИФИЦИРОВАННАЯ ХАРАКТЕРИСТИКА ГАПЛОГРУПП

Но настоящие трудности, по нашему опыту, начинаются при попытке внесения в единую таблицу не данных о популяциях, а разнородной информации о генотипах, изученных авторами разных публикаций. Не говоря уже о том, что разными коллективами используются различные наборы SNP и STR маркеров, нередко используются разные обозначения для одного и того же маркера (названия которых стремительно эволюционируют в ходе открытия новых SNP маркеров), а главное, большинство авторов публикует данные в виде частот гаплогрупп. Эта особенность принципиальна, поскольку номенклатура гаплогрупп на протяжении последних лет неоднократно менялась, и

простое объединение статей разных лет может привести не только к трудно сопоставимым, но и просто к неверным результатам.

При создании первой рабочей версии БД был использован традиционный подход - единицей хранения в ней выступала частота той или иной гаплогруппы в каждой популяции. Если разные авторы использовали разные маркеры, но было известно, что эти маркеры синонимичны (определяют одну и ту же гаплогруппу), данные этих авторов классифицировались как частоты одной и той же гаплогруппы в соответствующих популяциях. К сожалению, данный подход оказался трудоемким, поскольку интеграция каждой новой статьи в базу данных требовала принятия решений по «приравнению» гаплогрупп. И более того, он оказался бесперспективным – по мере уточнения филогенетического древа гаплогрупп оказывалось, что SNP маркеры, ранее рассматривавшиеся в литературе как синонимичные, в действительности находятся на разных иерархических уровнях ветвления филогенетического древа (один определяет ветвь, а другой – доминирующую субветвь той же ветви). С аналогичными трудностями, насколько нам известно, столкнулись и другие коллективы, составлявшие крупные сводки данных. В подавляющем большинстве случаев они создавались под конкретную задачу и через два-три года, после появления в научном обиходе новых SNP маркеров и изменений в обозначениях гаплогрупп, эти сводки становились малоприспособленными для пополнения и использования в других исследованиях.

Чтобы избежать этих осложнений и создать универсальную долгосрочную автоматическую инфосистему (АИС), при разработке ее нынешней версии нами в основу положен **принцип полного сохранения всей первичной информации о проанализированном образце**. Это позволяет дальнейшую обработку проводить автоматически специальными программными средствами базы данных, и при необходимости перенастраивать и автоматически повторять обработку образцов, но уже по новым условиям игры.

Поэтому единицей хранения в инфосистеме Y-base является отдельный проанализированный образец, а полями центральной таблицы служат все используемые в популяционных исследованиях SNP и STR маркеры (с возможностью добавления любых новых полей). Это означает, что инфосистема является, во-первых, не «популяционно-центричной», а «образец-центричной», а во-вторых, не «гаплогруппо-центричной», а «маркер-центричной». Эти два шага навстречу правильному построению инфосистемы (хранению максимально формализованных и первичных, а не производных данных) обеспечивают возможность анализа всех мировых данных о полиморфизме Y-хромосомы и долговременное использование инфосистемы вне

зависимости от возможных будущих изменений в номенклатуре гаплогрупп или топологии их иерархического древа.

Такой подход потребовал создания и программной реализации алгоритма для автоматического преобразования исходных данных о генотипах образцов в частоты гаплогрупп в соответствии с текущей версией их номенклатуры.

Y-BASE В ЭПОХУ ПОЛНОГО СЕКВЕНИРОВАНИЯ Y-ХРОМОСОМЫ

С 2013 года исследования Y-хромосомы вступили в совершенно новую фазу. Использование технологий секвенирования следующего поколения позволило охарактеризовать всю Y-хромосому (точнее, ее участки, поддающиеся надежному секвенированию) для многих образцов. И поскольку каждая семья несет свои собственные уникальные мутации, секвенирование десятка новых образцов из разных семей автоматически означает открытие десятка новых гаплотипов, которые могут претендовать на звание новой гаплогруппы. Однако новая гаплогруппа получает полное признание лишь после того, когда она встречена хотя бы у двух неродственных индивидов.

В результате были секвенированы сотни образцов и известное науке дерево Y-хромосомы за 3 последних года стало в 10 раз подробнее дерева, составленного за 10 предыдущих лет исследования. Понятно, что в этих условиях маркер-центричный принцип обозначения гаплогрупп является спасением – ведь буквенно-цифровые имена гаплогрупп могут измениться за один день. (Отметим в скобках, что в статье, обобщающей полногеномные ис-

следования Y-хромосомы [Karmin et al., 2015] предложен способ стабилизировать и имена гаплогрупп – именовать только ключевые, распространенные ветви на основных иерархических уровнях, а промежуточные ветви и их группы обозначать через знак объединения именованных ветвей).

Хотя Y-base может вобрать в себя информацию по частотам тысяч известных сейчас гаплогрупп и десяткам тысяч тех, которые будут несомненно открыты в ближайшие несколько лет, пока нужды в этом нет. Ведь для подавляющего большинства этих новых гаплогрупп их популяционные частоты неизвестны. Известно лишь, что гаплогруппа существует (данное сочетание SNP-маркеров встречено у пары образцов), но никто не изучал частоту этой гаплогруппы в популяциях. Тем самым все гаплогруппы резко делятся на два ранга: а) существующие лишь в виде ветви на дереве (частоты в популяциях не изучались); б) существующие и в виде ветви, и в виде географической карты частоты их распространения (частоты известны). И хотя раздел дерева Y-base может отражать все многообразие тысяч известных ветвей (ранг 1), ключевые разделы инфосистемы должны содержать данные лишь по тем гаплогруппам, для которых проведен популяционный скрининг и определены их частоты хотя бы в нескольких популяциях (ранг 2). В 2011 году – до эпохи полного секвенирования Y-хромосом – в мировой литературе мы обнаружили 242 такие гаплогруппы ранга 2. К 2015 году, насколько известно автору, в мировой копилке добавилось лишь около сотни новых открытых гаплогрупп (в основном в пределах «больших» гаплогрупп N, C, E, R1a, R1b), для которых проведен популяционный скрининг по новым субгаплогруппам.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ БД Y-BASE

СТРУКТУРА ИНФОСИСТЕМЫ Y-BASE

В качестве системы управления инфосистемы Y-base используется свободно распространяемая СУБД PostgreSQL. Инфосистема содержит такие основные разделы:

- таблица генотипов для изученных образцов (значения SNP и STR маркеров);
- таблица популяций (со справочниками народов, стран, регионов и др.), связанную с таблицей генотипов соотношением «один ко многим»;
- иерархически упорядоченное дерево SNP маркеров;
- справочник обозначений гаплогрупп и их соответствия одному или нескольким (синонимичным) SNP маркерам (номенклатура гаплогрупп);
- первоисточники данных (электронную библиотеку).

Логически работа с инфосистемой может быть разделена на два этапа:

- 1) ввод первичной информации (предусмотрены различные способы ввода);
- 2) проведение анализа хранимых данных для получения частот гаплогрупп Y-хромосомы в разных популяциях мира (предусмотрена возможность выбора и комбинирования различных критериев, формируемых пользователем системы).

Пользователь системы имеет возможность автоматического расчета частот гаплогрупп в популяциях мира. Для анализа данных сторонними программными средствами, реализована функция экспорта данных, с которыми в данный момент работает пользователь, в заданный формат электронной таблицы.

ЛОГИКА ОБРАБОТКИ ДАННЫХ

Основная сложность обработки данных состоит во взаимосвязи различных SNP маркеров, вытекающей из иерархической структуры «родословного дерева» Y-хромосомы.

Общепринято, что если для образца выявлено наличие определенного SNP маркера, то это однозначно доказывает принадлежность образца только к определенной ветви этого дерева. Одновременно это означает, что в образце заведомо отсутствуют SNP маркеры, специфичные для других крупных ветвей дерева Y-хромосомы.

Однако если выявленная ветвь Y-хромосомы имеет, например, пять субветвей, то каждая субветвь определяется еще и своим собственным SNP маркером. Поэтому допустим, что далее образец был проверен на SNP маркеры, специфичные только для трех из пяти известных субветвей данной ветви. Если оказалось, что он не относится ни к одной из трех проанализированных субветвей, то этот образец все еще может нести SNP маркеры на одну из двух оставшихся субветвей, но неизвестно, на какую именно. Однако – увы! – авторы исходных данных не генотипировали SNP-маркеры на оставшиеся две субветочки. Как поступать в этом случае?

В этом случае при подсчете частоты встречаемости всех ветвей (и субветвей) Y-хромосомы:

1) данный образец входит в подсчет частоты своей «крупной» ветви («+»);

2) входит с «отрицательным» значением «-» (как отсутствие данной ветви у данного образца) в подсчет частоты всех прочих ветвей-гаплогрупп;

3) входит с «отрицательным» значением «-» (как отсутствие данной субветви у данного образца) для тех трех субветвей своей ветви, к которым, как было выявлено, он не относится;

4) но этот образец никаким образом не входит в подсчет частоты двух оставшихся субветвей, о принадлежности его к которым из имеющихся данных нельзя ничего заключить («±»).

ЦЕНТРАЛЬНЫЙ МОДУЛЬ ОПРЕДЕЛЕНИЯ ГАПЛОГРУППЫ

Описанная обработка одного образца относительно легко осуществляется вручную грамотным специалистом, но обработка даже нескольких сотен образцов весьма трудоемка и уже ненадежна. Поэтому в инфосистему интегрирован специализированный модуль, который, исходя из статуса тех SNP маркеров, которые известны для данного образца, и иерархической организации дерева гаплогрупп, определяет, к какой именно ветви (гаплогруппе) или субветви относится данный образец.

Приведем пример работы модуля. Например, пользователь задал запрос на частоты гаплогрупп

G2 в разных популяциях мира. Модуль осуществит скрининг и сортировку всех записей (образцов) в отношении этого запроса на три типа.

В первый тип («-») попадут образцы, не относящиеся к данной гаплогруппе, а именно те образцы, которые несут SNP маркеры, маркирующие другие гаплогруппы (ветви) иерархического дерева гаплогрупп (SNP маркеров, исключаяющие их отнесение к гаплогруппе G2).

Во второй тип («+») попадут образцы, несущие SNP маркер, обозначаемый P15, который маркирует искомую гаплогруппу G2. В этот же тип попадут образцы, не проанализированные на маркер P15, но имеющие производный (derived) статус по SNP маркерам, подчиненным данному маркеру (маркирующие субветви в пределах ветви, маркируемой P15). В данном случае экспериментально подтвержденная принадлежность образца к субветви автоматически означает его принадлежность и к ветви в целом.

В третий тип («±») попадут образцы, относящиеся к той же суперветви, что и маркируемые P15, но положение которых внутри суперветви неизвестно. Для этих образцов нельзя вынести определенного суждения, относятся ли они к ветви G2 (P15) или нет.

Наконец, следующий модуль (расчета частот гаплогрупп) подсчитывает для каждой популяции долю образцов второго типа от общего числа образцов в данной популяции, но для популяций, имеющих хотя бы один образец третьего типа, будет отмечено, что частота гаплогруппы G2 неопределима. Эта информация – частоты гаплогруппы G2 во всех популяциях, где ее можно определить из имеющихся данных, и указание на невозможность такого определения для остальных популяций – и будет результатом искомого запроса.

Таким образом, применение модулей автоматического определения гаплогруппы и расчета частот гаплогрупп позволяет полностью автоматизировать наиболее трудоемкий и сложный этап обработки и преобразования первичных генотипических данных для решения каждой конкретной задачи. При этом возможность обновления дерева SNP-гаплогрупп обеспечивает эффективную работу инфосистемы для максимально широкого круга популяций: система не зависит от открытия новых ветвей иерархического дерева, введения в научный оборот новых SNP маркеров и даже реорганизации устоявшейся структуры иерархического дерева, поскольку все эти изменения можно вносить в дерево гаплогрупп, не меняя сами данные и процедуры их обработки.

Данный модуль инфосистемы является центральным в том смысле, что он преобразует первичную (экспериментально полученную) информацию о статусе SNP маркеров в данном образце в искомую информацию о том, к какой именно гаплог-

руппе принадлежит образец. Таким образом, происходит автоматическая обработка экспериментальных данных и помещение образца в одну из известных гаплогрупп. Данный модуль инфосистемы позволяет для каждого образца указать аллельное состояние каждого SNP маркера из полного перечня и, соответственно, проставить принадлежность или непринадлежность образца к каждой из сотен известных на данный момент гаплогрупп.

Крайне важный смысл этой операции состоит в том, что в таблице представлено аллельное состояние даже для тех SNP-маркеров, по которым данный образец не исследовался. Например, у образца, для которого установлено, что маркер M9 находится в состоянии derived (+), также можно предсказать состояние derived и по маркерам M89, SRY10831.1 и другим маркерам, предковым по отношению к мутации в маркере M9 (т.е. маркирующим более крупные ветви, на которых выросла эта более молодая ветвь), хотя в статье эти маркеры более крупных ветвей и не исследовались. Для сопоставления разных исследований, использующих разные панели маркеров и дифференцирующих ветви древа на разную глубину, такое указание маркеров для всех известных ветвей древа становится крайне полезным: мы можем автоматически отобразить для исследования любые более крупные ветви, даже если их маркеры не изучены в публикации и крупные гаплогруппы не указаны.

Результатом работы этого центрального программного модуля является таблица, в которой для каждого образца проставлены аллельные состояния не для 5–20 SNP маркеров, по которым для данного образца имеются экспериментальные данные, но по всем 700 маркерам из перечня известного на момент разработки модуля SNP маркеров. Обновление древа происходит в полуавтоматическом режиме, поэтому это число легко увеличивается до всех известных на данный момент маркеров. Конечно, разработка этого алгоритма программного модуля заняла месяцы, зато время обработки 1000 образцов тестовой программой, даже реализованной макросом MS Excel, заняло всего около 20 минут. Выполнение этой операции программой не только сокращает время работы на несколько порядков, но также исключает ошибки, вызванные человеческим фактором.

МОДУЛЬ РАСЧЕТА ЧАСТОТ ГАПЛОГРУПП

Исходя из данных о принадлежности каждого образца к той или иной гаплогруппе, этот модуль базы данных рассчитывает частоты выбранных пользователем гаплогрупп в выбранных популяциях. Тем самым модуль дает прямой ответ на наиболее типичный запрос к инфосистеме – каковы

частоты интересующей пользователя гаплогруппы в каждой из выбранных популяций?

Задача, которую решает этот модуль программы – непосредственное получение частот гаплогрупп для всех внесённых в базу данных популяций (или только избранных пользователем популяций или регионов) с использованием таблицы отнесения/исключения каждого образца к каждой из известных гаплогрупп, полученной в результате работы центрального модуля.

Алгоритм определения частоты гаплогруппы в популяции по очереди рассматривает каждую популяцию и каждый маркер, отдельно от других популяций и маркеров. Первый шаг – работа с «популяцией А», маркером 12f2.b, второй – с той же популяцией, но уже с маркером 12f2a, и т.д. пока программа не пройдёт все маркеры, выбранные для популяции А. Затем столько же шагов для популяции Б, столько же для В и т.д.

Каждый шаг – это выполнение следующей процедуры: программа оценивает взвешенную долю записей «D» в данном столбце для данной популяции и записывает её в строку «всего в популяции». Это число – доля образцов в популяции, которые имеют в данном маркере производное (derived) аллельное состояние, то есть относятся к гаплогруппе, которую данный маркер определяет. Например, взвешенная доля записей, имеющих в данной популяции измененное состояние («D») в локусе P14 – это частота гаплогруппы F в популяции А. Однако, иногда вычислить эту долю невозможно: когда хотя бы одна запись в данном столбце у данной популяции содержит «U», то есть неизвестное аллельное состояние. В этом случае вместо частоты пишется знак «H» - частота неопределима.

После того как программа проходит по всем популяциям и маркерам, алгоритм оставляет только записи, указывающие для каждого сочетания «популяция-маркер» долю образцов в данной популяции, имеющих производное состояние данного маркера. Теперь вспомним, что одна и та же гаплогруппа может определяться многими маркерами. Поэтому далее модуль объединяет все маркеры для одной гаплогруппы в одну запись (значительно сокращая число столбцов в итоговой таблице). Таким образом, итоговая таблица содержит в строках записи о популяциях, а в столбцах – о гаплогруппах. На пересечениях строк и столбцов – частоты гаплогрупп либо запись о невозможности их определения.

ДЕРЕВО ГАПЛОГРУПП Y-ХРОМОСОМЫ

На нашем сайте www.genofond.ru размещены три раздела базы данных: древо гаплогрупп, электронная библиотека и БД среднеэтнических частот (остальные разделы пока доступны только разработчикам и дружественным пользователям по их запросу).

Дерево гаплогрупп Y-хромосомы было оцифровано, внесено в базу данных и реализован графический веб-интерфейс для работы с ним. Иерархическая классификация гаплогрупп и определяющих их мутаций (маркеров) закодирована в виде таблицы, содержащей обозначение гаплогруппы, перечень дефенирующих ее маркеров (т.е. однозначно ее определяющих) и ссылку на маркеры, дефенирующие более крупную гаплогруппу («родительскую» по отношению к данной гаплогруппе).

Вспомогательная таблица представляет собой полный перечень маркеров с указанием их принадлежности к панели маркеров, определяющих одну и ту же филогенетическую ветвь, и со ссылкой на публикации с описанием данного маркера.

Раздел в настоящее время содержит 762 маркера, объединенных в 485 групп (соответственно, число гаплогрупп также составляет 485). Отметим,

что маркеры, являющиеся лишь «филогенетическими» синонимами (разные мутации, на данный момент определяющие одну гаплогруппу, но филогенетическое положение которых может стать различным после открытия дополнительных субветвей) перечисляются в панели маркеров через запятую и считаются разными маркерами. Маркеры же, являющиеся синонимами в полном смысле этого слова (одна и та же мутация, открытая независимо разными авторами и получившая поэтому разные обозначения), перечисляются через слеш (например, P220/S119) и рассматриваются инфосистемой как один маркер.

Большой объем современных знаний о топологии дерева (порядке возникновения мутаций и иерархической соподчиненности ветвей) делает невозможным удержание всего дерева в памяти даже интенсивно работающего в этой области спе-

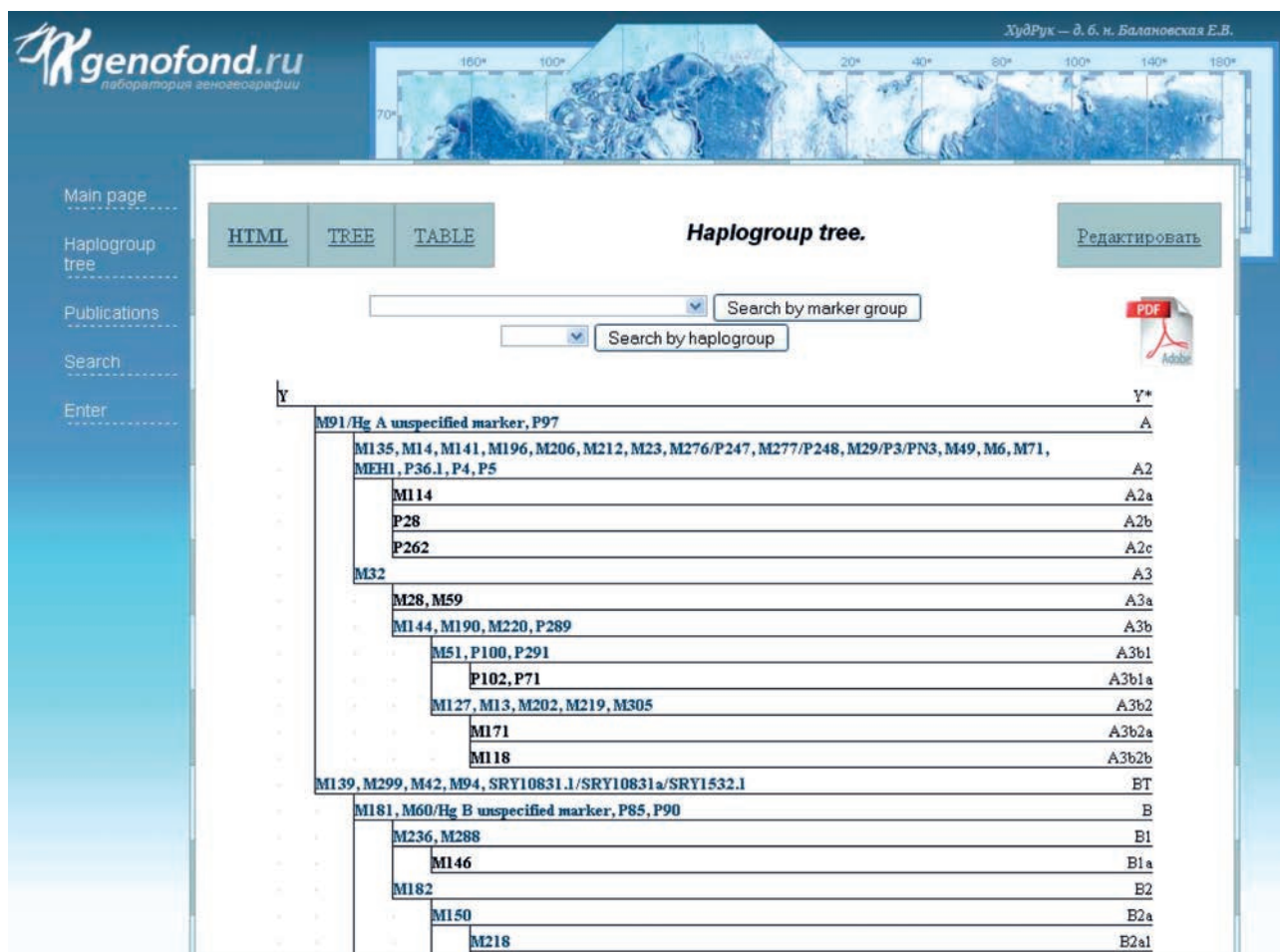


Рис. 1.1. БД Y-base. Раздел дерева гаплогрупп в html-режиме.

Обозначения: В верхней области расположены кнопки выбора режима (html, дерево, таблица), вход в режим редактирования (по паролю администратора) и кнопка печати выделенного фрагмента дерева в pdf файл.

Ниже находится фильтр по гаплогруппе (выбирается либо буквенное обозначение гаплогруппы из выпадающего списка, либо вводится один из маркеров-дефенаторов с клавиатуры).

Основная часть формы занята древом гаплогрупп, на котором нанесены все дефенирующие маркеры, а справа от ветвей – обозначения соответствующих гаплогрупп по текущей классификации.

Черным цветом показаны ветви нижнего уровня (не имеющие подчиненных субветвей). Щелчок по ветвям, маркеры которых подписаны синим шрифтом, раскрывает страницу с фрагментом дерева, корнем которого является выбранная ветвь.

циалиста. Поэтому удобный и быстрый просмотр интересующего фрагмента дерева является по-настоящему возникающей практической задачей. Созданная нами программа по своим возможностям (ускоренная навигация, поиск, получение справочной информации, печать в файл), по меньшей мере, не уступает основному зарубежному аналогу – сайту международного сообщества генетической генеалогии www.isogg.org.

Реализованы три режима работы с деревом для пользователей: режим просмотра html (представлен на рис. 1.1), режим дерева (рис. 1.2) и режим таблицы. Для администратора реализован режим редактирования, позволяющий обновлять как таблицу дерева целиком (загрузка нового дерева из файла), так и редактировать отдельные ветви. Администратор может удалять, добавлять и перепод-

чинять ветви, разделять группы маркеров, редактировать информацию по отдельным маркерам и т.д.

В целом, созданный раздел «дерево гаплогрупп» является эффективным инструментом и для организации хранимых данных о генотипированных образцах (привязка к известной гаплогруппе), и для расчета частот гаплогрупп, и удобным справочником, который может использоваться независимо.

ЭЛЕКТРОННАЯ БИБЛИОТЕКА ПУБЛИКАЦИЙ-ИСТОЧНИКОВ

Одним из важных элементов базы данных являются публикации – источники исходных данных. Легкий доступ к статьям внутри самой базы по-

The screenshot shows the 'Haplogroup tree' interface. At the top, there are tabs for 'HTML', 'TREE', and 'TABLE', with 'TREE' selected. A 'Редактировать' button is visible in the top right. The main content area displays a tree structure of haplogroups. The tree is partially expanded, showing levels A, BT, B, CFDE, CF, C, C-RBF2, C1, C2, C3, C4, C5, C6, F, and DE. Each level lists associated marker groups. Below the tree is a table of markers with the following data:

Marker	Source	Haplogroup	Marker group
1 P220/S119	Karafet et al (2008)	I2b1	L34, L36, L59, M223, P219/S24
2 M320.1	"private=single Druze individual: Sher	T1	M320.1
3 P85	Karafet et al (2008)	B	M181, M60/Ha B unspecified t

Рис. 1.2. БД Y-base. Раздел дерева гаплогрупп в режиме “tree”.

Обозначения: В верхней части экрана представлено раскрывающееся дерево гаплогрупп. Раскрытые ветви отображаются треугольником, направленным вниз, нераскрытые – треугольником, направленным вправо. Чтобы развернуть или свернуть ветвь, следует щелкнуть по соответствующему ей треугольнику. Ветви, не имеющие вложенных субветвей, обозначаются кружками.

В приведенном на рисунке примере раскрыты ветви первого уровня в пределах гаплогруппы С (показаны далее нераскрытые ветви С1 – С5, а также не имеющие вложений ветви С-RBF2 и С6).

Приводятся как обозначение гаплогруппы, так и состав определяющей ее группы маркеров.

В нижней части экрана приводится таблица маркеров. В ней можно осуществлять поиск, вводя искомые значения в поля, расположенные под заголовками соответствующих столбцов. При щелчке по гаплогруппе в верхней форме (дерево) в нижней части автоматически фильтруются все маркеры, определяющие данную ветвь.

зволяет быстро обратиться к первоисточнику за тем или иным уточнением и является эффективной дополнительной гарантией минимизации ошибок (например, при описании популяции).

База публикаций имеет и самостоятельную ценность как тематическая электронная библиотека (тематикой является изменчивость Y-хромосомы в популяциях человека). Хотя большинство статей присутствуют и в глобальных электронных библиотеках (например, PubMed), но далеко не для всех публикаций в свободном доступе имеются полнотекстовые версии; кроме того, в связи с огромным числом статей, Pubmed не всегда удобен для поиска публикаций по Y-хромосоме.

Интерфейс созданного раздела «Источники» показан на рис. 1.3 и рис. 1.4.

Пользователю предоставляется возможность фильтрации по любому из четырех основных полей (авторы, год издания, название статьи, выходные данные). При клике мышью на названии интересующей статьи открывается новая страница, на которой представлена подробная библиографическая ссылка (18 полей, включая аннотацию статьи). Также на этой странице пользователю предлагается возможность ознакомиться с текстом статьи (как правило, pdf файл), а также электронными приложениями (как правило, xls файлы с Y-хро-

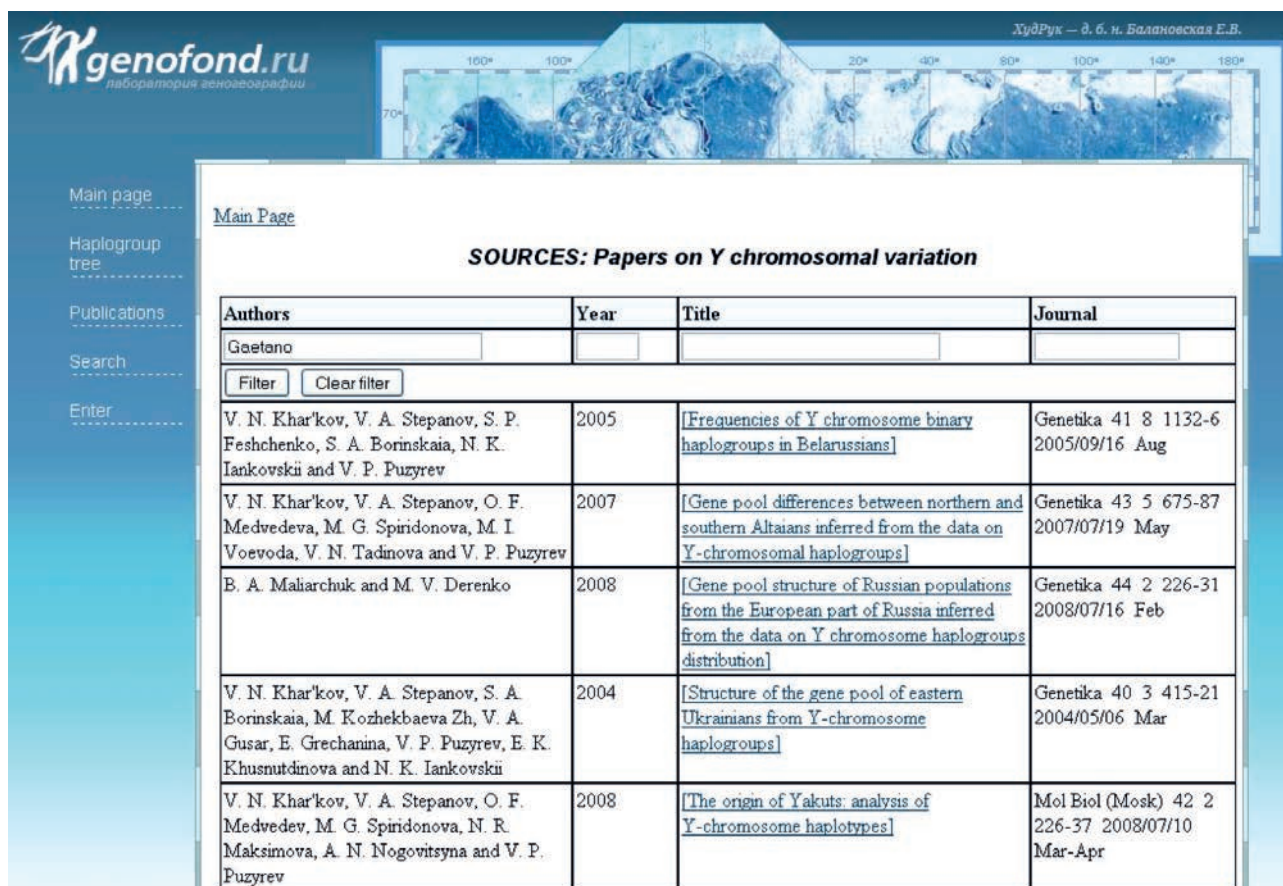
мосомными гаплотипами, которые как раз и явились основными источниками данных для наполнения инфосистемы).

Пополнение раздела осуществляется в режиме администратора. Заполнение библиографических полей предусмотрено в формате наиболее популярной среди генетиков библиографической программы Endnote, что значительно экономит время администратора (один из форматов экспорта Endnote является форматом для автоматического импорта в библиографический раздел Y-base). Ссылки на раздел «Источники» из других разделов инфосистемы осуществляются по ID публикации, а пользователю отображаются в формате «первый автор, год».

База «Источники» доступна на нашем сайте www.genofond.ru, раздел «Базы данных», далее «Y-хромосома» и далее «Publications». Интерфейс раздела создан на английском языке, поскольку разрабатываемая инфосистема ориентирована не только на российских пользователей.

БАЗА СРЕДНЕЭТНИЧЕСКИХ ЧАСТОТ

На сайте www.genofond.ru нами представлен, конечно же, крайне сокращенный, но пригодный к использованию аналог основной версии инфосис-



The screenshot shows the website www.genofond.ru with a search interface. The search results are displayed in a table titled "SOURCES: Papers on Y chromosomal variation". The table has four columns: Authors, Year, Title, and Journal. The search criteria include the author "Gaetano".

Authors	Year	Title	Journal
V. N. Khar'kov, V. A. Stepanov, S. P. Feshchenko, S. A. Borinskaia, N. K. Iankovskii and V. P. Puzyrev	2005	[Frequencies of Y chromosome binary haplogroups in Belarusians]	Genetika 41 8 1132-6 2005/09/16 Aug
V. N. Khar'kov, V. A. Stepanov, O. F. Medvedeva, M. G. Spiridonova, M. I. Voevoda, V. N. Tadinova and V. P. Puzyrev	2007	[Gene pool differences between northern and southern Altaians inferred from the data on Y-chromosomal haplogroups]	Genetika 43 5 675-87 2007/07/19 May
B. A. Maharchuk and M. V. Derenko	2008	[Gene pool structure of Russian populations from the European part of Russia inferred from the data on Y chromosome haplogroups distribution]	Genetika 44 2 226-31 2008/07/16 Feb
V. N. Khar'kov, V. A. Stepanov, S. A. Borinskaia, M. Kozhekbaeva Zh, V. A. Gusar, E. Grechanina, V. P. Puzyrev, E. K. Khusnutdinova and N. K. Iankovskii	2004	[Structure of the gene pool of eastern Ukrainians from Y-chromosome haplogroups]	Genetika 40 3 415-21 2004/05/06 Mar
V. N. Khar'kov, V. A. Stepanov, O. F. Medvedev, M. G. Spiridonova, N. R. Maksimova, A. N. Nogovitsyna and V. P. Puzyrev	2008	[The origin of Yakuts: analysis of Y-chromosome haplotypes]	Mol Biol (Mosk) 42 2 226-37 2008/07/10 Mar-Apr

Рис. 1.3. БД Y-base. Интерфейс раздела «Источники».

В качестве примера показан поиск статей с соавторством Gaetano.

genofond.ru
лаборатория геногеографии

Худрук – д. б. н. Балановская Е.В.

Main page
Haplogroup tree
Publications
Search
Enter

Main Page > SOURCES: Papers on Y chromosomal variation

Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome

Download	Size
Di Gaetano 2008 Differential Greek and northern African migrations to.pdf	321471 byte
Di Gaetano 2008 suppl.xls	46592 byte

Reference Type: Journal Article

Record Number: 28

Author: C. Di Gaetano, N. Cerutti, F. Crobu, C. Robino, S. Inturri, S. Gino, S. Guarrera, P. A. Underhill, R. J. King, V. Romano, F. Cali, M. Gasparini, G. Matullo, A. Salerno, C. Torre and A. Piazza

Year: 2009

Title: Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome

Journal: Eur J Hum Genet

Volume: 17

Issue: 1

Pages: 91-9

Epub Date: 2008/08/08

Date: Jan.

Рис. 1.4. БД Y-base. «Источники», просмотр отдельной записи.

Библиографические детали и файлы для скачивания для одной из найденных статей с соавторством Gaetano.

темы Y-base. Для этого мы определили среднеэтнические частоты, то есть характеристики не отдельных локальных популяций, а их этнических групп – народов, поскольку именно эта информация является наиболее часто востребуемой широкой аудиторией. Онлайн-вариант БД имеет простейшую структуру – таблицы Excel, в строках которой перечислены народы, столбцах – гаплогруппы, а на пересечении указывается частота данной гаплогруппы у данного народа. Такая онлайн-база данных не содержит STR гаплотипов, лишена возможностей автоматического пополнения, поиска по комбинации полей, верификации и других возможностей, реализованных в основной версии инфосистемы, однако она несет важнейшую информацию о частотах гаплогрупп Y-хромосомы у народов мира, и эта информация уже широко используется в исследованиях как нашего коллектива, так и других лабораторий.

База данных среднеэтнических частот гаплогрупп Y-хромосомы у народов мира содержит информацию о частотах 335 гаплогрупп Y-хромосомы у 252 народов и основана на суммарной выборке 55 024 человек.

Для удобства использования среднеэтническая база данных представлена в двух версиях – «полной» и «рекомендуемой». «Полная» версия базы данных содержит информацию о 335 гаплогруппах, что обычно является избыточным и затрудняет навигацию по базе. Поэтому в «рекомендуемой» версии дробные гаплогруппы объединены в 29 основных. Использование «полной» версии целесообразно лишь в тех редких случаях, когда нужно получить информацию по дробному субварианту той или иной гаплогруппы или детально разобраться, на чем основана указанная в рекомендуемой версии частота гаплогруппы. Во всех остальных случаях целесообразно пользоваться «рекомендуемой» версией базы данных.

В «полной» версии в данный момент содержится информация о частотах 335 гаплогрупп. Для удобства пользователей в связи с ограничениями одной из распространенных версий MS Excel (не более 256 столбцов) народы представлены в столбцах, а гаплогруппы – в строках. В ячейках содержатся частоты данной гаплогруппы у данного народа. Народы отсортированы справа налево в алфавитном порядке. Для каждо-

го народа указано его название, страна (или страны), где он был изучен, объем выборки, ссылка на источники информации.

В «рекомендуемой» версии содержится информация о частотах 29 основных гаплогрупп. Для удобства пользования народы представлены в строках, а гаплогруппы – в столбцах. Пустые ячейки означают, что для данного народа нет информации

о частоте данной гаплогруппы. Народы отсортированы сверху вниз в алфавитном порядке. Для каждого народа указано его название, страна (или страны) где он был изучен, объем выборки по данному народу, ссылка на источники информации. В «рекомендуемой» версии применено условное форматирование: ячейки окрашены разным цветом в зависимости от величины значения частоты.

ПРОГРАММА NAPLOMATCH ДЛЯ ПОИСКА СХОДНЫХ ГАПЛОТИПОВ

Совместно с созданием базы данных по Y-хромосоме нашим коллективом были разработаны еще две программы, помогающие обрабатывать массивы данных по Y-хромосоме – предиктор гаплогрупп Y-predictor и программа поиска совпадающих гаплотипов NaplomatcH. Обе программы размещены в свободном доступе (www.genofond.ru, раздел Базы данных / БД по Y-хромосоме).

СУТЬ ПРОГРАММЫ NAPLOMATCH

Задачи сравнения отдельных индивидуумов или целых популяций по спектру гаплотипов мало выполнимы вручную в силу их большой трудоемкости. Нами предложен алгоритм определения сходства STR гаплотипов, пригодный для анализа массовых выборок, и разработана компьютерная программа NaplomatcH, позволяющая находить гаплотипы, отличающиеся от заданного на любое количество мутационных шагов. Программа может функционировать в двух режимах: сравнение индивидов и сравнение популяций. Гибкость программы (возможность использования не встроеной, а любой внешней базы данных), удобство использования (работа происходит с таблицами MS Excel) и возможность распространения на гаплотипы других хромосом и других биологических видов может сделать ее новым полезным инструментом в популяционно-генетических, криминалистических и генеалогических исследованиях. Программа NaplomatcH написана на языке Pascal студентом-дипломником МАИ Ильей Ивановым под руководством автора.

НУЖДА В ПРОГРАММЕ ПОИСКА ГАПЛОТИПОВ

STR-гаплотипы Y-хромосомы являются популярным инструментом в исследованиях генофондов – в базе данных PubMed ежегодно фиксируется несколько десятков статей с их использованием. Не менее важную роль STR маркеры Y-хромосомы (Y-STR) играют и в судебно-медицинской практике: они являются второй по частоте использования (после аутосомных STR-маркеров) системой маркеров при проведении ДНК-экспертиз. Третьей областью применения Y-STR маркеров

является генетическая генеалогия – число коммерческих компаний, предлагающих своим клиентам анализ Y-STR для прослеживания мужской генеалогической линии, неуклонно растет и в России, и в мире.

Основными методами обработки данных о разнообразии гаплотипов Y-хромосомы являются: построение филогенетических сетей; подсчет гаплотипического разнообразия с последующей оценкой возраста гаплогрупп Y-хромосомы методом молекулярных часов; расчет R_{ST} генетических расстояний и построение на их основе графиков многомерного шкалирования; проведение анализа главных компонент; сравнение спектра гаплотипов в разных популяциях; анализ распространения конкретного гаплотипа в широком круге популяций. Программа NaplomatcH находится в русле дальнейшей разработки последнего из перечисленных методов – анализа совпадающих гаплотипов. В популяционно-генетических исследованиях такой подход зарекомендовал себя как один из надежных способов выявления исторических миграций. Например, экспансия финикийцев по Средиземноморью была генетически изучена именно в результате анализа наличия/отсутствия отдельных гаплотипов и их ближайших мутационных «соседей»-гаплотипов [Zalloua et al., 2008a]. Проследить миграцию крестоносцев на Ближний Восток удалось благодаря подробному анализу одного высокоинформативного гаплотипа [Zalloua et al., 2008b]. Для проведения судебно-медицинских экспертиз и для определения вероятного этногеографического происхождения также требуется поиск гаплотипов, совпадающих с заданным, в обширных базах данных. Однако широкое применение такого подхода и в популяционной генетике, и в криминалистике отчасти сдерживается отсутствием четкого алгоритма и специализированного программного обеспечения, поскольку проведение подобных операций вручную слишком трудоемко.

В наиболее популярной в популяционно-генетических исследованиях программе Arlequin можно проводить поиск только полностью совпадающих гаплотипов, т.е. поиск гаплотипов, отличающихся даже на один мутационный шаг, не предусмотрен. Другой программой, предоставляющей

возможность поиска одинаковых гаплотипов, является разработанный европейскими криминалистами онлайн ресурс YHRD <http://www.yhrd> [12]. Однако эта система, во-первых, работает только со своей внутренней, закрытой для пользователей базой данных, а во-вторых, проводит поиск только по отдельному гаплотипу, но не по их совокупности (популяционной выборке). Таким образом, есть острая нужда в программном обеспечении, которое позволило бы автоматизировать процесс сравнения гаплотипов, как между отдельным индивидом и обширной базой данных, так и между целыми популяциями.

АЛГОРИТМ

Алгоритм определения степени сходства гаплотипов основан на сумме различий по всем изученным микросателлитным локусам (Y-STR), т.е. по числу известных мутаций, разделяющих эти два гаплотипа. Сначала у отдельного исследуемого гаплотипа и в базе данных программа находит одинаковые локусы. Затем значения аллелей (число STR-повторов) в этих локусах вычитаются по модулю. И далее эти разности суммируются по всем локусам. Если сумма меньше или равна заданному максимальному шагу мутационных различий, то программа считает это совпадение весомым, и заносит в таблицу результатов.

Поясним вышесказанное простым примером. Если у исследуемого нами образца и i -го образца из базы есть отличие по локусу № 1 в один шаг и локусу № 4 в два шага, а по остальным локусам между ними нет отличий, то сумма мутационных шагов между ними равна трем (отличие в один шаг из первого локуса плюс отличие в два шага из четвертого локуса).

Программа Nhapmatch может функционировать в двух режимах: сравнение индивидов и сравнение популяций.

РЕЖИМ СРАВНЕНИЯ ИНДИВИДОВ

В этом режиме программа анализирует только один гаплотип данного индивида, находя в базе данных гаплотипы, совпадающие с ним полностью или частично – по числу заданных пользователем мутационных шагов. База данных является подгружаемой, поэтому пользователь может сам сформировать и обновлять требуемый массив данных для поиска. Программа производит поиск как гаплотипов, полностью совпадающих с заданным (0 отличий), так и частично совпадающих гаплотипов (отличающихся на 1, 2 и далее мутационных шагов), и указывает популяции, из которых они происходят. Этот режим оптимален при определении вероятного региона происхождения человека, что может использоваться в криминалистической прак-

тике и в генетической генеалогии, а в ряде случаев и в популяционной генетике.

РЕЖИМ СРАВНЕНИЯ ПОПУЛЯЦИЙ

В режиме сравнения популяций программа выполняет более сложную задачу сравнения с базой данных не отдельного гаплотипа, а целой выборки гаплотипов (представляющих интересующую нас популяцию, называемую далее *реперной*). Для этого программа вычисляет генетические расстояния (определяемые как доля сходных гаплотипов) между реперной популяцией и всеми прочими популяциями анализируемой базы данных. Поэтому этот режим оптимален для популяционно-генетических исследований. Программа последовательно проводит сравнение каждого гаплотипа, представленного в реперной популяции, с гаплотипами из популяций в подгружаемой базе данных и потом рассчитывает суммарные характеристики для всей реперной выборки в целом, т.е. находит количество совпадений с заданным пользователем числом мутационных шагов между исследуемой популяцией и каждой популяцией из базы данных. Как и в режиме сравнения индивидов, в режиме сравнения популяций доля общих гаплотипов между сравниваемыми популяциями определяется с разбивкой по 0, 1, 2 и далее мутационных шагов.

Выбор числа мутационных шагов, в пределах которого гаплотипы считаются совпадающими, определяется целями исследования. Если изучаются древние исторические события, то информативными оказываются не только полные совпадения (0 мутационных шагов), но и отличия в 1, 2 и 3 мутационных шага. Это вызвано тем, что гаплотип, который был общим в предковой популяции, мог мутировать, например, по одному разу в каждой из разделившихся популяций, и тогда при изучении современных популяций потомки этого общего гаплотипа будут различаться на два мутационных шага. В то же время имеет смысл рассматривать гаплотипы только в пределах ограниченного числа мутационных шагов, выбирая это число в зависимости от предполагаемой временной глубины родства сравниваемых популяций. Ведь, согласно «эволюционной» скорости мутирования STR маркеров Y-хромосомы, при изучении 17 STR-маркеров разница в 1 мутационный шаг между двумя индивидами означает, что их общий предок жил около 1500 лет назад, при использовании «генеалогической» скорости – около 500 лет. Поэтому для исследования популяционных событий в пределах последнего тысячелетия можно рекомендовать использовать гаплотипы полностью совпадающие или отличающиеся не более, чем на один шаг.

При работе, связанной со сравнением целых популяций, необходимо учитывать неравенство объемов выборок. Большое разнообразие гаплоти-

пов в нерекombинирующих генетических системах (мтДНК и Y-хромосоме) приводит к тому, что спектр обнаруженных гаплотипов зависит от объема выборки, и насыщение не достигается даже при многосотенных выборках. В результате в популяции, представленной большей выборкой, именно по этой причине может быть обнаружено большее число совпадающих гаплотипов, и генетическое сходство с этой популяцией будет завышено. Чтобы избежать этого, число совпадений должно быть нормировано на объемы выборок. Поэтому среди ряда показателей, вычисляемых программой, наиболее информативным является показатель «Отношение точных совпадений к размеру выборки». Эта возможность также является важным преимуществом данной программы.

РАБОТА ПРОГРАММЫ

Для работы программы Naplomatch требуется загрузить в неё два исходных файла: базу данных, содержащую гаплотипы, с которыми будет осуществляться сравнение, и файл с анализируемым гаплотипом (или выборкой гаплотипов). Для удобства пользователя, программа работает с файлами в форматах «.xls» и «.csv», совместимыми с MS Excel.

Файл базы данных содержит следующую информацию: популяция, к которой принадлежит образец; обозначение образца; маркер, определяющий гаплогруппу образца; гаплогруппа образца и количество повторов по каждому из исследованных STR локусов.

Файл с анализируемым гаплотипом содержит: обозначение образца, комментарий к нему (необязательное поле) и количество повторов по каждому из исследованных STR локусов. При наличии ошибок в базе или исходном файле, программа обратит на это внимание пользователя, выделив их красным цветом.

После загрузки данных необходимо задать максимальное число мутационных шагов (от 0 до любого задаваемого пользователем числа), в пределах которого гаплотипы обрабатываются как сходные. После этого можно начинать операцию сравнения всех гаплотипов. По окончании расчетов, на экране отобразятся все гаплотипы из базы данных, отличающиеся от анализируемого гаплотипа на число мутационных шагов, меньшее или равное заданному пользователем. Результаты организованы в таблицу, указывающую обозначения анализируемого гаплотипа и найденного гаплотипа из базы данных; количество совпавших локусов между этими гаплотипами; количество мутационных шагов, разделяющих гаплотипы: в каких локусах содержатся эти отличия; а также всю информацию о найденных гаплотипах, содержащуюся в базе данных. При нажатии кнопки «By population», полученные данные организуются в таком виде, что становится видно,

сколько и в какой популяции найдено совпадений с гаплотипами из исходных данных. Полученные результаты можно экспортировать в привычные форматы csv или xls.

Анализ быстродействия показал высокую скорость при обработке одиночных анализируемых гаплотипов и удовлетворительную – при анализе выборок гаплотипов. Программа эффективно работает и при загрузке баз данных с объемом 10 000 – 15 000 образцов, т.е. объем базы данных не является фактором, лимитирующим быстродействие. Что же касается объема файла с анализируемыми гаплотипами, то скорость оставалась оптимальной при одновременной обработке до 50 образцов. При использовании большего количества скорость работы программы значительно снижается, поэтому крупные выборки, превышающие 50 образцов, рекомендуется загружать в программу частями и затем суммировать число найденных совпадений.

ПРИМЕР В РЕЖИМЕ ИНДИВИДОВ

Приведем в данной главе только более простой пример использования в режиме сравнения индивидов, поскольку использование программы «в деле» – для реального исследования генофонда популяции донских казаков – можно найти в работе [Чухряева и др., 2015]. В режиме же индивидов мы проанализировали три произвольно выбранных индивида из той же выборки казаков; в качестве максимального был выбран шаг равный трем мутационным шагам. Один из выбранных индивидов принадлежал к гаплогруппе **I2a-P37**; второй к **R1a1-M198(xM458)** и третий к **R1a1a7-M458**. Результаты их сравнения с рядом популяций, которые могли быть связаны с происхождением популяции казаков, показаны в таблице 1.3. Из неё видно, что для двух образцов найдены точные совпадения: 4 совпадения для носителя гаплогруппы **I2a-P37** с восточными украинцами и 1 совпадение с западным украинцем, а для носителя **R1a1-M198(xM458)** – 3 совпадения с центральными русскими и 2 совпадения с южными русскими. Лишь для носителя гаплогруппы **R1a1a7-M458** в базе данных не найдено точных совпадений, однако есть совпадения с отличием в один шаг с целым рядом популяций: наибольшее количество совпадений обнаружено с центральными русскими, а также они имеются в популяции восточных украинцев, казанских татар и западных украинцев. Итак, для всех случайно отобранных трех образцов донских казаков, оказалось возможным очертить географическую зону вероятного происхождения прямого предка по мужской линии, что может представлять интерес для генетико-генеалогических исследований, а также при определении вероятного этногеографического происхождения образца ДНК при решении криминалистических задач.

Таблица 1.2. Результаты работы программы Haplomatch в режиме сравнения индивидов.

название популяции	N	Число точных совпадений	Число различий в один шаг	Число различий в два шага	Число различий в три шага
образец с гаплогруппой R1a1a-M458					
центральные русские	442	0	4	2	14
восточные украинцы	397	0	2	6	19
казанские татары	138	0	2	2	0
западные украинцы	586	0	1	8	28
литовцы	298	0	0	4	11
ногайцы	57	0	0	3	2
южные русские	115	0	0	3	2
шапсуги	97	0	0	0	1
грузины	169	0	0	0	1
абхазы	51	0	0	0	0
армяне	168	0	0	0	0
черкессы	129	0	0	0	0
калмыки	156	0	0	0	0
азовские греки	95	0	0	0	0
северные русские	121	0	0	0	0
образец с гаплогруппой R1a-M198					
центральные русские	442	3	8	24	23
южные русские	115	2	4	7	7
литовцы	298	0	3	12	18
ногайцы	57	0	1	2	1
казанские татары	138	0	2	5	8
западные украинцы	586	0	4	19	39
восточные украинцы	397	0	0	9	20
грузины	169	0	0	1	0
азовские греки	95	0	0	1	4
черкесы	129	0	0	2	3
абхазы	51	0	0	0	0
армяне	168	0	0	0	0
шапсуги	97	0	0	0	0
калмыки	156	0	0	0	0
северные русские	121	0	0	0	0
образец с гаплогруппой I2a1-P37					
восточные украинцы	397	4	5	11	20
западные украинцы	586	1	8	18	34
казанские татары	138	0	2	1	0
литовцы	298	0	0	2	1
центральные русские	442	0	0	4	2
ногайцы	57	0	0	1	0
южные русские	115	0	0	1	3
абхазы	51	0	0	0	0
армяне	168	0	0	0	0
черкесы	129	0	0	0	0
шапсуги	97	0	0	0	0
калмыки	156	0	0	0	0
грузины	169	0	0	0	0
азовские греки	95	0	0	0	0
северные русские	121	0	0	0	0

Для тех же трех образцов донских казаков мы попытались определить регион происхождения с помощью YHRD [12]. Точные совпадения были найдены для тех же двух образцов. Но по панели Y-filer, состоящей из 17 локусов, в YHRD становится возможным определить происхождение образца географически лишь крайне приблизительно – «Восточная Европа». Разумеется, для криминалистики и даже для генеалогии такие сведения являются чересчур общими и малоинформативными. Ресурс YHRD также предлагает поиск совпадений по минимальной панели, состоящей из 9 STR маркеров, по которой база данных гораздо обширнее, чем по Y-filer (17 STR маркеров). Разумеется, при поиске по минимальной панели совпадений было найдено на порядок больше. Но поскольку уменьшение количества анализируемых локусов уменьшает и разрешающую способность метода, то и в этом случае области генетического сходства остаются слишком широкими. А именно, для образца **R1a1a7-M458** она охватывает часть западной, восточную, южную Европу и центральную Россию; совпадения обнаружены также и в Юго-Западной Азии. Для образца **I2a-P37** область совпадений определена несколько точнее – в основном она приходится на Украину и центральную Россию, но и в центральной Европе обнаружено немало сходных гаплотипов. Но для образца, принадлежащего к гаплогруппе **R1a1-M198(xM458)** картина получается еще более запутанной – совпадения по минимальной панели найдены не только в России (где все же располагается наибольшее количество совпадающих образцов) и в Европе, но также и в Афганистане, Монголии, Китае и некоторых других странах. К тому же в YHRD данные по обнаруженным совпадениям невозможно получить в виде таблицы, где были бы указаны конкретные регионы происхождения образцов. Таким образом, единственный сходный по задачам ресурс YHRD не может рассматриваться в качестве замены разработанной нами программы Haplomatch. Существование, что Haplomatch позволяет пользователю подгружать любую базу данных, тогда как YHRD работает только с собственной встроенной, закрытой для пользователей базой данных. Также

Нарpmatch позволяет проводить сравнение не только индивидуальных гаплотипов, но и сравни-

вать генофонды популяций по выборкам гаплотипов.

ПРЕДИКТОР SNP-ГАПЛОГРУППЫ ПО STR-ГАПЛОТИПУ

СУТЬ ПРЕДИКТОРА

Программа-предиктор Y-Predictor нацелена на прогноз SNP гаплогруппы Y-хромосомы для каждого образца по данным об STR маркерах данного образца. В отличие от определения гаплогруппы по SNP маркерам, то есть по прямым данным, предиктор определяет (предсказывает, дает наиболее вероятный прогноз) гаплогруппу по косвенным данным, а именно по информации об STR маркерах. Основой для этого является общепризнанная высокая корреляция STR гаплотипа с гаплогруппой, определяемой по SNP маркерам. Хотя достичь стопроцентной правильности прогноза гаплогруппы по STR данным невозможно, для практического применения такая программа является чрезвычайно полезным инструментом. Создание алгоритма и программная реализация предиктора Y-Predictor выполнены Вадимом Урасиным, а организация работы, информационной и финансовой поддержки, а также проектирование интерфейса программы – автором данного исследования. Среди большого числа близких по задачам программ разработанный нами предиктор является одним из наиболее точных, и, пожалуй, наиболее удобным в использовании.

ОБЛАСТЬ ПРИМЕНЕНИЯ

Во-первых, для образцов, изученных как по SNP, так и по STR маркерам, модуль Y-Predictor дает второе независимое определение гаплогруппы, которое теоретически должно совпасть с первым – прямым определением по SNP маркерам. На практике же нередко возникают расхождения между результатами, полученными этими двумя методами, благодаря чему мы получаем дополнительную информацию.

Например, предположим, что образец по SNP маркерам относится к гаплогруппе **G**, но не был изучен на подчиненные SNP маркеры (субветви гаплогруппы), а прогноз по STR маркерам указывает на гаплогруппу **G2**. Очевидно, что в этом случае противоречия нет, а имеет место уточнение – пусть предположительное – SNP гаплогруппы благодаря применению STR-предиктора (переход с уровня над-ветви **G** на уровень ветви **G2**). Если же модуль-предиктор сообщает прогноз на иную гаплогруппу, например **R1b**, то, хотя такое противоречие может быть вызвано неверным прогнозом, оно также указывает пользователю на возможность ошибки в экспериментальных SNP данных по этому образцу и на необходимость относиться к све-

дениям из данной публикации с определенной осторожностью.

Во-вторых, для образцов, изученных по STR маркерам, но не изученным по SNP маркерам, модуль Y-Predictor является единственной возможностью определения гаплогруппы, пусть даже предположительного. Далее, уже в зависимости от конкретных задач, можно выбрать: а) пользоваться ли данными по образцам, гаплогруппы которых определены только предиктором; б) ограничиться данными по образцам, гаплогруппы которых определены прямым методом по SNP маркерам; в) или же (для задач, требующих особой тщательности) отобрать только те образцы, в которых гаплогруппы были определены обоими способами и их результаты совпали.

Третья область применения предиктора – упрощение определения гаплогрупп лабораторными методами. Это позволяет сначала провести определение STR маркеров. Затем предиктор дает прогноз гаплогруппы для данного образца. После этого проводится его тестирование лишь по тому конкретному SNP маркеру, которое предсказано предиктором. Таким образом, проводится проверка прогноза типированием одного SNP маркера вместо иерархического типирования множества SNP маркеров. Лишь в тех случаях, когда прогноз не совпал с прямым SNP тестированием, проводится тестирование всего ряда SNP маркеров.

ПРИНЦИП РАБОТЫ

Y-Predictor определяет Y-гаплогруппы для STR-гаплотипов (значений STR маркеров для тестируемого образца) с использованием филогенетических деревьев каждой отдельной ветви (гаплогруппы) и радиуса потомков основных узлов дерева. Филогенетические деревья, представляющие собой иерархическое древо эволюции STR маркеров, построены для каждой гаплогруппы в отдельности и получены перечни гаплотипов узлов дерева с радиусом потомков узла. Для минимизации расчетов используется только группа самых близких к корню дерева узлов (основных узлов), наиболее полно характеризующих разнообразие гаплотипов.

С этим набором заранее созданных деревьев сопоставляется тестируемый STR-гаплотип до момента обнаружения на одной из частей дерева такого гаплотипа, который наиболее сходен с тестируемым, но для которого уже точно известна принадлежность к конкретной гаплогруппе, благодаря параллельному анализу по SNP маркерам. В главном модуле программы Y-Predictor для тести-

руемого гаплотипа вычисляется вероятность происхождения от каждого основного узла дерева и выдвигается предположение о принадлежности к гаплогруппе.

Для обучения программы Y-Predictor использовались пятьдесят тысяч Y-STR гаплотипов из различных открытых источников и опубликованных научных статей по популяциям регионов всего мира, а также по большому массиву наших собственных неопубликованных данных о коренном населении Северной Евразии. Программой предлагается прогноз о принадлежности анализируемого гаплотипа к 76 основным гаплогруппам.

Перечень определяемых гаплогрупп: A-M91, A*-M91*(xP108,M32), A1-P108, A3-M32, B-M60, B2a-M150, B2b-M112, C-M130, C2-M38, C3-M217, D-M174, E-M40, E1a-M33, E1b1a-M2, E1b1b1-M35, E2-M75, F-M89, F*-M89*(xM201,M69,M523), G-M201, G1-M285, G2-P287, G2*-P287*(xP15,M377), G2a-P15, G2c-M377, H-M69, I1-M253, I1d-L22, I2-M438, I2*-M438*(xP37.2,S23), I2a-P37.2, I2b-S23, I2b1-M223, I2b2-S154, J1-M267, J1*-M267*(xM365,L136), J1b-M365, J1c-L136, J1c2-P56, J1c3-P58, J2-M172, J2a-M410, J2a*-M410*(xL26), J2a4-L26, J2b-M12, K-M9, K*-M9*(xM11, M214,

M45, M70), L-M11, L1-M27, L2-M317, N-M231, N1*-LLY22g*(xP43,Tat), N1b-P43, N1c-Tat, O-M175, O1-MSY2.2, O2-P31, O3-M122, P-M45, P*-M45*(xM242,M207), Q-M242, Q1a-MEH2, Q1b-M378, R1a*-M420*(xSRY10831.2), R1a1a-M198, R1a1a*-M198*(xM417), R1a1a1*-M417(xM458), R1a1a1g-M458, R1b-M343, R1b1*-P25*(xP297), R1b1b1-M73, R1b1b2-M269, R2-M124, T-M70, T*-M70*(xL162,L131), T2-L162, T3-L131.

Для прогноза учитываются значения 76 Y-STR маркеров: CDY-a, CDY-b, DYS 19/394, DYS 385a, DYS 385b, DYS 388, DYS 389-1, DYS 389-2, DYS 390, DYS 391, DYS 392, DYS 393, DYS 395S1a, DYS 395S1b, DYS 406S1, DYS 413a, DYS 413b, DYS 425, DYS 426, DYS 436, DYS 437, DYS 438, DYS 439, DYS 441, DYS 442, DYS 444, DYS 445, DYS 446, DYS 447, DYS 448, DYS 449, DYS 450, DYS 452, DYS 454, DYS 455, DYS 456, DYS 458, DYS 459a, DYS 459b, DYS 460, DYS 461, DYS 462, DYS 463, DYS 464a, DYS 464b, DYS 464c, DYS 464d, DYS 472, DYS 481, DYS 487, DYS 490, DYS 492, DYS 511, DYS 520, DYS 531, DYS 534, DYS 537, DYS 557, DYS 565, DYS 568, DYS 570, DYS 572, DYS 576, DYS 578, DYS 590, DYS 594, DYS 607, DYS 617, DYS 635, DYS 640, DYS 641,

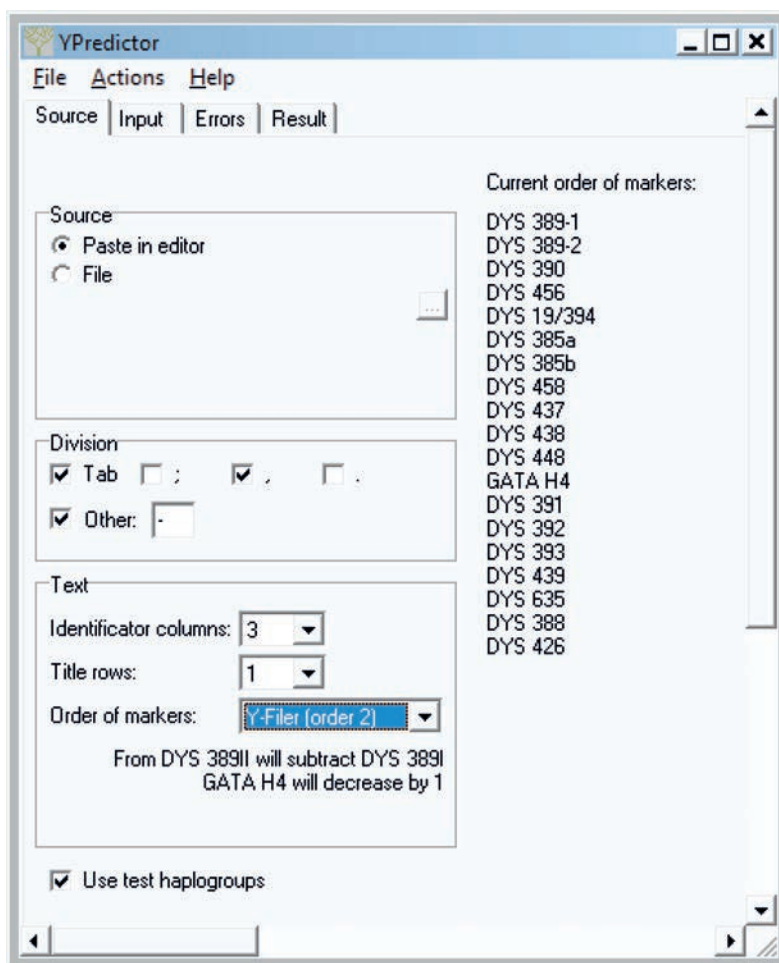


Рис. 1.5. Интерфейс оффлайновой версии программы предиктора.

GAAT1B07, GATA A10, GATA H4, YCA IIa, YCA IIb. В практике популяционных исследований, конечно, набор STR маркеров, типированных для образца, намного меньше (обычно от 7 до 17).

Основу предиктора составляют филогенетические деревья перечисленных гаплогрупп. При построении деревьев использовались четыре тысячи образцов, протестированных по SNP и широкой панели STR маркеров, максимально отражающие разнообразие гаплотипов внутри гаплогрупп и различия между гаплогруппами.

Для каждого узла деревьев произведена оценка возраста, что с учетом различий и совпадений в значениях маркеров позволяет вычислить вероятность того, что узел является предковым для анализируемого гаплотипа. Это вычисление производится для всех филогенетических деревьев, и таким образом для всех гаплогрупп вычисляется вероятность отнесения к ним гаплотипа.

ИНТЕРФЕЙС

Предиктор реализован как в виде отдельной запускаемой программы (рис. 1.5), так и в онлайн режиме (predictor.ydna.ru). Ввод данных осуществляется либо заданием STR гаплотипа в окне программы, либо указанием внешнего файла (таблица Excel) с выборкой, что удобно для обработки многих образцов. При анализе конкретного гаплотипа может быть выбрано число и порядок маркеров (например, на рисунке выбрано 17 маркеров, входящих в набор Y-filer). Вывод результата осуществляется как в окне программы, так и в файле.

Интерфейс для работы с программой позволяет обрабатывать не единичные гаплотипы, а до 200 гаплотипов одновременно, что является важным преимуществом данного предиктора перед зарубежными аналогами. По каждому гаплотипу выдается наиболее вероятная гаплогруппа и вероятность принадлежности к ней (от 50% до 100%).

1.2. БАЗЫ ДАННЫХ ПО мтДНК И АУТОСОМНЫМ МАРКЕРАМ

КРАТКАЯ ХАРАКТЕРИСТИКА БД MURKA

НЕОБХОДИМОСТЬ БАЗЫ

Митохондриальная ДНК (мтДНК) также является одной из наиболее изучаемых генетических систем. Хотя мтДНК быстро уступила пальму первенства более информативной Y-хромосоме, однако и сейчас ежегодно в мире публикуются десятки популяционно-генетических работ, посвященных анализу мтДНК в тех или иных популяциях. Данные по мтДНК содержатся также в статьях криминалистов, использующих мтДНК для целей идентификации личности. Чтобы весь этот огромный массив опубликованной информации приносил реальную пользу, исследователю необходимо иметь его под рукой в виде единой базы данных. Ведь в течение последних 25 лет в мировой науке сошла целая лавина статей, несущих сведения по огромному числу образцов мтДНК. Наша база данных «MURKA: база данных по митохондриальной ДНК и интегрированное программное обеспечение», создаваемая под руководством и при самом активном участии автора, росла также лавинообразно: в 2004 году в

базе имелась информация о 33 000 образцах мтДНК; в 2007 году – о 62 000 образцах; в 2008 году – о 95 000 образцах; в 2010 году – о 135 000 образцах, в 2011 году – о 168 000 образцах, на текущий момент – о 206 000 образцах (табл. 1.3).

СОДЕРЖАНИЕ БАЗЫ

Важной чертой базы данных MURKA является тщательная и подробная фиксация характеристик изученных популяций: не только их название и объем выборки, но и административная принадлежность (страна, провинция), этническая и лингвистическая принадлежность, географические координаты, ссылки на источник данных и т.д. База содержит как частоты гаплогрупп, так и сиквенсы (ГВС1, а если имеются данные, то и ГВС2 или полный сиквенс мтДНК) с указанием начальной и конечной позиции секвенирования (например: 16024; 16400). Ряд характеристик БД приведен в табл. 6. По представленным в таблице данным можно сделать вывод о превосходстве разработанной базы данных MURKA перед двумя наиболее полными зарубежными ана-

Таблица 1.3. Характеристика созданной базы данных по мтДНК

Параметр	БД MURKA	EMPOP (зарубежный аналог)
Объем (число образцов)	132 600	34 617
Число популяций	2 101	260
Число источников	497	Нет данных
Размещение	Локальная сеть	www.empop.org

логами. Так, общее число образцов составляет в БД MURKA 206 000. Из них 8 870 – полные мтДНК геномы, 15 000 – данные по отдельным гаплогруппам, 50 000 – данные общественных участников проекта «Генографик» (нет привязки к конкретным популяциям) и 132 600 образцов – обычные популяционные данные о коренном населении, которые и использовались в исследованиях, приведенных в последующих главах.

ИСТОРИЯ СОЗДАНИЯ

База данных создавалась при частичной поддержке ряда грантов РФФИ и РФФИ, а в 2009-2012 годах создание этой базы данных являлось одним из важных направлений темы НИР «Анализ распространения гаплотипов митохондриальной ДНК и Y-хромосомы у народов мира на основе создания геоинформационных систем», выполненной в Медико-генетическом научном центре РАМН под руководством автора. Хотя и эта многолетняя работа по созданию базы велась главным образом на чистом энтузиазме ее разработчиков, однако важна была также финансовая поддержка международного проекта «Genographic», и сейчас работа интенсивно продолжается все также под руководством автора в Институте общей генетики РАН.

Работа по сведению воедино всех опубликованных данных об изменчивости митохондриальной ДНК потребовала не одного года и продолжается по сей день. Первые варианты базы данных составлялись автором и затем под его руководством А.С.

Пшеничным в 2003–2005 годах. Программную реализацию основных функций в то время осуществил Ю.В. Пустовой. Современный вариант базы появился благодаря усилиям В.В. Запороженко, взявшего на себя труд не только собрать практически все опубликованные данные, но и провести определение гаплогрупп для каждого образца с помощью разработанного им оригинального программного обеспечения. Эта задача унификации данных разных авторов и определения гаплогрупп – едва ли не самая трудная при создании базы данных, поскольку для большинства опубликованных данных имеется информация только о некодирующей части мтДНК (ГВС1), тогда как надежное определение гаплогрупп возможно лишь при наличии достаточного набора данных по ПДРФ сайтам (SNP маркерам) кодирующей части мтДНК.

При создании текущей версии базы данных была проведена большая работа по подробной и стандартизированной характеристике охваченных популяций, включая обоснованное присвоение им определенных географических координат, для чего потребовалось обращение к статьям-первоисточникам и их внимательный разбор. В выполнении этой работы для двух тысяч популяций большую помощь оказали Антонина Кузнецова, Анастасия Агджоян, Алексей Балаганский, Андрей Шанько и Елена Баранова, каждый из которых описал более сотни популяций под руководством автора, которым также проведено исключение или объединение многих популяционных выборок для целей картографирования.

ПРОГРАММИРОВАНИЕ И ИНТЕРФЕЙС

ПРОГРАММНАЯ СРЕДА

База данных по мтДНК долгое время существовала лишь в виде системы таблиц MS Excel, затем была переведена в формат MS Access, средствами которого был реализован ряд функций поиска, ввода и выгрузки данных. Однако большой объем БД и сложность процедур обработки потребовали перевода базы в более мощную программную среду. Поэтому в 2010 году был осуществлен перевод базы данных в среду MS SQL Server, что позволяет обеспечивать целостность данных, надежность их хранения и ограничений доступа, а также совместимость с целым рядом уже написанных Windows приложений. Последний фактор оказался решающим в выборе именно данной СУБД, а не многочисленных СУБД, функционирующих в ОС Linux. Также существенна возможность использования MS Access в качестве клиента, и поэтому пользовательский интерфейс реализован в среде MS Access, интуитивно понятной большинству пользователей по опыту работы в родственных программах MS Excel и MS Word, но хране-

ние данных и протоколы их обработки реализованы в мощной среде MS SQL Server.

СТРУКТУРА ТАБЛИЦ

Реализована система из более чем 20 взаимосвязанных таблиц: генотипы, образцы, популяции, источники данных, народы, страны, провинции, стратегии формирования выборок, регионы мтДНК, пользователи, гаплогруппы и т.д.. Первые четыре таблицы являются центральными, соединенными по цепочке связью «один ко многим».

В таблице «генотипы» (Sequences) содержится нуклеотидная последовательность фрагментов мтДНК в сравнении с Кембриджской референсной последовательностью, приведен интервал секвенирования, указан номер доступа к этой последовательности в базе данных.ncbi (если имеется). Если один и тот же образец был изучен по разным фрагментам мтДНК (например по ГВС1 и ГВС2), эта информация приводится в разных записях таблицы сиквенсов.

Для объединения этих данных предусмотрена таблица образцов (Samples), в которой приводится идентификатор образца (ID) и указание на популяцию.

В таблице Populations приводится ссылка на источник данных и подробная характеристика популяции (регион мира, страна, народ, географические координаты, стратегия формирования выборок).

Многие вспомогательные таблицы (например, таблица стран) связаны с таблицей популяций и служат для ее формализованного заполнения и облегчения поиска по базе.

В последней из основных таблиц (References) указываются библиографические и иные сведения об оригинальных публикациях и неопубликованных данных нашего коллектива, послуживших источниками данных при наполнении БД.

Остальные таблицы содержат дерево гаплогрупп и различную вспомогательную информацию (например, тип популяционных данных: современные популяции, древняя ДНК, популяции, изученные только по одной гаплогруппе и т.д.).

ИНТЕРФЕЙС

Реализован пользовательский интерфейс, главная часть которого (форма поиска и отображения

результатов) представлена на рис. 1.6. В верхней части формы пользователь может задавать нужные параметры фильтра. Можно видеть (рис. 1.6), что предусмотрен поиск практически по всем полям базы. В приведенном примере задан поиск конкретного гаплотипа (отличия от Кембриджской последовательности в позициях 16224, 16304, 16311 молекулы мтДНК). Флажок «exact» не активирован, поэтому будет осуществляться поиск не только точных совпадений, но и всех гаплотипов, содержащих данные замены (разрешается наличие дополнительных мутаций).

В нижней части формы отображается информация, найденная в результате запроса. Пользователь может перемещаться между вкладками, просматривая, соответственно, информацию о найденных сиквенсах (вкладка Sequences), секвенированных образцах (вкладка Samples), о популяциях, из которых происходят эти образцы (вкладка Populations), и об источниках, в которых была опубликована информация о популяциях (вкладка References). На рис. 1.6 приведен пример с просмотром популяций, а на рис. 1.7 – с просмотром сиквенсов для того же запроса.

Также реализована возможность просмотра и копирования таблиц источников, популяций, образцов, сиквенсов и гаплогрупп.

The screenshot shows the MURKA Database search interface. At the top, there are search filters for Ref. Type, Pop. Type, and various demographic fields. A search query for the haplotype '16224-16304-16311' is entered in the 'Contig' field. The search results summary indicates that 45 samples were found across 23 populations and 17 references. The 'Populations' tab is active, displaying a table with the following data:

Pop	Population Name	Population Type	Status	Reference N.	World Region	Country	Ethnos
1859	Egypt of FBI Laboratory	POP	Valid	FBI Laboratory	N AFRICA		Egyptians
1852	Portugal Northern of Gonzalez	POP	Valid	Gonzalez 2003	W EUROPE		Portuguese
6246	Cyprus of Irwin	POP	Valid	Irwin 2007a	N EAST	Cyprus	Greeks
2327	Slovaks of Koledova						Slovaks
1972	Egyptians northern a,b						Egyptians
2109	Iranians (southwest, Iran) of Metspalu						Iranians S
2441	Portugal Central of Pereira					Portugal	Portuguese
1697	Algerians of Plaza						Algerians
1703	Catalans of Plaza	POP	Valid	Plaza 2003	W EUROPE		Spain
2181	Kurds of Richards	POP	Valid	Richards 2000	N EAST		Kurds
1651	Croatian Islanders, Hvar	POP	Valid	Tolk 2001	BALCANS		Croatians

Рис. 1.6. Поисковая форма базы данных MURKA. Активна вкладка «Популяции».

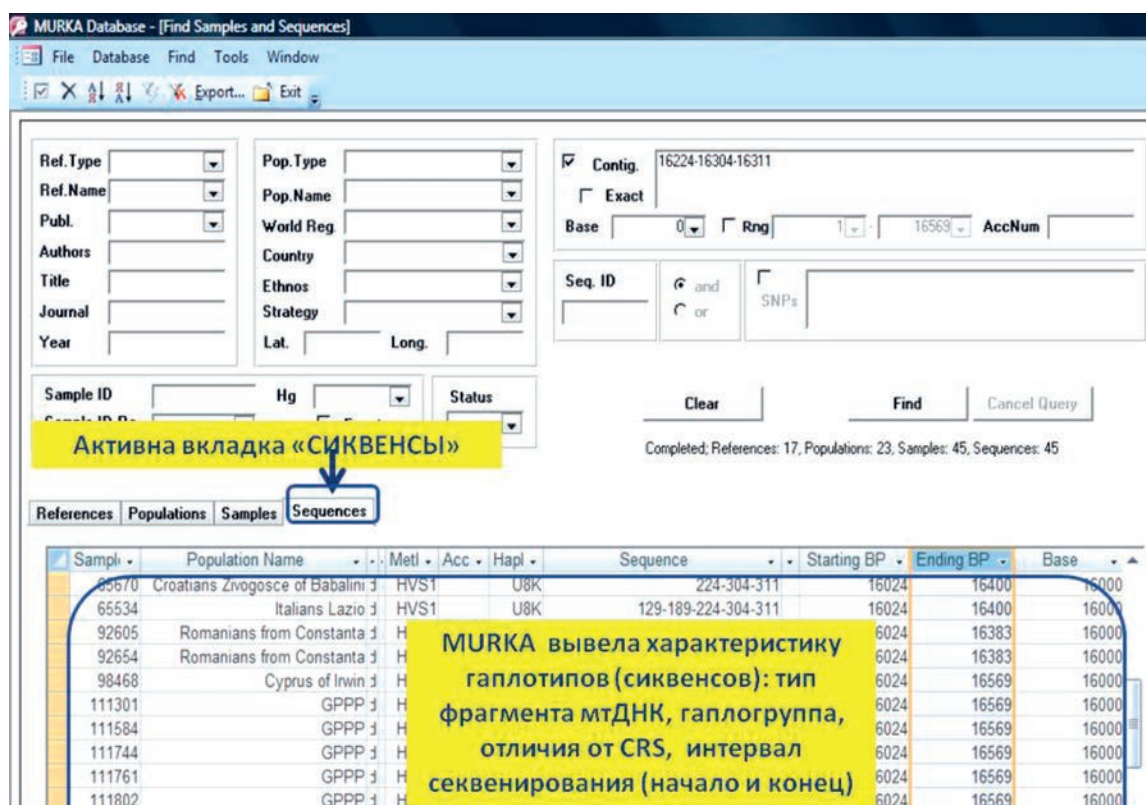


Рис. 1.7. Поисковая форма базы данных MURKA. Активна вкладка «Сиквенсы».

БАЗЫ ДАННЫХ ПО АУТОСОМНЫМ МАРКЕРАМ

Аутосомные маркеры традиционно являлись основными генетическими системами, используемыми в популяционной генетике. Сначала они были представлены так называемыми классическими маркерами (иммунологическими, физиологическими и биохимическими). Затем большинство исследователей сосредоточились на использовании отдельных аутосомных SNP и STR маркеров, причем среди последних особое место занимают маркеры, широко используемые при ДНК-идентификации в криминалистической практике. На современном этапе, хотя продолжается использование и всех перечисленных типов маркеров, акцент сместился на широкогеномные маркеры. И поскольку наш коллектив – российская школа геногеографии – с восьмидесятых годов прошлого века ведет исследования закономерностей в структуре генофондов, его все эти годы отличала традиция сбора больших массивов данных и их формализация в банках данных.

БАНК ПО КЛАССИЧЕСКИМ МАРКЕРАМ

Первым стал банк данных GENE POOL, разработанный под руководством Е.В. Балановской и включающий данные по частотам нескольких десятков классических маркеров в нескольких тысячах популяций СССР и, в меньшей степени, мира. Банк реализован в СУБД FoxPro и обладает разви-

тыми системами фильтрации, ввода и экспорта данных, процедурами верификации данных и т.д. Схема данных представляет собой центральную таблицу частот, связанную с крупными таблицами популяций и библиографических источников и большим числом вспомогательных таблиц. Во время создания банка (90е годы XX века) системы создания баз данных были далеко не так разработаны, как сейчас, поэтому практически все процедуры разработчикам приходилось проектировать и программировать самостоятельно, и создание столь полного, продуманного и эффективно организованного банка было огромным научным достижением. Именно на использовании этого банка были основаны серии основополагающих работ по исследованию структуры генофонда народонаселения СССР в целом, генофондов отдельных народов и географических регионов, выявлению селективной структуры генофонда мира и т.д. В частности, на информации этого банка данных основана генетическая часть нашего исследования русского генофонда, и поэтому в книге [Балановская, Балановский, 2007] также уделено внимание характеристике банка данных GENE POOL и опубликована в виде таблиц та его часть, которая относится к русским популяциям.

Когда в обиход популяционной генетики стали массово входить отдельные диаллельные и мульт-

тиаллельные ДНК-маркеры, автором данной книги было проведено наполнение банка данных GENEPOOL информацией о полиморфизме этих аутосомных ДНК-маркеров в популяциях Восточной Европы и некоторых других регионов Северной Евразии. Информация, извлеченная из этого банка и проанализированная автором методами геногеографии, легла в основу, например, монографии [Лимборская и др., 2002] и ряда других работ. Таким образом, с точки зрения содержания был создан, пополнялся и поддерживался банк данных о полиморфизме ДНК-маркеров, но с точки зрения программной оболочки и возможностей это была база данных GENE POOL, расширенная за счет новых ДНК маркеров.

БАЗА ДАННЫХ ПО МАРКЕРАМ ДНК-ИДЕНТИФИКАЦИИ

В настоящее время из панелей аутосомных ДНК маркеров наиболее массовые исследования проводятся по STR-маркерам, входящим в популярные панели маркеров ДНК-идентификации, которые чрезвычайно широко используются в криминалистике. А поскольку для определения точности ДНК-идентификации надо знать частоты аллелей этих маркеров в популяции, откуда происходит анализируемый образец, в результате усилий многих лабораторий эти частоты были определены для множества популяций по всему миру. Пожалуй, маркеры ДНК-идентификации являются одной из наиболее подробно изученных генетических систем, хотя и уступающих по числу охваченных популяций группам крови АВ0 (безусловному лидеру), а также мтДНК и Y-хромосоме. Под руководством автора была составлена база данных частот маркеров ДНК-идентификации в популяциях Евразии.

Проведенная работа по разработке этой базы данных основывалась на сборе и унификации литературных данных. Первая сводка данных была составлена в ходе подготовки статьи [Степанов и др., 2011] – она включала и обширную экспериментальную информацию, полученную в этом исследовании коллективом, и литературные данные по популяциям России и некоторых стран Европы. Объем этой базы составлял 8600 образцов. Затем база данных была значительно увеличена и унифицирована нами в ходе выполнения проектов РФФИ 11-04-01867а и 12-06-12002-офи_м. Общий объем накопленных в базе данных составляет на данный момент 18 528 образцов из 84 популяций. Общее число обнаруженных аллелей достигает почти шести сотен (593 аллеля). Общее число локусов (изученных хотя бы в одной статье-источнике) составило 33 STR маркера. Географическое положение популяционных выборок, включенных в базу, показано на рис. 1.8.

В базе заполняются следующие поля:

- источник данных (ссылка на публикацию);
- страна обследования;
- провинция (в пределах страны);
- народ;
- название популяции;
- географические координаты (широта, долгота и какая именно точка бралась за ориентир при определении координат);
- объем выборки;
- обозначения локусов и аллелей.

Основное содержание базы данных - частоты каждого аллеля для каждой популяции, в которой данный локус был изучен. Для каждого локуса в каждой популяции рассчитаны также и внесены в БД значения его гетерозиготности.

МАССИВЫ ШИРОКОГЕНОМНЫХ ДАННЫХ

Особенностью широкогеномных массивов данных является очень большое число изученных маркеров (сотни тысяч) при очень небольшом числе изученных популяций (как правило, единицы, в редких случаях – один-два десятка). А поскольку и число популяционно-генетических работ, в которых публикуются широкогеномные данные, пока не так велико, то даже суммарное число популяций мира, в которых они изучены, исчисляется пока не тысячами, как для мтДНК и Y-хромосомы, а всего лишь десятками.

Другой особенностью широкогеномных данных является то, что единицей анализа выступает не популяция, а индивид, при этом число изученных индивидов суммарно по всем статьям составляет уже несколько тысяч. Поэтому ранее разработанные инфосистемы для широкогеномных маркеров не подходят ни по техническим параметрам, ни по идеологии.

В настоящее время отсутствует единая инфосистема, объединяющая широкогеномные данные. Но биоинформатики, анализирующие их, обычно не встречают технических затруднений в стыковке массивов данных из разных опубликованных работ. По мере накопления данных будет востребована инфосистема, содержащая сведения не столько о самих генотипах, сколько об изученных популяциях и индивидах. Ведь, имея их исчерпывающий список и удобную систему фильтрации популяций и индивидов, можно будет быстро сформировать список интересующих индивидуальных образцов, а затем уже получать данные по их генотипам из оригинальных публикаций. Создание такой инфосистемы коллективом под руководством автора пока только начато, поэтому широкогеномные данные, использованные в этой работе, представляют собой отдельные массивы данных, описанные в соответствующих разделах по ходу их изложения.

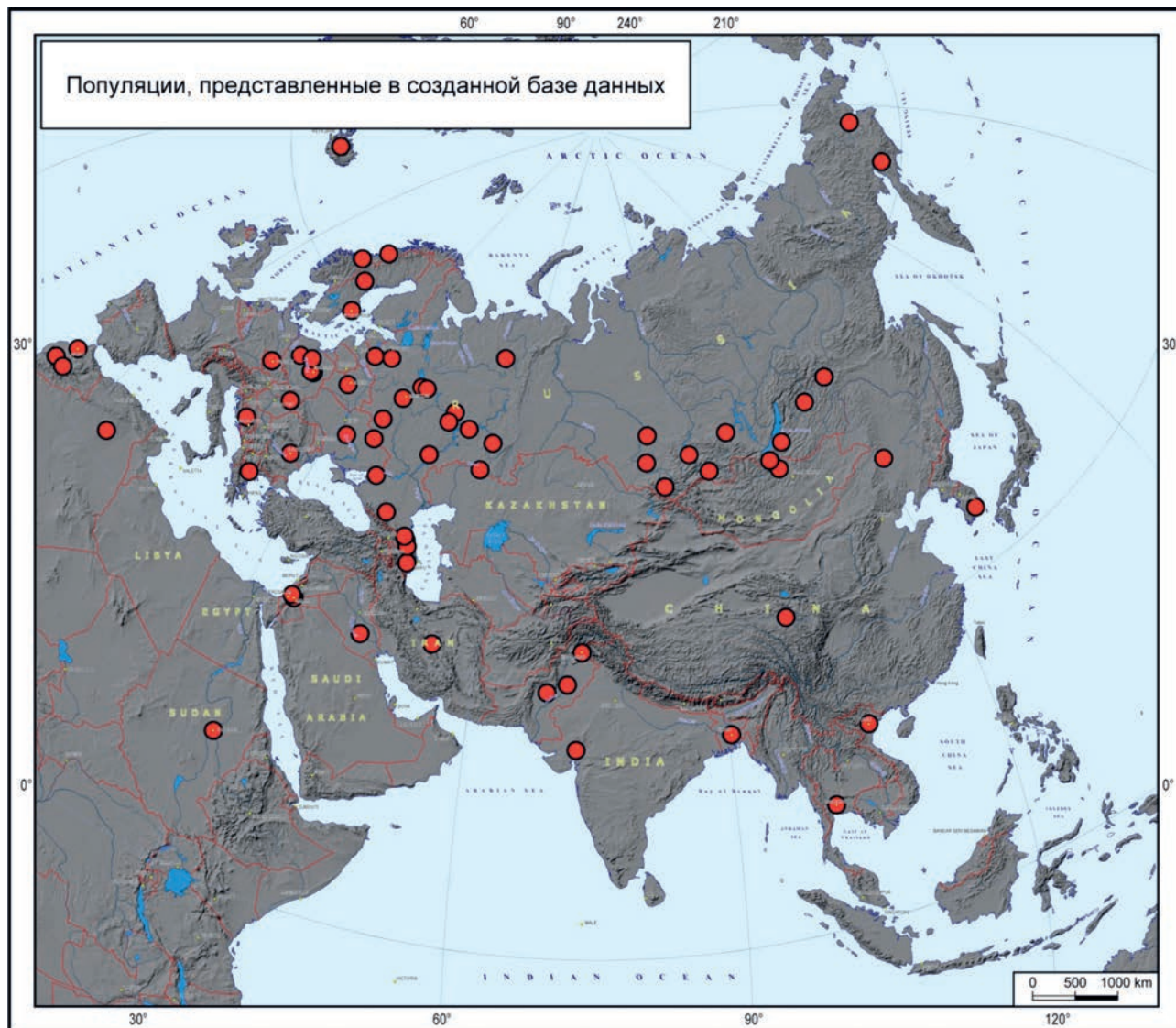


Рис. 1.8. Популяции в базе данных частот аутосомных STR маркеров ДНК-идентификации.

ПРИМЕНЕНИЕ СОЗДАНЫХ БАЗ ДАННЫХ

Разработанные базы данных изменчивости Y-хромосомы и митохондриальной ДНК не только послужили одним из столпов, на которых выстроено исследование, приведенные в последующих главах, но имеют и самостоятельное значение. То же относится и к созданным картографическим атласам, о которых речь пойдет дальше.

Во-первых, созданные базы данных и атласы охватывают коренное население всех континентов мира, а данная монография сосредоточена на населении Европы в контексте генофонда Евразии. Во-вторых, базы данных имеют значение не только для геногеографических исследований, но и для решения ряда других научных и практических задач. В данном разделе на нескольких конкретных примерах обрисовывается область применения созданных баз данных в научных исследованиях.

ИСПОЛЬЗОВАНИЕ БАЗЫ ДАННЫХ MURKA

Созданная база данных по изменчивости мтДНК в популяциях мира MURKA широко применялась как автором и его коллегами, так и известными зарубежными учеными, обращавшимися к нам за данными по мтДНК. База данных использовалась как минимум в пяти международных публикациях, четыре из которых вышли в ведущих журналах с наиболее высокими импакт-факторами (от 9 до 13). Также она цитировалась и в многочисленных русскоязычных публикациях нашего коллектива. Ниже представлен краткий обзор зарубежных публикаций, применявших и цитировавших нашу базу данных (отечественные публикации, как менее важные для оценки признания базы данных, здесь не рассматриваются).

БАЗА ДАННЫХ МИТОТИПОВ НЕКОРЕННОГО НАСЕЛЕНИЯ МИРА

В 2007 году в журнале PLoS Genetics опубликована статья [Behar et al., 2007], в которой представлена база данных митотипов «общественных участников» проекта «Генография» и программа-предиктор митохондриальных гаплогрупп. В этой работе, на основе наиболее обширной (78 590 об-

разцов) базы данных по некоренному населению США, Европы и других стран, анализируются общие закономерности изменчивости и мутирования мтДНК. Созданная нами база данных MURKA (по коренному населению) использована в этом исследовании для целей сравнения (с базой данных по «общественным участникам») и для верификации предложенного алгоритма автоматического определения гаплогрупп мтДНК.

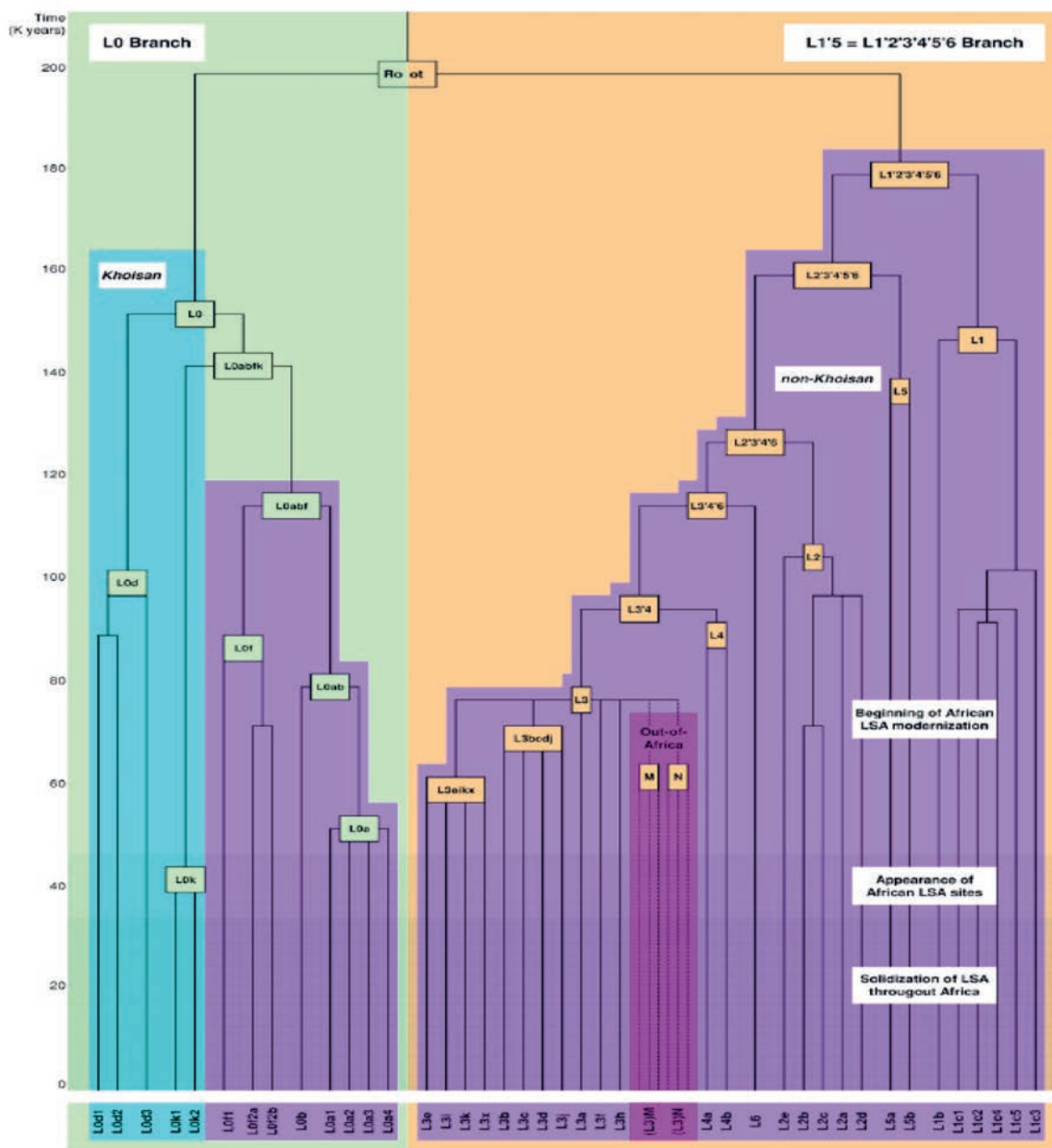


Рис. 1.9. Филогенетическое древо митохондриальной гаплогруппы L – разнообразие мтДНК генофонда коренного населения Африки [Behar et al., 2008].

Зеленая часть графика соответствует «койсанской» макроветви древа мтДНК, а рыжая – макроветви остального человечества. Заливка внутренних прямоугольников: голубая маркирует гаплогруппы, распространенные у койсанов, сиреневая – у не койсанских популяций Африки, розовая – у популяций вне Африки. Видно, что часть гаплогрупп «койсанской» ветви встречается и у других популяций Африки. Также наглядно видно, какая небольшая доля существующего в Африке разнообразия была вынесена из нее частью человечества, заселившей все остальные континенты.

АНАЛИЗ ПОЛНЫХ СИКВЕНСОВ ДРЕВНЕЙШЕЙ ГАПЛОГРУППЫ L мтДНК

В 2008 году в журнале *American Journal of Human Genetics* были опубликованы результаты анализа полных нуклеотидных последовательностей митохондриальной ДНК [Behar et al., 2008]. Эта трудоемкая работа была необходима для ответа на вопрос – каковы были самые первые этапы микроэволюции *Homo sapiens*. Главным результатом этой работы стало уточнение филогенетического древа человечества (рис. 1.9). Укажем две важнейшие его черты. Во-первых, разделение древа на два крупных ствола (койсаны – и все остальное человечество) произошло по данным о мтДНК более 140 000 лет назад. Вторая черта – уже известная из более ранних работ, но от того не менее удивительная. На древе человечества почти все разнообразие приходится на народы Африки. Лишь малая часть ветвей общего древа (показанных на розовом фоне) дала побеги на всех остальных континентах – Евразии, Америки, Австралии. Это древо хорошо иллюстрирует общий принцип отслеживания миграций – расселяющиеся популяции, оторвавшиеся от исходного массива, забирают с собой в путь лишь малую часть ветвей, малую часть имеющегося генетического разнообразия. Дальнейшая микроэволюция приводит к росту вторич-

ных субгаплогрупп в разных регионах планеты, позволяя проследить все более поздние миграции.

РАСХОЖДЕНИЕ ГЕНОФОНДОВ САМОГО НИЗКОРОСЛОГО И САМОГО ВЫСОКОРОСЛОГО НАСЕЛЕНИЯ ПЛАНЕТЫ – ПИГМЕЕВ И БАНТУ

В 2008 году в известном журнале *PNAS* опубликована статья [Quintana-Murci et al., 2008]. В этом исследовании митохондриальный генофонд пигмеев экваториальной Африки изучен в сопоставлении с окружающими африканскими генофондами. Эта информация по окружающим популяциям взята из нашей базы данных. Этот пример показывает, что база данных MURKA действительно является глобальной и подробной, если даже известные французские ученые, в течение многих лет изучающие африканские генофонды, не имеют более подробных источников о генофонде Африки, чем наша база данных. И это при том, что база создавалась московскими специалистами, никогда специально Африкой не интересовавшимися, но зато проводившими подробный скрининг информации, опубликованной по всем континентам (в том числе и по Африке).

В статье анализируется древнейшее и важное событие в истории населения Африки и челове-

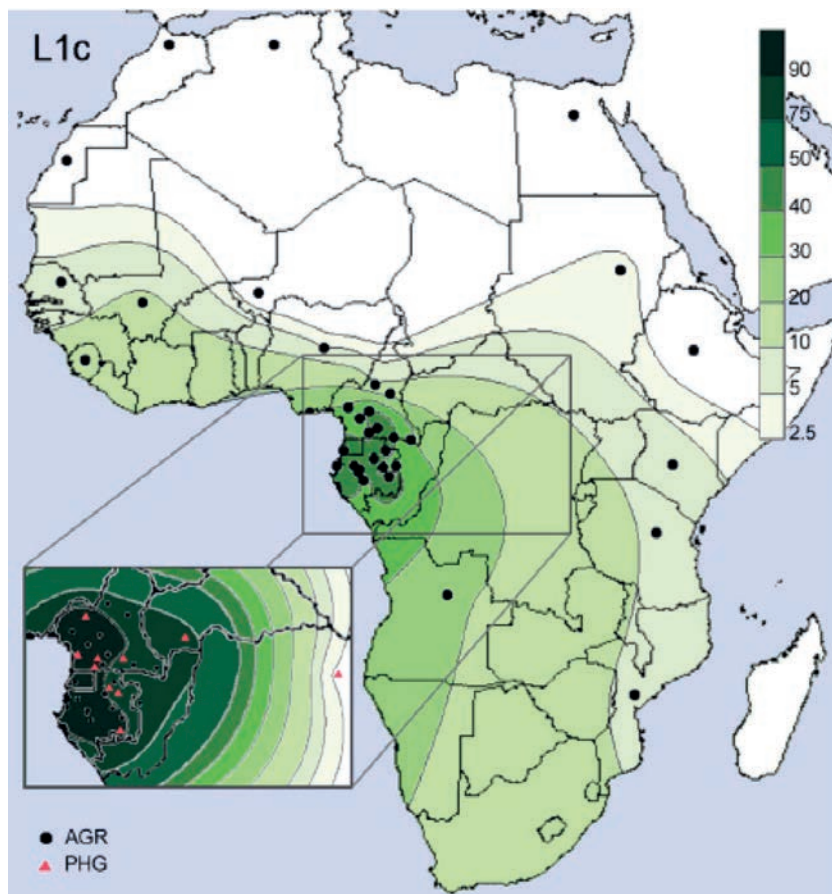


Рис. 1.10. Распространение гаплогруппы L1c [Quintana-Murci et al., 2008].

ства в целом – расхождение генофондов самого низкорослого и самого высокорослого населения планеты – пигмеев и бантуязычных народов. По данным о мтДНК около 70 000 лет назад существовала их единая общность [Quintana-Murci et al., 2008]. Их разделение, видимо, было связано с климатическими кризисами в истории нашей планеты. Ледниковые периоды в истории земли имели для Африки не менее катастрофичные последствия, чем для Европы. Это было время иссушения планеты – исчезали леса, их место в Африке занимали саванны и пустыни, разделившие предков пигмеев и банту. Только много тысяч лет спустя, когда обе популяции уже приобрели своеобразные антропологические и генетические черты, их ареалы вновь стали соприкасаться. Но, как показано, поток генов между ними был строго односторонним: от пигмеев к банту (рис. 1.10). Народы банту брали в жены маленьких женщин пигмеев, приносивших свои гаплогруппы мтДНК. Но у пигмеев не прослеживаются линии мтДНК бантуязычных народов.

НЕОЛИТИЗАЦИЯ ЕВРОПЫ

В 2009 году опубликована глава в зарубежной монографии, посвященная рассмотрению основных итогов изучения генофонда Европы [Balapovsky, 2009]. Она включает как обзор основных исследований за последние 20 лет, так и результаты собственных исследований с использованием баз данных по Y-хромосоме и мтДНК. Основной акцент в работе сделан на возможности применения генетических данных для изучения вопросов неолитизации Европы и на сравнении полученных нами результатов с археологическими и лингвистическими данными по этой крупной научной проблеме.

АНАЛИЗ МЕТИСИРОВАННЫХ ПОПУЛЯЦИЙ

В 2010 году в ведущем генетическом журнале *American Journal of Human Genetics* вышла статья международного авторского коллектива [Quintana-Murci et al., 2010]. В этом исследовании проведен генетический анализ маркеров мтДНК и Y-хромосомы в популяции цветного населения Южной Африки – крайне гетерогенной популяции, сформировавшейся в результате смешения пришлых европейских и индийских колонистов с аборигенным банту- и койсаноязычным населением Африки. Благодаря изучению маркеров, наследующихся по материнской и отцовской линиям, удалось реконструировать ход сложения этой популяции и выявить асимметричные генные потоки со стороны мужчин и женщин четырех популяций – источников миграций.

Весь статистический анализ в этом исследовании выполнен благодаря привлечению данных о

популяциях Европы, Индии, Индонезии, койсанскому населению Африки, бантуязычному населению Африки по маркерам мтДНК и Y-хромосомы. При этом вся генетическая информация по населению этих регионов мира взята из разработанных нами баз данных Y-base и MURKA.

Исторические сведения и последние данные по анализу аутосомных ДНК маркеров показывают, что южно-африканское цветное население является уникальной, в высокой степени метисированной популяцией, возникшей в результате смешения различных народов Африки, Европы и Азии. Однако оставались неизвестными пути формирования этой популяции. В данной работе с участием автора проведен детальный филогеографический анализ митохондриальной ДНК и Y-хромосомы. Показано, что эта популяция образовалась путем взаимодействия, по крайней мере, пяти различных родительских популяций: койсанов, банту, Европы, Индии и Юго-Восточной Азии (рис. 1.11). Однако они внесли различный вклад в ее образование. Более того, обнаружены значительные различия между генетическими вкладами предковых популяций в мужскую и в женскую части популяции. Общая картина гендерно-специфичной метисации показывает, что данная популяция возникла в основном в результате смешения мужчин из европейских, африканских и азиатских популяций с коренными койсанскими женщинами.

АНАЛИЗ ДРЕВНЕЙ ДНК

В 2010 году в одном из наиболее престижных научных журналов *PLOS Biology* опубликована статья [Naak et al., 2010]. Эта статья является одним из итогов многолетней совместной работы автора с Австралийским центром древней ДНК. Полученные этим центром данные о митохондриальном генофонде одной из первых неолитических культур Европы были сопоставлены с нашей обширной базой данных о мтДНК современных популяций Евразии. Оказалось, что генофонд неолитической популяции Европы, несмотря на свое выраженное своеобразие, кластеризуется с популяциями Ближнего Востока, а не с популяциями Европы. Это позволило выдвинуть компромиссную теорию между двумя конкурирующими гипотезами – гипотезой «демической» диффузии неолитических земледельцев и гипотезой чисто «культурной» диффузии навыков земледелия (без замены генофонда). Полемика между этими гипотезами является, пожалуй, наиболее широко обсуждаемой проблемой в популяционной генетике человека на протяжении последних двадцати лет. Предложенный компромисс заключается в том, что, хотя (как показывают наши данные) миграция земледельцев с Ближнего Востока имела место (постулат гипотезы «демической» диффузии), но смена генофон-



Рис. 1.11. Графики многомерного шкалирования: родство между южно-африканским цветным населением и предполагаемыми родительскими популяциями [Quintana-Murci et al., 2010].

(А) график, основанный на частотах гаплогрупп мтДНК; (В) график, основанный на частотах гаплогрупп Y-хромосомы.

да Европы не произошла, и далее земледелие распространялось путем культурных заимствований (постулат гипотезы «культурной» диффузии), поскольку эта мигрировавшая из Передней Азии группа была немногочисленной и растворилась в генофонде Европы, передав им навыки земледелия. Важно отметить, что весь статистический анализ в этом крупном исследовании основывался на информации из нашей базы данных. Подробнее эти результаты рассматриваются в главе 8.

АССОЦИАЦИЯ СЕРДЕЧНО-СОСУДИСТОЙ ПАТОЛОГИИ С ГАПЛОГРУППАМИ мтДНК

В 2011 году в журнале *Annals of Human Genetic* опубликована статья с нашим участием [Haber et al., 2011]. Она служит примером использования разработанных баз данных в медико-генетичес-

ких исследованиях, поскольку в ней демонстрируется ассоциация сердечно-сосудистой патологии с гаплогруппами **A** и **W** мтДНК в ливанской популяции. В литературе имелись предположения о возможном механизме такой ассоциации. Однако для повышения статистической достоверности ассоциаций, а также для выявления географического распространения линий мтДНК, демонстрирующих такую ассоциацию, было необходимо привлечь данные не только по Ливану, но и по всем окружающим регионам. Использование базы данных MURKA позволило изучить положение этих линий мтДНК в общем контексте изменчивости соответствующих гаплогрупп и определить связь линий, обнаруживающих ассоциацию, с историческими миграционными потоками в Ливан.

1.3. СОБСТВЕННЫЕ ДАННЫЕ

В соответствии с основной задачей монографии – охарактеризовать генофонд народонаселения Европы и его взаимодействие с окружающими гено-

фондами на текущем срезе наших знаний – исходным материалом послужили созданные под руководством автора базы данных, объединившие соб-

ственные данные с опубликованными данными других исследователей. В то же время собственные данные в ряде случаев составляли основную часть всего анализируемого массива или даже, чтобы обеспечить методологическое единообразие массива анализируемых популяций, анализировались отдельно. Поэтому собственные данные – точнее, данные, полученные нашим коллективом на протяжении последних 15 лет – требуется охарактеризовать в сколько-нибудь подробном и систематическом

виде. Эта характеристика и представлена в данном разделе. В нем сначала рассматривается стратегия формирования выборок, используемая нашим коллективом во всех экспедициях, затем перечисляются все охваченные к настоящему времени популяции, в том числе рассматриваются активно вовлеченные в данное исследование популяции Европы, Кавказа и Азии, и в конце дается характеристика процедур генотипирования мтДНК и Y-хромосомы, включая ее полное секвенирование.

СТРАТЕГИЯ СБОРА ВЫБОРОК

Биологический материал собран в ходе многолетних экспедиций (1998–2014 гг.), проведенных под руководством проф. Е.В. Балановской и автора данной работы. Экспедиции проводились коллективами лаборатории популяционной генетики человека МГНЦ РАМН и лаборатории геномной географии ИОГен РАН как самостоятельно, так и в сотрудничестве с Адыгейским, Белгородским, Кемеровским, Харьковским государственными университетами, научными учреждениями Армении, Белоруссии, Казахстана, Киргизии, Монголии и Таджикистана в рамках грантов РФФИ, РГНФ и международного проекта «Genographic». Автор лично участвовал в проведении экспедиций по обследованию шапсугов (1998 г.), русских Архангельской области (1999 г., 2001 г.), Костромской обл. (2001 г.), Смоленской обл. (2002 г.), Тверской обл. (2003 г.), Псковской обл. (2004 г.), народов Дагестана (2005 г.), Осетии (2006 г.), Таджикистана (2006 г.), Афганистана (2006 г.), Татарстана (2006 г.), Калмыкии (2007 г.), донских казаков (2008 г.), Казахстана (2009 г.). В проведении всех остальных экспедиций автор участвовал на этапах их планирования и организации (собирались образцы венозной крови, изредка – слюна).

ПРАВИЛА СБОРА ОБРАЗЦОВ

Во всех экспедициях при обследовании генофондов коренного населения использовалась единая стратегия формирования выборок, заключающаяся в строгом следовании следующим общим правилам.

Правило 1: «Три поколения» (генетически эффективные миграции). Анализировались только индивиды (старше 18 лет), у которых родители и все бабушки и дедушки (т.е. на протяжении трех поколений) родились в данной популяции и относились к данной этнической группе. Если более далекие предки индивида пришли в данный регион из других областей, но оставили в этой популяции потомство на протяжении более трех поколений, то такая миграция считалась генетически эффективной и такие индивиды включались в обследование.

Правило 2: «Не родственники» (нескоррелированность генотипов). Родственники (до третьей степени родства) из выборки тщательно исключались, поскольку наличие родственных индивидов не только уменьшает эффективный размер выборки, но и смещает частоты, создавая иллюзию миграции, если в выборку попала группа родственников – потомков единичного мигранта. Для отсева родственников, во-первых, использовались информаторы (люди, хорошо знающие родословные своих односельчан). Во-вторых, обследуемому предъявлялся список всех, уже прошедших обследование, и предлагалось найти в списке своих родственников. Если родственники находились, и степень родства была ближе третьей, то новый претендент не обследовался. В-третьих, для каждого обследуемого создавалась родословная, также помогающая выявлять родственные связи.

Правило 3: «Информированное согласие». Все обследованные давали письменное информированное согласие на предоставление образца крови для популяционно-генетического анализа. Тексты информированного согласия одобрены Этической комиссией МГНЦ РАМН, под контролем которой проводились все экспедиционные обследования. Большинство обследованных являются людьми среднего или старшего возраста. Если в порядке исключения (обычно по настоятельной просьбе родителей) брались образцы у детей, то в этом случае требовалось письменное согласие одного из родителей.

Правило 4: «Только мужчины» (максимум генетической информации). В большинстве выборок мужчины составляют более 95%, что обеспечивает возможность анализа и Y-хромосомы, и мтДНК, и аутосомных маркеров. Также такие выборки облегчают анализ X-хромосомы, которая у мужчин находится в гемизиготном состоянии, снимая этим проблему выявления гаплотипов, остро стоящую при анализе гомологичных хромосом.

Правило 5: «Образцы крови» (максимум генетической информации). Для обеспечения большого количества ДНК и, следовательно, максимума генетической информации о каждом обследованном

дуюмом, в качестве биологического образца практически всегда брались образцы венозной крови (от 5 мл до 12 мл с помощью системы вакуумных пробирок Vacuette). Только в редких случаях собирались образцы слюны, и в редчайших – образцы буккального эпителия (например, было принято решение не забирать венозную кровь при формировании выборки в Афганистане).

Правило 6: «Субпопуляции» (охват генетической структуры популяции). Большинство

популяций исследовалось не в одной, а в нескольких географических локальностях, охватывающих максимальное разнообразие популяции или этноса. Например, изучались два района в пределах одной области для русских популяций, или два-три горных ущелья для кавказских. В этом случае собранные выборки более полно характеризуют данную популяцию в целом, а не ее часть, генофонд которой может оказаться своеобразным в результате дрейфа генов или миграций.

БАНК ОБРАЗЦОВ «КОРЕННОЕ НАСЕЛЕНИЕ СЕВЕРНОЙ ЕВРАЗИИ»

В результате интенсивных экспедиций, начатых в 1998 году и продолжающихся – по несколько экспедиций в год – вплоть до настоящего времени, нашим коллективом создан наиболее обширный и наиболее качественно сформированный биобанк образцов от коренного населения Северной Евразии. Биобанк находится в собственности проф. Е.В. Балановской и автора данной работы, определенные права на образцы имеют также международный проект «The Genographic project» и руководители тех экспедиций, которые проводились хотя и по нашей инициативе, нашей методологической программе и зачастую при нашем финансировании, но силами региональных российских и международных коллективов, с которыми ведется сотрудничество. Биобанк широко используется в научных исследованиях в России и хорошо известен российским и зарубежным специалистам как “Elena Balanovska’s and Oleg Balanovsky’s samples from North Eurasian indigenous populations”. Объем биобанка кратко можно охарактеризовать четырьмя параметрами: образцы от 24 тысяч человек, охвачено 256 популяций, представляющих 90 народов и население 16 стран. Качество биобанка определяется строгим следованием стратегии сбора образцов, описанной в предыдущем пункте. Более подробная информация о биобанке представлена на рис. 1.12 и в табл. 1.4, а также в списке охваченных этнических групп. Отметим, что благодаря частым экспедициям коллек-

ция ежегодно увеличивается на 2–3 тысячи образцов, поэтому, например, в 2017 году она уже может превысить 30 тысяч образцов.

ОХВАЧЕННЫЕ ЭТНИЧЕСКИЕ И СУБЭТНИЧЕСКИЕ ГРУППЫ

Перечислим их в алфавитном порядке: абазинцы, абхазы, аварцы, адыгейцы, азербайджанцы, алтайцы, андийцы, армяне, балкарцы, башкиры, белорусы, бесермяне, болгары, буряты, вепсы, гагаузы, греки крымские (азовские), грузины-менгрелы, грузины-лазы, грузины-имеретинцы, даргинцы, дунгане, езиды, ингуши, ишкашимцы, кабардинцы, казахи, кайтагцы, калмыки, караимы, каракалпаки, караногайцы, карачаевцы, карелы, киргизы, кубачинцы, кумыки, курды, лакцы, лезгинцы, литовцы, молдаване, монголы-баяд, монголы-дурвэд, монголы-захчин, монголы-уулд, монголы-халха, мордва-мокша, мордва-эрзя, мордва-шокша, нанайцы, негидальцы, ненцы, нивхи, ногайцы астраханские, ногайцы ставропольские, орочи, осетины, русские, русские-казаки, рушанцы, табасаранцы, таджики, татары астраханские, татары казанские, татары крымские, татары кряшены, татары мишари, татары сибирские, тувинцы, туркмены, удмурты, удэгейцы, узбеки, украинцы, украинцы-казаки, ульчи, хакасы, хамнегане, черкесы, чеченцы, чувашаи, шорцы, шугнанцы, эвенки, ягнобцы, якуты.

Таблица 1.4. Охват географических регионов биобанком коренного населения Северной Евразии

Восточная Европа	69 популяций	6 269 образцов
Кавказ	55 популяций	6 113 образцов
Волго-Уральский регион	42 популяций	2 866 образцов
Центральная Азия	51 популяций	5 433 образцов
Сибирь и Дальний Восток	39 популяций	3 397 образцов
ИТОГО	256 популяций	24 078 образцов

ИЗУЧЕННЫЕ ПОПУЛЯЦИИ ВОСТОЧНОЙ ЕВРОПЫ, КАВКАЗА, ЦЕНТРАЛЬНОЙ АЗИИ И СИБИРИ

В данной монографии использованы многие, но не все выборки, представленные в нашем биобанке. В соответствии с планом исследования соб-

ственные данные по Y-хромосоме охватывали три группы популяций: славянские (и близкие к ним балтские) народы (N=1353) (рис. 1.13), популяции

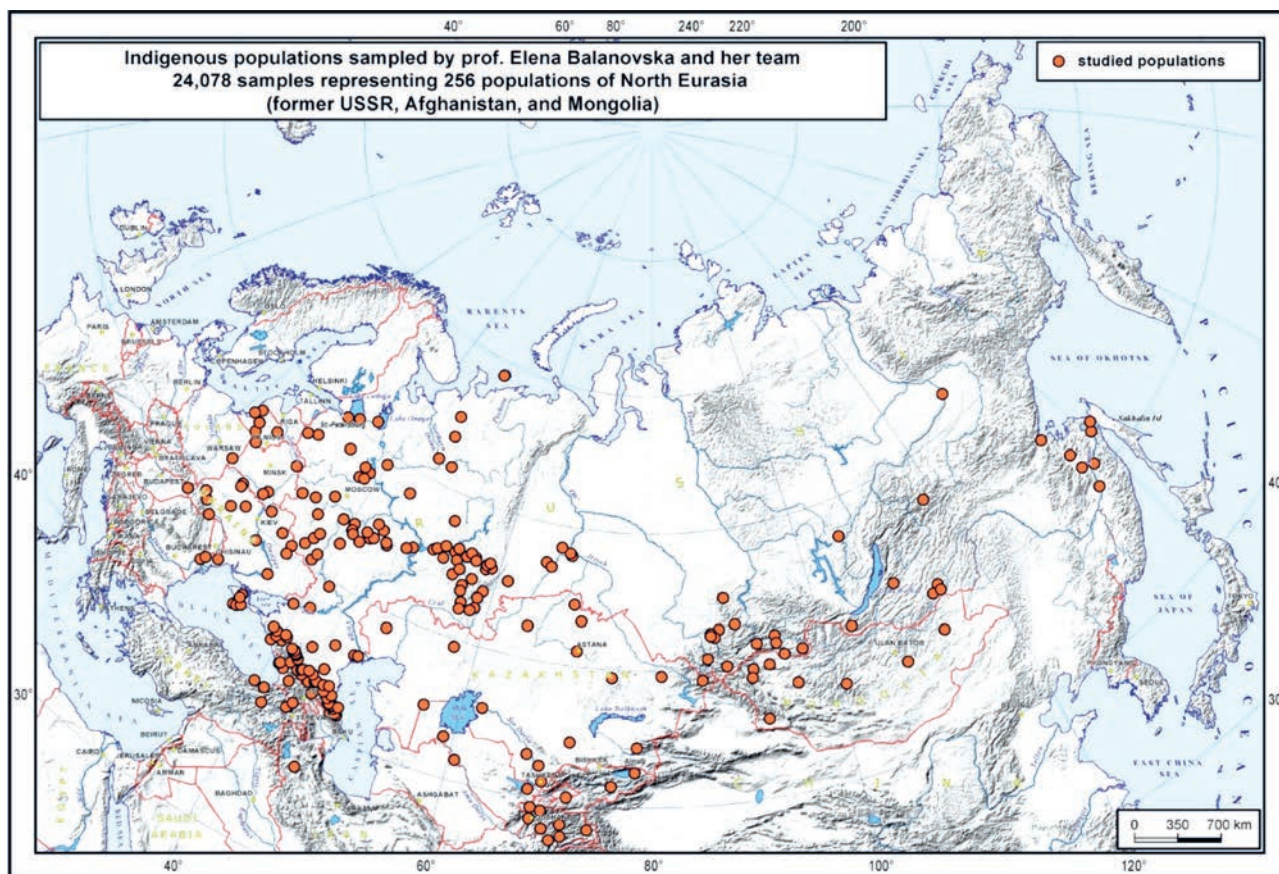


Рис. 1.12. Карта популяций биобанка коренного населения Северной Евразии.

Показано географическое положение популяций, обследованных в экспедициях нашего коллектива.

Кавказа ($N=1695$), популяции Сибири и Центральной Азии ($N=1241$); суммарная выборка превысила четыре тысячи образцов (табл. 1.1). По митохондриальной ДНК были охарактеризованы русские, украинские, белорусские и северокавказские популяции, суммарно составляющие более трех тысяч образцов (табл. 1.2).

ПОПУЛЯЦИИ ВОСТОЧНОЙ ЕВРОПЫ

При изучении народонаселения Европы в целом (главы 2–5) наши собственные данные входили в анализируемый массив данных лишь как его составная часть. Однако анализ генофонда балто-славянских народов (глава 6) проведен, уже опираясь преимущественно на собственные данные, и, в частности, характеристика восточнославянских генофондов практически целиком основана на собственных данных нашего научного коллектива. Это изучение восточных славян проведено по большим выборкам из основных региональных групп русских, украинцев, белорусов. Нами исследованы пятнадцать русских популяций по маркерам Y-хромосомы (табл. 1.5) и тринадцать русских, пять украинских* и пять белорусских популяций по мтДНК (табл. 1.6).

Принципиальной особенностью анализируемого материала является не столько его объем, сколько качество сформированных выборок и их география. Под «качеством» подразумевается следование описанной выше стратегии формирования выборок, обеспечивающей их высокую репрезентативность. Под «географией» понимается изучение разнообразия генофонда на всей обширной территории расселения восточных славян, для чего требуется анализ многочисленных популяций. Без детальной характеристики ДНК полиморфизма в различных региональных группах русского, украинского и белорусского народов корректное изучение их генофонда было бы невозможно.

ВОСТОЧНЫЕ СЛАВЯНЕ В ИХ ИСТОРИЧЕСКОМ АРЕАЛЕ

При этом наше исследование намеренно проведено только в пределах исторического ареала восточнославянских народов и не включило восточнославянское население Урала, Сибири и Средней Азии, сформировавшееся относительно поздно. Такая логика объясняется необходимостью изучить сначала генофонд русского народа и остальных двух

* При этом две популяции из Западной Украины рассматриваются суммарно, т.к. увеличение объема объединенной выборки обеспечивает более высокую статистическую достоверность результатов.

Таблица 1.5. Популяции, изученные по маркерам Y-хромосомы.

ПОПУЛЯЦИЯ	НАРОД	ГЕОГРАФИЧЕСКАЯ ЛОКАЛИЗАЦИЯ	N	Долгота	Широта
Пинега	русские	Архангельская обл., Пинежский р-н	114	46.53	63.43
Мезень	русские	Архангельская обл., Лешуконский р-н	54	45.74	64.9
Красноборск	русские	Архангельская обл., Красноборский и Ленский р-ны	91	45.94	61.56
Вологда	русские	Вологодская обл., разные р-ны	121	39.9	59.23
Кашин	русские	Тверская обл., Кашинский р-н	73	37.61	57.36
Порхов	русские	Псковская обл., Порховский р-н	57	29.56	57.77
Остров	русские	Псковская обл., Островский р-н	75	28.32	57.35
Смоленск	русские	Смоленская обл., Рославльский р-н	107	32.88	53.93
Кострома (Унжа)	русские	Костромская обл., Мантуровский и Межевской районы	52	44.77	58.33
Русские Белгород	русские	Белгородская обл., Яковлевский, Красненский, Прохоровский р-ны	143	36.48	50.78
Ливны	русские	Орловская обл., Ливненский р-н	110	37.59	52.4
Пристен	русские	Курская обл., Пристенский р-н	45	36.71	51.23
Репьевка	русские	Воронежская обл., Репьевский р-н	96	38.65	51.08
Кубанские казаки	русские	Адыгея, Майкопский р-н	90	40.17	44.51
Терские казаки	русские	Кабардино-Балкария, Майский, Прохладненский р-ны	125	44.0	43.7
Шапсуги	шапсуги	Краснодарский край, Туапсинский, Лазаревский р-ны	100	39.1	44.2
Абхазы	абхазы	Абхазия, разные р-ны	58	41.10	43.1
Черкесы	черкесы	Карачаево-Черкесия, Хабезский, Прикубанский р-ны	142	41.7	43.8
Осетины	осетины	Северная Осетия, Дигорский, Ирафский, Алагирский, Правобережный, Пригородный р-ны; Южная Осетия.	359	44.0	42.8
Ингуши	ингуши	Ингушетия, Малгобекский р-н	143	45.0	43.1
Чеченцы	чеченцы	Ингушетия, Малгобекский р-н; Чечня, Ачхой-Мартановский р-н; Дагестан, Казбековский, Хасавюртовский, Новолакский р-ны	330	46.0	43.2
Аварцы	аварцы	Дагестан, Унцукульский, Гунибский р-ны	115	46.9	42.5
Даргинцы	даргинцы	Дагестан, Акушинский, Дахадаевский р-ны	101	47.2	42.2
Кубачинцы	кубачинцы	Дагестан, Дахадаевский р-н	65	47.6	42.1
Кайтагцы	кайтагцы	Дагестан, Кайтагский р-н	33	47.63	42.2
Лезгины	лезгины	Дагестан, Ахтынский р-н	81	47.70	41.5
Армяне	армяне	Краснодарский край и Адыгея, разные р-ны	115	40.1	44.1
Армяне-амшены	армяне	Краснодарский край (мигранты из Турции)	53	-	-
Кумандинцы, челканцы, тубалары	северные алтайцы	Алтайский край, Солтонский р-н Алтай, Турочакский и Чойский р-ны	202	87.42	52.2
Алтай-кижи, теленгиты	южные алтайцы	Алтай, Онгудайский, Улаганский, Кош-Агачский, Чойский, Турочакский р-ны	207	88.50	50.03
Шорцы (горные и абаканские)	шорцы	Кемеровская обл., Таштагольский р-н Хакасия, Таштыпский р-н	139	88.00	52.75
Качинцы, сагайцы, койбалы, кзылыльцы	хакасы	Хакасия, Аскизский, Бейский, Орджоникидзевский, Таштыпский, Ширинский р-ны	185	90.00	53.00
Казахи Алтая	казахи	Казахстан, Восточно-Казахстанская обл., Катон-Карагайский р-н	122	85.60	49.16
Киргизы северные	киргизы	Киргизия, Нарынский, Таласский р-ны	140	76.00	41.42
Киргизы Памира	киргизы	Гаджикистан, Горно-Бадахшанская обл., Мургабский р-н.	106	73.95	38.15
Монголы Алтая	монголы	Монголия, аймаки Байан-Улгий, Ховд, Гоби-Алтай	140	96.00	48.00
Суммарно	20 этносов	6 стран, 39 областей, 81 район	4289	-	-

восточнославянских народов на исконной территории их формирования. Это позволит при необходимости провести и корректное изучение славянских популяций, переселившихся в XVI–XX веках за

Урал. Генофонд этих «зауральских» популяций отражает не столько историю формирования восточнославянских народов, сколько историю их позднейшей экспансии, как было показано в наших работах

Таблица 1.6. Популяции, изученные по маркерам митохондриальной ДНК.

ПОПУЛЯЦИЯ	НАРОД	ГЕОГРАФИЧЕСКАЯ ЛОКАЛИЗАЦИЯ	N	Долгота	Широта
Пинега	русские	Архангельская обл., Пинежский р-н	144	46.53	63.43
Смоленск	русские	Смоленская обл., Рославльский и Ершицкий р-ны	147	32.88	53.93
Кострома (Унжа)	русские	Костромская обл., Мантуровский и Межевской р-ны (р. Унжа)	79	44.77	58.33
Белгород	русские	Белгородская обл., Яковлевский, Красненский, Прохоровский р-ны	148	36.48	50.78
Боровский	русские	Калужская, Боровский р-н	70	36.50	55.00
Брятинский	русские	Калужская обл., Брятинский р-н	75	34.53	54.00
Болховский	русские	Орловская обл., Болховский р-н	76	36.00	53.43
Черемисиновский	русские	Курская обл., Черемисиновский р-н	62	37.25	51.87
Михайловский	русские	Рязанская обл., Михайловский р-н	82	39.00	54.22
Спаск-Рязанский	русские	Рязанская обл., Спаск-Рязанский р-н	86	40.53	54.40
Петровский	русские	Тамбовская обл., Петровский р-н	76	40.25	52.63
Кубанские казаки	русские	Адыгея, Майкопский р-н	132	40.17	44.51
Терские казаки	русские	Кабардино-Балкария, Майский, Прохладненский р-ны	124	44.00	43.70
Столин	белорусы	Брестская обл., Столинский р-н	89	26.77	51.97
Светлогорск	белорусы	Гомельская обл., Светлогорский р-н	71	29.72	52.62
Витебск	белорусы	Витебская обл.	100	29.00	55.00
Брест	белорусы	Брестская обл.	104	24.00	52.00
Гомель	белорусы	Гомельская обл.	121	30.00	52.00
Украинцы Белгород	украинцы	Белгородская обл., Красногвардейский и Грайворонский р-ны	95	35.75	50.40
Черкассы	украинцы	Черкасская обл.	179	32.07	49.43
Хмельницкая	украинцы	Хмельницкая обл., Старо-Константиновский р-н	179	27.30	49.70
Украинцы западные	украинцы	Львовская и Ивано-Франковская обл., разные р-ны	157	24.00	49.25
Лезгины	лезгины	Дагестан, Ахтынский р-н	69	47.70	41.50
Кайтагцы	кайтагцы	Дагестан, Кайтагский р-н	32	47.63	42.20
Кубачинцы	кубачинцы	Дагестан, Дахадаевский р-н	55	47.60	42.10
Аварцы	аварцы	Дагестан, Унцукульский, Гунибский	81	46.90	42.50
Даргинцы	даргинцы	Дагестан, Акушинский, Дахадаевский	95	47.20	42.20
Чеченцы	чеченцы	Чечня, Ачхой-Мартановский р-н; Дагестан, Казбековский, Хасавюртовский, Новолакский р-ны; Ингушетия, Малгобекский р-н	239	46.00	43.20
Осетины	осетины	Северная Осетия, Дигорский, Ирафский, Алагирский, Правобережный, Пригородный р-ны; Южная Осетия.	267	44.00	42.80
Суммарно	10 этносов	23 области, 40 районов	3234	-	-

по анализу структуры фонда фамилий [Балановская, Балановский, 2007]. Та же цель – выявить не сиюминутный портрет генофонда, а его исторически устойчивые черты – определила и подход к формированию выборок, то есть описанную стратегию, направленную на выявление «коренного», наименее метисированного генофонда.

ОХВАТ ВСЕГО РАЗНООБРАЗИЯ ВОСТОЧНЫХ СЛАВЯН

При планировании пунктов обследования учитывались данные геногеографии (предшествовавших работ), лингвистики, этнографии, антропологии.

Например, несмотря на сравнительно небольшую территорию, население Западной Украины обязательно должно было быть представлено в общем ряду выборок. В противном случае (если все украинские выборки происходили бы из иных регионов Украины), терялась бы существенная часть разнообразия генофонда, поскольку антропологические и лингвистические данные указывают, что основное разнообразие украинского населения сосредоточено на западе его ареала.

Аналогично, этнографическое подразделение Белоруссии на северную и южную нашло отражение в изучении обеих частей в нашем исследовании.

Для русских представлены обширные выборки казаков и северных русских, а также целый ряд популяций Центральной России. Учитывая, что одним из основных факторов формирования русского генофонда явилась ассимиляция славянскими племенами автохтонного финно-угорского населения, изучены и сопоставлены две группы русских популяций. Первая – популяции, для которых (по антропологическим и историческим данным) можно ожидать наибольшую долю финно-угорского субстрата (население части районов Архангельской, Вологодской, Костромской областей), другая – предположительно потомки исходно славянского населения (Псковская, Курская, Смоленская области).

СЛОИ КАВКАЗСКОГО ГЕНОФОНДА

Народонаселение Кавказа представляет собой напластование множества миграций, часть которых мы обнаруживаем в современном языковом разнообразии народов Кавказа:

- 1) древние миграции, сформировавшие пласт народов северокавказской языковой семьи;
- 2) миграции ираноязычных народов Евразийских степей ныне представлены только ираноязычными осетинами, но предположительно ранее включали карачаевцев и балкарцев;
- 3) ранние миграции тюркоязычных народов, приведшие к смене языка у балкарцев и карачаевцев, а также, возможно, у кумыков и азербайджанцев;
- 4) поздние миграции степных тюркоязычных народов, отраженные в популяциях ногайцев и караногайцев;
- 5) несколько волн миграций славянского населения, отраженных последовательно в терских, затем кубанских казаках, далее – в многочисленном современном русском населении Кавказа.

Именно поэтому из имеющихся в нашем распоряжении обширных ДНК коллекций, включающих все перечисленные народы Кавказа, мы выбрали для генотипирования те народы, которые позволяют наиболее полно выявить процессы формирования народонаселения Кавказа: многочисленные народы северокавказской языковой семьи и ираноязычные осетины, а также народы, представляющие Закавказье, тюркские и славянские популяции.

ИЗУЧЕННЫЕ ПОПУЛЯЦИИ КАВКАЗА

В пределах Северного Кавказа выделяют три географических региона, занимающих территорию от Черного моря до Каспийского: Западный, Центральный и Восточный Кавказ. Нами изучены народы, представляющие все три географических региона.

Среди современного населения Северного Кавказа представлены шесть лингвистических ветвей:

абхазо-адыгская, нахская и дагестанская ветви северокавказской лингвистической семьи, иранская и славянская ветви индоевропейской семьи и тюркская ветвь алтайской семьи. Все эти шесть ветвей представлены в нашей работе: абхазо-адыгская ветвь (шапсуги, абхазы, черкесы), иранская ветвь (осетины, 6 субэтнических групп), нахская ветвь (ингуши, чеченцы из Ингушетии, Чечни и Дагестана), дагестанская ветвь (аварцы, даргинцы, кубачинцы, кайтагцы, лезгины), армянская (суммарная армянская и армяне-амшены), славянская (кубанские и терские казаки) и тюркская (азербайджанцы, балкарцы, карачаевцы, караногайцы, кумыки, ногайцы).

Впрочем, народы славянской и тюркской групп стоят на особом счету и рассматриваются отдельно. Во-первых, потому, что они представляют собой довольно поздние волны миграций на Кавказ. Во-вторых, тюрки Кавказа по данным антропологии представляют «Евразию в миниатюре» – от европеоидных балкарцев и карачаевцев до монголоидных караногайцев – и потому требуют специальных подходов и отдельного исследования.

Каждая лингвистическая группа представлена репрезентативной выборкой, что позволяет быть уверенными в надежности их генетических портретов: абхазо-адыгская N=300 образцов, иранская N=359, нахская N=473, дагестанская N=395 образцов; каждая группа представлена 3–6 этническими или субэтническими популяциями.

Для народов, представляющих генофонды, окружающие основные группы народонаселения Кавказа (Закавказье, тюркские народы и славянские народы), выборки также достаточно представительны (тюркская N=870, армянская N=168, славянская N=222). Тюркская группа представлена всеми шестью тюркоязычными народами Кавказа. Исследованное нами армянское население Северного Кавказа формировалось в XVIII–XIX вв. за счет выходцев из Ирана и Турции, а в конце XX в. пришла новая волна армянских переселенцев из Абхазии, Азербайджана, Грузии, Армении, Нагорного Карабаха. Поэтому первая изученная выборка включила суммарную группу армян Адыгеи и Краснодарского края. Вторая выборка представляет армян-амшенов, говорящих на особом диалекте. Они являются потомками длительно изолированной группы, проживавшей на юго-восточном побережье Черного моря (на территории Турции), часть которых недавно переселилась на территорию России. Лингвистически армяне относятся к армянской группе индоевропейской семьи. На геногеографических картах обе популяции армян размещены на территории Армении.

В общей сложности для характеристики народов Кавказа по маркерам Y-хромосомы изучено 2780 образцов. Они включают 1527 образцов коренного населения Кавказа, а также 780 образцов

тюркоязычных народов, и 383 образцов армян и казаков Северного Кавказа (рис. 1.13).

НЕСКОЛЬКО ПОПУЛЯЦИЙ НА НАРОД

В отличие от большинства предыдущих исследований генофонда Кавказа, в данной работе крупные народы представлены не одной, а несколькими популяциями. Так, чеченцы изучены в трех республиках (Ингушетия, Чечня, Дагестан), соответствующих трем группам в пределах исконного ареала чеченцев. Субэтническое подразделение осетин проведено в соответствии с исторически сложившимися «обществами» (традиционными группами) – дигорцы-христиане (Дигорский район), дигорцы-мусульмане (Ирафский район), алагирцы, куртатинцы, тагаурцы, туалыцы, кударцы, джавцы, ксанцы. Армяне и казаки также представлены не одной, а двумя географически и исторически различными популяциями (армяне-амшены и суммарная группа армян Краснодарского края и Адыгеи; казаки кубанские и терские).

Для выявления роли дрейфа генов в населении Кавказа были изучены не только крупные народы, представляющие собой сложную подразделенную популяцию, но и три малых изолированных этноса. Один из них представляет западные популяции Кавказа – это причерноморские шапсуги, до Кавказской войны бывшие ядром черкесской общности («адыге»). Другие два изолированных этноса – кубачинцы и кайтагцы – представляют Восточный Кавказ.

Важнейшей особенностью сбора образцов было тщательное формирование выборок. Требовалось

избежать последствий современной метисации со славянскими народами, так как она может создать иллюзию давних контактов с населением Европы. Необходимо было также избежать включения недавней метисации между коренными народами Кавказа, так как это также исказит веками сложившуюся структуру народонаселения Кавказа. Именно поэтому в работе мы вынуждены ориентироваться лишь на наши собственные коллекции ДНК, которые тщательно собирались в течение многих лет по единой технологии и потому позволяют с одинаковой степенью надежности анализировать народы по всей протяженности Кавказа.

ИЗУЧЕННЫЕ ПОПУЛЯЦИИ АЗИИ

Для выявления степени и характера влияния азиатских миграций на формирование генофонда Европы и решения вопросов о характере исторического взаимодействия генофондов Западной и Восточной Евразии был изучен ряд народов Центральной Азии и Южной Сибири. Их ареалы охватили пять государств – Россию, Казахстан, Киргизию, Монголию и Таджикистан. Поскольку по историческим данным наиболее мощное воздействие могли оказывать миграции по степной полосе, основной акцент исследования был сделан на населении Алтае-Саянского региона (алтайцы, казахи, шорцы, хакасы), а также Тянь-Шаня и Памира (киргизы). Собран и проанализирован 1241 образец представителей 13 малых народов и субэтнических групп крупных этносов. В Южной Сибири многие народы региона имеют статус «коренных

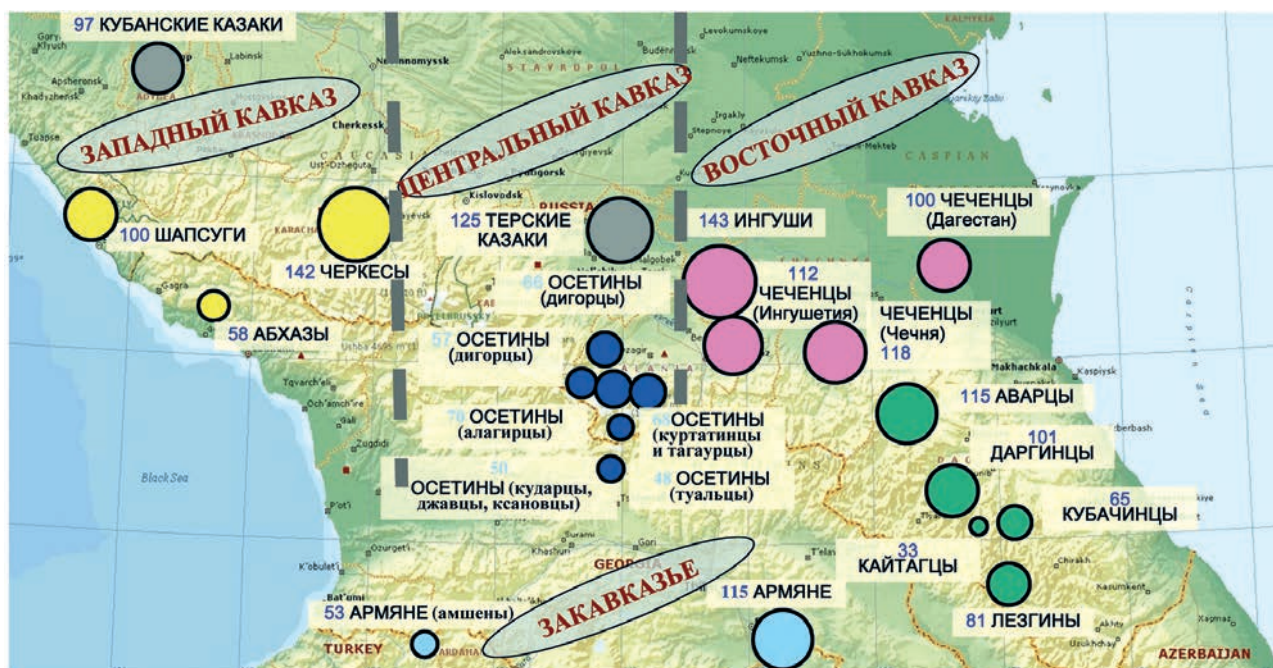


Рис. 1.13. Географическое положение и этнолингвистическая принадлежность изученных популяций Кавказа. Размер кружка пропорционален объему выборки. Цветом указана принадлежность популяции к лингвистической группе: желтый цвет – абхазо-адыгская группа; синий цвет – иранская группа; зеленый цвет – нахская группа; красный цвет – дагестанская группа; фиолетовый цвет – армянская группа; серый цвет – славянская группа.

малочисленных народов Российской Федерации» [Постановление Правительства РФ от 24 марта 2000 г. N255 «О едином перечне коренных малочисленных народов РФ»]. Эта малочисленность коренного населения и интенсивная метисация разных групп как с пришлым славянским населением, так и друг с другом, потребовала особых усилий для формирования репрезентативных выборок.

Суммарный ареал изученных народов велик: от степных районов Западной Сибири, где равнина постепенно сменяется северными предгорьями Алтая и цепью Саянских хребтов, до альпийских лугов Тянь-Шаня и высокогорного Памирского плато. На Алтае исследованием охвачены практически все

коренные народы: шорцы (горные), северные алтайцы (челканцы, кумандинцы, тубалары), южные алтайцы (алтай-кижи, теленгиты), казахи южных предгорий Алтая (Казахстан). В Саянах изучены хакасы (качинцы, сагайцы, койбалы, кызыльцы) и абаканские шорцы. На Тянь-Шане и Памире были изучены контрастные группы киргизов – соответственно, нарынских и мургабских. Каждая этническая группа представлена несколькими локальными популяциями. Лучше всего изучены народы Алтая (представлены семь малых народов) и Саян (три крупных объединения хакасов и выборка абаканских шорцев). Этнический состав, объемы выборок и регионы их сбора представлены в таблице 1.5.

АНАЛИЗ ДРЕВНЕЙ ДНК

Анализ древней ДНК – заманчивый путь изучения истории формирования генофондов. И хотя на нем подстерегают опасности получения неверных экспериментальных результатов (из-за контаминации древних образцов с современной ДНК) и неверных выводов (из-за крошечных объемов выборок и немногочисленности изученных популяций), некоторым научным коллективам удается пройти мимо этих сцилл и харибд и получить новые и при этом надежные результаты.

В таблице 1.7 представлен перечень тех популяций древней ДНК, которые были изучены при участии автора. Они охватывают все основные эпохи – палеолита, мезолита, неолита, раннего металла, античности и Нового времени и образцы из раз-

ных регионов Европы. Большинство этих исследований было инициировано автором и им же проведена значительная часть популяционно-генетического анализа полученных экспериментальных данных. Генотипирование этих образцов по маркерам мтДНК проведено в Австралийском центре древней ДНК, в том числе при участии автора. В остальных случаях автор настоящей книги лишь был одним из соавторов результирующих публикаций (это относится к полногеномным данным). Данный перечень (табл. 1.7) представляет значительную часть от всех изученных в мировой науке древних популяций Европы и ее окрестностей, и систематическое описание этих результатов составило основу главы 8.

Таблица 1.7. Изученные популяции древней ДНК.

Популяция	Эпоха	Датировка	Локализация	N	Генотипирование
Мальта, Афонтова гора	палеолит	17-24 тыс. лет назад	Россия, Иркутская обл., Красноярский край	2	Полные геномы
Южный Олений остров, Попово	мезолит	7,5 тыс. лет назад	Россия, Карелия	11	мтДНК
Лошбор	мезолит	8,0 тыс. лет назад	Люксембург	1	Полные геномы
Мотала	мезолит	8,0 тыс. лет назад	Мотала, Швеция	7	Полные геномы
Деренбург, культура линейно-ленточной керамики	неолит	7,0-7,5 тыс. лет назад	Германия, Деренбург	21	мтДНК
Штуттгарт, культура линейно-ленточной керамики	неолит	7,0 тыс. лет назад	Германия, Штуттгарт	1	Полные геномы
Большой Олений остров	эпоха раннего металла	3,5 тыс. лет назад	Россия, Мурманская обл.	23	мтДНК
Скифы Дона	античное время	2,5 тыс. лет назад (VI-III вв. до н.э.)	Россия, Ростовская обл.	16	мтДНК
Чальмны-Варрэ	новое время	0,2 тыс. лет назад (XVIII век н.э.)	Россия, Мурманская обл.	42	мтДНК

ОСОБЕННОСТИ ГЕНОТИПИРОВАНИЯ Y-ХРОМОСОМЫ И МТДНК

В данном подразделе кратко описывается технология генотипирования, с помощью которой получен тот большой массив собственных данных, который был описан в предыдущих подразделах. Эти методические вопросы имеют значение постольку, поскольку унификация процедур генотипирования обеспечивает возможность сравнительного анализа данных по самым разным популяциям. Без такой унификации возникали бы проблемы и разных наборов гаплогрупп Y-хромосомы, изученных для разных популяций, и вопросы надежности полученных экспериментальных данных, и многие другие. Поэтому стоит подчеркнуть, что собственные данные нашего коллектива унифицированы как на этапе сбора образцов по единой стратегии и единым правилам, так и на этапе генотипирования этих образцов.

ФОРМИРОВАНИЕ ДНК-КОЛЛЕКЦИЙ

Выделение ДНК из практически всех коллекций образцов венозной крови проводилось методом фенол-хлороформной экстракции, гарантирующим высокий выход и качество ДНК, а также стабильность при многолетнем хранении. Для очень немногих популяций в качестве исключения применялся метод выделения ДНК высаливанием или ряд хорошо себя зарекомендовавших коммерческих наборов. Для каждого образца сформированы коллекционная, рабочая и архивная (резервная) аликвоты. Проведено измерение концентрации ДНК, причем не только спектрофотометрически, но определялась и эффективная концентрация методом ПЦР в реальном времени с использованием набора Human DNA Quantifiler kit (Applied Biosystems). Измерение концентрации каждого образца проводилось в двух независимых реакциях, а при высокой дисперсии результатов анализ повторялся еще в двух повторностях. Сформированы коллекции нормализованной (до 2 нг/мкл) ДНК в 96-луночных планшетах, из которых по мере необходимости приготавливались планшеты с сухими образцами ДНК. Именно ДНК, нормализованная до единой концентрации, и использовалась в большинстве видов анализа.

ПАНЕЛЬ SNP МАРКЕРОВ Y-ХРОМОСОМЫ

Для анализа полиморфизма Y-хромосомы использованы обе системы маркеров: однонуклеотидные полиморфные локусы SNP и высоковариабельные микросателлитные STR маркеры.

Исследование SNP позволяет отнести исследуемый образец к той или иной ветви родословного древа Y-хромосомы – гаплогруппе. В начале популяционных исследований Y-хромосомы на ней

было известно лишь около десятка гаплогрупп, далее это число, благодаря специальным исследованиям, постепенно увеличивалось, и к началу эпохи массового полного секвенирования Y-хромосомы (т.е. к 2013 году) на филогенетическом древе Y-хромосомы было известно около 500 гаплогрупп, из которых примерно для половины проводился анализ ее распространения хоть бы в некоторых популяциях хотя бы одной лабораторией в мире. При этом и число, и набор маркеров, генотипированных в разных исследованиях, различались драматически. Поэтому перед нами стояла задача сформировать набор гаплогрупп, по которым проводить генотипирование наших коллекций.

Гаплогруппы Y-хромосомы географически высоко специфичны, поэтому нам для изучения популяций Северной Евразии не было большой нужды включать африканские, американские или австралийские гаплогруппы, но евразийские гаплогруппы надо было включить почти все (за исключением разве что приуроченных только к Южной и Юго-Восточной Азии). Ведь в генофонде изучаемых народов из удаленных частей Северной Евразии можно встретить максимальное разнообразие спектра гаплогрупп – как западно-евразийских, так и восточно-евразийских. При этом вставал вопрос и о глубине филогенетического разрешения, поскольку анализировать все до одной субгаплогруппы, известные в пределах какой-либо гаплогруппы, было нереально с точки зрения финансового обеспечения. Поэтому приходилось включать в панель лишь субгаплогруппы, наиболее частые в Северной Евразии, причем сама частота нередко выяснялась в ходе как раз наших исследований.

В итоге была разработана панель из 44 основных и 32 дополнительных гаплогрупп. Основные гаплогруппы генотипировались практически без исключений для всех изученных нами популяций. Дополнительные представляли собой региональные субтипы, информация по которым имела значение лишь при изучении данного региона, а генотипирование их в остальных популяциях не было столь уж необходимым. Например, субтипы гаплогруппы **O3** было важно изучить в популяциях Сибири, но их генотипирование в популяциях Кавказа было избыточным – гаплогруппа **O3** настолько редка на Кавказе, что для интерпретации ее присутствия у считанных образцов вполне достаточно знать, что это **O3**, и не так важно, в каком именно субтипе **O3** они относятся (появление того или иного субтипа становится уже игрой случая при таких низких частотах).

В этом перечне гаплогруппы «основной панели» указаны жирным шрифтом, а «дополнительные» маркеры в пределах основных гаплогрупп даны обычным шрифтом и в скобках. Вот этот перечень:

C-M130, C-M217, C-M48 (C-M407), D-M174, E-M96, E-M35, E-M78 (E-V13, E-M123), F*-M89, G*-M201, G-M285, G-P15, G-P16, G-P18, G-M406, G-P303 (G-M377), H-M69, I-M170, I-M253, (I-P215), I-P37, I-M223, J-M304, J-M267, J-P58, J-M172, J-M12 (J-M47), J-M67, J-M92, K-M9, L-M20 (L-M27, L-M317, L-M357), N-M231 (N-LLY22g), N-M128, N-P43, N-M178, O-P186 (O-M119, O-P31), O-M122 (O-P201, O-M159, O-M134, O-M7), Q-M242, R-M207, R-M198, R-M458, R-M343, R-M73, R-M269 (R-L23, R-L11, R-U106, R-U198, R-L48, R-P312, R-M65, R-M153, R-M167, R-U152, R-M126, R-M160, R-L2, R-L20, R-M37, R-M222), R-M124, T-M70.

В 2014-2015 годах к нашим «основной» и «дополнительным» панелям Y-SNP маркеров добавились еще две панели маркеров, открытых на гаплогруппах **N1c-M178** и **C3-M217** и включающих суммарно 31 маркер. Подробнее эти панели описаны в тех разделах книги, где рассматриваются полученные с их помощью результаты.

МЕТОД ГЕНОТИПИРОВАНИЯ SNP МАРКЕРОВ

Генотипирование проводилось методом ПЦР в реальном времени с использованием TaqMan зондов на SNP маркеры (эти зонды были разработаны по нашему заказу компанией Applied Biosystems). Пример результирующего графика приведен на рис. 1.14. Можно видеть, что этот метод дает надежные и объективные результаты. Важно, что этим методом возможно объединение на одном графике результатов всех постановок одного и того же маркера, даже сделанных в разные годы, что обеспечивает унифицированный подход при отнесении образцов к одному или другому кластеру. Также существенно, что график всегда можно пересмотреть повторно, если возникают сомнения в правильности определенного генотипа. Наконец, формат автоматического экспорта результатов в электронные таблицы, предусмотренный в программах обработки результатов TaqMan, насуточно необходим при массового анализа тысяч образцов. Генотипирование Y-хромосомы проведено в основном в лаборатории популяционной генетики человека МГНЦ РАМН и в лаборатории геномной географии ИОГен РАН под руководством и при участии автора, за исключением русских популяций, проанализированных лично автором на базе Эстонского биоцентра [Balanovsky et al, 2008]. SNP маркеры в этих русских популяциях анализировались не с помощью TaqMan зондов, а более старым методом ПДРФ, как описано в [Balanovsky et al., 2008].

При типировании популяций каждого региона (славяне, Кавказ, Южная Сибирь и Средняя Азия) вначале подбирались панель гаплогрупп, наиболее частых в данном регионе по информации нашей базы данных Y-base. Эта панель маркеров геноти-

пировалась для каждого образца. Затем спектр используемых SNP маркеров увеличивался, но они уже генотипировались только для тех образцов, которые, по результатам первого раунда генотипирования, могли относиться к соответствующим гаплогруппам.

Для наиболее эффективного и быстрого определения SNP-гаплогруппы применялись две стратегии (по отдельности или совместно). Во-первых, «иерархическое типирование» - анализ на все субгаплогруппы, дочерние по отношению к гаплогруппе, определенной при предыдущем этапе анализа образца. Во-вторых, «предикция»: на основании данных по STR гаплотипу определялась ожидаемая гаплогруппа для образца, которая далее подтверждалась типированием соответствующего SNP-маркера (предикция проводилась с помощью разработанной нашим коллективом программы-предиктора, описанной в разделе 1.1).

ГЕНОТИПИРОВАНИЕ STR МАРКЕРОВ

Анализ высокополиморфных STR маркеров Y-хромосомы позволяет выявить в пределах каждой гаплогруппы множество родственных гаплотипов. Различия между индивидами по числу коротких tandemных повторов (STR – short tandem repeats) возникают гораздо чаще, чем по SNP из-за более высокой скорости мутирования STR – быстрого изменения числа tandemных повторов. Однако по каждому отдельному STR маркеру легко могут возникать одинаковые аллели в разных гаплогруппах. Поэтому, не зная, к какой гаплогруппе принадлежит данный образец, мы не можем с уверенностью сказать, действительно ли он родственен другому образцу с похожим набором STR маркеров.

Именно поэтому наиболее эффективно параллельное исследование обеих генетических систем – SNP и STR. Их образно сравнивают с часовой и минутной стрелкой: первая (SNP) позволяет понять положение образца на родословном древе в «глобальном» масштабе, а вторая (STR) – определить его более точное положение. Исследование STR маркеров на фоне данных об SNP маркерах позволяет датировать возникновение отдельных кластеров гаплотипов (т.е. субветвей родословного древа Y-хромосомы).

Для получения данных о STR гаплотипах амплифицировались 17 STR локусов Y-хромосомы (DYS389I, DYS389II, DYS390, DYS456, DYS19, DYS385a, DYS385b, DYS458, DYS437, DYS438, DYS448, GATA_H4, DYS391, DYS392, DYS393, DYS439, DYS635) с использованием коммерческого набора Y-filer PCR Amplification Kit (Applied Biosystems). Фрагментный анализ амплификатов проводили на капиллярном секвенаторе Applied Biosystems 3130xl Genetic Analyzer. Ряд образцов анализировался также с использованием второго

мультиплексного набора (Genographic custom multiplex, Applied Biosystems), включавшем два STR локуса, DYS388 и DYS426, вместе с шестью SNP локусами (M17, M60, M91, M139, M175, M186).

ПОЛНОЕ СЕКВЕНИРОВАНИЕ Y-ХРОМОСОМЫ

За последнее десятилетие полногеномный анализ стал мощным методом обнаружения новых SNP-маркеров. Это особенно актуально для Y-хро-

мосомы, с ее максимальной межпопуляционной изменчивостью и географической специфичностью аллелей (гаплогрупп). Полное секвенирование каждой очередной Y-хромосомы фактически открывает новую гаплогруппу (новую веточку на филогенетическом дереве) и несколько специфичных для нее SNP-маркеров.

В конце 2013 года компании FamilyTreeDNA и BGI независимо создали долгожданные техноло-

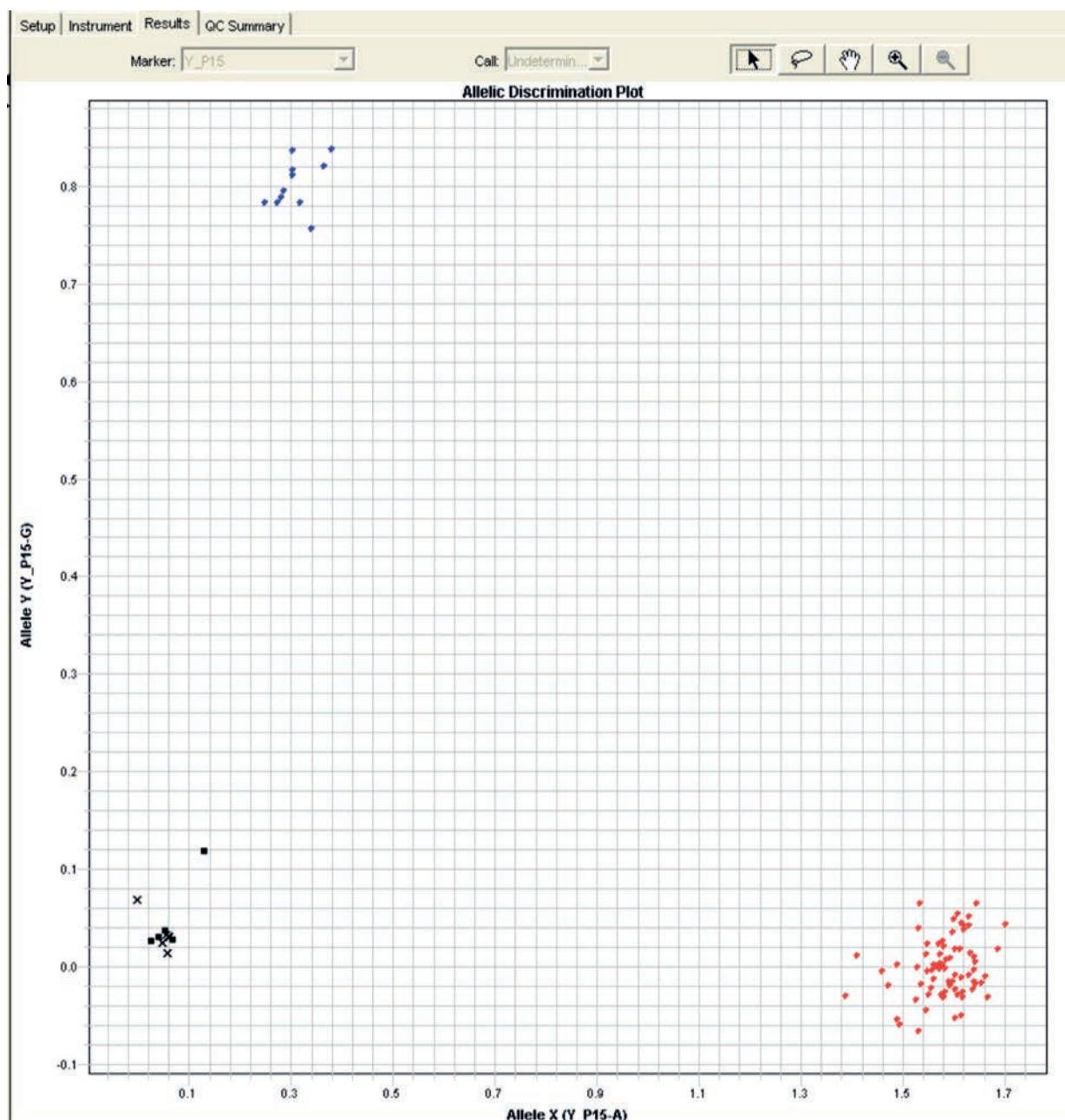


Рис. 1.14. Генотипирование SNP-маркеров методом TaqMan.

Приведен результирующий график накопления двух флуоресцентных меток после ПЦР-реакции. Представлены результаты генотипирования образцов шапсугов по маркеру P15, определяющему гаплогруппу **G2a** Y-хромосомы. Каждый образец показан отдельной точкой. Распределение образцов по Y-и X оси в соответствии с наличием или отсутствием соответствующей метки (а значит, соответствующего нуклеотида) позволяет разделить их на три группы: преобладание сигнала от аллеля P15-A (образцы отмечены красными точками), преобладание сигнала от аллеля P15-G (синий цвет), отсутствие обоих сигналов (реакция не прошла, черный цвет). Данный пример иллюстрирует эффективность и удобство метода TaqMan для генотипирования SNP-маркеров.

гии Y-capture (отдельный анализ Y-хромосомы), оптимальные по соотношению «цена-информативность» и потому даже более перспективные для изучения генофонда, чем полные геномы человека. Лидирующие коллективы мира тотчас приступили к анализу генофондов на основе Y-capture. Поскольку наш коллектив обладает самым обширным во всем мире биобанком по народонаселению Северной Евразии и обширным опытом исследований Y-хромосомы, это дало нам возможность успешной конкуренции в этом новейшем и перспективном направлении исследований в мировой науке.

По самой логике подобных популяционных исследований, работа естественным образом распадается на две части: обнаружение новых маркеров на Y-хромосоме и скрининг их распространенности в популяциях.

«ПОЛНЫЕ ИГРЕКИ»: ДЕРЕВО

Для обнаружения новых маркеров необходимо проводить секвенирование Y-хромосомы. Эти сиквенсы можно получить либо в качестве побочного продукта полного секвенирования генома (стоимость не менее 3 тысяч долларов США), либо физическим выделением Y-хромосомы (необходимы клеточные линии, и в целом это едва ли оправданный путь), либо секвенированием только Y-хромосомы после избирательного обогащения полногеномных библиотек данной хромосомой. Именно этот путь является наиболее целесообразным по соотношению цена/результат. Технология избирательного секвенирования Y-хромосомы разработана двумя компаниями – китайской лидирующей компанией полногеномного секвенирования BGI (цена выше 1000 долларов) и американской компанией FamilyTreeDNA (цена 700 долларов). Хотя технология американской компании позволяет прочитать чуть более короткую часть Y хромосомы, но число проанализированных образцов важнее длины секвенирования, поэтому нами был выбран именно этот вариант. Хотя полное секвенирование Y-хромосомы на порядок дешевле полного секвенирования генома, но и такие исследования требуют заметных финансовых затрат. Поэтому «полногеномная» часть наших исследований Y-хромосомы была бы невозможна без полученного автором гранта РНФ (№ 14-14-00827) и в особенности без сотрудничества с лабораторией исторической генетики МФТИ, взявшей на себя расходы по секвенированию образцов из тех гаплогрупп, для анализа которых финансирования РНФ оказалось недостаточно.

Целью является выявление новых ветвей филогенетического дерева и обнаружение специфичных для них генетических маркеров, а эффективность такого выявления зависит от двух факторов:

от числа секвенированных образцов и от степени представленности различных ветвей дерева в этих образцах. Конечно же, принадлежность образца к ветви заранее неизвестна (выяснить ее – это и есть цель секвенирования), однако охват географически контрастных популяций и генетически контрастных образцов (по уже известным для них маркерам Y-хромосомы) значительно повышает вероятность представленности филогенетически контрастных ветвей в выборке.

После получения сиквенсов необходимо провести биоинформационную обработку, и уже на основании обработанных сиквенсов построить филогенетическое дерево. Первоначальная обработка (выравнивание на геном, отсеивание ложных сигналов и т.п.), завершающаяся списком маркеров, обнаруженных у каждого образца, предлагается секвенирующей компанией, хотя и этот этап исследователю полезно параллельно проводить самостоятельно. Последующие этапы – приведение сиквенсов всех образцов к единому списку генотипированных маркеров, разработка алгоритма их сведения в единую таблицу – должны проводиться самим исследователем.

«ПОЛНЫЕ ИГРЕКИ»: ПОПУЛЯЦИОННЫЙ СКРИНИНГ

Наконец, выбор маркеров для скрининга (из большого числа маркеров, которые, как ожидается, будут открыты в результате секвенирования) в связи с высокой ценой этого этапа целесообразно проводить с максимальной эффективностью, включая лишь те маркеры, которые гарантированно маркируют не частные «семейные» линии Y-хромосомы, а достаточно распространены хотя бы в некоторых популяциях.

Для скрининга распространенности ветвей, определяемых выбранными маркерами, необходимо охватить максимально широкий географический ареал, в идеале полностью покрывающей территорию изучаемой гаплогруппы, и при этом необходима высокая плотность покрытия ареала изученными популяциями. Для этого наша коллекция образцов Северной Евразии является оптимальным решением.

К настоящему времени с использованием такого подхода мы провели исследование уже трех гаплогрупп (**C3-M217**, **N1c-M178** и **G1-M285**) и завершается исследование восточноевропейских ветвей **R1b**. Для этих гаплогрупп проведено полное секвенирование Y-хромосомы более 100 образцов, построены филогенетические деревья, обнаружено большое количество новых субветвей, подавляющее большинство которых характеризуются высокой географической специфичностью. Для каждой из этих ветвей выбрано по одному SNP-маркеру и созданы их панели, позволяющие про-

водить массовый скрининг распространенности этих новых открытых ветвей Y-хромосомы в популяциях. В эти панели включены маркеры, определяющие все основные субветви, известные на данный момент в пределах этих гаплогрупп (как полученные другими коллективами, так и наши собственные).

Ко времени написания этой книги нами широко используются панели для гаплогрупп **N1c-M178** и **C3-M217**, включающих 8 и 23 маркера. По этим панелям генотипированы почти все образцы из наших коллекций, относящиеся к соответствующим гаплогруппам (суммарно более двух тысяч образцов). Также готовится панель по восточноевропейским и азиатским субгаплогруппам **R1b**, и можно ожидать, что в ближайшие пару лет нами будут разработаны и войдут в обиход и панели для остальных важнейших североевразийских гаплогрупп.

По мнению автора, именно эта работа по резкому увеличению филогеографического разрешения Y-хромосомы путем не только открытия новых гаплогрупп в результате полного секвенирования, но и популяционного скрининга на эти новые маркеры является сейчас наиболее перспективной в мировой популяционной генетике. Подробнее ход работы, используемые подходы, биоинформационные методы и первые полученные результаты систематически описаны в разделе 7.4. а также привлекаются по ходу изложения в некоторых других частях книги.

ГЕНОТИПИРОВАНИЕ МИТОХОНДРИАЛЬНОЙ ДНК

Митохондриальные маркеры используются в исследованиях генофонда популяций человека благодаря их высокой скорости мутирования и отсутствию рекомбинации (что методологически превращает мтДНК в такой мультиаллельный локус, в котором можно восстановить порядок происхождения аллелей друг от друга). Поэтому мтДНК, как и другая нерекombинирующая система в геноме человека – Y-хромосома – идеальны для филогенетических исследований, основанных на реконструкции дерева происхождения аллелей (гаплотипов) друг от друга. Митохондрии передаются детям от их матери и происходят от тех митохондрий, которые оказываются в яйцеклетке в момент оплодотворения. Ключевым вопросом для филогенетических исследований с помощью мтДНК является возможность рекомбинации мтДНК отца и матери. Считается, что митохондрии сперматозоидов (вместе с отцовской мтДНК) деградируют после оплодотворения. Указания на наличие рекомбинации между отцовской и материнской мтДНК, произошедшие в результате передачи мтДНК от отца, подробно разбирались и были опровергнуты в критическом обзоре [Bandelt et al., 2005].

ЭВОЛЮЦИЯ РЕФЕРЕНСНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ мтДНК

Для характеристики каждой секвенированной мтДНК указывают ее отличия от последовательности, избранной к качеству референсной. Первым референсом была первая полностью секвенированная мтДНК, так называемая «кембриджская» последовательность (CRS – Cambridge reference sequence), впервые опубликованная в [Anderson et al., 1981]. При повторном исследовании того же образца почти двадцать лет спустя [Andrews et al., 1999], обнаружился ряд небольших неточностей в первоначальном определении последовательности. Исправленный вариант CRS обозначают как rCRS (revised CRS).

В отношении гипервариабельных сегментов, которые исследуются наиболее часто, CRS не отличается от rCRS. Однако этот референс был неудобен тем, что находится на самом конце одной из ветвей филогенетического дерева, характерной именно для Европы. Однако в мировом генетическом разнообразии и глобальной истории популяций человека Европа, конечно, является окраинным, периферийным регионом. Поэтому для описания гаплогрупп других регионов мира (и даже других европейских ветвей) при использовании rCRS в качестве референса приходилось сначала отсчитывать мутации от rCRS до корня мирового дерева, а потом от корня до изучаемой гаплогруппы. Такой двойной отсчет хотя и стал для исследователей привычным за двадцать лет использования, но был неудобен и приводил иногда к ошибкам.

Поэтому в работе [Behar et al., 2012] было предложено использовать в качестве референсной последовательности реконструированный «корень» филогенетического митохондриального дерева человечества, которых получил обозначение RSRS – Reconstructed Sapiens Reference Sequence. В данный момент основной часто обновляемый ресурс по митохондриальному дереву – www.phylotree.org – предлагает версии дерева, ориентированные как на rCRS, так и на RSRS.

ДВА РЕГИОНА мтДНК

Функционально мтДНК разделяют на так называемые кодирующий регион (включающий нуклеотиды в мтДНК от номера 00578 до номера 16023) и контрольный регион (включающий остальные нуклеотиды, 16024-00574). В кодирующем регионе расположены 37 генов. В середине контрольного региона расположен сайт инициации репликации H-цепи (от которого и ведется отсчет номеров нуклеотидов в мтДНК) и сайты инициации транскрипции обеих цепей. Контрольный регион реплицируется в последнюю очередь и долго существует в виде нереплицированного (общего для двух новых цепей) участка – так называемой D-

петли (D-loop). Внутри D-петли существует 3 гипервариабельных сегмента (ГВС). Накопление мутаций в них происходит особенно быстро - основная часть полиморфизма мтДНК сосредоточена в первом гипервариабельном сегменте (ГВС1) контрольного региона. На D-петлю приходится около 23% всего полиморфизма мтДНК у человека при ее общей длине 6,8% от всей мтДНК [Horai et al., 1995].

Многие мутации, возникающие в D-петле ведут себя как условно нейтральные. Мутации, возникающие в кодирующем регионе, наоборот, потенциально подвержены жесткому естественному отбору в связи с большой значимостью функционирования митохондрий для жизнедеятельности организма.

КАКОЙ РЕГИОН мтДНК ИЗУЧАТЬ?

Для исследования мтДНК можно использовать полное секвенирование всех 16,5 тысяч нуклеотидов, и такой подход в последние 2–3 года стал наиболее распространенным. Но подавляющее большинство образцов мтДНК, изученных к настоящему времени, изучены не столь полно, поскольку основной массив был накоплен в предшествующие годы, когда трудоемкие технологии секвенирования позволяли проводить полный анализ мтДНК лишь для немногих образцов. Большинство образцов поэтому изучались либо лишь по ГВС – гипервариабельным сегментам (одному или нескольким), либо по гипервариабельным сегментам плюс по нескольким (филогенетически наиболее информативным) SNP маркерам кодирующей части. Необходимость исследования маркеров кодирующего региона обусловлена тем, что в гипервариабельных сегментах мтДНК нередко параллельно возникают одинаковые мутации в разных гаплогруппах, и поэтому анализ только ГВС часто оказывается недостаточным для правильного определения гаплогруппы [Kivisild et al., 1999]. Как правило,

первый способ (анализ только ГВС) типичен для статей криминалистов, для которых главным является совпадение/несовпадение образцов, а второй способ (дополнение ГВС филогенетически информативными маркерами кодирующего региона) – для популяционно-генетических работ, для которых принципиально определение принадлежности к конкретной гаплогруппе.

ГЕНОТИПИРОВАНИЕ мтДНК

Этот второй подход – параллельный анализ и контрольного, и кодирующего регионов – был использован и в нашем исследовании. Генотипирование мтДНК включило секвенирование ГВС1 между позициями 16024 и 16400. Секвенирование проводилось на капиллярных секвенаторах ABI 377, ABI 3130xl, Megabace 2000 с использованием секвенирующих наборов DYEnamic™ ET terminator или BigDye Terminator v3.1. Определение одонуклеотидных замен в кодирующем регионе молекулы мтДНК проводилось методом ПДРФ-анализа или методом TaqMan; анализировались маркеры на гаплогруппы **H, V, HV, U, U4, K, R, T, J, J1, N, N1b, W, X, I, M, A, C, Z, D, D5**. Для контроля правильности определения гаплогруппы использовались также характерные сочетания мутаций (мотивы) в ГВС1 мтДНК, которые также применялись для уточнения субгаплогрупп (например, **U5a, U5b, U2e** и другие варианты в пределах гаплогруппы **U** определялись по мотивам ГВС1). Перечень характерных мотивов для ряда гаплогрупп опубликован с участием автора в работе [Behar et al., 2007].

Генотипирование мтДНК большинства популяций проведено автором на базе Эстонского биоцентра, оставшаяся часть (главным образом южнорусские и северокавказские популяции, табл. 1.6) была генотипирована под руководством автора в лаборатории популяционной генетики человека МГНЦ РАМН.

1.4. КАРТОГРАФИЧЕСКИЙ И СТАТИСТИЧЕСКИЙ АНАЛИЗ

Он любил три вещи на свете:
За вечерней пенье, белых павлинов
И стертые карты Америки.
Не любил, когда плачут дети,
Не любил чая с малиной
И женской истерики.
...А я была его женой.

Анна Ахматова

Быстрокрылых ведут капитаны,
Открыватели новых земель,
Для кого не страшны ураганы,
Кто изведаль мальстремы и мель,
Чья не пылью затерянных хартий, —
Солью моря пропитана грудь,
Кто иглой на разорванной карте
Отмечает свой дерзостный путь

Николай Гумилев

В анализе генофондов различного масштаба (от всей Евразии до популяций Русского Севера) и различной географической локализации (от Европы до Сибири), вошедших в эту книгу, применя-

лись единые подходы и единый набор методов. Поэтому разумно изложить их однажды в начале книги, чтобы не повторять потом многократно методические подробности. Использованные методы

можно разделить на статистические, филогенетические и картографические. Статистические большей частью основаны на частотах аллелей (для Y-хромосомы и мтДНК-гаплогрупп), филогенетические – на анализе гаплотипов (включая расчет возраста гаплогрупп по накопленному разнообразию гаплотипов), а картографические – показывают распространение статистических или филогенетических показателей в пространстве и позволяют проводить дальнейший статистический анализ уже самих полученных карт.

ЛОГИКА КАРТОГРАФИЧЕСКОГО АНАЛИЗА

Целью картографического анализа является «проявление» структуры генофонда. Поэтому карта – это не иллюстрация, а инструмент исследования. Не картинка, а научная модель. Не фотография, а образ генофонда. Как всякий хороший инструмент, карта непроста в использовании и требует правильного обращения. Как всякая модель, она может существовать во множестве вариантов – в зависимости от задаваемых параметров. Как всякий образ, карта может быть прокомментирована, но не передана словами. Карта иногда красива, и всегда логична, проверяема, математически точна.

На первом этапе, согласно логике картографического анализа генофонда, строятся карты распространения отдельных генов (или других признаков). Эти карты содержат рассчитанные значения признака в каждом из тысяч узлов карты, равномерной сетью покрывающих территорию. И нужно подобрать такие параметры картографирования, чтобы полученный набор значений оптимально моделировал распространение признака и соответствовал целям конкретного исследования.

На втором этапе эти рассчитанные значения подвергаются статистическому анализу. Простейший пример – среднее значение по всем узлам карты. Сложнейший – корреляция картографированного признака с неким другим признаком, рассчитанная для каждого узла карты (причем в одной части территории корреляция может оказаться положительной, а в другой – сменить плюс на минус).

На третьем этапе карты разных признаков комбинируются друг с другом – создаются обобщен-

ные карты. Поскольку карты – это математические матрицы значений, с ними можно проводить любые математические операции – складывать, делить друг на друга, извлекать корни и брать производные... Поэтому любой популяционно-генетический параметр, который обычно рассчитывается для данных в исходных изученных популяциях, можно рассчитать и для картографированных значений, тем самым построив карту этого параметра. Обратное неверно: бывают ситуации, когда по карте рассчитать какой-то параметр можно, а вот без карты – нельзя. Например, если разные анализируемые признаки изучены по разному набору популяций, то многие виды многомерного статистического анализа становятся невозможными, тогда как картографический анализ, позволяющий рассчитать значения всех признаков в каждом из узлов сетки карты, становится универсальным инструментом изучения генофонда, позволяя проводить и те виды статистического анализа, которые без карты провести никак нельзя.

На всех этих этапах карта представляет собой файл с данными, не воспринимаемый толком человеческим глазом. Но этот файл всегда можно визуализировать, показав разные группы значений в разных узлах разными цветами и выведя их в тех частях карты, где находится соответствующий узел. Тем самым статистическая модель предстает в обличье привычной географической карты, на которую можно наносить любую дополнительную информацию – от береговой линии до мест археологических раскопок.

В итоге можно получить карту распространения любого аллеля, или карту любого параметра генофонда, провести ее последующий анализ для уточнения закономерностей в географической изменчивости, связать с другими картами и создать в итоге точный и выразительный геногеографический образ генофонда.

Картографический способ представления данных о генофонде оптимален еще и тем, что допускает использование генетических результатов не только генетиками, но и специалистами в смежных областях: карта является универсальным средством, понятным представителям многих научных дисциплин, чего нельзя сказать о генетической терминологии.

ПРОГРАММА GENE GEO

ПРОГРАММА GENE GEO – ОБЩАЯ ХАРАКТЕРИСТИКА

Теория, правила и алгоритмы такого картографического анализа разрабатывались еще в 90е годы прошлого века рядом коллективов. Наиболее последовательно и полно они были разработаны российской школой геногеографии, по преимуществу

в работах Е.В. Балановской. Для их практического применения под ее руководством был последовательно разработан ряд картографических пакетов, из которых наилучшим и наиболее широко использовавшимся был GGMAG. Но в начале текущего столетия этот пакет, разработанный под DOS, с одной стороны, начал терять совместимость с со-

временными операционными системами, и в него все труднее стало вносить усовершенствования, а с другой стороны, стремительное развитие средств программирования делало возможным решение тех задач (особенно в плане визуализации), которые ранее были нереальными. Поэтому назрела насущная необходимость в разработке нового, современного программного обеспечения для картографо-статистического анализа генофонда. Им стала программа GeneGeo.

Как видно из его названия, программа разрабатывалась специально для целей геногеографии. GeneGeo разрабатывается под руководством автора данного исследования совместно с сотрудником кафедры картографии МГУ к.г.н. С.М. Кошелем (вместе с которым разрабатывался и GG MAG). Программа GeneGeo соответствует современным мировым стандартам картографических программ, в том числе в разделе графического отображения карт и совместимости с другими картографическими пакетами (Surfer, ArcGIS). Для возможности международного использования интерфейс программы представлен на английском языке. С точки зрения программирования, авторских прав, использования в публикациях и т.д., программа GeneGeo, является новым программным продуктом, написанным практически «с нуля». При этом с точки зрения реализуемых методов, GeneGeo является продолжением описанной выше геногеографической традиции, представленной за рубежом работами L.L. Cavalli-Sforza и его школы, а в России – работами Ю.Г. Рычкова, Е.В. Балановской и их учеников, включая и автора данного исследования.

Программа GeneGeo разрабатывается нашим коллективом в течение ряда лет (с 2007 года по настоящее время). В связи с обширностью и трудоемкостью задачи, разные этапы этой работы выполнялись в разное время и при частичной поддержке ряда грантов, а в 2009–2012 годах создание этой программы являлось одним из направлений темы НИР «Анализ распространения гаплотипов митохондриальной ДНК и Y-хромосомы у народов мира на основе создания геоинформационных систем», выполняемой в Медико-генетическом научном центре РАМН под руководством автора.

ИСПОЛЬЗОВАНИЕ ПРОГРАММЫ GeneGeo

Картографическая программа GeneGeo широко используется не только в отечественных, но и в ряде зарубежных исследований. Программа использовалась (и, соответственно, цитировалась) в следующих международных и отечественных статьях:

1) в статье [Haak et al., 2010], опубликованной в журнале PLoS Biology (импакт-фактор 13);

2) в статье [Balanovsky et al., 2011], опубликованной в журнале Molecular Biology and Evolution (импакт-фактор 10);

3) в статье [Knappskog et al., 2014], опубликованной в журнале Oncotarget (импакт-фактор 6,6);

4) в статье [Der Sarkissian et al., 2014], опубликованной в журнале PLOS One (импакт-фактор 3,5);

5) в статье [Kushniarevich et al., 2015], опубликовано в журнале PLOS One (импакт-фактор 3,5);

6) в статье [Balanovsky et al., 2015], опубликованной в журнале PLOS One (импакт-фактор 3,5);

7) в статье [Haber et al., 2011], опубликованной в журнале Annals of Human Genetics (импакт-фактор 2,6);

8) в статье [Pshenichnov et al., 2013], опубликованной в журнале American Journal of Physical Anthropology (импакт-фактор 2,5);

9) в статье [Балановский и др., 2010], опубликованной в журнале «Медицинская генетика»;

10) в статье [Дибирова и др., 2010], опубликованной в журнале «Медицинская генетика»;

11) в статье [Балаганская и др., 2011], опубликованной в журнале «Вестник Московского университета. Серия XXIII. Антропология»;

12) в статье [Балановская и др., 2011], опубликованной в журнале «Вестник Московского университета. Серия XXIII. Антропология»;

13) в статье [Балановский и др., 2011], опубликованной в журнале «Генетика» (импакт-фактор 0,25).

14) в статье [Степанов и др., 2011] «Характеристика популяций Российской Федерации по панели пятнадцати локусов, используемых для ДНК-идентификации и в судебно-медицинской экспертизе», опубликованной в журнале «Acta Naturae».

Весь картографический анализ, представленный в настоящей книге, выполнен именно с использованием GeneGeo. И в связи с ключевой ролью картографического анализа в данном исследовании полезно описать, как основные этапы анализа реализованы в этой программе.

ЭТАП 1: ПОСТРОЕНИЕ КАРТ ОТДЕЛЬНЫХ ПРИЗНАКОВ

Для построения карты отдельного признака достаточно исходной информации о значениях этого признака в популяциях и о географических координатах этих популяций. Кроме этой минимальной информации в формате файла исходных данных для каждой популяции предусмотрены: указание ее названия (на английском и русском язы-

ках), объема выборок из них, ссылки на источники данных, исследователя, указавшего эту информацию, и даты указания, два столбца комментариев. Для всего массива данных в начале файла приводится его «паспорт», включая анализируемый регион, цель картографического анализа, исследователей, создавших и модифицировавших файл, даты

его создания и последней модификации. Для удобства пользователей все эти исходные данные представляются в виде csv файла, прекрасно совместимого с Excel.

При запуске GeneGeo, прежде всего, нужно указать картографическую основу, то есть картографируемый регион в заданной картографической проекции, для которого на основу нанесены моря и суша, береговая линия, границы государств и их провинций, главные и второстепенные реки, главные и второстепенные озера, столицы, надписи названий географических объектов, гидрография и прочие объекты. Пользователь может выбирать, какие именно из этого широкого спектра объектов отобразить при построении конкретной карты. Картографические основы разрабатываются отдельно, но к настоящему времени их для GeneGeo разработано уже достаточно много, чтобы проводить анализ любого региона мира, а для Северной Евразии созданы отдельные основы и для многих ее субрегионов. В данном исследовании использованы картографические основы: «Евразия», «Европа», «Восточная Европа», «Кавказ», «Степная полуса Евразии» и некоторые другие.

После того, как основа выбрана, нужно загрузить файл с исходными данными, указать параметры сетки узлов будущей карты (рекомендуется использовать заданные по умолчанию для данной картографической основы, иначе карту нельзя будет анализировать совместно с другими, уже построенными на этой картографической основе) и, самое главное, задать параметры картографирования. При разных параметрах на основе одних и тех же исходных данных получатся разные карты, поэтому выбор параметров является важной и ответственной задачей. К настоящему времени в GeneGeo реализован только алгоритм средневзвешенной интерполяции, хотя начата реализация кригинга и ряда других алгоритмов. Средневзвешенная интерполяция – использованная и в GGMAg, и во многих зарубежных геногеографических работах – состоит в следующем. Значение признака в каждом узле карты рассчитывается как среднее значений во всех опорных точках (т.е. изученных популяций, значения признака для которых взяты из файла исходных данных), в пределах заданного радиуса от рассчитываемого узла; при этом значение в каждой опорной точке берется с весом, обратно пропорциональном расстоянию от нее до рассчитываемого узла, и это расстояние возведено в заданную степень.

Таким образом, для построения карты требуется задать, во-первых, радиус, за пределами которого опорные точки уже не участвуют в расчетах значения признака для данного конкретного узла сетки. Во-вторых, требуется задать степень весовой функции: если задать степень 1, то вес опорных точек будет обратно пропорционален рассто-

янию до них; если задать степень 2, он будет обратно пропорционален квадрату расстояния; если степень 3 – кубу; и так далее. Понятно, что чем больше степень, тем меньше роль более далеких опорных точек. Поэтому из двух параметров важнейшим является именно степень – ведь значения в далеких точках, пусть даже и попавшие в радиус, будут иметь очень небольшой вес при высоких степенях весовой функции.

Радиус обычно подбирается из предположений исследователя о том, на каком расстоянии возможен поток генов в изучаемом регионе и для того масштаба времени, за который предположительно сформировались изучаемые закономерности. Также нужно учитывать плотность распределения опорных точек по карте – если радиус задать слишком маленьким, для каких-то узлов на слабо изученных частях территории в радиус может не попасть ни одной опорной точки (тогда GeneGeo отобразит в узле лишь «белое пятно», это, впрочем, может быть полезным, если часть картографируемого региона действительно совсем не представлена в исходных данных). Если радиус необоснованно мал, в него может попасть очень мало опорных точек, и тогда расчет по ним будет ненадежным. Например, для карт мира в данном исследовании использован радиус в 3 тысячи километров. А для карт Европы – 1 тысяча. Радиус можно задавать как в километрах, так и в числе опорных точек – в этом случае радиус для каждого узла будет своим – программа будет расширять его до тех пор, пока в него не попадет заданное пользователем число опорных точек.

Выбор весовой функции, как говорилось, важнее. При ее малых степенях значения в удаленных точках будут «затирать» локальные флуктуации значений признака: если в одной опорной точке значения значительно ниже, чем в окрестных, и этих окрестных опорных точек много, а степень мала, тогда даже в узле, ближайшем к этой опорной точке с низким значением, может перевесить влияние окружающих точек, и тогда значение в узле будет лишь немного ниже, чем в окружении. При больших степенях, наоборот, карта будет повторять все флуктуации в исходных данных. Поэтому если важен основной тренд карты, и локальные флуктуации мало интересуют, то малые степени полезны. Отметим, что эти локальные флуктуации в каких-то опорных точках могут не соответствовать реальности (например, быть следствием ошибки выборки), и в этом случае «затирание» такой ложной особенности соседними точками полезно. Но если исследователь доверяет значениям в подавляющем большинстве опорных точек и хочет получить карту, отражающую все особенности пространственного распределения признака, следует задавать более высокие степени весовой функции. Основной способ выбрать оптимальную степень – это построить серию карт



Рис. 1.15. Процесс построения карты отдельного признака в GeneGeo.

Видна картографическая основа «Евразия» с нанесенными на нее изученными популяциями, показанными красными опорными точками. В окне слоев выведен слой опорных точек и (неактивированный) регион отсечения, других слоев пока нет. В окне создания сеточного слоя указано имя слоя с исходными данными, выбранный метод интерполяции, число узлов сетки карты и ряд других параметров. Также указано имя файла шкалы интервалов. В окне параметров интерполяции указана степень весовой функции 4 и радиус 900 тысяч километров (900 000 метров).

на одних и тех же данных, но с разными параметрами, проанализировать различия, повторив это по возможности для карт других таких же исходных признаков, и оценить, при каких параметрах построения карты лучше соответствуют целям конкретного картографического исследования. Есть и количественные методы оценки оптимальной степени весовой функции, но они не учитывают всех деталей карт. По опыту работы автора в GeneGeo, малые степени - это 1 и 2, а более высокие - 3 и 4. Большинство карт, представленных в данной книге, построены со степенью 3.

GeneGeo отображает карту по принципу слоев. На картографическую основу (слой по умолчанию) последовательно накладываются слои с опорными точками и сеточные слои (со значениями признаков). Слоям можно задать степень прозрачности и результирующее картографическое изображение будет наложением всех слоев. На практике в окне слоев обычно загружено несколько сеточных слоев, с которыми ведется работа. Но все

они непрозрачны, и поэтому в выводимое на экран изображение входит один сеточный слой (самый верхний в списке) и один или несколько слоев с опорными точками.

На рис. 1.15 приведен реальный пример, как задаются параметры построения карты.

После этого, чтобы получить карту на рис. 1.16, автор только щелкнул мышью по кнопке Start в окне создания сеточного слоя.

Карта на рис. 1.17 отличается только заданием другой весовой функции (на рис. 1.16 степень 4, а здесь степень 1) и другим радиусом (на рис. 1.16 радиус 900 км, а здесь радиус 2 000 км). Сравнивая карты на рис. 1.18 и 1.19, можно наглядно убедиться в справедливости сказанного выше о важности параметров интерполяции и особенностях карт, построенных при малых и больших степенях весовой функции.

На рис. 1.16 показана карта отдельного признака, построенная в результате выполнения процедур, заданных на рис. 1.15.

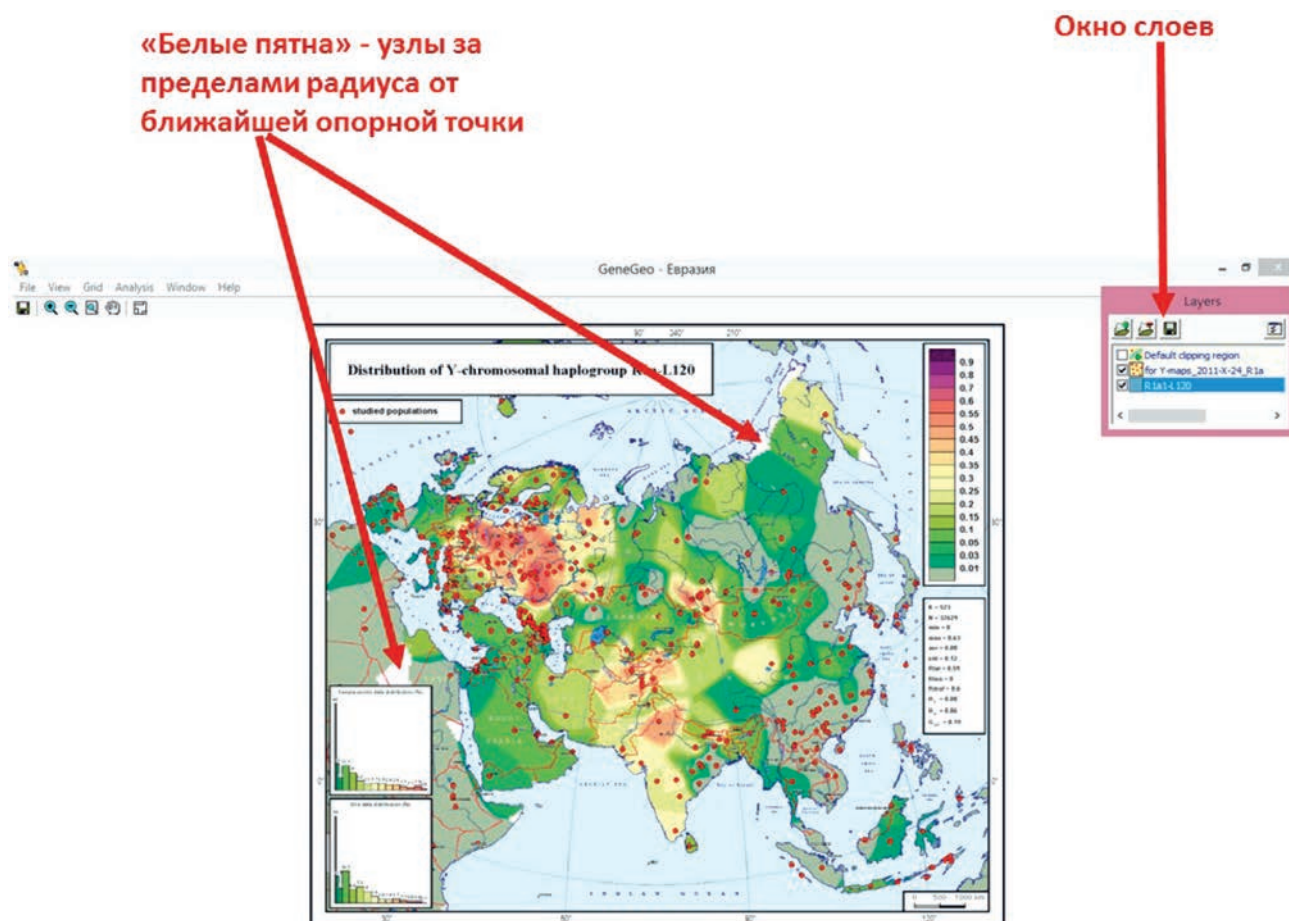


Рис. 1.16. Результат построения карты отдельного признака в GeneGeo.

Карта построена в результате выполнения процедур, заданных в примере на рис. 1.17. Видны «белые пятна» в Африке, на крайнем северо-востоке Сибири и на юге Камчатки – узлы сетки на этих территориях находятся далее, чем 900 км от любой из опорных изученных популяций (показанных красными точками). В окне слоев добавился сеточный слой (слой с интерполированными значениями). Видимая карта представляет собой наложение слоя с опорными точкам поверх сеточного слоя.

Важно отметить, что при картографическом анализе генофондов, как правило, анализируется не один, а много генетических маркеров, и для каждого нужно построить карту. Поскольку параметры картографирования для всех однотипных маркеров в одном исследовании должны быть одинаковыми, в GeneGeo реализован пакетный режим

построения карт: пользователь один раз задает параметры картографирования и указывает файл с исходными данными, содержащий значения разных признаков в разных столбцах csv файла, и программа строит сразу все карты. Опция серийного построения карт в пакетном режиме многократно ускоряет проведение картографического анализа.

ЭТАП 2: АНАЛИЗ ОТДЕЛЬНОЙ КАРТЫ

ЭТАП 2 – АНАЛИЗ ОТДЕЛЬНОЙ КАРТЫ

Разнообразные виды анализа, которые можно провести с картой отдельного признака, можно сгруппировать в четыре типа:

- расчет статистических параметров карты,
- анализ ее закономерностей в плавущем окне
- анализ корреляции с другими картами
- статистические трансформации карт.

Статистические параметры, рассчитываемые по умолчанию, выводятся в статистическом окне легенды карты (рис. 1.16).

АНАЛИЗ В ПЛЫВУЩЕМ ОКНЕ состоит в том, что каждый узел сетки карты поочередно становится центром окна заданного размера. Окно, таким образом, постепенно проплывает по всей карте. И пока оно плывет, рассчитывается какой-либо ста-

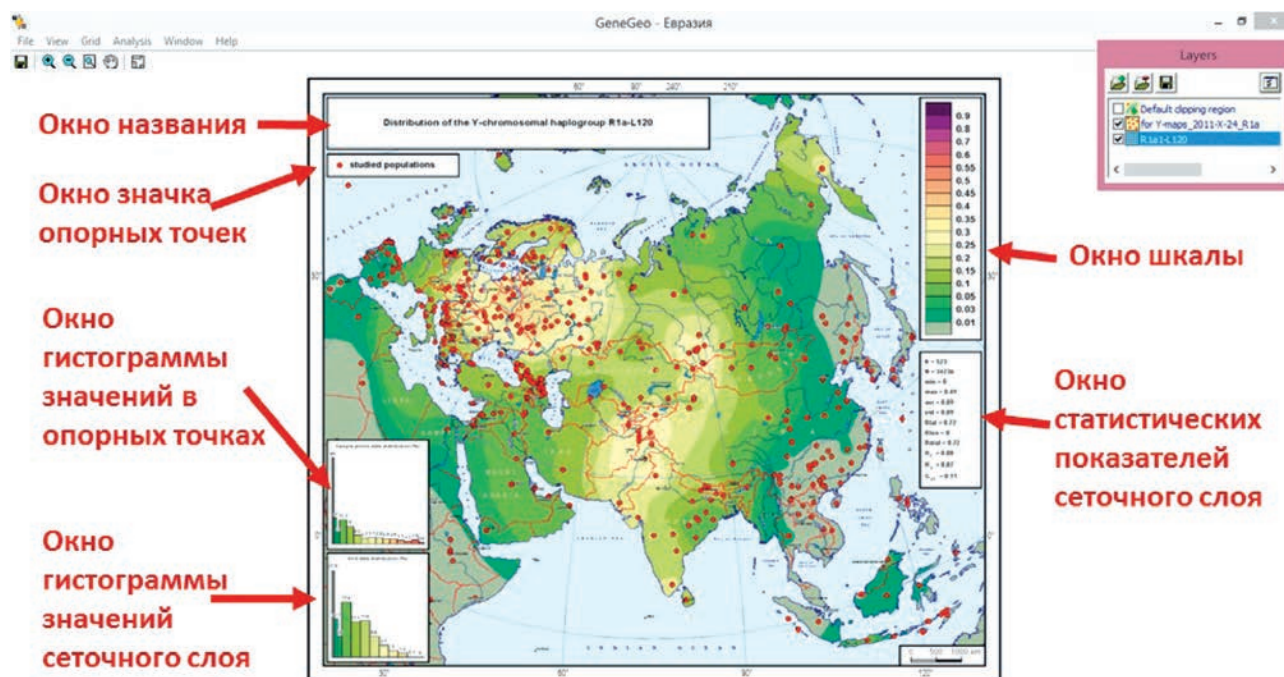


Рис. 1.17. Значение параметров интерполяции: карта, построенная по тем же исходным данным, что и на предыдущем рисунке, но с другими параметрами.

Карта построена при следующих параметрах картографирования: степень весовой функции 1, радиус влияния 2 000 км. В результате увеличения радиуса белые пятна исчезли, распределение признака стало намного более сглаженным, а сравнение гистограмм исходных и картографированных значений показывает, что наиболее высокие значения признака исчезли с карты за счет влияния более далеких, но многочисленных узлов с низкими значениями. На карту выведены 6 окон легенды: 1) Окно названия; 2) Окно значка, которым показаны опорные точки; 3) Окно гистограммы исходных значений – показано распределение значений в исходных точках по интервалам шкалы; 4) Окно гистограммы интерполированных значений – показано распределение значений в узлах сеточного слоя по интервалам шкалы; 5) Окно шкалы – указаны граничные значения интервалов и цвет интервалов; 6) Статистическое окно. В нем указаны число опорных точек, число узлов сетки карты, а также статистические параметры, рассчитанные по значениям во всех узлах сеточного слоя: – минимальное, максимальное и среднее значения, дисперсия; – коэффициенты корреляции с широтой, долготой и коэффициент множественной корреляции с географическими координатами; – показатели общей гетерозиготности, средней ожидаемой внутрипопуляционной гетерозиготности и межпопуляционного разнообразия; при их расчете картографируемый признак рассматривается как аллель диаллельного локуса. Поскольку в данном случае картографируется одна из гаплогрупп Y-хромосомы, интерпретация показателей гетерозиготности затруднена, но показатель межпопуляционного разнообразия весьма информативен, особенно при сравнении с другими гаплогруппами или другими генетическими системами.

статистический показатель по всем узлам, находящимся в текущий момент времени в пределах окна, и это значение присваивается текущему центральному узлу плывущего окна. В результате для каждого узла рассчитан показатель по значению всех его окрестных узлов, попавших в окно заданного размера. То есть создан новый сеточный слой, новая карта, содержащая уже не те значения, которые были рассчитаны первоначально, а значения какого-либо показателя, рассчитанные по узлам, близким к данному. Этот метод в корне отличается от статистического метода анализа генофонда, поскольку в его основе лежит географическое расположение популяций с разными значениями.

Размер плывущего окна задается пользователем, форма его может быть квадратной или круглой, а рассчитываемый показатель в теории лю-

бым, а на практике в GeneGeo предусмотрен широкий спектр алгебраических, статистических, тригонометрических и других математических операторов.

Простейшим способом применения плывущего окна является усреднение всех значений, попавших в окно. Но и этот способ имеет большое значение – он позволяет нивелировать мелкие колебания значений признака между соседними популяциями, вычлняя и делая явным глобальный тренд в их изменчивости.

В плывущем окне можно вычислять и величину межпопуляционного разнообразия, что является примером использования более сложного математического оператора. При этом мы получаем карту, показывающую, на каких территориях межпопуляционное разнообразие велико, а на каких

территориях все популяции генетически более или менее похожи друг на друга. Этот подход лежит в основе выявления и картографирования генетических границ.

АНАЛИЗ КОРРЕЛЯЦИИ С ДРУГИМИ КАРТАМИ можно провести двумя способами.

Первый способ состоит в расчете корреляций между последовательностями значений во всех узлах первой карты и второй карты. (Понятно, что карты должны быть созданы на одной и той же картографической основе и с одинаковым числом узлов сетки). При этом способе географический аспект сведен к минимуму, поскольку анализируются по сути два статистических ряда, а географическое положение членов ряда игнорируется. Результатом этого вида анализа является не новая карта, а просто показатель корреляции между сравниваемыми картами.

Второй способ состоит в использовании плавущего окна – корреляция для каждого узла рассчитывается между значениями в окрестных узлах

на первой карте и в тех же узлах второй карты. Результатом является создание карты корреляций – в каждом узле записаны значения корреляции двух исходных карт в данной географической области. Поразительно, что такие карты часто выявляют географическую неоднородность корреляционной связи. То, что при первом способе расчета предстало, например, как слабая положительная корреляция, на корреляционной карте иногда выглядит как сильная положительная корреляция в одной части ареала, и нулевая или даже отрицательная – в другой.

Для анализа в плавущем окне также предусмотрен пакетный режим – пользователем один раз задаются параметры и указывается список исходных карт, а программа проводит анализ для всех них.

Последний тип анализа отдельных карт – ее статистическая трансформация – состоит в применении одного и того же математического оператора к каждому узлу. Например, можно вычесть значения в узле из единицы и тем самым получить карту альтернативного аллеля (для диаллельных систем).

ЭТАП 3: РАСЧЕТ ОБОБЩЕННЫХ КАРТ

Как корреляционные карты построены по данным двух карт исходных признаков, так обобщенные карты построены по данным множества исходных карт. Простейшим примером обобщенных карт являются суммарные карты. Например, можно суммировать карты гаплогрупп **C**, **D** и **O** Y-хромосомы (эти гаплогруппы считаются восточноевразийскими) и получить карту суммарной частоты восточноевразийских гаплогрупп. При построении суммарных карт значение в данном узле первой карты суммируется со значением в том же узле на второй карте, со значением того же узла на третьей карте и так далее для всех суммируемых карт. Эта процедура проводится для каждого узла. Но суммарные карты слишком просты и не слишком интересны. Более широко в геногеографии применяются обобщенные карты других типов: карты генетических расстояний; карты генетических границ (межпопуляционного разнообразия); карты главных компонент;

карты внутривнутрипопуляционного разнообразия (гетерозиготности, гаплотипического разнообразия). Создание всех этих типов карт предусмотрено в GeneGeo.

Например, специальный модуль, вызываемый из меню Analysis, посвящен построению карт генетических расстояний. При этом необходимо задать файл с частотами аллелей в реперной популяции (от которой будут рассчитываться генетические расстояния) и набор карт распространения этих аллелей по изучаемой территории. Программа рассчитывает расстояния от реперной популяции для каждого узла сетки карты и поместит эти значения в соответствующие узлы. Тем самым каждый узел, то есть каждая точка географического пространства будет заявлять о себе, насколько популяция в этой точке генетически сходна с реперной популяцией, причем в среднем по всему набору анализируемых аллелей.

СОЗДАНИЕ КАРТОГРАФИЧЕСКОГО ОБРАЗА

Важным направлением работы по разработке GeneGeo является постоянное совершенствование графического представления создаваемых карт. На рисунке 1.18 приведено сравнение одной и той же карты, построенной в старом пакете GGMAG (разработана под руководством Е.В. Балановской и Ю.Г. Рычкова в 1990-х годах), и той же карты, построенной в пакете GeneGeo. Легко заметить графические преимущества нового пакета: современный

дизайн, возможность переключения слоев карты, гибкая цветовая шкала, отображение рельефа отмывкой, интуитивный интерфейс и т.д.

Как из одних и тех же исходных данных при разных параметрах интерполяции получаются разные сеточные слои (карты в виде матриц), так и из одного и того же сеточного слоя при разных параметрах визуализации получаются разные картографические изображения (карты в обычном понимании).

ШКАЛА. Важнейшим параметром визуализации является шкала интервалов. Ведь узлы сеточного слоя отображаются на карте точками разного цвета в зависимости от величины значения в узле. Шкала как раз и задает, какой интервал значений отображается каким цветом. В книге [Балановская, Балановский, 2007] мы уже описывали, какими должны, и главное, какими не должны быть шкалы, чтобы полученный картографический образ передавал объективную информацию, а не попытки составителя карты навязать читателю свои априорные представления о географии признака. Чтобы не повторяться, укажем лишь, что составителю карты можно рекомендовать пользоваться равномерными шкалами, а читателю карты – доверять лишь тем чертам карт, которые основаны на различиях хотя бы в два интервала шкалы, и не считать действительно различающимися территории, значения в которых попадают в соседние интервалы. Пример важности шкалы можно видеть в нашей недавней работе [Balanovsky et al., 2015] где мы сначала привели карту в неравномерной шкале, подчеркивающей то, что для нас было всего важнее (повышенные частоты приурочены к древнему ареалу ираноязычных популяций), но потом для объективности привели ту же карту и в нашей стандартной шкале, которая первоначально создает несколько иное впечатление, и лишь при внимательном рассмотрении в ней проявляется указанная выше закономерность, хотя и видно, что она не является единственно возможной интерпретацией карты.

УНИВЕРСАЛЬНАЯ ШКАЛА. Большинство карт, представленных в этой книге, построены в «универсальной» шкале, разработанной автором для картографирования гаплогрупп мтДНК и Y-хромосомы. Эта шкала (см., например, рис. 16 и 17) состоит из трех равномерных отрезков: от 1% до 5% интервалы идут с шагом 2%, от 5% до 60% интервалы идут с шагом 5%, а от 60% до 100% интервалы идут с шагом 10%. Это позволяет сохранить традиционный шаг в 5% на основной части возможного диапазона варьирования, при этом выделить более подробно закономерности в части низких частот (большинство миграций приносят не слишком большой процент новой гаплогруппы), но менее подробно показать различия между сверхвысокими частотами (потому что редко когда имеет значение, составляет частота в какой-то популяции 82% или 88% – и то и другое значение слишком велики и оба являются редкостью). Частота ниже 1% в универсальной шкале отображается нейтральным серым цветом, выбивающимся из разноцветной (от темно-зеленых через зеленые и желтые к красноватым и бордовым тонам) цветовой гаммы остальной части шкалы. Это сделано потому, что гаплогруппы мтДНК и Y-хромосомы

географически специфичны. Большинство из них распространены только в каком-то одном регионе мира, а на большей части планеты отсутствуют или встречаются лишь спорадически (частота ниже 1%). Поэтому на картах в этой шкале почти весь мир предстает в серых тонах отсутствия гаплогруппы. И на этом фоне ярко проявляется не серым, а цветным ареалом, тот регион, где гаплогруппа распространена, пусть и с невысокими частотами.

СОСТАВЛЯЮЩИЕ КАРТОГРАФИЧЕСКОГО ОБРАЗА. Кроме выбора шкалы, при создании картографического изображения также можно:

- выбрать выводимые на карту географические объекты из числа предусмотренных картографической основой;
- задать цвет, размер и форму значков изученных популяций;
- пользуясь файлами такого же формата, как файл опорных точек, вывести любые другие значки-обозначения;
- выбрать язык надписей карты (английский или русский);
- задать частоту следования линий градусной сетки;
- выбрать отображения интервалов изолиниями, заливкой, или и тем, и другим;
- задать шрифты и ввести сами тексты надписей в заголовке карты и в шкале интервалов;
- настроить отображение географического рельефа (рельеф отображается современным способом аналитической отмывки – как бы тени, падающие от гор – при этом можно задавать угол падения света и интенсивность отображения рельефа);
- выбрать выводимые окна легенды и задать места их размещения;
- и еще многое другое.

В целом, GeneGeo является прекрасным инструментом не только для построения геногеографических карт и не только для их разностороннего анализа, но и для их отображения в виде выразительных, наглядных, радостных образов. А еще программа очень удобна в работе, множеством способов экономит время и безотказна – настоящий друг геногеографа.

КАРТОГРАФИЧЕСКАЯ БАЗА ДАННЫХ. В связи с тем, что созданные карты исчисляются многими сотнями, и с тем интересом, который они (судя по запросам от специалистов) представляют как для коллег-генетиков, так и для специалистов смежных гуманитарных дисциплин, под руководством автора создана общедоступная онлайн-версия картографической базы данных, а точнее, базы карт.

Эта база решает как задачу публикацию созданных карт на сайте, так и задачу легкого поиска, просмотра и скачивания необходимой карты. База данных доступна на разрабатываемом нашим коллек-

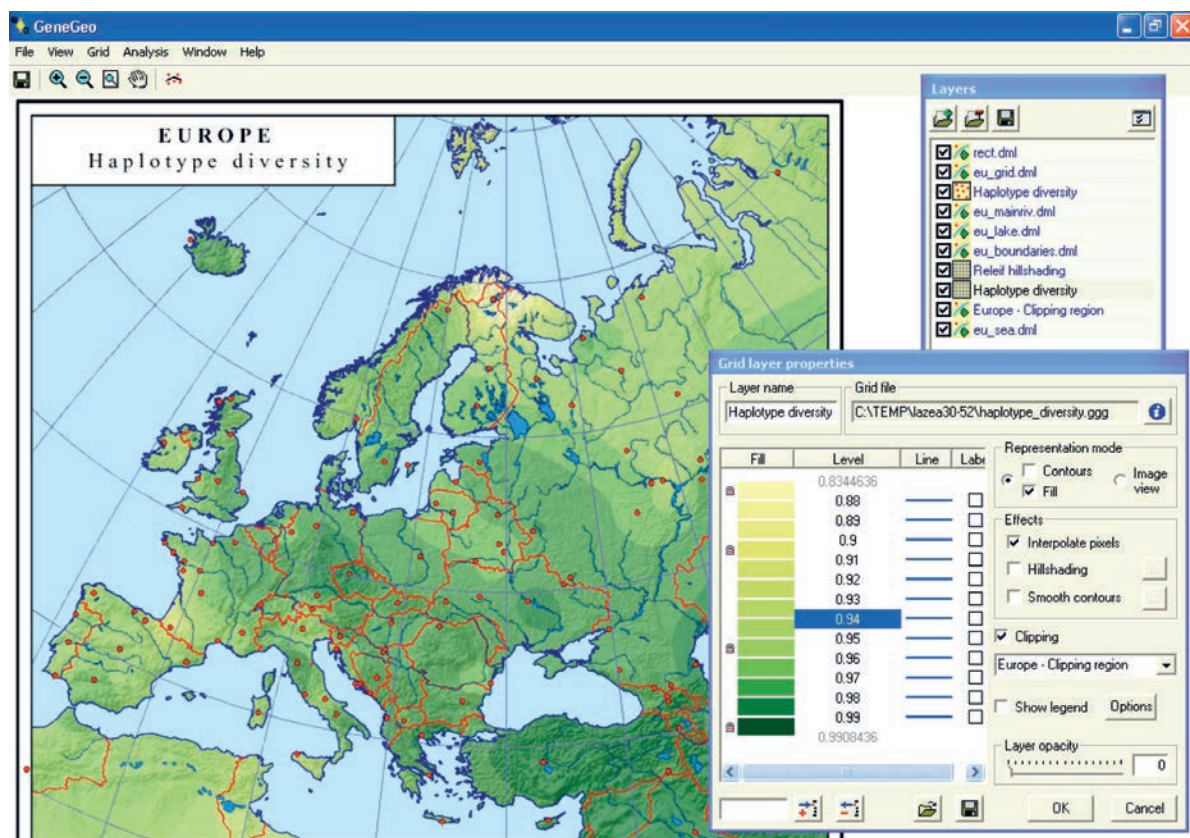
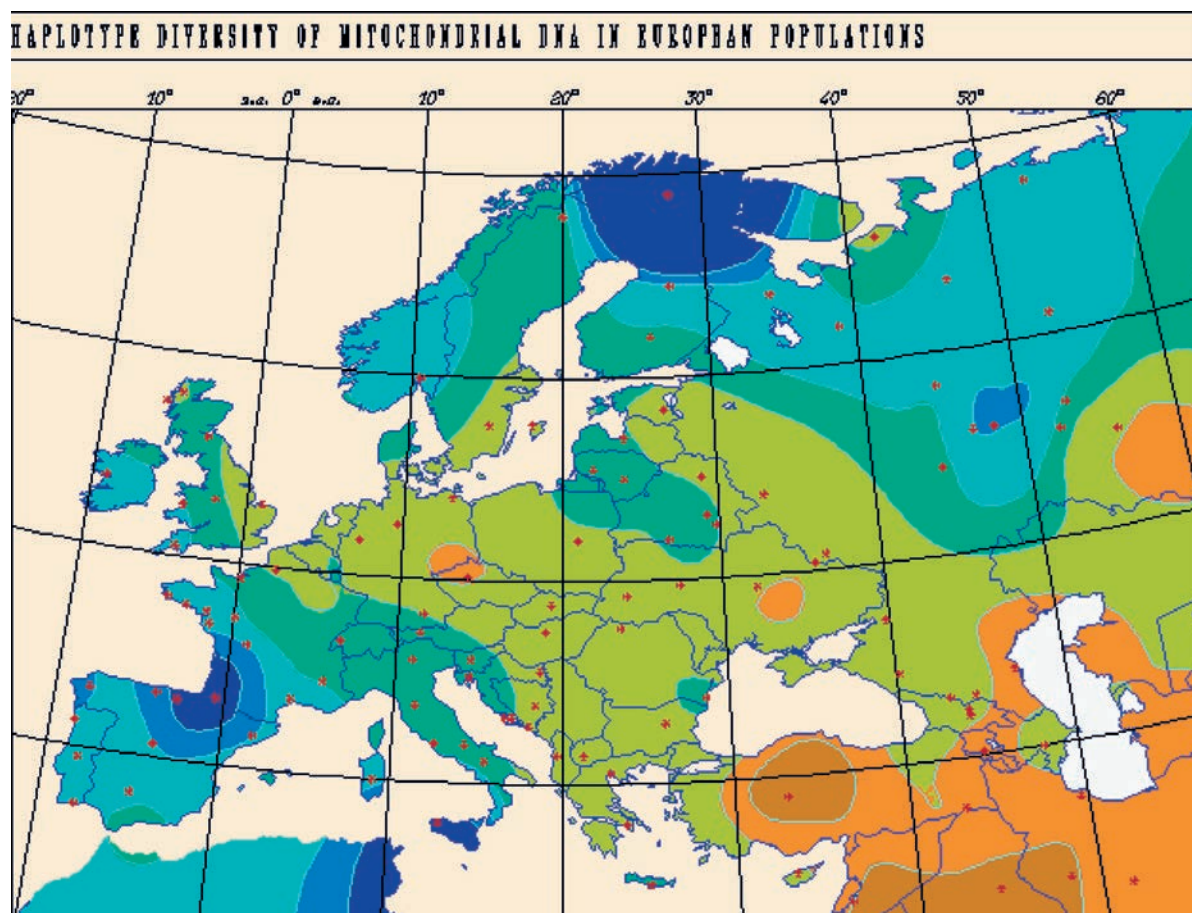


Рис. 1.18. Сравнение графических возможностей отечественного пакета GGMAG (вверху) и пакета GeneGeo, разработанного под руководством автора (внизу).

На обеих картах отображена одна и та же исходная информация.

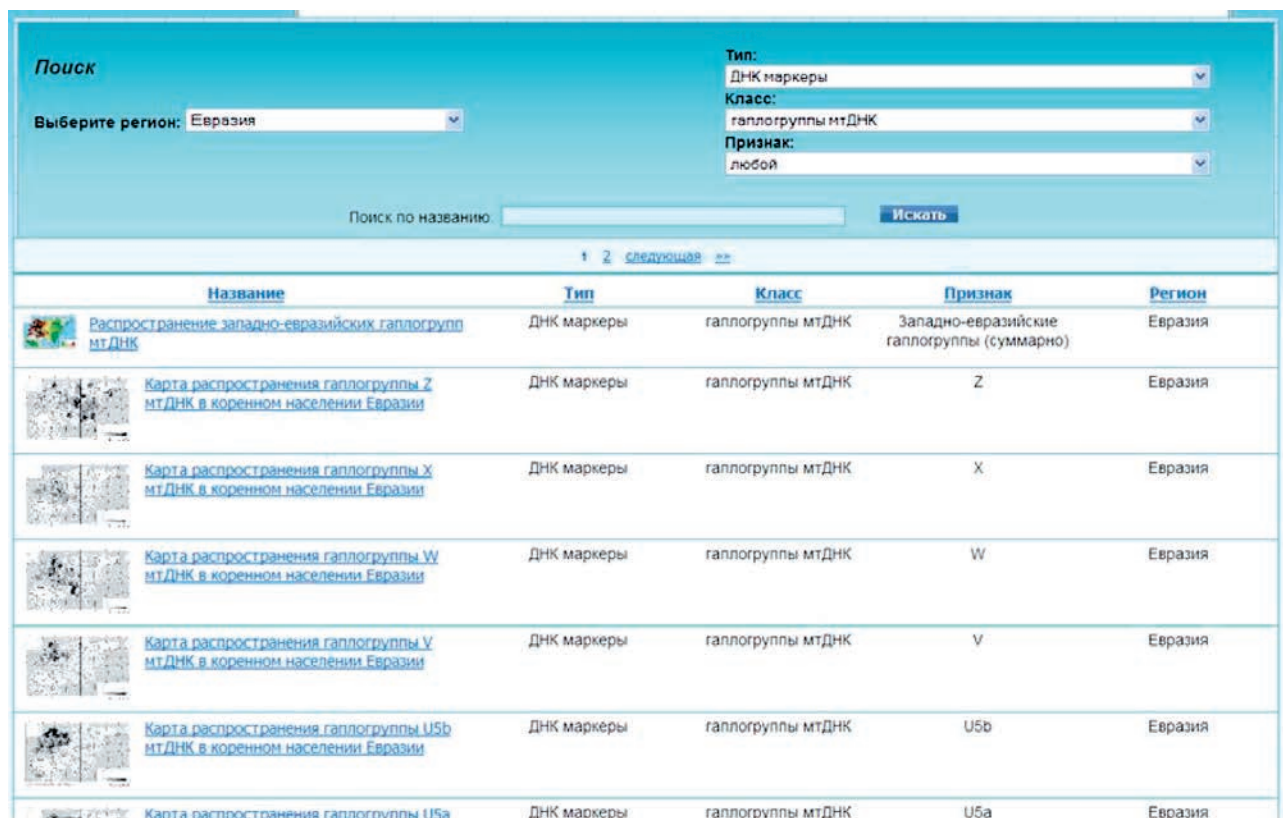


Рис. 1.19. Интерфейс системы фильтрации картографической базы данных (www.genofond.ru).

тивом сайта www.genofond.ru в разделе «Атласы». Интерфейс базы данных позволяет осуществлять поиск нужных карт одновременно в двух направлениях – по типу картографируемого признака и по региону, для которого создана карта.

Поиск (фильтрация) по признаку позволяет последовательно выбрать тип признака (например, ДНК-маркеры), класс признака (например, митохондриальная ДНК) и собственно признак (например, конкретные гаплогруппы мтДНК). Пользователь может остановиться на любом уровне общности признаков, не указывая подробно признак, возмо-

жен одновременный поиск по группе признаков.

Поиск (фильтрация) по региону позволяет пользователю выбрать, карту какого региона требуется найти (весь мир, Евразия, Европа, степная полоса Евразии и т.д.).

По завершении поиска пользователю выводится список найденных карт с их уменьшенными изображениями. При щелчке мышью по любой из карт отображается ее увеличенное изображение с сопровождающей информацией по карте. Пример работы системы фильтрации приведен на рисунке 1.19.

МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА

РАСЧЕТ ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ

По таблицам частот гаплогрупп рассчитывались генетические расстояния по Нею [Nei, 1975] с использованием разработанной нашим коллективом программы DJgenetic [Balanovsky et al., 2008]. Полученные матрицы расстояний использовались четырьмя способами:

1) визуализировались методами многомерного шкалирования с помощью программы Statistica 6.0 [StatSoft. Inc, 2001];

2) в ряде случаев для целей визуализации матриц генетических расстояний, в дополнение к мно-

гомерному шкалированию, использовался и кластерный анализ. Дендрограммы строили в программе Statistica 6.0 [StatSoft. Inc, 2001] с применением метода Уорда (Ward's method);

3) по матрицам генетических расстояний рассчитывались средние попарные расстояния в группе популяций для характеристики межпопуляционного разнообразия внутри этой группы;

4) В тех случаях, когда расстояния между конкретными популяциями представляли особый интерес, рассматривались и эти значения в отдельных ячейках большой матрицы «все популяции от всех».

АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТ ДЛЯ ЧАСТОТ ГАПЛОГРУПП

Частоты гаплогрупп использовались не только для расчета генетических расстояний, но и для расчета главных компонент, AMOVA и ряда других методов. Анализ главных компонент проводился в той же программе Statistica 6.0, а для совмещения на одном графике расположения и популяций, и гаплогрупп (по факторным нагрузкам на главные компоненты) использовали программу SPSS.

МЕТОД AMOVA ДЛЯ ОЦЕНКИ ГЕТЕРОГЕННОСТИ

Параллельно с этими методами, нацеленными на визуализацию данных, проводился анализ иерархической изменчивости (AMOVA) с целью выявить наиболее гетерогенные и наиболее гомогенные группы популяций. Исходным материалом служили частоты гаплогрупп, и для анализа использовали программу Arlequin 3.11 [Schneider и др., 2000]. Расчет межпопуляционного разнообразия проводили на разных иерархических уровнях изученных популяций. Определялась генетическая гетерогенность как отдельных этносов (различия по частотам гаплогрупп локальных популяций в пределах этносов), так и лингвистических групп (различия по частотам гаплогрупп популяций, относящихся к различным ветвям лингвистических групп).

G_{ST} И ГЕНЕТИЧЕСКИЕ РАССТОЯНИЯ ДЛЯ ОЦЕНКИ ГЕТЕРОГЕННОСТИ

Гетерогенность этносов и их групп на разных иерархических уровнях также оценивалась и двумя другими методами – как средние попарные генетические расстояния между популяциями (упомянуты выше как третий способ использования матриц расстояний) и через расчет G_{ST} в оригинальной программе MEGERA (мера генетического разнообразия).

МЕТОД AMOVA ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ КЛАССИФИКАЦИЙ

Метод AMOVA также использовался для сравнительного анализа эффективности генетической, географической и лингвистической классификаций популяций в определении сходства генофондов.

ТЕСТ МАНТЕЛЯ ДЛЯ ОЦЕНКИ ЭФФЕКТИВНОСТИ КЛАССИФИКАЦИЙ

В программе Arlequin 3.11 выполнялся тест Мантеля для оценки общих и частных коэффициентов корреляций между матрицами генетических, географических и лингвистических расстояний. Матрицы лингвистических расстояний, полученные по стословному списку М. Сводеша [Swadesh, 1955] с исключением заимствованных слов по методу С.А. Старостина [Старостин, 1989], были лю-

безно предоставлены член-корр. РАН А.В. Дыбо и проф. О.А. Мудраком.

РАСЧЕТ ГЕОГРАФИЧЕСКИХ РАССТОЯНИЙ

Необходимые для последних двух видов анализа географические расстояния рассчитывались по географическим координатам популяций (табл. 1, 2) в разработанной нашим коллективом программе DistGeo (www.genofond.ru).

АНАЛИЗ ГЕНЕТИЧЕСКИХ ГРАНИЦ (WOMBLING)

Выявление генетических границ на географической карте (зоны резких изменений частот гаплогрупп) проводили как статистическими, так и картографическими методами. Статистический способ реализован в программе Barrier 2.2 [Manni and Guerard, 2004], а картографический описан в соответствующем разделе.

АНАЛИЗ ШИРОКОГЕНОМНЫХ ДАННЫХ

Исследования обширных панелей, включающих сотни тысяч SNP-маркеров – полногеномные, а точнее, широкогеномные (genome-wide) маркеры, проводятся как методами анализа частот аллелей, так и методами анализа гаплотипов. Эти маркеры в подавляющем большинстве аутомсомные (а немногие присутствующие X-хромосомные и Y-хромосомные маркеры исключаются из анализа) и, хотя они являются рекомбинирующими, но на близких расстояниях на хромосоме гаплотипы (сочетания SNP-маркеров) не успевают разбиваться рекомбинацией за исторические промежутки времени, а число разбиений может даже использоваться для оценки времени. В связи с очень большим числом маркеров тот же анализ главных компонент проводится в специально разработанных программах, в нашем случае PLINK [Purcell et al., 2007]. Из методов анализа гаплотипов в нашем исследовании использован IBD анализ – выявление в разных популяциях гаплотипов, идентичных по происхождению, и подсчет их доли как показателя родства популяций. Эти методы подробнее описаны в тех разделах, где они применяются.

РАСЧЕТ ГАПЛОТИПИЧЕСКОГО РАЗНООБРАЗИЯ мтДНК

Простейшим показателем, рассчитываемым по данным об отдельных гаплотипах мтДНК и Y-хромосомы, является их разнообразие. Гаплотипическое разнообразие рассчитывали как дополнение до единицы суммы квадратов частот всех гаплотипов, встреченных в популяции, такой расчет легко провести даже в Excel. Например, если в популяции обнаружен только один гаплотип (все обследованные члены популяции одинаковы по мтДНК) – по-

казатель выявленного гаплотипического разнообразия равен нулю; если все обследованные различаются по мтДНК (ни один из гаплотипов не встречен дважды) – показатель гаплотипического разнообразия составляет единицу. Можно видеть, что этот показатель, рассчитываемый для гаплоидных систем, идентичен гетерозиготности, рассчитываемой для обычных диплоидных маркеров. Ведь квадрат частоты гаплотипа (аллеля) – это частота гомозигот, сумма таких квадратов для всех аллелей – гомозиготность (доля гомозигот), а величина, дополняющая ее до единицы – и есть доля ге-

терозигот, то есть гетерозиготность популяции. Ту же мысль можно выразить и иначе: гаплотипическое разнообразие оценивает вероятность того, что два случайно выбранных из популяции гаплотипа окажутся разными. Но ведь и гетерозиготность оценивает то же самое: какова вероятность, что два аллеля из популяции, объединившиеся в одной зиготе, будут разными. Подчеркнем, что при расчете гаплотипического разнообразия не учитывается степень родства гаплотипов друг с другом, зато все прочие методы их анализа основаны как раз на этом родстве.

1.5. ДИСКУССИИ О ДАТИРОВКАХ

Всеобщая грамотность, открывшая людям доступ к научным текстам, создает иллюзии, что чтение равнозначно пониманию.

В.А. Шнирельман

В возможности датировок миграций и других событий истории популяций заключается важное преимущество нерекombинирующих («однородительских») генетических систем перед аутосомными (рекомбинирующими) генетическими системами. В тоже время получение надежных датировок является большой проблемой, ставшей предметом многих числа научных, околонучных и даже лженаучных дискуссий. Поэтому на этом вопросе стоит остановиться подробнее.

Предположим, мы обнаружили монофилетичный кластер гаплотипов. Раз он «монофилетичный», то по определению представляет одну линию происхождения, значит все гаплотипы кластера произошли от одного исходного гаплотипа. И предположим, что мы знаем, от какого именно гаплотипа. С ходом времени в результате мутаций в пределах кластера будут возникать все новые дочерние гаплотипы (отличающиеся на один мутационный шаг), а от них, в свою очередь, будут возникать внучатые гаплотипы (отличающиеся

на два шага) и так далее. И чем больше пройдет времени, чем больше новых гаплотипов возникнет, тем на большее число мутационных шагов они будут отличаться от исходного, то есть тем более разнообразные гаплотипы мы выявим в этом кластере. Следовательно, верно и обратное: чем большее разнообразие гаплотипов мы видим в пределах кластера, тем дольше он существует. То есть по разнообразию гаплотипов в пределах кластера мы можем судить о возрасте этого кластера. Эта логика проиллюстрирована на рис. 1.20. По сути, это и есть принцип молекулярных часов, где время отсчитывается происходящими мутациями. Если скорость мутаций постоянна, то часы будут работать. Обо всем этом спора нет (оговорки, которые мы делали по ходу, говоря «предположим» и «если», разобраны в конце этого подраздела). Но споры начинаются, когда приходится решать, как именно подсчитывать разнообразие гаплотипов и как именно переводить его в годы.

МЕТОДЫ ФИЛОГЕНЕТИЧЕСКОГО АНАЛИЗА

ПОСТРОЕНИЕ ФИЛОГЕНЕТИЧЕСКИХ СЕТЕЙ

Родство рассматриваемых гаплотипов принято отображать в виде филогенетического дерева. А когда однозначное дерево реконструировать не удается, и остаются равновероятные возможности, что превращение одного гаплотипа в другой шло тем или иным путем, то это отображается как слияние нескольких ветвей, то есть дерево превращается в сеть. Методология построения и анализа филогенетических сетей была хорошо разработана в популяционной генетике в 90е годы прошлого века. Филогенетические сети были построены в програм-

ме Network 4.1.1.2 (Fluxus Technology-Ltd.) (www.fluxus-engineering.com) на основе алгоритма reduced median с порогом редукции, равным 1. Полученные графики затем редактировались и снабжались легендами в программе Network Publisher (Fluxus Engineering, Clare, U.K.). Построение филогенетических сетей мы проводили и по гаплотипам мтДНК, и по STR гаплотипам Y-хромосомы.

ВЫЯВЛЕНИЕ КЛАСТЕРОВ ГАПЛОТИПОВ

На большинстве филогенетических сетей обнаруживается ряд более или менее компактных кла-

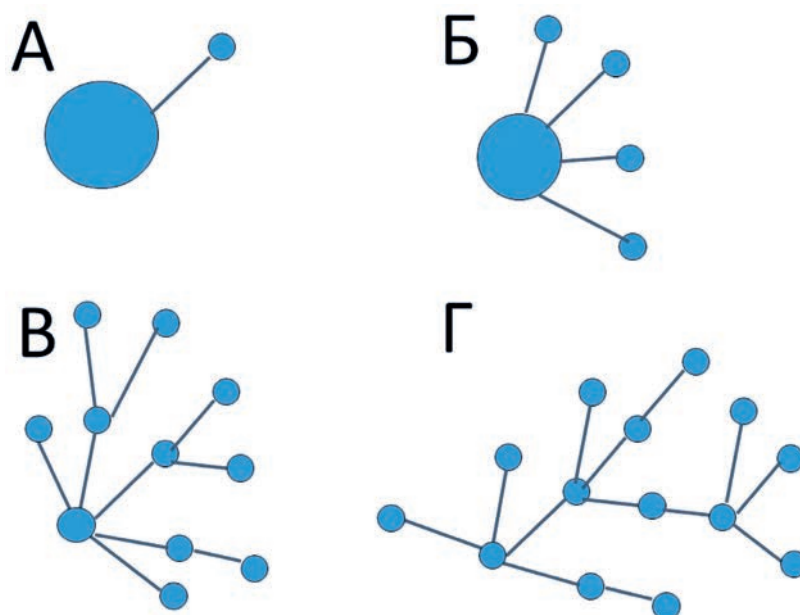


Рис. 1.20. Молекулярные часы: нарастание дерева гаплотипов.

Показан постепенный рост дерева гаплотипов: возникновение новых гаплотипов в результате мутаций:

А. Один исходный гаплотип (встречен у многих образцов, поэтому большой кружок) и один производный (только у одного образца, маленький кружок). Б. Возник уже целый ряд производных гаплотипов. В. Из «дочерних» гаплотипов начали возникать уже «внучатые», а частота исходного гаплотипа уже почти не отличается от любого из производных. Г. Дерево приобрело сложную структуру, и уже нельзя однозначно понять, какой именно гаплотип был исходным.

стеров гаплотипов, выявление и датировка времени их происхождения стало важным аспектом многих современных исследований Y-хромосомы [Zhivotovsky et al., 2004; Rootsi et al., 2007; Derenko et al., 2007; Haber et al., 2010 и многие другие работы]. При этом нет алгоритма, автоматически выявляющего кластеры гаплотипов на сети, и определение числа и границ кластеров оставляется на усмотрение исследователя. Таким образом, проведение этой процедуры неизбежно оказывается до некоторой степени произвольным. Поэтому, чтобы минимизировать эту произвольность, при выявлении кластеров мы применяли следующие формализованные правила.

1) Поскольку большинство филогенетических сетей имеют четко выделяющуюся центральную зону (вероятный корень, от которого происходит гаплогруппа), мы считали кластерами только те группы гаплотипов, которые связаны с этим корнем через один и тот же узловой гаплотип (иными

словами, кластерами считались только строго монофилетические ветви на реконструированной сети).

2) Этот узловой гаплотип рассматривался в качестве предкового гаплотипа-основателя (founder) для своего кластера. Выбор гаплотипа-основателя важен при расчете возраста с использованием показателя s .

3) Рассматривались только кластеры, содержащие 10 или более образцов, чтобы избежать ошибок в расчете возраста из-за малых объемов выборок.

4) В качестве популяционно-специфичных кластеров рассматривались только те группы гаплотипов, специфичность которых для данной популяции или группы популяций была выше 80%. Это означает, что в популяционно-специфичном кластере суммарная доля образцов из всех других популяций не может составлять более 20%.

ДАТИРОВКИ ПО ГАПЛОТИПАМ

РАСЧЕТ ИЛИ МОДЕЛИРОВАНИЕ?

Все предложенные методы можно поделить на методы расчета и методы моделирования.

Из методов моделирования в нашем исследовании использованы только BATWING (как один из способов датировки кластеров на Кавказе, раз-

дел 7.3) и BEAST (как один из методов датирования ветвей филогенетического дерева, полученного по полным сиквенсам Y-хромосомы, раздел 7.4). BATWING кратко описан ниже в этом разделе, параметры применения BEAST – в разделе 7.4. Поскольку методы моделирования используются реже и бурных дискуссий не вызывают, в этом разделе

мы будем говорить не о них, а только о методах расчета.

ТРИ СОСТАВНЫХ ЧАСТИ РАСЧЕТА ВОЗРАСТА

Расчет возраста кластера очень прост: надо накопленное разнообразие поделить на скорость возникновения мутаций и умножить на длину поколения. Например, если разнообразие у нас оценивается в 40 условных единиц, а скорость мутаций 2 единицы за поколение, то понятно, что это разнообразие накопилось за 20 поколений. А если новые поколения людей нарождаются, допустим, каждые 25 лет, то возраст кластера будет 500 лет: $(40/2)*25=500$. Но такая теоретическая простота наталкивается на практические сложности: как количественно выразить разнообразие, какова скорость мутирования, какова длина поколения?

С длиной поколения все относительно понятно – ее оценки определяются в демографических исследованиях популяций, а оценки, применяемые для анализа Y-хромосомы, варьируют не слишком сильно, от 25 до 33 лет. С методами чуть сложнее: в научной литературе наиболее широко используются два метода – показатели s и ASD, а в генетической генеалогии иногда используется и метод расчета по доле исходного гаплотипа. Хуже всего

со скоростью мутирования: для STR гаплотипов Y-хромосомы (на которых в основном и рассчитывался возраст гаплогрупп до начала массового полного секвенирования Y-хромосомы в 2013 году) были предложены две скорости: «эволюционная» и «генеалогическая», и они различались в три раза! Рассмотрим эти три составные части последовательно, начав с методов.

МЕТОД ДАТИРОВКИ С ПОМОЩЬЮ ПОКАЗАТЕЛЯ s .

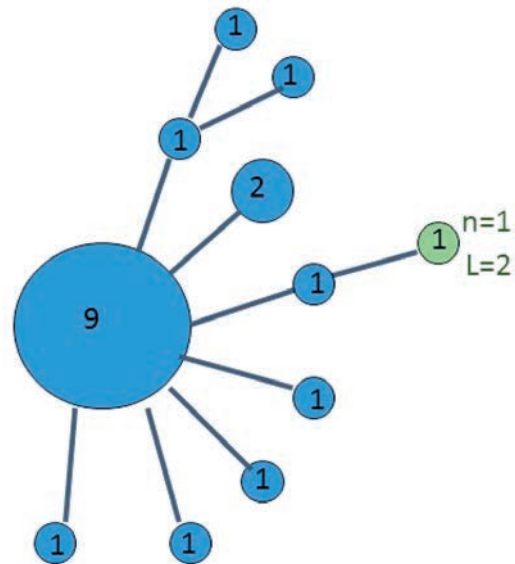
Этот классический метод [Forster et al., 1996] основан на подсчете среднего числа мутаций, накопившихся в пределах кластера (это число и обозначается s). Для этого метода необходимо иметь филогенетическое дерево, показывающее происхождение всех гаплотипов кластера друг от друга. Также необходимо знать, какой из гаплотипов является исходным (гаплотип-основатель – founder haplotype), т.е. первый гаплотип, из которого произошли все остальные. Расчет прост: поочередно рассматривается каждый гаплотип, для гаплотипа определяется число мутационных шагов, отделяющих его от исходного гаплотипа-основателя, и если такой гаплотип встречен более чем у одного образца, то это число умножается на число образцов. Когда такой расчет проведен для каждого гаплотипа, полученные величины суммируются и де-

ρ – это среднее число мутаций от исходного гаплотипа до всех встреченных гаплотипов

N – число образцов с данным гаплотипом

L – длина ветви, т.е. число мутационных шагов от данного гаплотипа до исходного

$$\rho = (\sum nL) / N$$



$$\rho = (6*1 + 1*2 + 3*2 + 9*0) / 20 = 0.7$$

Рис. 1.21. Расчет возраста кластера гаплотипов с помощью показателя s .

Показано дерево родства гаплотипов: один исходный (большой кружок, встречен у 9 образцов) и 10 производных гаплотипов, из них девять гаплотипов встречены только у одного образца каждый, а десятый гаплотип встречен у двух образцов. Для гаплотипа, выделенного зеленым, подписаны показатели n (встречен у одного образца) и L (отстоит от исходного гаплотипа на два мутационных шага).

Проведен расчет показателя s – суммировано число образцов в гаплотипах, умноженное на число шагов от данного гаплотипа до исходного: шесть гаплотипов отстоят на один шаг и представлены одним образцом каждый ($6*1$), один гаплотип отстоит на один шаг и представлен двумя образцами ($1*2$), три гаплотипа представлены одним образцом и отстоят на два шага ($3*2$), исходный гаплотип представлен 9 образцами и отстоит сам от себя на ноль шагов. Величина s в данном примере составляет 0,7.

лятся на общее число образцов. Можно видеть, что по сути подсчитывается число мутационных шагов, на которое изученные образцы в среднем отстоят от исходного гаплотипа. Пример применения этого метода показан на рис. 1.21. Для простых филогенетических деревьев расчет легко проводится вручную, но он также автоматизирован в программе Network. В работе [Saillard et al., 2000] описана возможность расчета статистической ошибки показателя s , а возможные погрешности обсуждаются в [Cox, 2008].

ДАТИРОВКИ С ПОМОЩЬЮ ПОКАЗАТЕЛЯ ASD

Для использования этого второго метода не требуется знать ни гаплотип-основатель, ни схему возникновения из него остальных гаплотипов. Метод состоит в определении средних квадратичных различий (average squared difference – ASD) между STR-гаплотипами [Sengupta et al., 2006]. То есть вновь как бы рассчитывается среднее расстояние от исходного гаплотипа до каждого гаплотипа в выборке, но за исходный гаплотип по каждому STR-маркеру берется не предполагаемый гаплотип-основатель, а принимается средневзвешенное значение этого STR-маркера во всех изученных образцах. Накопленное разнообразие оценивается также иначе – по дисперсии значений в разных образцах вокруг этого среднего. Проводится расчет отдельно для каждого STR-маркера, а затем результаты по всем маркерам усредняются. Этот расчет несложно провести вручную, а удобнее всего – просто в Excel.

ДАТИРОВКИ ПО ДОЛЕ ИСХОДНОГО ГАПЛОТИПА

Этот метод требует только знания того, какой гаплотип является исходным. Определяется его частота, и из нее рассчитывается прошедшее время: чем больше прошло времени, тем меньше становится эта частота, поскольку исходный гаплотип постепенно мутирует в производные. Убывание доли исходного гаплотипа описывается логарифмической формулой, используемой во многих областях науки: например, для характеристики радиоактивного распада (доля нераспавшихся ядер убывает со временем) или для характеристики словарного состава (доля не замещившихся слов из списка Сводеша убывает со временем по тому же закону). Этот метод используется только в генетической генеалогии и в ДНК-генеалогии и не был опубликован в признанных научных журналах. Поэтому я затрудняюсь привести точную ссылку – но среди пользователей метода общепринято считать, что он был разработан Дмитрием Адамовым и Анатолием Клесовым.

СРАВНЕНИЕ МЕТОДОВ РАСЧЕТА

Ранжируя методы от простого к сложному, начнем с доли исходного гаплотипа.

Его достоинством является то, что дерево филогенетических отношений гаплотипов – которое на практике нелегко реконструировать надежно – знать не требуется. Но тем же достоинством обладает и метод ASD. Недостатком – и весьма существенным – метода доли исходного гаплотипа является то, что он применим только для самых простых филогенетических схем, лишь для самых начальных этапов истории новой гаплогруппы. Например, в примерах на рис. 1.20 этот метод был бы применим только в случаях А и Б, а уже в случае В погрешность при его применении становилась бы огромной: ведь стоит в выборке случайно оказаться не двум, а трем или одному образцам с исходным гаплотипом, и его доля изменится в полтора-два раза, а значит, кардинально изменится и оценка возраста. От этого недостатка свободны и метод расчета ASD, и метод расчета s .

О двух достоинствах метода ASD уже сказано выше. Недостатком его является чрезмерная упрощенность модели: ведь «среднестатистический» гаплотип вовсе не обязательно является предковым, и, главное, усреднение значений по всем локусам как бы игнорирует, что в действительности они не являются независимыми и существуют лишь в полном сцеплении друг с другом в виде конкретных гаплотипов.

Поэтому наиболее обоснованным является показатель s . Но и он не свободен от недостатков. Ведь для его использования требуется – не всегда известное – древо родства гаплотипов.

Резюмируем: когда древо родства гаплотипов реконструировано надежно, оптимален метод s , в противном случае целесообразно применить ASD, а расчет по доле исходного гаплотипа не оптимален ни в каком случае.

СКОРОСТЬ МУТИРОВАНИЯ

Для определения скорости мутирования любых участков генома возможны два подхода: прямой подсчет и калибровка.

Первый подход – прямой подсчет – состоит в сравнении генотипов родителей и их потомства. Хотя мутации случаются редко, но при больших выборках можно обнаружить достаточное их количество. Частота встречаемости мутантных аллелей в поколении потомков по сравнению с поколением родителей и будет скоростью возникновения мутаций. Такие исследования для STR-маркеров Y-хромосомы были проведены неоднократно – в основном благодаря тому, что STR-маркеры Y-хромосомы часто генотипируются при определении отцовства. В результате таких определений накапливаются многие сотни случаев, когда отцовство несомненно подтверждается (в том числе по другим маркерам), а значит, имеются генотипы отцов и сыновей, и можно определить, как часто в них

происходили мутации. Таких исследований было проведено немало [Gusmao et al., 2005; Sánchez-Diz et al., 2008; Ge et al., 2009], каждый раз на все больших объемах выборок. Результаты всех этих работ совпали, что является важнейшим подтверждением корректности этой оценки частоты мутаций. Средняя скорость мутирования изученных STR-маркеров (обычно использовался 17-ти маркерный набор Y-filer) составляет $2.1 \cdot 10^{-3}$ на локус за поколение. То есть для среднестатистического STR-маркера вероятность смутировать при передаче от отца к сыну составляет 0,0021 – примерно двадцать один шанс из десяти тысяч. Поскольку эта скорость мутирования определена при прямом подсчете числа мутаций в известных родословных, она получила название «генеалогической».

Второй подход – калибровка – состоит в определении разнообразия гаплотипов, накопленного популяцией за время ее существования. Для этого нужно изучить популяцию, для которой известно время ее «основания», и найти в ней кластеры гаплотипов, возникшие за все время существования популяции и роста ее численности от небольшого числа основателей до современной численности. Понятно, что этот подход намного сложнее, и у него больше потенциальных источников погрешности. В то же время подход калибровок является общепринятым в науке – достаточно привести пример калибровок в радиоуглеродном методе датирования. При этом определяется доля распавшихся атомов для образца с известным возрастом и отсюда вычисляется, с какой скоростью распадаются атомы. И эта скорость потом применяется для расчета возраста других образцов, для которых тоже определена доля распавшихся атомов, но возраст неизвестен. Скорость мутирования STR-маркеров Y-хромосомы была определена в работе [Zhivotovsky et al., 2004] на двух примерах. Ими послужили кластеры гаплотипов, накопленные в популяции маори, сформировавшейся в результате миграции полинезийцев на дотолее необитаемую Новую Зеландию не позднее 800 лет назад, и в популяции цыган Бол-

гарии, сформировавшихся в результате исторически датированной миграции предков цыган из Индии и их разделения на эндогамные группы внутри Европы около 900–1000 лет назад. К сожалению, других работ по калибровке скорости мутирования Y-STR маркеров не было проведено, поэтому отсутствуют независимые подтверждения (как, впрочем, и опровержения) этой скорости. При оценке этим методом для среднестатистического STR-маркера вероятность смутировать при передаче от отца к сыну составляет 0,0007 – примерно семь шансов из десяти тысяч. Эта скорость получила название «эволюционной», поскольку определена для эволюционирующей популяции.

Как видим, «эволюционная» и «генеалогическая» скорости различаются в три раза (!) – семь шансов или двадцать один шанс из десяти тысяч. Поэтому и возраст гаплогрупп, рассчитанный при использовании той или иной скорости, будет различаться в три раза. Научные дискуссии по этой проблеме описаны чуть ниже.

ДЛИНА ПОКОЛЕНИЯ

Для перехода от датировок в числе поколений к датировкам в годах требуется задать длину поколения. В данном исследовании во всех методах нам пришлось пользоваться двумя разными величинами длины поколения для двух мутационных скоростей.

При использовании эволюционной скорости мутирования, длина поколения принималась равной 25 лет, поскольку эволюционная скорость была первоначально определена в годах и пересчитана в поколения при использовании именно такой длины поколения [Zhivotovsky et al., 2004].

При использовании генеалогической скорости мутирования, длина поколения принималась равной 30 годам, поскольку данная величина примерно соответствует длине поколения для мужчин, определенной в демографических исследованиях [Fenner, 2005], и та же величина была определена для кавказских популяций при их генетико-демографическом исследовании [Почешхова, 2008].

ДАТИРОВКИ ПО ПОЛНЫМ СИКВЕНСАМ Y-ХРОМОСОМЫ

В самые последние годы в изучении Y-хромосомы произошла «полногеномная революция» – стали стремительно накапливаться данные не по нескольким десяткам STR-маркеров, а по нескольким миллионам позиций на Y-хромосоме, каждая из которых представляет собой потенциальный SNP-маркер. И хотя скорость мутирования SNP маркеров намного меньше, чем STR, но за счет их огромного числа теперь каждый секвенированный образец характеризуется своим собственным набором из встреченных только у него SNP-марке-

ров, и каждая ветвь на дереве тоже несет несколько специфичных для нее маркеров. Тем самым решается проблема неоднозначной реконструкции дерева – дерево теперь обычно реконструируется однозначно. Предковый гаплотип тоже четко определяется по дереву. То есть снимаются основные проблемы, свойственные применению ρ -показателя. Метод расчета возраста гаплогрупп по полногеномным данным столь прозрачен, что обычно даже не обсуждается, но по сути это именно ρ – среднее число мутаций от предкового узла-гаплог-

типа до гаплотипов-потомков служит мерой возраста кластера, происходящего от этого предкового узла. Мерой возраста, выраженного в числе мутаций, – но ведь нужно знать также скорость мутирования и длину поколения?

Скорость мутирования для полного секвенирования Y-хромосомы уже определена во многих исследованиях. Причем, в отличие от скорости STR, и прямой подсчет, и калибровки получены в нескольких работах, причем для калибровок использованы самые разные подходы (рис. 1.22). На этом рисунке показаны величины скорости, полученные во всех этих исследованиях. Видно, что первоначально (статьи 2009–2012 годов) разброс был почти двукратный, но почти все последующие исследования (можно надеяться, более основательные) дали близкие оценки, и центр тяжести определений скорости явно находится около величины $0,8 * 10^{-9}$ мутаций на нуклеотид в год. То есть для среднестатистического

нуклеотида на Y-хромосоме вероятность смутировать за один год составляет примерно восемь шансов из 10 миллиардов. Разумеется, для одного отдельно взятого нуклеотида это весьма редкое событие, но если секвенируются, скажем, 10 миллионов нуклеотидов, то вероятность мутации хотя бы одного за один год составляет восемь шансов из ста. А за сто лет – уже восемьдесят шансов из ста. То есть одна мутация произойдет в среднем за 120 лет. Значит, за тысячу лет (это небольшой срок для объектов популяционных исследований) произойдет уже целых восемь мутаций, то есть по числу мутаций можно оценить возраст гаплогруппы с приемлемой статистической погрешностью.

Важно, что полногеномная «эволюционная» скорость (полученная по калибровкам заселения Сардинии [Francalacci et al., 2013] и Америки [Poznik et al., 2013]) совпала с полногеномной «генеалогической» скоростью, полученной по много-

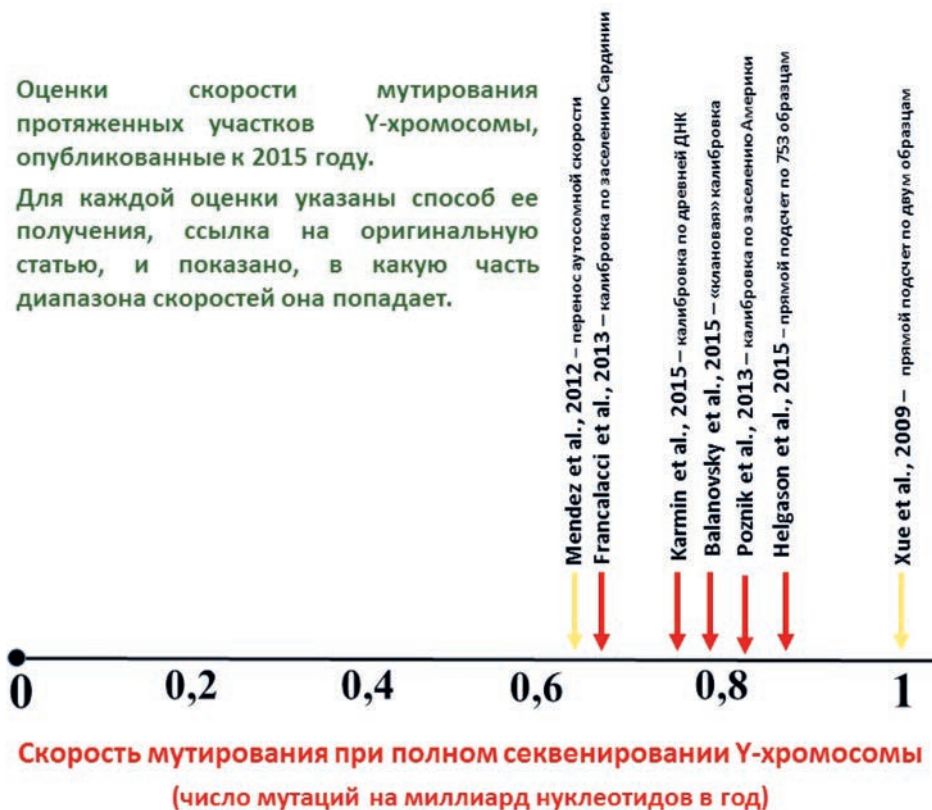


Рис. 1.22. Оценки «полногеномной» скорости мутирования на Y-хромосоме.

[Xue et al., 2009] – скорость $1 * 10^{-9}$ мутаций на нуклеотид в год. Проведен прямой подсчет мутаций по 1 родословной между 2 образцами, разделенными 13 поколениями. [Mendez et al., 2012] – скорость $0,62 * 10^{-9}$ мутаций на нуклеотид в год. На Y-хромосому просто перенесена скорость, принятая для аутосом. [Poznik et al., 2013] – скорость $0,82 * 10^{-9}$ мутаций на нуклеотид в год. Калибровка по заселению Америки, принимая дату равной 15 тысячам лет. [Francalacci et al., 2013] – скорость $0,65 * 10^{-9}$ мутаций на нуклеотид в год. Калибровка по археологическим датам заселения Сардинии. [Balanovsky et al., 2015] – скорость $0,78 * 10^{-9}$ мутаций на нуклеотид в год. Калибровка по времени жизни генеалогического предка изученных представителей рода казахов-аргынгов (справедливость в данном случае генеалогических преданий подтверждается совпадением с реконструированным генетическим деревом). [Karmin et al., 2015] – скорость $0,74 * 10^{-9}$ мутаций на нуклеотид в год. Калибровка по древней ДНК, датированной радиоуглеродным методом, и современным образцам – ее потомкам. [Helgason et al., 2015] – скорость $0,87 * 10^{-9}$ мутаций на нуклеотид в год. Проведен прямой подсчет мутаций по 274 родословным между 753 образцами, разделенными 2449 поколениями. Желтыми стрелками показаны две первые опубликованные оценки, а красными – результаты последних исследований.

численным исландским родословным [Helgason et al., 2015]. Расхождение между этой и более ранней генеалогическими оценками [Xue et al., 2009; Helgason et al., 2015] составляет лишь 13% и несомненно, что предпочтение следует отдать более поздней и основанной на несравненно большем объеме данных «исландской» скорости. Конечно, любая из этих оценок не бесспорна: например, использованная дата заселения Америки (15 тысяч лет) весьма условна, поскольку, например, при калибровке скорости мутирования митохондриальной ДНК [Forster et al., 1996] та же дата принималась равной 25 тысячам лет.

Способ оценки скорости мутирования, использованный в нашей собственной работе [Balanovsky et al., 2015], мы назвали «клановым», потому что он является промежуточным между калибровкой и подсчетом по родословной: скорость определена по разнообразию, накопленному в популяции (что роднит с калибровкой), но изученная казахская популяция аргынов, как выяснилось, происходит преимущественно от одного основателя-родоначальника (что роднит с подсчетом по родословной).

НАУЧНЫЕ ДИСКУССИИ

Трехкратные различия между «эволюционной» и «генеалогической» скоростями мутирования представляют собой научную проблему, ставшую предметом оживленных дискуссий [Zhivotovsky et al., 2004; Di Giacomo et al., 2005; Zhivotovsky et al., 2005; Zhivotovsky et al., 2006]. Естественно, что наибольшее число статей по этому вопросу опубликовано автором «эволюционной» скорости профессором Л.А. Животовским, искавшим причины таких различий. Последняя из перечисленных работ подводит итоги дискуссии и предлагает объяснение различий между скоростями.

Объяснение [Zhivotovsky et al., 2006] заключается в том, что гаплотипы возникают в популяции в таком количестве, как следует из «генеалогической» скорости, но часть из этих гаплотипов затем элиминируется дрейфом генов, и потому при рассмотрении сохранившихся гаплотипов (а только их, естественно, и может изучить исследователь в реальной популяции) обнаруживается меньшее разнообразие, чем разнообразие, возникшее по ходу истории популяции, а потому и рассчитываемая по этим данным скорость мутирования оказывается меньше. Это предположение было подтверждено компьютерным моделированием, и был сделан важный вывод, что соотношение «генеалогической» и «эволюционной» скоростей может быть не только трехкратным, а практически любым, в зависимости от конкретных демографических параметров популяции, а также в зависимости от – во многом стохастической – де-

Наконец, в работе [Karmin et al., 2015] использован еще один подход: подсчет числа мутаций, по которым современные гаплотипы отличаются от секвенированного древнего образца ДНК, который филогенетически может считаться их предком (точнее, расположен на филогенетическом дереве очень близко к своему общему предку с современными гаплотипами).

Что же касается вопроса о длине поколения, то при оценках «полнохромосомных» скоростей в качестве вводной информации использовались даты в годах (время заселения Сардинии и Америки, время жизни основателя клана, датировка древнего образца), поэтому и скорости указаны сразу в мутациях на год, и нет необходимости промежуточного пересчета через длину поколения.

Итак, в отличие от скоростей мутирования Y-STR-маркеров, в которых конкурируют две оценки, различающиеся в три раза, скорости мутирования при датировках по полным сиквенсам Y-хромосомы определены с приемлемой точностью, и существенных разногласий в этом вопросе нет.

мографической истории конкретной гаплогруппы. Иного объяснения трехкратным различиям скоростей в научной литературе предложено не было, и объяснение элиминацией части гаплотипов вошло в обиход.

В то же время оставалось непонятным, какой все же скоростью пользоваться на практике. Многие исследователи полагали, что поскольку эволюционная скорость была получена не для семей, а для популяций, то именно ее логично использовать для датировок истории гаплогрупп в популяциях. Некоторые другие авторы предпочитали генеалогическую скорость, а во многих работах приводились расчеты по обеим скоростям. Последний вариант особенно наглядно показывал реально существующую неопределенность в оценках возраста большинства гаплогрупп. Поэтому большинство специалистов предпочитали не основывать свои выводы на столь шатком фундаменте, как генетические датировки, и предпочитали другие основания, такие как географические закономерности в распространении гаплогрупп, градиенты их разнообразия, сравнение с лингвистическими и археологическими данными и другие подходы.

К вопросу сравнения двух скоростей мы вернемся в главе 7, в которой покажем, что генетические датировки, полученные для генофондов Кавказа с использованием «генеалогической» скорости, оказались куда в более хорошем согласии с лингвистическими и историческими датировками, чем полученные с помощью «эволюционной» скорости.

ти. Другим указанием в пользу «генеалогической» скорости являются данные полного секвенирования Y-хромосомы: в них «эволюционная» скорость не отличается существенно от «генеалогической», а значит, и для STR-гаплотипов «эволюционная» должна бы быть близка к твердо установленной (независимо показанной во многих исследованиях) «генеалогической» скорости.

В работе [Karmin et al., 2015] для нескольких десятков гаплогрупп были сопоставлены датировки, полученные по STR-маркерам и по данным полного секвенирования. Оказалось, что для молодых гаплогрупп (в пределах 5-10 тысяч лет) «эволюционная» скорость сильно занижает возраст, и для соответствия полногеномным результатам скорость мутирования STR-маркеров должна быть близка к «генеалогической» или даже превосходить ее. Зато для более древних гаплогрупп «эволюционная» скорость подходит лучше, а начиная при-

мерно возрастов порядка 30 тысяч лет, скорость мутирования STR-маркеров, требующаяся для совпадения результатов с полногеномными, фактически совпадает с «эволюционной» скоростью.

Итак, по мнению автора, окончательной ясности в этом вопросе нет, но большинство аргументов указывают на обоснованность применения в большинстве случаев «генеалогической» скорости мутирования при обработке данных по Y-STR маркерам. Впрочем, этот вопрос потерял свою остроту, поскольку возраст гаплогрупп теперь достаточно надежно определяется по результатам полного секвенирования Y-хромосомы. А для датировок тех совсем молодых кластеров в пределах этих гаплогрупп, где требуется использование быстро мутирующих STR маркеров, сама их молодость (приближающая их временной диапазон к диапазону известных генеалогий) обосновывает применение к ним «генеалогической» скорости.

ОКОЛОНАУЧНЫЕ И ЛЖЕНАУЧНЫЕ ДИСКУССИИ

ОКОЛОНАУЧНЫЕ ДИСКУССИИ

Словом околонатура – надеюсь, не обидным – назовем исследования в области генетической генеалогии, поскольку их в основном проводят люди, которые сами себя называют «любителями», противопоставляя «профессионалам» – обычным популяционным генетикам. Это та сфера, что на Западе называется *citizen science* – наукой, которой занимаются не профессиональные ученые, а обычные люди. Эти исследования действительно находятся *около* науки – являясь научными по предмету исследований, они не всегда следуют строгой научной методологии. Они находятся около науки и формально, поскольку не интегрированы в сложившуюся структуру науки, хотя иногда взаимодействуют с ней. Настойчивость, энтузиазм и несомненная талантливость помогают многим представителям генетической генеалогии получить значимые результаты, часть из которых со временем входит и в академическую науку. Однако отсутствие навыков критического подхода, поверхностность познаний и склонность к ангажированным интерпретациям приводят некоторых других представителей генетической генеалогии к выводам, имеющим с наукой мало общего.

Вообще же главное, но легко преодолимое, расхождение между генетической генеалогией и популяционной генетикой кроется в причинах психологических. Представителям науки нелегко принять форму, в которой генетическая генеалогия подает свои результаты (например, в виде серии не всегда связанных между собой кратких «постов» на интернет-форумах). Представители же генетической генеалогии, похоже, склонны компенсиро-

вать естественно возникающее в их положении «любителей» ощущение неполноценности острой критикой в адрес популяционных генетиков-«профессионалов».

Излюбленным предметом этой критики является как раз скорость мутирования.

Неудивительно, что генетическая генеалогия выбрала для своих расчетов «генеалогическую» скорость мутирования, которая, действительно, подтверждается не только на парах «отец-сын», но и на более глубоких генеалогиях. Но удивительно, с каким энтузиазмом критикуется скорость «эволюционная». Такая острота характерна для вопросов принципиальных, а скорость мутирования вовсе не является ключевым вопросом для популяционной генетики, ведь датировки по изменчивости STR гаплотипов Y-хромосомы являются очень незначительным ее разделом.

Генетики, как правило, сознают трудность надежного определения возраста гаплогруппы – ведь даже если решить проблему скорости, останется вопрос неполноты выборки гаплотипов (многие реально существующие подветви в анализируемой выборке могут отсутствовать), неточности реконструированного дерева (даже при использовании большого числа маркеров на дереве остаются – пусть не всегда отображаемые – ретикуляции, превращающие ее в запутанную сеть), и даже точно определенный возраст гаплогруппы не всегда помогает достичь цели – датировать события в истории популяций. К тому же в ряде случаев трудность в определении абсолютного возраста гаплогрупп не скажется существенно на выводах, поскольку для решения многих задач достаточно относительного возраста, то есть важно определение после-

довательности появления разных гаплогрупп в изучаемом регионе – какая гаплогруппа распространилась первоначально, какие потом, а какие гаплогруппы маркируют наиболее поздние миграции.

Поэтому популяционные генетики относятся к расхождениям в датировках философски, а к самим датировкам – с изрядной долей скепсиса. Но генетические генеалоги отчего-то кладут вопрос датировок в сердцевину своего мировоззрения и убеждены если не в точности каждой конкретной даты, то в принципиальной достижимости и принципиальной важности точных дат. Поэтому в отношении генетических генеалогов к генетике создается некая аберрация зрения: хотя расчеты возраста по STR-гаплотипам являются очень небольшой частью исследований Y-хромосомы, проводимых популяционным генетиками, и генетические генеалоги прекрасно знают эти исследования, генеалоги искренно убеждены, что «неправильная эволюционная скорость» является для работ популяционных генетиков краеугольным камнем и сводит на нет всю их ценность.

Нужно отметить, что в интернет-дискуссиях на эту тему нередко смешиваются совершенно различные вещи. Например, скорости мутирования и метод расчета возраста – хотя очевидно, что любой метод можно скомбинировать и с той, и с другой скоростью. Любопытно, что смешивая скорость и метод, многие генеалоги начисто отрицают эволюционную «скорость Животовского», но при этом пользуются методом ASD [Sengupta et al., 2006], хотя в этой статье Л.А. Животовский был вторым соавтором и, учитывая его квалификацию математика, скорее всего, занимался как раз разработкой метода.

Накал страстей поражает воображение. Не будем приводить здесь цитаты – желающие легко найдут их даже на наиболее основательном из русскоязычных интернет-сайтов по генетической генеалогии www.molgen.org, не говоря уже о сайтах, для которых не так характерны взвешенные высказывания. Не будем и развивать далее тему о проблемах взаимодействия популяционной генетики и генетической генеалогии. Лучше бегло упомянем те совместные достижения, которые получены в рамках сотрудничества представителей генетической генеалогии и одного лишь нашего коллектива. Это база данных Y-base, в наполнении которой большую помощь оказал Роман Сычев; база данных MURKA, текущая версия которой разработана Валерием Запороженко; программа-предиктор, разработанная по нашему заказу Вадимом Урасиным; наши совместные статьи-диалоги с Вадимом Вереничем и Евгением Матюшонком на сайте Генофонд.рф; моя статья в «Российском вестнике генетической генеалогии»; помощь ряду генеалогов, исследующих гаплотипы Кавказа и Закавказья, передача им групп фамилий (не индиви-

дуальных, что относится к конфиденциальной информации, а именно групп) – для серий изученных нами гаплотипов; полезные совместные обсуждения и многое, многое другое. Можно надеяться, что эти реальные совместные достижения перевесят мифические противоречия, и русскоязычные генетики и генеалоги догонят и перегонят своих англоязычных коллег по части взаимопонимания, взаимопомощи и сотрудничества друг с другом.

ЛЖЕНАУЧНЫЕ ДИСКУССИИ

«Лженаука» – сильное слово, но ДНК-генеалогия в варианте, пропагандируемом А.А. Клесовым, заслуживает его в полной мере. Это было давно понятно специалистам, а после публикации коллективной статьи 24 ведущих российских ученых – генетиков, лингвистов, археологов, историков, генетических генеалогов и других (Троицкий вариант-наука, №1(170), 13 января 2015), «ДНК-генеалогия Клесова» получила статус лженауки, так сказать, официально. Обсуждению лженауки было бы, наверное, не место в этой научной монографии, если бы А.А. Клесов не распространял ДНК-генеалогии настолько широко и агрессивно в русскоязычном интернете, что многие ученые, не являющиеся профессиональными генетиками, поневоле знакомятся с ней. Не будучи знакомы с реальной картиной генетических исследований истории народов – представленной в интернете весьма скудно – они иногда поддаются под очарование запредельного упрощения А.А. Клесовым генетических данных, методов и результатов. Но здесь мы рассмотрим лишь учение ДНК-генеалогии о расчете возрастов гаплогрупп по разнообразию гаплотипов.

От генетической генеалогии, в русле которой А.А. Клесов начинал свою деятельность на генетическом поприще, он унаследовал: (1) веру в непогрешимость «генеалогической» скорости мутирования и (2) в непогрешимость метода расчета возраста по доле исходного гаплотипа (разработанного им совместно с генетическим генеалогом Дмитрием Адамовым). За годы самостоятельной деятельности он: (3) объявил «генеалогическую» скорость своим личным изобретением, (4) составил инструкцию расчета возраста и справочную таблицу, (5) ввел поправку на обратные мутации, (6) назвал все это единственно верным способом датировок, и более того – (7) новой наукой, не имеющей отношения к популяционной генетике.

Прокомментируем по пунктам.

(1) «Генеалогическая» скорость мутирования, действительно, неплохо обоснована. Забавно, что автор этих строк, который едва ли не больше чем, кто-либо другой из популяционных генетиков, доказывал преимущество именно «генеалогической»

скорости мутирования перед «эволюционной», в интернет-публикациях А.А. Клесова фигурирует как убежденный приверженец «эволюционной» скорости. Должен сказать, что я не так уж убежден и в абсолютной правильности скорости «генеалогической», но именно она подтвердилась в двух исследованиях, которые мы проводили в разное время и на разных генофондах: это было одним из основных выводов и обобщающей статьи по Кавказу (Balanovsky et al., 2011), и статьи по калибровке полногеномной – и заодно основанной на STR – скорости мутирования (Balanovsky et al., 2015). Итак, вера в непогрешимость генеалогической скорости, хотя и имеет ненаучные истоки, вряд ли приведет к научным ошибкам, тем более на тех интервалах времени, с которыми имеет дело ДНК-генеалогия. Поэтому на этом первом пункте клесовское учение ДНК-генеалогии не вызывает нареканий.

(2) Метод расчета по доле исходного гаплотипа, как указывалось выше, правомочен, но имеет ряд ограничений и в целом наименее оптимален. Для датировок молодых кластеров с простой филогенетической структурой он приемлем, но для старых и сложных филогений, скорее всего, определит возраст хуже других методов. Вероятно, А.А. Клесов это знает, поскольку часто пользуется общепринятым расчетом по числу накопленных мутаций, а достоинством своего метода (доли исходного гаплотипа) называет не точность, а лишь простоту расчетов. Итак, и здесь особых проблем нет (хотя нет и преимуществ), и клесовская ДНК-генеалогия остается в рамках правомочных с точки зрения популяционной генетики методов.

(3) Объявление А.А. Клесовым «генеалогической» скорости своим изобретением требует сразу трех комментариев. Во-первых, им, насколько можно судить, действительно проводились исследования шотландских родословных и была получена, естественно, та же «генеалогическая» скорость мутирования, которая получалась всеми: была известна во многих исследованиях пар отец-сын, недавно получена нами на казахской родословной [Balanovsky et al., 2015], а также получалась и в ряде исследований в русле генетической генеалогии. Во-вторых, этот результат – возможно, вполне научный – не был никогда А.А. Клесовым опубликован в признанных научных журналах, и даже в его интернет-публикациях методы описаны недостаточно полно для их воспроизведения. Именно поэтому мне и пришлось выше ставить оговорку «насколько можно судить». В-третьих, если А.А. Клесовым и была независимо подтверждена известная задолго до него и многократно подтвержденная «генеалогическая» скорость, это не делает ее некой третьей «клесовской» скоростью – она остается все той же, следуя устоявшемуся термину, «генеалогической».

(4) Пошаговая инструкция расчета возраста – вещь, несомненно, полезная для начинающих любителей генетической генеалогии или ДНК-генеалогии, но не более того. Справочная таблица, в которой скорости мутирования приведены не только для обычного для популяционной генетики 17-ти маркерного набора STR-маркеров, но и для других наборов маркеров, предлагаемых компанией коммерческого ДНК-тестирования ftDNA, тоже полезна, хотя такие расчеты проводились и многими представителями генетической генеалогии. Итак, и на этом этапе противоречий с подходами популяционной генетики нет.

(5) Поправка на обратные мутации правильна, но только для наименее оптимального метода расчета по доле исходного гаплотипа и при внешне заданной скорости мутирования. Ведь действительно, могут происходить и обратные мутации производных гаплотипов в исходный, и они занижат наблюдаемую долю производных гаплотипов по сравнению с долей реально возникших производных гаплотипов. Но для других методов расчета и при использовании калиброванных скоростей мутирования введение такой поправки было бы неверным, ведь обратные мутации могли происходить и в примере, послужившем основой калибровки, а значит уже в неявном виде учтены. Итак, и тут в методе расчета ДНК-генеалогии нет ошибок? Возможно. Хотя область применения сужается с каждым новым шагом до предельно простых случаев.

(6) Но ошибки точно начинаются с этапа, когда этот «клесовский» способ расчета объявляется единственно верным. Во-первых, он не «клесовский», а разработан популяционными генетиками и использовался задолго и независимо от А.А. Клесова. Во-вторых, этот способ не единственно верный: и расчет по числу мутаций, и особенно расчет по доле исходного гаплотипа имеют свои недостатки, да и «генеалогическая» скорость мутирования не бесспорна.

(7) Область деятельности (ДНК-генеалогия Клесова), основанная на позаимствованном из популяционной генетики методе датировок кластеров гаплотипов, объявляется новой наукой и популяционной генетике противопоставляется. Тут уже комментарии излишни – поскольку ДНК-генеалогия самим ее гуру выводится за пределы академической науки на просторы лженауки.

Стоит добавить, что этот вывод за пределы науки строится на странной нелепице, с завидным постоянством внедряемой в умы почитателей клесовской ДНК-генеалогии. Им внушается нелепое представление, что сфера исследований генетики не изменилась со времен монаха Грегора Менделя, и что генетика занимается только генами – теми фрагментами наследственного материала, которые отвечают за фенотипические и иные признаки. А весь основной массив ДНК, о котором

мы не знаем его связи с внешним проявлением, остается якобы бесхозным, и именно его присваивает себе ДНК-генеалогия. Нелепый девиз: «Гены – генетике, а ДНК – ДНК-генеалогии», как ни удивительно, находит отклик в умах почитателей, имеющих среднее образование. Казалось бы, даже школьное образование обеспечивает понятие о генетике. Генетика – это отнюдь не «наука о генах». Это наука о наследственности и изменчивости. А наследственность, даже если рассматривать только связанную с ДНК, включает в себя множество самых различных типов нуклеотидных последовательностей, среди которых гены составляют меньшинство. Как термин «атом» (т.е. «неделимый») используется и ныне, но означает совсем иное, чем для философов античности, так и

термин «ген» используется и ныне, но означает совсем иное, чем для Иоганнсена, который ввел его более ста лет назад. Поэтому генетикам даже нелегко разъяснять понятия, известные школьникам, что генетика работает со всем наследственным материалом, со всей ДНК, как и развеивать прочие невежественные постулаты ДНК-генеалогии. (Например, столь же невежествен постулат А.А. Клесова о том, что популяционная генетика якобы занимается путем от гена к признаку, хотя этим занимаются совсем иные разделы генетики и других биологических наук). Но именно с помощью наслаивания одних невежественных постулатов на другие и формируется лженаука ДНК-генеалогия, которые зарубежные коллеги назвали «генетической астрологией».

ДАТИРОВКИ В ЭТОЙ КНИГЕ

Датировки по данным полного секвенирования Y-хромосомы используются в разделе 7.4, а по данным о STR-гаплотипах наиболее последовательно и систематически применялись при исследовании генофонда Кавказа (раздел 7.3). Для датировки STR-кластеров в пределах гаплогрупп использовались три метода:

1) Показатель ρ использовался для датировки кластеров, как описано в [Forster et al., 1996; Saillard et al., 2000].

2) Метод BATWING [Wilson et al., 2003], основанный на моделировании возникновения гаплотипов, был использован для получения независимых оценок возраста тех же кластеров.

3) Показатель ASD [Sengupta et al., 2006] использовался для определения времени, необходимого для накопления наблюдаемой изменчивости внутри популяции для всей гаплогруппы (а не для отдельных кластеров внутри гаплогрупп, в отличие от первого и второго методов анализа).

Хотя подавляющее большинство статистических расчетов в данном исследовании выполнены лично автором, но анализ методом BATWING проведен совместно с сотрудником Нью-Йоркского исследовательского центра IBM Дэном Платтом (Dan Platt), а расчет ASD – вместе с научным координатором проекта «Генография» Давидом Сорией (David Soria) в рамках совместного исследования [Balanovsky et al., 2011].

ДАТИРОВКИ С ПОМОЩЬЮ МЕТОДА BATWING

BATWING позволяет определить временные границы, между которыми могла возникнуть мутация, впоследствии не повторявшаяся (unique event polymorphism – UEP). Обычно в качестве UEP используются SNP маркеры, но этот подход может быть расширен за счет использования «виртуальных мутаций», маркирующих кластеры STR гап-

лотипов, как описано в работах Scrusiani с соавторами [Scrusiani et al., 2004, 2006]. Выделялись популяции, в которых присутствуют эти кластеры, и расчеты методом BATWING включали все образцы из этих популяций с целью датировать UEP. Мы использовали те же параметры распределения, что и в работе [Xue et al., 2008].

BATWING также применялся для моделирования последовательности популяционных ветвлений и датировки точек ветвления. Кластеры гаплотипов, определенные на филогенетических сетях, никоим образом не учитывались в этих расчетах. Таким образом, эти результаты предоставляют методически независимую оценку времени разделения популяций.

ОГОВОРКА 1: МОНОФИЛЕТИЧНОСТЬ

Все методы датировок кластеров исходят из предположения, что рассматриваемые гаплотипы реально происходят от одного предка. Но для этого нужно точно знать дерево происхождения гаплотипов друг от друга. При использовании полного секвенирования Y-хромосомы это достигается, но при использовании данных по STR-гаплотипам получаемое дерево является лишь приближенной реконструкцией реально происходившей эволюции гаплотипов. Конечно, чем больше STR-маркеров анализируется, тем точнее выделяются гаплотипы и лучше реконструируется их родство друг с другом, но даже при использовании и нескольких десятков, и более сотни STR-маркеров остаются неоднозначности в дереве.

Публикуемые схемы родства гаплотипов, даже если они отображаются в виде сетей (ретикуляции на сетях как раз предназначены для отображения разных возможных путей эволюции), часто создают иллюзию почти полной однозначности дерева. В действительности же эти схемы определяются

порогами, задаваемыми программе при построении дерева, или внутренними алгоритмами самой программы. При использовании наиболее популярной программы Network сеть при отображении большинства возможных путей эволюции выглядит как спутанный клубок. Но при задании жестких условий исключения всех ветвей, кроме наиболее вероятных, клубок можно превратить в красивое звездообразное дерево, ретикуляции на котором остаются лишь там, где решительно невозможно отдать предпочтение одному из двух возможных путей. Более того, даже в названии статьи, описывающей метод построения медианных сетей [Bandelt et al., 2000], присутствует оборот «жадная редукция» (greedy reduction), указывающий на то, что исключение менее вероятных ветвей является неотъемлемой – и вполне разумной – особенностью метода. Но ведь эти не отображаемые на дереве пути эволюции тоже вполне возможны (особенно при жестких порогах). А значит та группа гаплотипов, которая на полученном дереве показана как происходящая из одного гаплотипа-основателя, на самом деле может включать гаплотипы, имевшие в действительности других предков, пусть скорее всего в какой-то мере и родственных приписанному им предку-гаплотипу.

Эта погрешность, неизбежная при использовании Y-STR данных, является наиболее серьезным препятствием для надежного определения возрастов кластеров – в ряде случаев не менее серьезным, чем трехкратные различия при использовании разных скоростей мутирования. Для минимизации этой погрешности можно только внимательно относиться к построению филогенетических схем, сравнивать схемы, полученные при разных параметрах и порогах, тщательно рассматривать детали топологии и брать в работу лишь те кластеры, которые представляются действительно монофилитическими группами.

По этому пути мы шли в анализе филогенетических сетей гаплогрупп на Кавказе, и очень полезным оказалось следование некоторым формальным правилам. Например, при выделении какой-либо группы на сети в качестве кластера, все гаплотипы сети, дистальные по отношению к гаплотипу-основателю, обязательно должны быть включены в кластер.

ОГОВОРКА 2: НАЙТИ ОСНОВАТЕЛЯ

Выделение этого гаплотипа-основателя представляет собой отдельную проблему. В идеальном случае на сети выявляется четкая звездообразная структура с одним гаплотипом в центре, представленном у большого числа образцов и расходящимися от него по лучам гаплотипами-потомками (каждый из которых представлен у меньшего числа образцов, чем основатель) и потомками потом-

ков (рис. 1.20 Б, В). Такая структура действительно иногда выявляется в реальных данных. Но куда чаще приходится работать с филогенетической сетью сложной топологии, на которой просматриваются группы гаплотипов, лишь отдаленно напоминающие звездообразную структуру. В этом случае предковым стоит считать тот гаплотип, который, во-первых, представлен у большого числа образцов, а во-вторых, имеет большое число потомков. При противоречии между этими правилами предпочтение стоит отдавать «правилу потомков», потому что частота предкового гаплотипа могла уменьшиться, а частота одного из потомков случайно вырасти. Но если у гаплотипа на сети совсем мало потомков, то и у нас совсем мало оснований считать его исходным гаплотипом кластера.

Еще одно правило, тоже выработанное при изучении кавказских гаплотипов: предковый гаплотип обязательно должен быть в своем кластере ближе всех к центру сети (общему предку всех вообще гаплотипов на сети). Иными словами, мы запретили себе допускать «мутации назад» при интерпретации филогенетической сети, и включать в кластер гаплотип, который согласно сети, является «предком предка» (или, если уж его включать, то его и принимать за предка). Это правило введено из тех соображений, что достаточно уже той неизбежной погрешности, которая внесена при построении сети объективным методом, и не стоит увеличивать погрешность за счет субъективных предположений, что родство гаплотипов могло быть не таким, а чуть иным. Правильнее принимать построенную сеть как данность и выделять кластеры, строго следуя ее топологии. Многочисленные примеры выделения гаплотипов-основателей приведены в [Balanovsky et al., 2011] и в разделе 7.3.

ОГОВОРКА 3: ПОСТОЯНСТВО СКОРОСТИ

Метод молекулярных часов основывается на том, что скорость возникновения мутаций постоянна и не меняется ни во времени, ни в пространстве. То есть что вероятность возникновения мутации в данном гаплотипе одинакова, независимо от того, принадлежит ли этот гаплотип палеолитическому сибиряку или современному негру. В отличие от других сделанных оговорок, эта не создает реальных затруднений. Ведь действительно нет оснований полагать, что скорость унаследованных мутаций меняется во времени или пространстве. Даже такие события, как ядерные взрывы в Хиросиме или Челябинске-40, судя по большинству исследований, не увеличили частоту мутаций, унаследованных следующими поколениями (всплеск соматических мутаций у самих облученных людей, естественно, к этому вопросу не относится). А для остальной истории человечества нет

оснований предполагать условия, стимулирующие или замедляющие мутации в мере, хотя бы приближающейся к этим двум печальным примерам.

Другое дело, что скорость мутаций теоретически может несколько различаться в зависимости от генетического окружения на хромосоме, и в некоторых исследованиях были найдены указания на несколько различающиеся скорости мутирования для разных гаплогрупп мтДНК и Y-хромосомы. Достоверно показано и то, что частота мутаций на Y-хромосоме увеличивается с возрастом отца. Поэтому в принципе возможно, что в популяциях с различной длиной мужских поколений скорость накопления мутаций будет несколько различаться. Однако на обычно исследуемых промежутках времени длина поколений относительно постоянна для всего человечества. Итак, на практике всеми этими сомнениями в универсальности скорости мутаций можно пренебрегать – особенно по сравнению с проблемой привязывания данных по датировке кластеров (гаплогрупп) к событиям истории популяций.

ОГОВОРКА 4: ВОЗРАСТ КЛАСТЕРА И ВОЗРАСТ ПОПУЛЯЦИИ

Автор этих строк никогда не сможет забыть слышанный им на заре использования мтДНК научный доклад, в котором возраст гаплотипов, встреченных у славян, был определен в 20 тысяч лет, и тут же был сделан вывод, что славянская популяция 20 тысяч лет назад и сформировалась. Даже сделанное докладчику указание на наличие километровой толщи льда двадцать тысяч лет назад (эта дата попадает на период последнего оледенения) в месте, где была собрана выборка и где поэтому предполагалось возникновение славян, не поколебало его уверенности. Ведь вывод был сделан на основе данных генетики, которая имеет репутацию почти что точной науки.

Возможность допустить ту же ошибку, пусть не в таких масштабах, дамкловым мечом постоянно нависает над очень многими и любителями, и даже профессионалами, не исключая, возможно, и самого автора. Уж очень велик соблазн связать историю гаплогруппы, встреченной в популяции, с историей самой этой популяции. И несомненно, что демографическая история популяции действительно влияет на историю содержащихся в ней гаплогрупп. Вот только эти две истории редко когда могут быть приравнены друг к другу.

В случае с подледным возникновением славян ошибка очевидна: генетическое разнообразие, пусть даже действительно двадцатитысячелетнего возраста, хотя и было обнаружено у славян, начало накапливаться еще у их далекой предковой популяции. И если в какой-то популяции обнаружена некая величина генетического разнообразия

в пределах некой гаплогруппы, то часть этого разнообразия, действительно, накоплена популяцией за время ее истории, но другая часть разнообразия была у популяции «еще на старте» весьма отдаленном. И даже если какая-то гаплогруппа встречается только в одной популяции, и возникновение гаплогруппы хорошо датировано, и популяция имеет логически обоснованное «начало» (допустим, миграция на остров), то возраст популяции вовсе не обязательно совпадает с возрастом гаплогруппы. Ведь гаплогруппа могла возникнуть задолго до формирования популяции, но не сохраниться в других местах, а сохраниться лишь в популяции, мигрировавшей на остров, и тогда гаплогруппа будет старше популяции. Но могла она возникнуть и после формирования популяции на каком-то этапе ее истории, и тогда гаплогруппа будет моложе популяции. Второй вариант иллюстрируется, например, тем же исследованием генофонда Кавказа (раздел 7.3), которое не раз уже приводилось в качестве примера. В нем для многих популяций были обнаружены специфичные для них кластеры гаплотипов. Раз эти кластеры встречаются в одной популяции, но отсутствуют в других, то датировки кластеров указывают на время отделения этой популяции от родственных ей групп. Но это время является нижней (наименьшей, более близкой к современности, более поздней) оценкой времени выделения популяции из родительской прапопуляции, поскольку кластер мог возникнуть как сразу после разделения популяций, так и в любой момент времени после этого события.

ОГОВОРКА 5: РАЗНЫЕ СКОРОСТИ ДЛЯ РАЗНЫХ МАРКЕРОВ

При рассмотрении скоростей мутирования упоминалось, что они относятся к «среднестатистическому» маркеру. Но частота мутирования, например, разных STR-маркеров отличается на один-два порядка. К счастью, частота мутирования конкретного локуса обычно не важна, поскольку всегда анализируются их наборы. Поэтому нужно знать скорость мутирования именно набора локусов (то есть вероятность того, что смутит какой-то любой – не важно, какой именно – локус из набора). Конечно, скорости мутирования разных наборов могут различаться, если в один набор включены быстромутирующие, а в другой – медленно мутирующие маркеры. Но поскольку используемые в анализе гаплогрупп наборы состоят из нескольких десятков локусов (от 17 до 111), и они, за редким исключением, не подбирались специально по скорости мутирования, то случайное включение быстрых и медленных маркеров взаимно компенсируется, и скорости мутирования разных наборов мало отличаются друг от друга. В целом, хотя эту

особенность и стоит иметь в виду, она имеет меньшее значение, чем другие источники погрешностей.

**ОГОВОРКА 6: ИСКЛЮЧЕНИЕ ИЗ РАСЧЕТОВ
ОТДЕЛЬНЫХ STR ЛОКУСОВ**

Поскольку дублированные маркеры DYS385a и DYS385b типизируются на одной паре праймеров,

и поэтому нельзя установить, какой из аллелей относится к какому из двух локусов, эти локусы принято исключать из анализа при построении филогенетических сетей. Соответственно, эти маркеры были исключены и из расчетов с использованием показателя s , а также при анализе методом BATWING. Что касается расчетов показателя ASD, то они проводились в двух вариантах: с включением и с исключением DYS385a и DYS385b.

ГЛАВА 2.

СТРУКТУРА ГЕНОФОНДА ЕВРОПЫ В ЗЕРКАЛЕ Y-ХРОМОСОМЫ

Бурное развитие популяционной генетики человека, ознаменовавшееся появлением таких новых направлений, как филогеография, молекулярная антропология, этногеномика, археогенетика, во многом связано с широким использованием двух информативных, но своеобразных генетических систем – Y-хромосомы и митохондриальной ДНК. Это сформировало своеобразный «двусистемный» способ мышления – следуя ему, исследователь ожидает, что большинство миграций должны оставить след в географическом распространении обеих систем: и мтДНК (отражающей материнскую линию наследования), и Y-хромосомы (отражающей отцовскую линию наследования). По изменчивости этих двух систем опубликовано множество работ, охвативших население практически всех стран Европы, а в последние годы к ним добавились и исследования полногеномных панелей маркеров. Поскольку европейский генофонд одновременно яв-

ляется и наиболее гомогенным, и наиболее изученным в мире по всем трем упомянутым генетическим системам (Y-хромосомы, мтДНК, полногеномных панелей), именно на его примере можно усовершенствовать методологию анализа генофонда. Если удастся выявить пространственную и этнолингвистическую структурированность даже гомогенного европейского генофонда – это означает, что применяемые подходы и во всех иных, более гетерогенных, частях света позволят обнаружить как важные особенности их пространственной структуры, так и проследить траектории этногенеза. Отметим, что предыдущее обобщающее исследование, преследовавшее ту же цель [Rosser et al., 2000] проводилось более десяти лет назад по выборке (N=3 616), в десять раз меньшей нашей (N=38 000). Пожалуй, пришло время провести новый обобщающий анализ генофонда Европы на новом витке его изученности.

2.1. ПРЕДШЕСТВОВАВШИЕ ИССЛЕДОВАНИЯ

ОБОБЩАЮЩИЕ И РЕГИОНАЛЬНЫЕ РАБОТЫ

РАННИЕ ОБОБЩАЮЩИЕ РАБОТЫ

Исследования изменчивости Y-хромосомы в Европе начались в 1990е годы. И почти одновременно с плодотворной обобщающей статьей по изменчивости мтДНК в Европе [Richards et al., 2000], появились две крупные работы, суммирующие данные об изменчивости в Европе Y-хромосомы [Semino et al., 2000; Rosser et al., 2000]. Обе статьи основывались на внушительных объемах данных, и, хотя были написаны в различной манере, выявили сходные черты европейского генофонда.

Россер с коллегами [Rosser et al., 2000] следовали *феноменологическому* подходу, давая описание обнаруженным ими паттернам географической изменчивости Y-хромосомы. Они показали удивительно четкую клинальную изменчивость в распространении всех гаплогрупп Y-хромосомы и статистически обосновали, что генетическое сходство популяций определяется их географическим соседством, а не лингвистическим сходством их языков.

Что же касается статьи Семино с коллегами [Semino et al., 2000], то они следовали *ассоциативному* подходу, заключив, что наблюдаемые географические закономерности современного генофонда могли быть сформированы такими факторами со сходным географическим распространением, которые известны для периода верхнего палеолита. (Тем самым авторы этой статьи присоединились к сторонникам «палеолитического» времени формирования генофонда Европы и, следовательно, «культурной» теории неолитизации – без миграции самих носителей неолитической культуры в Европу). Легко заметить, что, делая такой вывод, авторы интерпретируют географические закономерности по *ассоциации* с событиями, имевшими сходную географию. Строго говоря, это логически не обоснованно, поскольку наличие корреляции между двумя явлениями еще не доказывает причинно-следственную связь между ними.

И, действительно, один из наиболее убежденных последователей альтернативной гипотезы «де-

мической диффузии», т.е. миграции в Европу не только культуры, но и самих носителей неолитической культуры [Chikhi et al., 2002], проведя независимый анализ того же самого массива данных, что и Семино с коллегами [Semino et al., 2000], пришел к противоположным выводам, интерпретируя те же закономерности как сформировавшиеся в неолитический период.

Одна из проблем «ассоциативного» подхода состоит в том, что он не учитывает, что миграции населения во многом определяются географическими барьерами и географическим ландшафтом, и потому маршруты миграции повторяются вновь и вновь в разные эпохи. Это не позволяет на основе «ассоциаций» различать время экспансии. Именно поэтому и в обзоре [Barbujani, Bertorelle, 2001] отмечено, что палеолитическая и неолитическая миграции в Европу шли, вероятно, близкими маршрутами.

В целом, эти два первых обобщающих исследования изменчивости Y-хромосомы в Европе [Semino et al., 2000; Rosser et al., 2000] выявили, прежде всего, ярко выраженную географическую дифференциацию европейского населения по маркерам Y-хромосомы. Это сразу ярко отличило Y-хромосомные данные и привлекло к ним большое внимание, поскольку по данным об аутосомных маркерах и мтДНК Европа выглядела довольно однообразным массивом популяций. В этих и последующих работах было показано:

- а) что в западноевропейских популяциях преобладает гаплогруппа **R1b-M343**;
- б) тогда как в Восточной Европе высока частота гаплогруппы **R1a1-SRY1532**;
- в) популяции Балканского полуострова характеризуются высокой частотой гаплогруппы **I2a-P37**;

г) тогда как в Скандинавии резко увеличена доля гаплогруппы **I1-M253**;

д) а гаплогруппы **J2-M172** и **E1b1b1-M35** приурочены к Южной Европе.

ПОСЛЕДУЮЩИЕ РЕГИОНАЛЬНЫЕ РАБОТЫ

После выхода этих ранних обобщающих исследований подавляющее большинство последующих работ концентрировалось на более дробных регионах [Malyarchuk et al., 2002, 2003, 2004, 2006; Behar et al., 2003; Харьков и др., 2004, 2007, 2008, 2011; Zakharov et al., 2004; Brion et al., 2005; Cinnioglu et al., 2004; Di Giacomo et al., 2003; Flores et al., 2003, 2004; Tambets et al., 2004; Alonso et al., 2005; Goncalves et al., 2005; Kayser et al., 2005; Pericic et al., 2005; Capelli et al., 2006, 2007; Lappalainen et al., 2006, 2008; Adams et al., 2008; Balanovsky et al., 2008, 2011; Battaglia et al., 2008; Bertocini et al., 2012; Fechner et al., 2008; Marchani et al., 2008; Mirabal et al., 2009; Varzari et al., 2009; King et al., 2011; Yunusbaev et al., 2012; Хуснутдинова и др., 2012; многие другие работы] или на отдельных гаплогруппах [Di Giacomo et al., 2004; Cruciani et al., 2007, 2010; Myres et al., 2010; Underhill et al., 2010, 2015; Mendez et al., 2011; Onofri et al., 2008, Derenko et al., 2006, 2007, 2010; Rootsi et al., 2004, 2007, 2012; Tofanelli et al., 2009 и другие работы]. Обобщающий анализ в масштабе всей Европы более не проводился. Поэтому не потерял своей актуальности сделанный еще в 2000 году и описанный выше вывод, что обнаруженная клинальная изменчивость гаплогрупп Y-хромосомы объясняется скорее географическими закономерностями, чем характером расселения лингвистических групп [Rosser et al., 2000].

ГОМОГЕННОСТЬ И ГЕТЕРОГЕННОСТЬ ЕВРОПЫ

Если главной чертой, обнаруженной в митохондриальном генофонде Европы, является его гомогенность [Simoni et al., 2000], то изменчивость Y-хромосомы, напротив, характеризуется исключительной гетерогенностью. Так же, как и для мтДНК, для Y-хромосомы выделены несколько основных гаплогрупп, преобладающих в населении Европы. Но если частоты основных митохондриальных гаплогрупп практически одинаковы во всех уголках Европы, то распределение гаплогрупп Y-хромосомы демонстрирует четкие географические градиенты. Ни аутосомные классические маркеры и ДНК-маркеры (до полногеномных исследований), ни гаплогруппы мтДНК никогда не обнаруживали столь элегантные тренды. Следовательно, Y-хромосома оказалась новым и весьма эффективным инструментом для популяционной генетики населения Европы.

Конечно, необходимо отдавать себе отчет, что генофонд Европы не может быть одновременно и гомогенным, и гетерогенным. Величина разнообразия генофонда Европы, конечно же, только одна. Вопрос лишь в том, в какой степени разные маркеры способны выявить эту реально существующую степень изменчивости генофонда. Например, имея в своем распоряжении данные по нескольким десяткам аутосомных маркеров, «классические» популяционные генетики достигли хорошего уровня разрешения в изучении генетических различий между *народами* Европы [Cavalli-Sforza et al., 1994; Рычков, Ящук, 1983]. А вот при использовании данных по митохондриальной ДНК различия между отдельными народами Европы почти не выявляются, и генетикам удается успешно оперировать только на более высоком иерархическом уровне, выявляя *географические регионы* [Richards et al.,

2002] или *лингвистические группы* [Balanovsky, 2009]. Y-хромосома значительно лучше фиксирует межпопуляционную дифференциацию и выявляет генетические различия даже между *локальными популяциями*, относящимися к одному и тому же народу, как будет показано ниже. А исследования, основанные на широкогеномных маркерах позволяют различать генетическое положение *индивидуумов* среди популяций [Novembre et al., 2008]. Таким образом, разные типы маркеров в разной степени выявляют гетерогенность генофонда.

Напомним, что под «гетерогенностью» генофонда в этой книге всегда подразумевается величина межпопуляционной изменчивости в пределах генофонда. Например, гетерогенность генофонда Европы – это величина различий между популяциями Европы; гетерогенность генофонда немцев – это величина различий между популяциями немцев и т.д. Но то, что разные маркеры в разной сте-

пени выявляют гетерогенность, не мешает сравнивать генофонды по уровню их гетерогенности, лишь бы это сравнение проводилось отдельно для каждого типа маркеров. Например, гетерогенность Европы по любому типу маркеров всегда будет ниже гетерогенности Сибири, и это одно из кардинальных различий между этими генофондами. Нельзя лишь сравнивать величины гетерогенности двух генофондов, если эти величины получены по разным типам маркеров. А если – в идеальном случае – все сравниваемые генофонды изучены по всем типам маркеров, тогда важным инструментом становится степень согласованности показаний разных типов маркеров. Если при переходе от генофонда к генофонду величины их гетерогенности по разным маркерам изменяются согласованно, то такая согласованность служит наилучшим подтверждением реальности выявленных различий в гетерогенности генофондов.

ПРЕДЫДУЩИЕ ИССЛЕДОВАНИЯ ОСНОВНЫХ ГАПЛОГРУПП

ГАПЛОГРУППЫ R1b и R1a

Возвращаясь к обзору работ по изменчивости Y-хромосомы в Европе, отметим вывод о том, что две основные гаплогруппы, суммарно вбирающие почти половину Y-генофонда Европы, распределены в направлении запад-восток. Из них гаплогруппа **R1b-M343** достигает частоты более 50% в Западной Европе и убывает к востоку, а гаплогруппа **R1a-M198** почти столь же часта в Восточной Европе и убывает к западу. Анализ другого типа маркеров Y-хромосомы – микросателлитных STR маркеров – также обрисовал западный и восточный домены как основные структуры в генофонде Европы [Roewer et al., 2005].

Как отмечалось выше, интерпретации «по ассоциации» должны делаться с большой осторожностью. Поэтому ассоциацию этих двух доменов с двумя основными мезолитическими рефугиумами (юго-западным «припиренейским» и юго-восточным «причерноморским») лучше считать одной из возможных гипотез, имеющей право на существование, но не доказанной. Например, можно было бы предложить альтернативную гипотезу, интерпретирующую эти два домена как наследников культур не палеолита, а позднего неолита - культуры колоковидных кубков и культуры шнуровой керамики. Подробный анализ разнообразия гаплогрупп в пределах гаплогруппы **R1b-M269**, формирующей западный домен, позволил предложить и третью гипотезу, возводящую формирование этого разнообразия ко времени неолита [Balaresque et al., 2010]. Можно выдвинуть и другие гипотезы, основанные на географических ассоциациях.

ГАПЛОГРУППЫ I1 и I2a

Две основные европейские гаплогруппы - **R1b** и **R1a** – распространены также в населении Средней Азии, Ближнего Востока, Индии, Северной Африки. Но две другие гаплогруппы – **I1-M253** и **I2a-P37** (ранее они обозначались **I1a** и **I1b**) – встречаются практически только в Европе, где они, вероятно, и возникли. Если **R1b-M343** и **R1a-M198** занимают запад и восток Европы, то **I1-M253** и **I2a-P37** преобладают, соответственно, на севере и юге Европы. Гаплогруппа **I1-M253**, распространенная в Скандинавии и на всех побережьях Балтийского моря, привлекает несколько меньшее внимание исследователей по причине сравнительно позднего заселения этого региона - только после окончательного отступления ледника. «Балканская» гаплогруппа **I2a-P37** является основной гаплогруппой на юго-востоке Европы, но редка на Ближнем Востоке, и поэтому ее нельзя считать принесенной оттуда на Балканы первыми неолитическими земледельцами (этот вопрос подробно рассмотрен ниже).

Три оставшиеся частые гаплогруппы **E-M96**, **J-M304** и **N1c-M178** встречаются не по всей Европе, а приурочены к ее субрегионам. Поэтому считается, что они маркируют более поздние и не столь масштабные волны миграций в Европу, которые не затрагивали население всего европейского субконтинента.

ГАПЛОГРУППА N1c

Гаплогруппа **N1c** (она же ранее обозначалась **N3** или **TAT**) ограничена северо-востоком Европы,

и главным образом встречается в популяциях, говорящих на языках финно-угорской группы. Но она также очень часта и в коренном населении Сибири. В течение почти десяти лет оставалось неясным, свидетельствует ли такая география о миграции из Европы на восток в Сибирь или же о миграции из Сибири на запад, в Европу.

В 2007 Рутси с соавторами [Rootsi et al., 2007] показали, что корни этой гаплогруппы находятся в кластере восточноазиатских линий Y-хромосомы, и поэтому наличие этой гаплогруппы в Европе следует интерпретировать как влияние азиатских генофондов. Авторы этой работы [Rootsi et al., 2007] предполагают несколько последовательных шагов этой растянувшейся на многие тысячелетия миграции из Северного Китая в Восточную Европу, которая началась в раннем голоцене и прошла через вторичную «экспансию» (резкое увеличение численности носителей гаплогруппы) на этом долгом пути.

В том же году Деренко с соавторами [Derenko et al., 2007] изучили изменчивость микросателлитных локусов в пределах этой гаплогруппы более подробно и дали оценки ее возраста. Они выделили два субварианта. Один из них был принесен в Европу, по их оценкам, около 6–10 тысяч лет назад, тогда как второй (менее частый) вариант был принесен в ходе меньшей по масштабу и более поздней миграции, имевшей место около 2–4 тысяч лет назад. Конечно же, к этим, как и к любым абсолютным оценкам времени, полученным по молекулярно-генетическим данным, следует относиться с осторожностью.

К мнению о миграции носителей гаплогруппы **N1c-M178** из Азии в Европу (а не наоборот), присоединяется и В.Н. Харьков [Харьков, 2012]. В проведенном им подробном анализе гаплотипов выделены кластеры, специфичные для отдельных народов и регионов Сибири, в частности показано своеобразие кластера, характерного для якутов (у которых частота этой гаплогруппы максимальна). Отметим, что существование этноспецифичных кластеров в пределах гаплогруппы **N1c-M178** было выявлено независимо ранее и в нашем исследовании [Балаганская и др., 2011]. В.Н. Харьков, в согласии с предыдущими исследованиями, рассматривает Восточную Европу как зону вторичной экспансии этой гаплогруппы, а высокое разнообразие гаплотипов объясняет меньшим действием, по сравнению с Сибирью, дрейфа генов, при большей интенсивности метисации.

Таким образом, все три исследования [Rootsi et al., 2007; Derenko et al., 2007; Харьков, 2012] согласуются друг с другом в том, что население северо-востока Европы имеет значительный (или даже преобладающий) генетический компонент, происходящий из верхнепалеолитической или мезолитической Южной Сибири и Центральной Азии.

Вывод об азиатских, а не европейских корнях гаплогруппы **N1c**, подтверждается и ее подробным филогенетическим деревом, полученным уже по результатам полного секвенирования Y-хромосомы [Karmin et al., 2015].

Это создает проблему для «двусистемного подхода» (совместное рассмотрение результатов по Y-хромосоме и мтДНК), являющегося как бы сокращенным вариантом развиваемого нами «полисистемного подхода». Дело в том, что по данным о мтДНК сибирские (восточноевразийские) гаплогруппы встречаются в Европе с крайне низкой частотой и только на самых восточных ее рубежах. Эти восточноевразийские гаплогруппы не образуют сколько-нибудь значимого генетического пласта ни в одном из регионов к западу от Урала. Например, когда типичная восточноевразийская гаплогруппа мтДНК **F** была обнаружена на хорватских островах (и с низкой частотой также на основной территории Хорватии), это рассматривалось как парадокс, и специальная статья [Tolk et al., 2001] была посвящена объяснению этого факта возможными миграциями, связанными с восточными торговыми путями Венецианской республики, в которую тогда входили эти острова. Данные широкогеномных маркеров также показывают лишь очень ограниченную представленность сибирского (североазиатского) и тем более восточноазиатского компонентов в генофонде Европы (глава 4). Вот почему наличие восточноевразийского компонента в генофонде значительной части Европы, столь явно выраженное в высокой частоте Y-гаплогруппы **N1c-M178**, является одним из немногих крупных противоречий между данными по Y-хромосоме и данными по мтДНК.

Представляется, что эту проблему можно снять, приняв во внимание, что генетическая граница между Европой и Азией находится намного восточнее географической границы (Уральских гор). Западная часть Средней и Центральной Азии, Южной Сибири (в особенности Алтай) и значительная часть Западной Сибири могут рассматриваться как генетически промежуточный регион в наше время [Балановская, Балановский, 2007] и как регион с преобладанием западноевразийского (можно сказать европеоидного) компонента в прошлом [Грехова и др., 1996; Балановская и др., 1997]. С этих позиций выглядит естественным, что Y-хромосомная гаплогруппа **N1-LLY22** (включая дочернюю гаплогруппу **N1c-M178**), несмотря на ее возникновение в Восточной Азии 20–30 тысяч лет назад, к неолитическому времени являлась типичной гаплогруппой и для европеоидного населения евразийских степей к западу от Алтая, и для монголоидного населения евразийских степей к востоку от Алтая. Именно из западной половины своего тогдашнего ареала гаплогруппа **N1c-M178** могла распространиться к северо-западу в ходе многочис-

ленных миграций, известных для этой территории. Тогда становится понятным, почему эти миграции не принесли в Европу восточноевразийские митохондриальные гаплогруппы: на территории к западу от Алтая, откуда шли эти миграции, даже в наши дни преобладают западноевразийские варианты мтДНК. Преобладание западных генетических вариантов на данной территории было тем более справедливо для того времени, поскольку тюркоязычные монголоидные народы начали расселяться на эти земли только около 2000 лет назад.

ГАПЛОГРУППЫ Е И J

Последние две частые в Европе гаплогруппы Y-хромосомы обозначаются **Е-М96** и **J-М304**. Обе они преобладают в населении Северной Африки и Ближнего Востока, а в Европе приурочены главным образом к Средиземноморью. Впрочем, далеко не все варианты этих гаплогрупп достигли Европы, а преимущественно только одна ветвь гаплогруппы **Е** (а именно **Е-V13** – основной субвариант гаплогруппы **Е-М35**) и две ветви гаплогруппы **J2-M172 (J2-M410 и J2-M241)**.

Если распространение гаплогруппы **J2-M410** следует особому тренду, две остальные гаплогруппы (**Е-М35**, **J2-M241**, а также рассмотренная выше гаплогруппа **I2a-P37**), сосредоточены почти исключительно на Балканах и не встречены со сколь угодно существенными частотами в соседних регионах. В работе [Battaglia et al., 2008] даны следующие оценки возраста этих гаплогрупп: для **Е-V13** – от 4 до 7,5 тысяч лет; для **J-M241** – от 3,5 до 6 тысяч лет; для **I2a-P37** – от 5,5 до 10 тысяч лет. Конечно, эти оценки были основаны на данных о STR-маркерах, и сейчас уже есть куда более надежные оценки по результатам полного секвенирования Y-хромосомы. Но во время выхода работы [Battaglia et al., 2008] других оценок не было, а эти оценки, пусть приблизительно, но соответствовали времени неолитизации юго-восточной части Европы, где в наши дни распространены эти гаплогруппы. Поэтому авторы этой работы предполагали, что автохтонное мезолитическое население Балканского полуострова (генетически восходящее к палеолиту) восприняло неолитический культурный комплекс (главным образом, земледелие) от своих ближневосточных соседей, сохранив генофонд палеолитического населения. Переход к заимствованным неолитическим технологиям привел к быстрому росту численности этих теперь уже земледельческих популяций, которые распространились по всему Балканскому полуострову и далее передали культурный неолитический комплекс (но не свой генофонд) мезолитическим популяциям соседних частей Европы. Такая модель хорошо объясняет, почему эти три гаплогруппы приурочены к территории Балкан, почему их частота убы-

вает по направлению к соседним регионам и почему их возраст близок к времени неолитизации Балкан (быстрый рост численности как раз и приводит к накоплению микросателлитного разнообразия, из которого вычисляется возраст гаплогруппы).

Однако внутренняя логика этой модели прямо противоположна той, которая была применена Ричардсом с соавторами [Richards et al., 2000] при анализе данных по митохондриальной ДНК. Ведь Ричардс с соавторами показали, что митохондриальные гаплогруппы, изменчивость которых была накоплена в Европе *in situ*, имеют *палеолитический* возраст. На основе этого они сделали вывод, что современные европейцы являются потомками *палеолитического* населения Европы [Richards et al., 2000]. А спустя восемь лет Баттаглия с соавторами показали, что Y-хромосомные гаплогруппы, изменчивость которых была накоплена в Европе *in situ*, имеют *неолитический* возраст. Но из этого противоположного факта они сделали прежний вывод – что современные европейцы являются потомками опять-таки *палеолитического* населения Европы [Battaglia et al., 2008].

Оба исследования опираются на значительный объем данных, и результаты обоих выглядят обоснованными. Но этот пример наглядно показывает необходимость применения более надежных и универсальных способов логической интерпретации молекулярно-генетических данных, по крайней мере, когда речь идет о столь сложных процессах, как неолитизация.

Что же касается данного конкретного случая, то возможный логический компромисс, на мой взгляд, состоит в следующем. Модель, предложенная в работе [Battaglia et al., 2008] в действительности подразумевает действие обоих процессов – и культурной, и демической диффузии (и распространение одной лишь традиции земледелия, и распространение самих земледельцев). Хотя авторы и не формулируют этого в явном виде, но их модель подразумевает, что культурные заимствования земледелия без миграции земледельцев имели место только *между* регионами (сначала между Малой Азией и Балканами, затем между Балканами и другими регионами Европы). При этом *внутри* регионов разумно предполагать наличие демической диффузии (распространение земледелия путем расселения увеличивающейся в численности популяции земледельцев).

Завершая на этом обзор основных предшествовавших работ по изменчивости Y-хромосомы в Европе, нужно отметить, что среди перечисленных и еще большего числа лишь упомянутых работ почти отсутствуют исследования, посвященные Европе в целом. По сути, лишь в работе [Rosser et al., 2000] анализируются закономерности структуры европейского генофонда, а все остальные работы

сосредоточены на отдельных регионах или отдельных гаплогруппах. Но за 15 лет, прошедших после статьи [Rosser et al., 2000], массив данных по изменчивости Y-хромосомы вырос во много раз – и по числу охваченных популяций и по числу выде-

ляемых гаплогрупп. Поэтому на последующих страницах предпринята попытка обобщающего анализа накопленных в мировой науке данных о структуре генофонда Европы, как она видится в зеркале Y-хромосомы.

2.2. ГЕНОГЕОГРАФИЯ ГАПЛОГРУПП Y-ХРОМОСОМЫ В ЕВРОПЕ

Анализ генофонда Европы начнем с рассмотрения геногеографии отдельных гаплогрупп. Отчасти это является продолжением предыдущего раздела, в котором кратко суммировалось то, что было известно о распространении гаплогрупп Y-хромосомы до нашего исследования. Но теперь мы рассмотрим современные, подробные карты распространения всех основных европейских гаплогрупп и тех их субвариантов, для которых уже известны частоты в ряде популяций Европы.

В нашей базе данных Y-base содержится почти вся опубликованная информация о частотах гаплогрупп Y-хромосомы в мире (и в Европе в частности), но далеко не вся эта информация пригодна для изучения географического распространения гаплогрупп. Как и при создании других геногеографических атласов, для создания карт распространения Y-хромосомы в Европе данные были тщательно отобраны и создан массив данных для карт Y-хромосомы в Европе. Рассмотрим, по каким критериям этот массив данных формировался.

ОРГАНИЗАЦИЯ ДАННЫХ

МАССИВ ДАННЫХ ДЛЯ КАРТ: ПОПУЛЯЦИИ

Для надежного определения частоты однородительских генетических маркеров (Y-хромосомы и мтДНК) требуется объем выборки в два раза больший, чем для аутосомных маркеров – поэтому нижняя граница выборки должна быть не менее 70-100 образцов. Маленькие выборки или объединялись с географически соседними выборками того же народа, или (если объединение было невозможно) не включались в анализ. В результате средний объем выборки в массиве данных для карт составляет $N=136$ образцов. Но делались и исключения из правила. По некоторым народам или же региональным группам в пределах народа, которые для полноты картины хотелось обязательно представить на карте, были доступны лишь небольшие выборки. Тогда, скрепя сердце, приходилось включать в массив для карт и выборки меньшего размера. Но такие исключения делались редко – из 251 картографированных популяций лишь десятая часть (29 популяций) имеют объем выборки ниже $N=50$, а подавляющее большинство (192 выборки) имеют объем $N=70$ и выше.

Если для какого-либо народа были данные по его региональным популяциям, то для картографирования не использовались выборки этого народа, для которых географическая привязка неизвестна (если авторы при публикации указали лишь принадлежность к народу).

В итоге массив для карт включил данные по 251 популяциям (выборкам), и суммарное число образцов во всех выборках составило 34 294 образцов.

Стоит отметить, что значительная часть этих данных получена благодаря исследованиям нашего коллектива под руководством автора этих строк: 45 популяций и 4 880 образцов. Это наши данные по русским популяциям [Balanovsky et al., 2008], украинским, белорусским и некоторым дополнительным русским популяциям [Kushniarevich et al., 2015], популяциям Северного Кавказа [Balanovsky et al., 2011], а также наши еще неопубликованные данные по популяциям крымских и казанских татар, мокши и эрзи. (Данные по частотам гаплогрупп из трех опубликованных статей приведены в таблицах 2.1, 2.2, 2.3.) Получается, что из всего массива данных по изменчивости Y-хромосомы в Европе каждая пятая популяция (на рис. 2.1 обозначены синими кружками) и каждый седьмой образец изучены нашим коллективом, хотя исследованиями в этой области занималось и занимается несколько десятков лабораторий по всей Европе. Если добавить, что Европа – самый изученный регион мира, то это отчасти объясняет, почему именно нашим коллективом предпринят обобщающий анализ генофонда Европы. Для построения карт использованы также данные Y-base не только по европейским популяциям, но и по популяциям близлежащих регионов, входящих в картографируемый ареал.

Кроме 251 популяций Европы, изученных по панелям гаплогрупп, охватывающих весь их спектр, еще около 200 популяций были изучены по субвариантам только одной какой-то гаплогруппы. В основном это статьи [Myres et al., 2010; Underhill et al., 2010, 2014], посвященные гаплогруппам **R1b** и **R1a**, и ряд статей по гаплогруппам

N, E, и J. Как правило, выборки, использованные в этих статьях для углубленного генотипирования отдельных гаплогрупп, раньше уже изучались по всему спектру гаплогрупп. Рис 2.1. дает представление о плотности покрытия Европы и смежных регионов популяциями, изученными не по одной гаплогруппе, а по всему их спектру с обычным уровнем филогенетического разрешения.

МАССИВ ДАННЫХ ДЛЯ КАРТ: ГАПЛОГРУППЫ

Это понятие «обычного филогенетического уровня» изученности гаплогрупп очень условно, и уровень этот растет с каждым годом. Если в 2000 году число гаплогрупп в обычной статье не дотягивало до десятка, то в 2010е годы большинство статей используют панели из 40–70 гаплогрупп. Конечно, начиная примерно с 2013 года, число известных гаплогрупп, обнаруженных благодаря полному секвенированию, исчисляется уже тысячами.

Но обнаружение их – только первый шаг, который требует множество следующих шагов и усилий по анализу большого массива самых разных популяций на наличие новых гаплогрупп. И пока не проведено определение их частот в широком спектре популяций, значение этих новооткрытых гаплогрупп для исследований генофонда невелико.

Объединяя в одной таблице данные из десятков статей разных лет, неизбежно оказывается, что некоторые гаплогруппы изучены в очень большом числе популяций, а некоторые – лишь для немногих популяций. К счастью, для картографирования это не представляет большой проблемы (уж в двух-трех десятках популяций изучены почти все отобранные для картографирования гаплогруппы, а такого числа опорных точек обычно достаточно для выявления основных трендов). Но все же при рассмотрении карт стоит иметь в виду, что подробность исходных данных для разных гаплогрупп неодинакова. Стоит иметь в виду и то, что любая

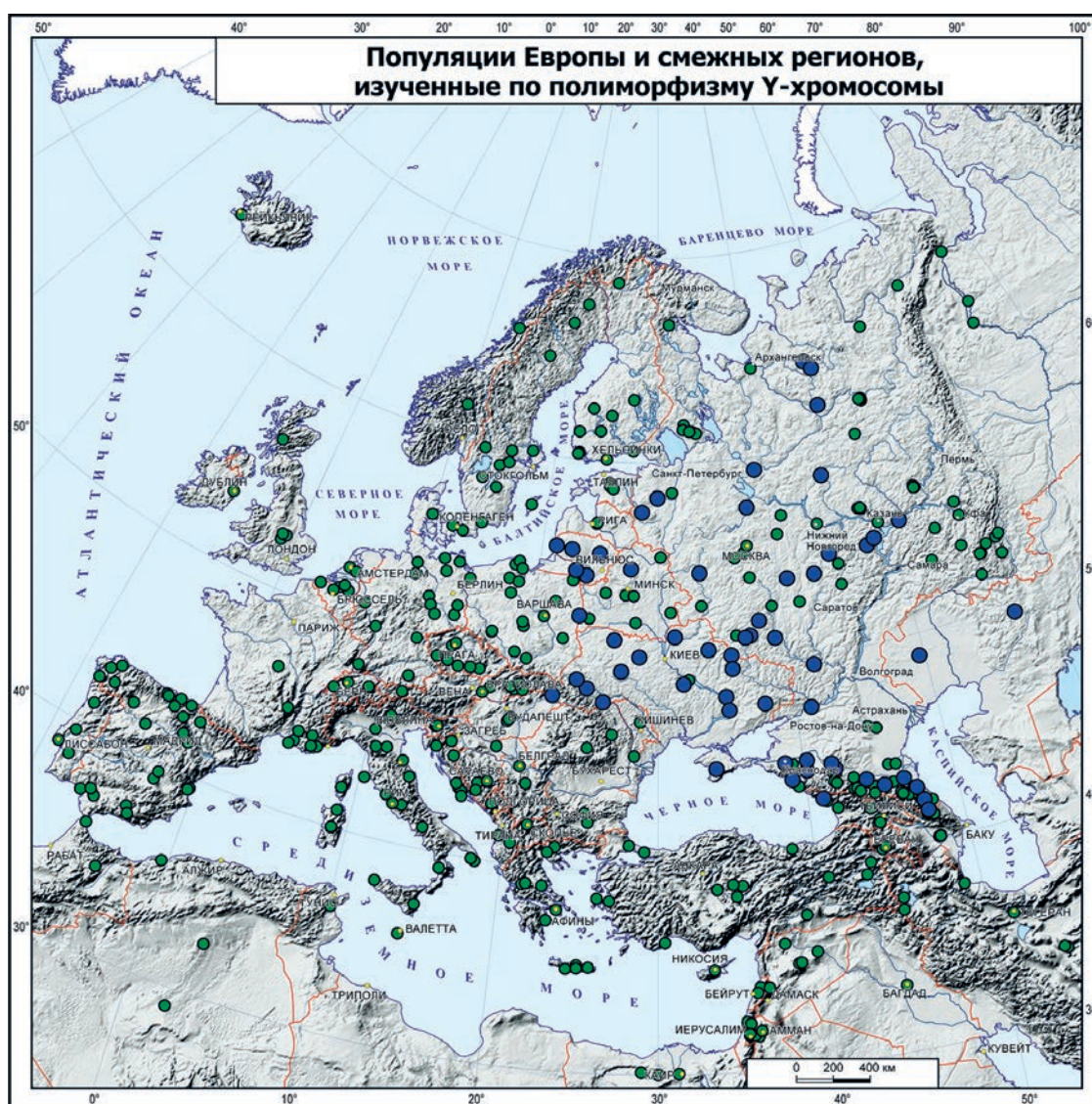


Рис. 2.1. Географическое положение популяций Европы и смежных регионов, изученных по полиморфизму Y-хромосомы (изученные нами популяции обозначены синими кружками).

Таблица 2.1. Наши данные о частотах гаплогрупп Y-хромосомы в русских популяциях (опубликованы в [Balanovsky et al., 2008]).

E-	E-	F* ¹	G1* ²	G2a	I*	I1	I2a1	I2a2a	J1	J2	K* ³	T	N* ⁴	N1b	N1c	Q	R1a	R1b1a1	R1b1a2	R2
M78	M123	M89	M285	P15	M170	M253	P37	M223	M267	M172	M9	M70	M231	P43	TAT	M242	SRY1532	M73	M269	M124
0	0	1.9	0	0	0	0	0	0	0	0	0	0	0	7.4	46.3	0	44.4	0	0	0
0	0	0	0	0	0	0.9	4.4	0	0.9	0	0	0	0	15.8	23.7	0.9	39.5	0	14	0
0	0	1.1	0	2.2	3.3	12.1	9.9	0	0	5.5	0	0	0	3.3	36.3	0	19.8	0	6.6	0
0.8	0	0	1.7	0.8	0	11.6	8.3	1.7	0	0.8	0	0	0	3.3	35.5	1.7	33.1	0	0.8	0
0.2	0	0.8	0.4	0.8	0.8	6.2	5.7	0.4	0.2	1.6	0	0	0	7.5	35.5	0.7	34.2	0	5.4	0
5.8	0	1.9	0	0	0	11.5	11.5	3.8	3.8	3.8	0	0	0	0	13.5	0	32.7	1.9	9.6	0
4.1	0	1.4	0	0	0	2.7	8.2	4.1	0	4.1	0	0	0	0	11	1.4	56.2	0	6.8	0
1.8	1.8	0	0	0	1.8	3.5	10.5	0	0	0	3.5	0	1.8	1.8	15.8	0	52.6	0	5.3	0
4	0	0	0	0	1.3	6.7	9.3	0	1.3	1.3	0	0	0	0	28	0	45.3	0	2.7	0
7.5	0	0.9	0	0	0	1.9	10.3	0	0	2.8	1.9	2.8	0	0.9	13.1	0	45.8	0	11.2	0
4.6	0.4	0.8	0	0	0.6	5.3	10	1.6	1	2.4	1.1	0.6	0.4	0.5	16.3	0.3	46.5	0.4	7.1	0
0.9	0	0	0	0	0	8.2	13.6	0.9	0	0.9	0	2.7	0	0.9	4.5	0.9	62.7	0	3.6	0
2.2	0	0	0	0	0	4.4	17.8	0	0	2.2	0	2.2	0	0	13.3	0	55.6	0	2.2	0
1	0	1	0	2.1	1	3.1	16.7	0	2.1	1	0	0	0	0	6.3	0	59.4	0	5.2	1
0.7	0	0	0	0.7	0	3.5	12.6	0.7	0	4.2	0.7	1.4	0	0.7	11.9	0	59.4	0	2.8	0
3.3	0	0	0	1.1	1.1	4.4	16.5	2.2	0	4.4	0	1.1	0	1.1	6.6	1.1	47.3	0	8.8	0
2.4	0	0	0	19.2	0	5.4	7.6	0	3.2	5.6	4.0	0.8	0	0	6.4	0.8	38.4	0	5.6	0
1.75	0	0.166	0	3.9	1.3	4.8	14.2	0.6	0.9	3.0	0.8	1.4	0	0.5	8.2	0.5	53.8	0	3.8	0.16

Примечания:

¹ К гаплогруппе F* относятся образцы с производным аллелем маркера M89, но с предковыми аллелями по маркерам M9, M201, M170, 12f2.² К гаплогруппе G1* относятся образцы с производным аллелем маркера M285, но с предковым по маркеру P20.³ К гаплогруппе K* относятся образцы с производным аллелем по маркеру M9, но с предковыми по маркерам M70, 92R7, M231.⁴ К гаплогруппе N* относятся образцы с производным аллелем по маркеру M231, и с предковыми по маркерам TAT, P43, M128.

гаплогруппа состоит из многих – в том числе еще неизвестных – субветвей. Поэтому показанная на карте частота, как обычно в геногеографии, сформирована не каким-то одним, а суммой множества частных популяционных событий.

Для картографирования отбирались гаплогруппы, составляющие заметную часть генофонда хотя бы в некоторых популяциях Европы и при этом изученные, как только что было сказано, в нескольких десятках популяций. Этим критериям удовлетворили 40 гаплогрупп: **E1b1a1-DYS271, E1b1b1a1b-**

L142.1, E1b1b1a1c-V22, E1b1b1b1-M81, E1b1b1c-M123, G2a1-P16, G2a3a-M406, G2a3b1-P303, I1-L118, I2a-P37.2, I2b-L35, J1-L255, J1c3-P58, J2-L228, J2a-L152, J2a3b-M67, J2a3h-L207.1, J2b-L282, L-M11, N1b-P43, N1c-M46, O3-M122, Q-M242, R1a1-L120, R1a1a1g-M458, R1a-M558, R1a-Z284, R1a-Z93, R1b1a1-M478, R1b1a2-L265, R1b-L10, R1b1a2a1a1a-M405, R1b1a2a1a1a1-M467, R1b1a2a1a1b-P312, R1b1a2a1a1b2a-M153, R1b1a2a1a1b2b1-M167, R1b1a2a1a1b4b-M222, R1b1c-V88, R2a-L266, T-L206.

ГАПЛОГРУППА R1a И ЕЕ ВАРИАНТЫ

ГАПЛОГРУППА R1a (БЕЗ ПОДРАЗДЕЛЕНИЯ НА СУБВЕТВИ)

Как указывалось в обзоре предшествовавших работ, гаплогруппа **R1a** является одной из двух наиболее частых гаплогрупп Европы. Более того, она является и наиболее широко распространенной гаплогруппой в Евразии в целом. Для ее генотипирования используются разные маркеры – M198, M17, SR1542, филогенетическая разница между которыми почти отсутствует, и известно множество

других маркеров, определяющих эту гаплогруппу, поэтому для ее обозначения может использоваться любая из этих них. В массиве данных для карт она фигурирует как **R1a-L120**.

На рис. 2.2 можно видеть четкий тренд географического распространения **R1a-L120**: хотя она распространена по всей Европе, но обширная зона максимальных частот находится в Восточной Европе, причем в этих популяциях она составляет почти половину генофонда. На востоке зону максимума ограничивает Волга, причем на всем ее

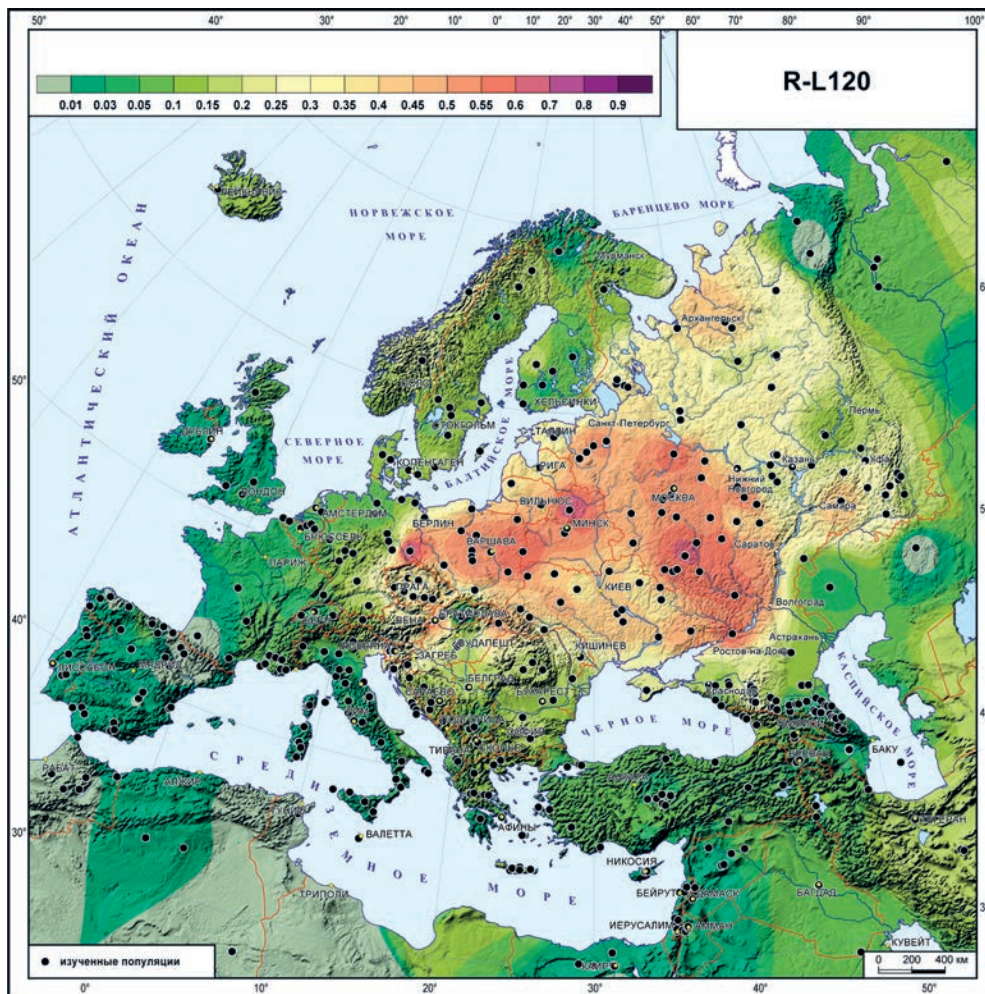


Рис. 2.2. География гаплогруппы R1a-L120 (R1a) Y-хромосомы в Европе.

протяжении – от Верхней Волги до низовий. На юге зона максимума доходит почти до Черного моря, на северо-западе – до Балтийского моря, а на юго-западе ограничивается Карпатами и Альпами. Но наиболее любопытны границы этой области на западе: хотя никаких географических барьеров там нет, зона высоких частот **R1a** все равно довольно четко ограничена, и падение частоты на каких-то двухстах-трехстах километрах достигает почти 30% (от примерно 35–45% у поляков и сорбов до 10–15% у немцев). Анализу этой «польско-немецкой генетической границы» были посвящены несколько статей [Ploski et al., 2002; Kayser et al., 2005; Rebala et al., 2013]¹.

Я не упомянул еще границу зоны максимальных частот **R1a** на севере – хотя перепад частот там не так выражен, как на западе (не от красных к зеленым, а только от красных к желтым тонам, то есть процентов на пятнадцать), но северная граница тоже выявляется на карте вполне рельефно. В данном случае она проходит внутри ареала одного народа – русского – маркируя отличия генофонда Русского Севера от остальных русских популяций. Эти желтые тона умеренных частот занимают не только Русский Север, но и все пространство Восточной Европы между Волгой и Уралом, и лишь за Уральским хребтом частоты **R1a** падают до столь же низких значений (зеленые тона на карте), что и в Западной Европе, Скандинавии, Балканах и Передней Азии.

СУБВАРИАНТЫ ГАПЛОГРУППЫ **R1a**

В пределах такого огромного ареала гаплогруппы **R1a-L120**, несомненно, существует почти столь же огромное число субвариантов этой гаплогруппы, причем многие из них могут иметь четкие и сравнительно небольшие географические ареалы. Но в течение долгого времени субтипы в пределах **R1a** были неизвестны. Пока, благодаря работам лаборатории Peter Underhill в Стэнфордском университете, не были открыты сначала первые два, а потом и еще несколько субвариантов. Конечно, в настоящее время число известных субвариантов **R1a** исчисляется сотнями: каждый образец, для которого проведено полное секвенирование всего генома или Y-хромосомы, образует свою ветвь на филогенетическом древе. Но пока хотя бы для некоторых из этих вариантов не проведен скрининг их распространенности в популяциях и не определены частоты их встречаемости в разных популя-

циях Европы, построение геногеографических карт невозможно. Сейчас об ареалах распространения этой субветви можно рассуждать лишь гадательно, основываясь на порой зыбких сведениях о местах происхождения тех единичных образцов, по которым каждая из этих сотен ветвей была открыта. Интерес к этой гаплогруппе столь велик, что, конечно, такой популяционный скрининг будет проведен разными лабораториями мира (в том числе и нашим коллективом) в самые ближайшие годы, и тогда геногеография субвариантов **R1a** станет на порядок подробнее. Но на данный момент мы ограничены теми маркерами, по которым проведен массовый популяционный скрининг в статьях [Underhill et al., 2010; 2014] и в работах нашего коллектива: это три основных маркера европейской ветви **R1a** и маркер азиатской ветви **R1a**.

“АЗИАТСКАЯ ВЕТВЬ” **R1a-Z93**

Деление на европейскую и азиатскую ветви проявляется на филогенетическом древе **R1a** очень четко [Karmin et al., 2015]. Маркером азиатской ветви является Z93. В Европе эта ветвь приурочена только к ее юго-восточным окраинам (Предкавказье, Поволжье) и почти отсутствует на основной территории Европы (рис. 2.3; благодарю А. Штрунова и В. Таганкина, заметивших на карте ошибку, которая теперь исправлена). Перейдем теперь к рассмотрению европейских ветвей.

“ЕВРОПЕЙСКАЯ ВЕТВЬ” **R1a-M458**

Первый из них, **R-M458**, выделяет центрально-европейский субвариант в общем массиве гаплогруппы **R1a**, занимающей половину Европы. Поэтому после открытия этого маркера мы специально генотипировали его во многих славянских и северокавказских популяциях, что позволило впервые построить подробную карту его распространения с учетом Восточной Европы и Кавказа (рис. 2.4).

Карта демонстрирует наиболее высокие частоты этого маркера у западных славян (поляков и чехов), где он составляет четверть генофонда, и несколько сниженные частоты в окружающих популяциях. Распространение этого маркера на запад ограничено прилежащими популяциями немцев (вероятно, ассимилировавших западнославянские группы, ранее населявшие эти территории). Зато на восток зона высоких частот гаплогруппы **R-M458** распространяется далеко, захватывая

¹ Отметим в скобках, что объективная приуроченность данной гаплогруппы к славянским популяциям на данной весьма ограниченной территории, к сожалению, стала одним из поводов к столь распространенному на любительских форумах в интернете навешиванию «славянского» ярлыка на всю гаплогруппу **R1a**. Но такая привязка гаплогруппы – на всем протяжении ее евразийского ареала в пространстве и несоизмеримого со славянством протяженного существования во времени – к одной лишь славянской лингвистической группе, увы, уже не только граничит со лженаукой, но и переходит эту грань.

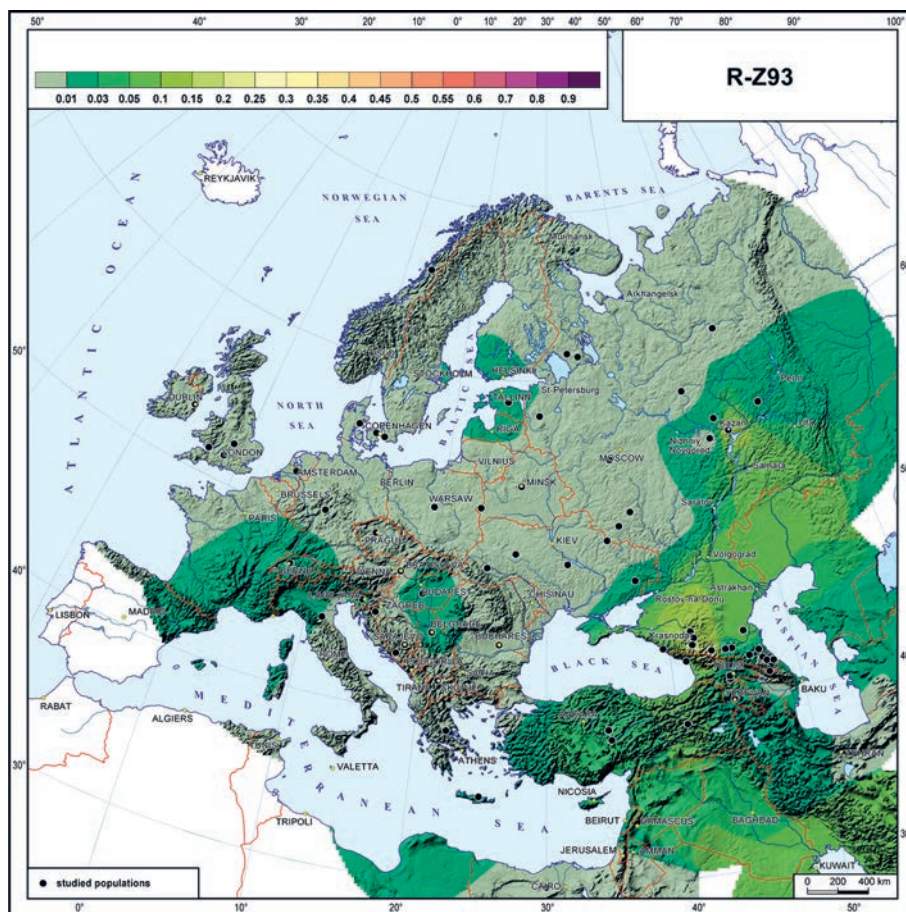


Рис. 2.3. География гаплогруппы R1a-Z93 Y-хромосомы в Европе.

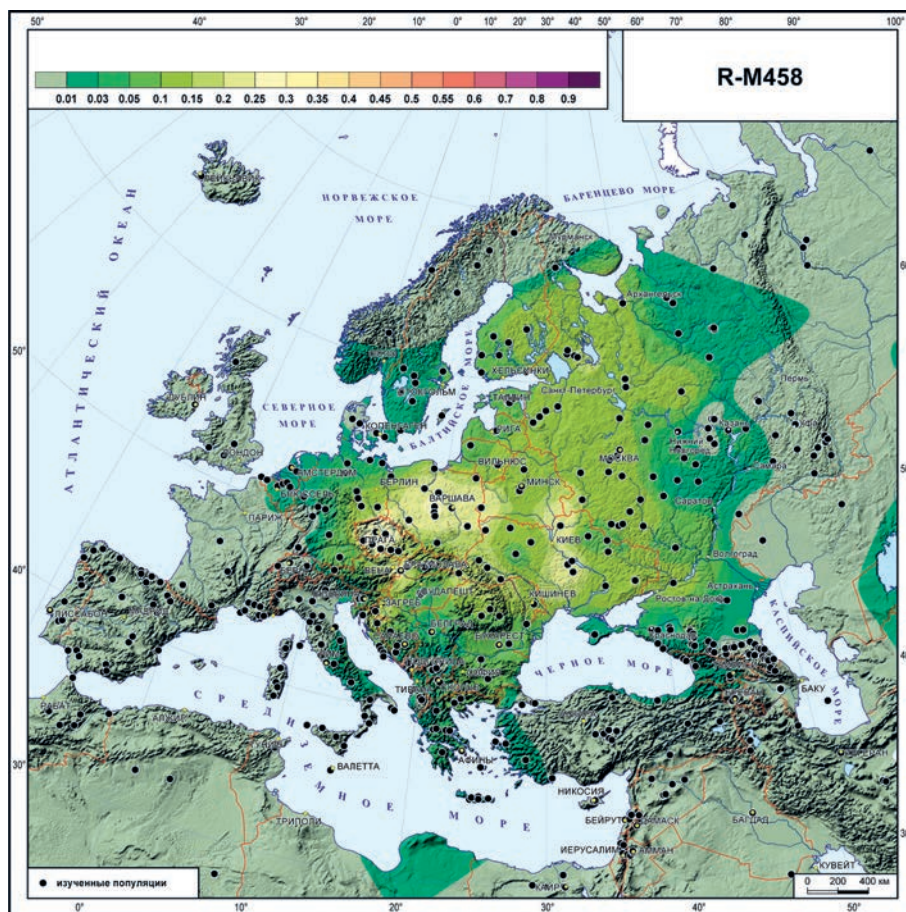


Рис. 2.4. География гаплогруппы R1a-M458 Y-хромосомы в Европе.

весь ареал восточных славян. Характерно, что эта восточная экспансия более выражена в лесостепной зоне (у украинцев), чем в лесной (белорусы и русские). Еще более любопытным представляется ограниченность экспансии на юг: у южных славян и неславянских народов Балкан (венгры, румыны) частота этой гаплогруппы мала, причем прослеживается соответствие генетической «границы» высоких частот **R-M458** и географической границы Карпат. К северу от этого горного хребта расположена зона максимальных частот гаплогруппы, а к югу от Карпат частота резко снижается. По-видимому, Карпаты выступили в роли географического барьера для потока генов, препятствуя распространению этой гаплогруппы на Балканы. Отметим, что в отличие от **R1a** в целом, встречающейся хотя бы с низкими частотами в любом регионе Европы, **R-M458** распространена только по всей Восточной Европе, соседних территориях Западной Европы и на Балканах, но практически полностью отсутствует на большей части Западной Европы – Пиренейском полуострове, Франции, Британских островах, значительной части Аппенинского и Скандинавского полуостровов, а также отсутствует за пределами Европы – и в Передней Азии, и в Сибири.

“ЕВРОПЕЙСКАЯ ВЕТВЬ” R1a-M558

Второй европейский субвариант **R1a** обозначается по названию маркера **R-M558** (рис. 2.5.). Филогенетически он братский (для мтДНК в таких случаях принято говорить «сестринский», но тут речь идет о Y-хромосоме) для **R-M458**. Оба этих маркера распространены сходным образом: **R-M558** тоже распространен преимущественно в Восточной Европе, тоже реже встречается на Балканах, заходит с низкими частотами в соседние области Западной Европы, и так же, как и **R-M458**, практически отсутствует за пределами Европы. Ареалы этих субтипов перекрываются и по зоне их максимальных частот у западных и восточных славян. Основное различие между паттернами распространения **R-M458** и **R-M558** состоит только в большей приуроченности к западной или восточной части их общего ареала.

R-M458, в отличие от **R-M558**, имеет склонность к западу ареала. У западных славян встречается в среднем с частотой 25% (т.е. четверть генофонда), а у восточных славян в среднем встречается с частотой 15%, хотя и повышается до 23% у западных белорусов и у некоторых украинских популяций (но не западных, а центральных).

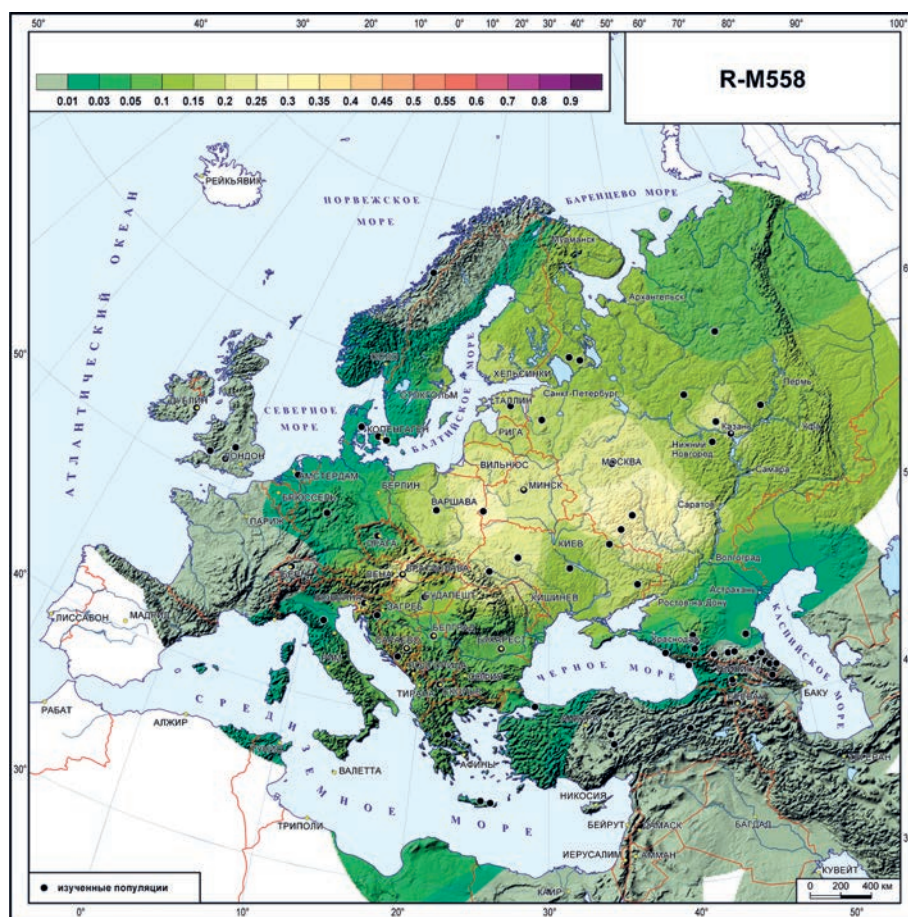


Рис. 2.5. География гаплогруппы R1a-M558 Y-хромосомы в Европе.

R-M558 имеет склонность к востоку их общего ареала. Он встречается с частотой выше 30% в обеих изученных по этому маркеру популяциях южных русских; составляет 20–25% у белорусов, поляков, словаков, западных украинцев; снижается ниже (или значительно ниже) 20% у центральных русских, чехов, центральных и восточных украинцев и чехов. Впрочем, **R-M558** изучен в значительно меньшем числе популяций, чем **R-M458**, и поэтому карта его распространения может еще значительно уточниться в будущем. Мы видим, что даже это основное различие между паттернами распространения **R-M458** и **R-M558** выражается в небольших различиях частот и имеет довольно сложную географию.

Этот пример ясно показывает, что повышение филогенетического разрешения далеко не всегда приводит к повышению географического разрешения: хотя в пределах европейской ветви **R1a** и удалось выделить два субтипа, но их географическое распространение очень сходно. По всей видимости, они возникли в родственных популяциях и распространялись совместно, в ходе одних и тех же миграций. Или же, возникнув даже в разных частях ареала популяций, интенсивно обменивавшихся миграциями друг с другом, они распространились в пределах всего этого ареала. Конечно, повышать филогенетическое разрешение необходимо, и, как правило, выделяемые субтипы имеют более четкую географическую привязку, чем ветвь в целом; субтипы следующего уровня имеют еще более узкую зону распространения в пределах зон субтипов предыдущего уровня и так далее. Но эта географическая четкость зависит не только и даже не столько от уровня филогенетического разрешения, сколько от структуры генофонда, от исторически заданного соотношения изоляции и метиса-

ции, то есть, в терминах популяционной генетики, от баланса дрейфа генов и миграций.

Если этот баланс сдвинут в сторону изоляции (как на Кавказе или в Сибири), то частоты гаплогрупп будут резко различаться между популяциями, гаплогруппы будут иметь географически узкие ареалы, мало перекрывающиеся друг с другом, и гаплогруппы будут достигать очень высоких частот в «своих» популяциях. Если же баланс сдвинут в сторону миграций, то частоты будут различаться не столь драматически, ареалы будут довольно широкими и перекрывающимися друг с другом, а субгаплогруппы достигать не слишком высоких частот, потому что остальная часть генофонда будет представлена другими гаплогруппами, ареалы которых заходят на эту же территорию. Все эти черты хорошо видны на примере **R-M458** и **R-M558**, потому что сдвиг баланса в сторону не изоляции, а миграций особенно характерен именно для Европы.

ЕВРОПЕЙСКАЯ ВЕТВЬ **R1a-Z284**

Зато еще один субвариант **R1a** – гаплогруппа **R-Z284** (рис. 2.6.) – имеет совершенно другой паттерн, который можно назвать «скандинавским» или «североевропейским». По имеющимся на настоящий момент данным [Underhill et al., 2014] его частота составляет 20% в Норвегии, 7% в Дании, и по 3% в Англии и Швеции, а за пределами Северной Европы **R-Z284** встречена лишь у единичных образцов. Эти частоты наверняка скорректируются в ходе изучения новых популяций и увеличения объемов выборок уже изученных народов, но общая приуроченность к популяциям Северной Европы несомненна уже сейчас.

ГАПЛОГРУППА **R1b** И ЕЕ ВАРИАНТЫ

ГАПЛОГРУППА **R1b** (В ЦЕЛОМ)

Гаплогруппа **R1b**, «братская» по отношению к гаплогруппе **R1a** и вторая среди наиболее частых в Европе, имеет тоже широкий ареал (рис. 2.8), однако основная зона её максимальных частот приходится на Западную Европу. Важным вкладом в изучение генофонда Европы стало открытие новых информативных SNP маркеров в пределах одной из ветвей гаплогруппы **R1b** – гаплогруппы **R-M269** (на картах обозначена с помощью филогенетически близкого маркера как **R-L265**). Это масштабное исследование выполнено большим международным коллективом с участием автора и опубликовано в [Muges et al., 2010].

Карта распространения гаплогруппы **R1b** в целом (**R-L10**, рис. 2.8) ярко показывает, что это основной компонент, составляющий более поло-

вины Y-хромосомного генофонда Западной Европы. Лишь население южной части Аппенинского полуострова несет сниженные частоты этой гаплогруппы, поскольку у них преобладают средиземноморские гаплогруппы, да еще Скандинавия по географии гаплогрупп относится не к Западной Европе, а формирует самостоятельный северный домен, в котором преобладает иная гаплогруппа – **I1**.

Гаплогруппа **R1b** даже в большей степени свойственна Западной Европе, чем **R1a** свойственна Восточной, поскольку **R1b** во многих популяциях (Британские острова, баски и ряд других популяций Пиренейского полуострова) составляет даже не половину, а более двух третей генофонда. В целом, тренд гаплогруппы **R1b** – максимальные частоты у басков и постепенное снижение частоты к востоку от Пиренейского полуострова – очень напоминает географию классического маркера **Rh-**

d (резус-фактор). Всплеск частоты на востоке Европы отражает повышенную частоту **R1b** у некоторых популяций башкир, известную из работы [Лобов, 2009]. Впрочем, нашим коллективом сейчас завершается подробное исследование разнообразия Y-хромосомы в популяциях башкир (более 1000 образцов). В этом исследовании выявлено, что частота **R1b** велика только в немногих родовых группах, тогда как для большинства родовых групп башкир характерны другие гаплогруппы, а частоты **R1b** у них невелики. Поэтому всплеск **R1b** на востоке Европы отражает не общеевропейскую закономерность, а лишь локальные закономерности происхождения отдельных родовых групп степных популяций Евразии.

Хотя **R1b** достигает высоких частот в основном в Западной Европе, карта (рис. 2.8) показывает, что она (как и **R1a**) распространена повсеместно, занимая весь картографированный ареал и уходя далеко за его пределы по Евразии. Считается, что **R1b** имеет переднеазиатское происхождение, и разные ее ветви распространились по различным уголкам Старого Света, но особенно «повезло» лишь той ветви, которая распространилась в Европе и достигла в ней столь значительных частот.

СУБВАРИАНТЫ ГАПЛОГРУППЫ R1b

Карты на рис. 2.9, 2.10 и 2.11 показывают три основные ветви **R1b** (Рис. 2.7), которые – со множеством оговорок – можно назвать африканской, степной евразийской и западноевропейской гаплогруппами. Действительно, когда мы смотрим на карту **R1b** в целом (рис. 2.8), то видим ее наличие в том числе и в Северной Африке.

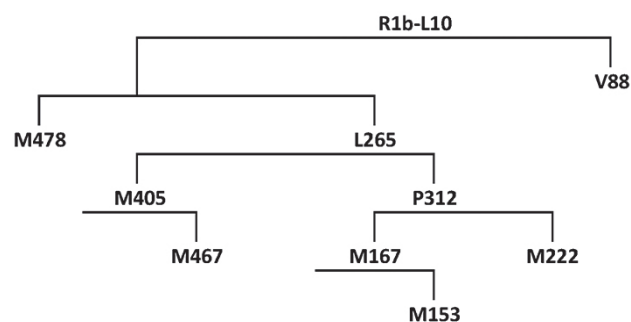


Рис. 2.7. Филогенетическое дерево гаплогруппы R1b-L10. На рисунке показаны только ветви, которые подробно рассматриваются с помощью геногеографических карт; структура схемы соответствует полному дереву гаплогруппы R1b-L10 по версии ISOGG на октябрь 2015 г.

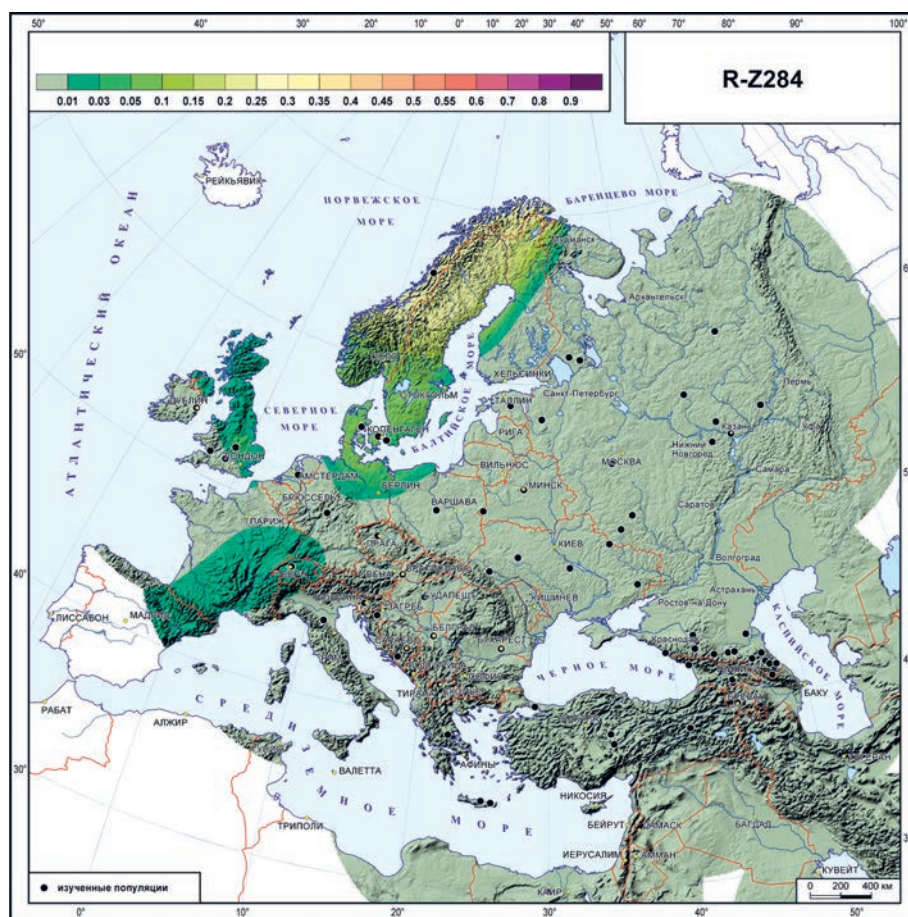


Рис. 2.6. География гаплогруппы R1a-Z284 Y-хромосомы в Европе.

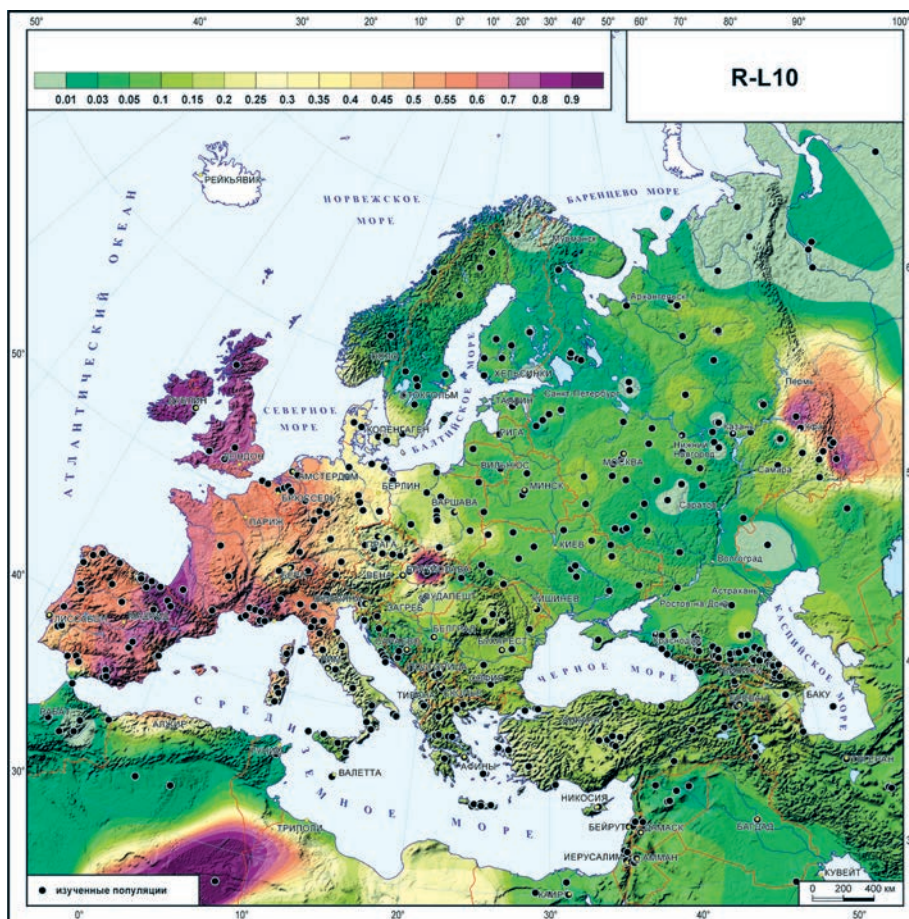


Рис. 2.8. География гаплогруппы R1b-L10 Y-хромосомы в Европе.

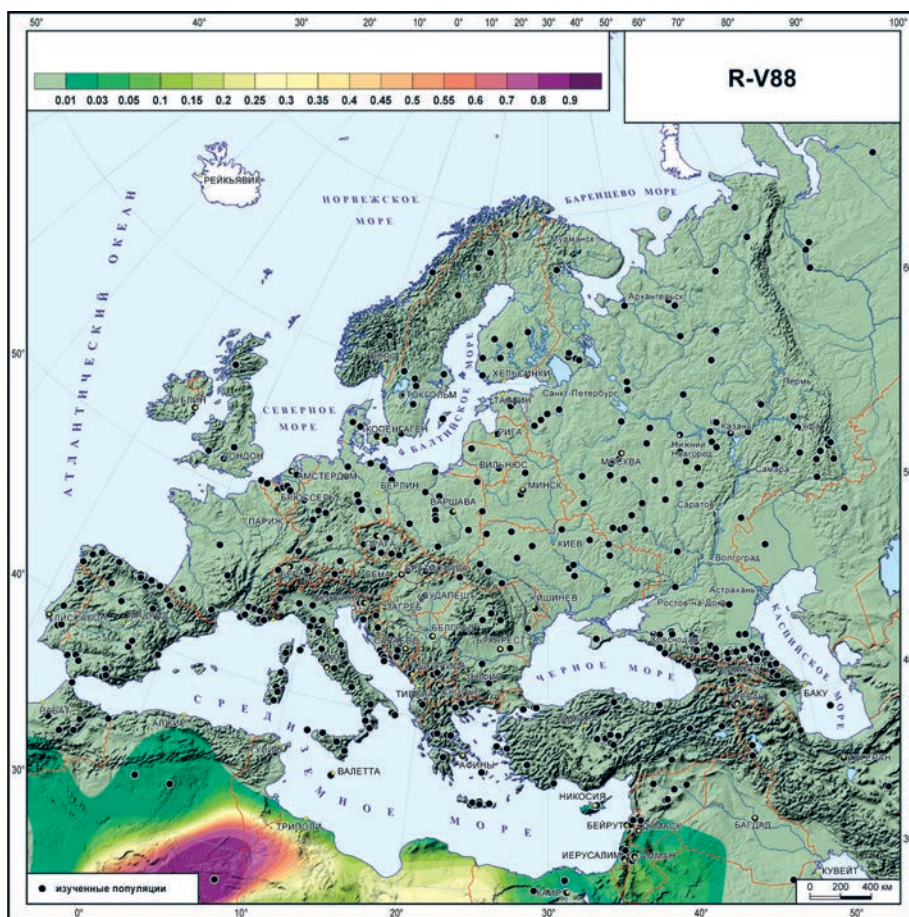


Рис. 2.9. География гаплогруппы R1b-V88 Y-хромосомы в Европе.

“АФРИКАНСКАЯ ВЕТВЬ” R1b-V88

Карта **R-V88** (рис. 2.9) как раз и показывает частоты пусть и невысокие (1–6%, с повышением до 26% только в одной популяции), но зато по всей Северной Африке (от Марокко до Египта), и почти полное отсутствие этой гаплогруппы в Европе и в Азии. То есть «африканская» часть **R1b** – это почти исключительно ветвь **R-V88**. Своего максимума она достигает далеко за пределами картографируемого ареала – в Центральной Африке [Stuciani et al., 2010].

“СТЕПНАЯ ВЕТВЬ” R1b-M478

Карта следующей ветви – **R-M478** – показывает ее приуроченность к степным популяциям Евразии (рис. 2.10). Зона ее максимальных частот также находится по большей части за пределами картографируемого ареала – на пространствах от Урала до Алтая. Впрочем, как можно видеть на карте, отдельными «языками» эта гаплогруппа доходит и до Кавказа, оставаясь приуроченной к степным областям Южного Урала и Предкавказья, и составляя только 1–10%, редко в каких популяциях до 20% генофонда.

“ЗАПАДНОЕВРОПЕЙСКАЯ ВЕТВЬ” R1b-L265

Третья из основных ветвей гаплогруппы **R1b** – гаплогруппа **R-M269**, или, как её здесь обозначаем на картах, **R-L265** – составляет основную (по частоте и ареалу) часть **R1b**. Поэтому карта ее распространения (рис. 2.11) в целом вторит карте **L10**, лишь без африканской части ее ареала. Эта гаплогруппа подразделена на множество субгаплогрупп, многие из которых специфичны для той или иной территории в пределах ареала ветви. Остановимся на двух основных ветвях внутри **R-L265**: центральноевропейской гаплогруппе **R-M405** и её пиренейско-британском «брате» **R-P312** со всеми их «сыновними» ветвями (рис. 2.12–2.17).

“ЦЕНТРАЛЬНОЕВРОПЕЙСКАЯ ВЕТВЬ” R1b-M405
В ЦЕЛОМ

Гаплогруппа **R-M405** (рис. 2.12) встречается практически по всей Западной, Центральной и Северной Европе (за исключением южных областей Пиренейского и Аппенинского полуостровов, а также Балканского полуострова – на юге, Финляндии и Кольского полуострова – на севере), а также и в

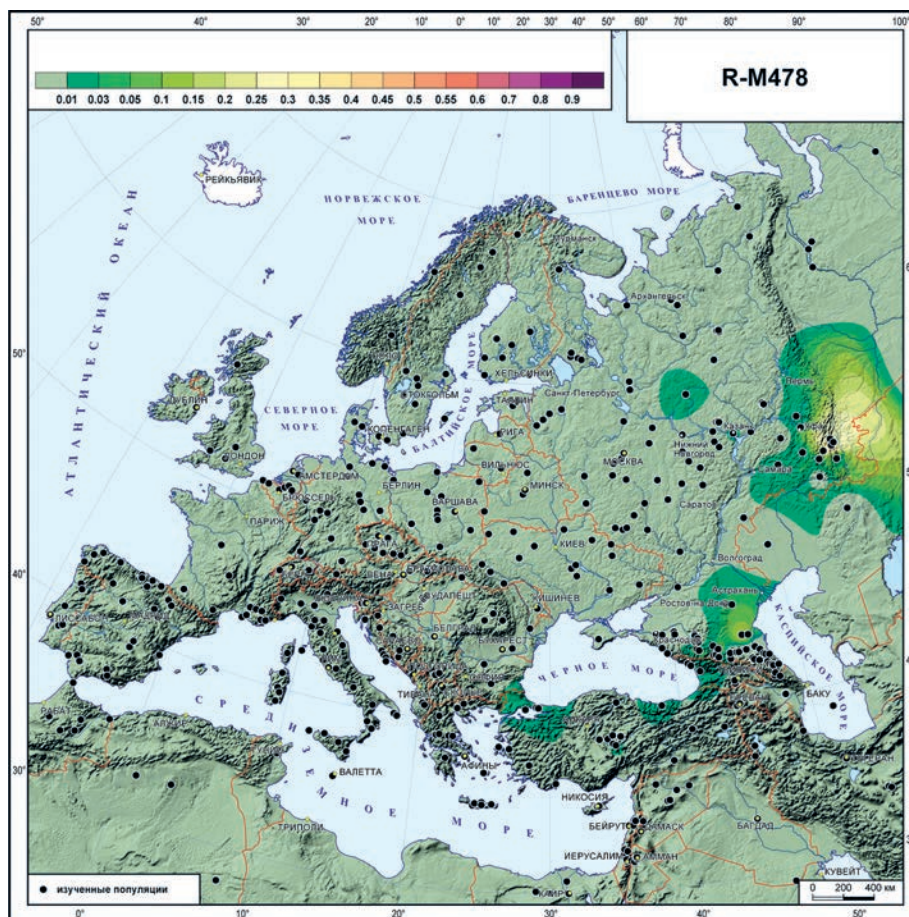


Рис. 2.10. География гаплогруппы R1b-M478 Y-хромосомы в Европе.

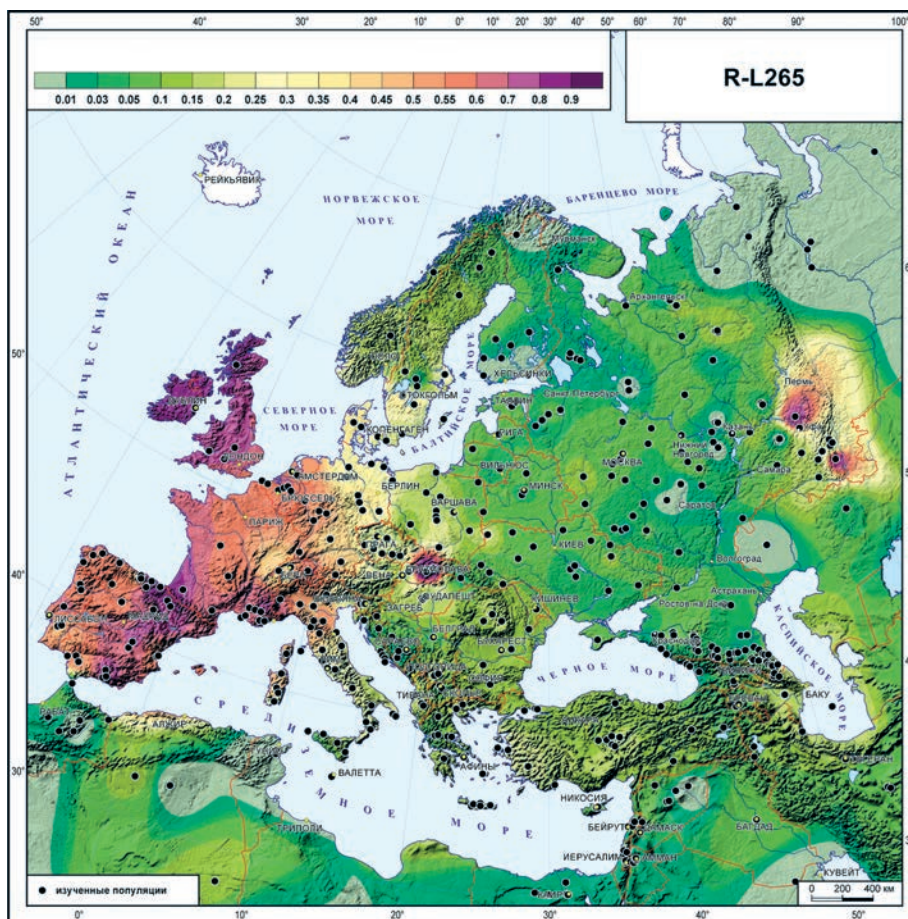


Рис. 2.11. География гаплогруппы R1b-L265 Y-хромосомы в Европе.

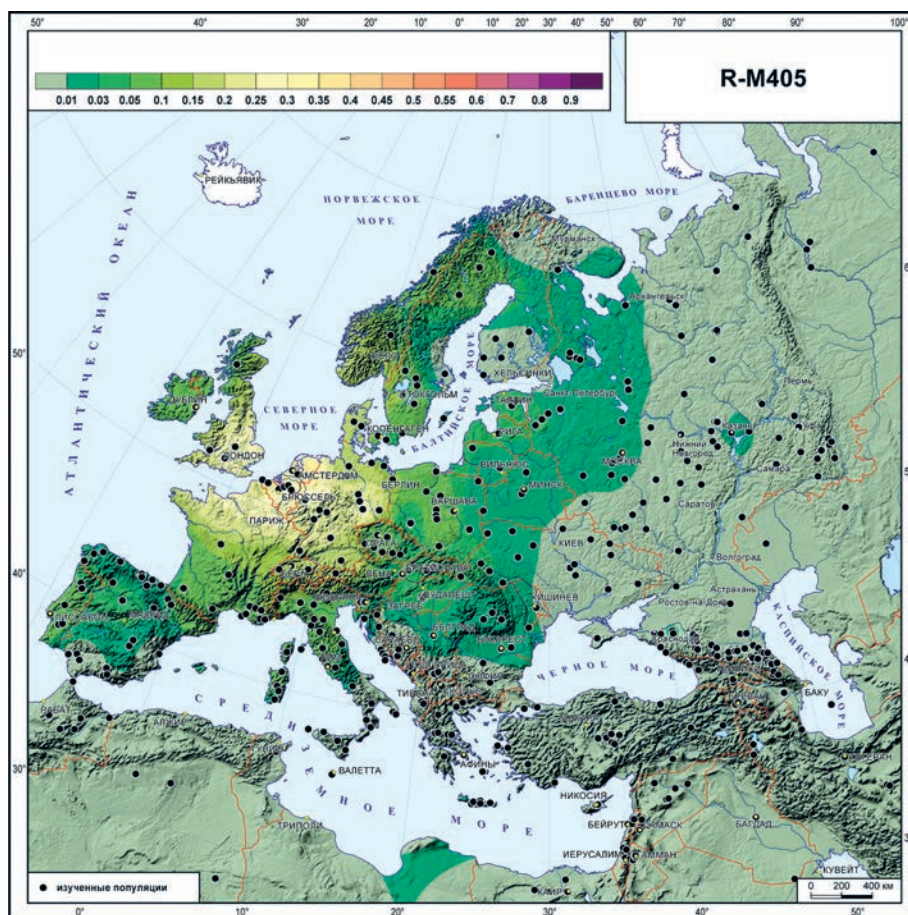


Рис. 2.12. География гаплогруппы R1b-M405 Y-хромосомы в Европе.

западной части Восточной Европы. Однако с максимальными частотами (18–36%) гаплогруппа **R-M405** встречается преимущественно в населении Центральной Европы (на территории Германии, Швейцарии, Бельгии, Голландии, Дании и юге Британии). Датировки как возникновения, так и выделения внутренних субвариантов гаплогруппы **R-M405** по данным www.yfull.com составляют около 5 тыс. лет назад (в интервале 4,4–5,3 тыс. лет назад), т.е. приходится на бронзовый век. Похоже, что уже в это время происходили и миграционные процессы, приводящие к распространению гаплогруппы **R-M405** в разные части Европы. Не исключено, что данные по гаплогруппе **R-M405** указывают на один из эпизодов распространения носителей индоевропейских языков на территории Европы. На одно из направлений распространения носителей гаплогруппы **R-M405**, возможно, указывает география одной из её субветвей – гаплогруппы **R-M467**.

«ГЕРМАНО-БРИТАНСКАЯ» ВЕТВЬ R1b-M467

Скромное распространение гаплогруппы **R-M467** (рис. 2.13) – как по её частотам (от максимума 4% в восточной Германии через значения около 3% в северной и южной Германии, Голландии и на юге Британии до 0,5% во Франции), так и по гео-

графии (не самые обширные, по европейским масштабам, территории вокруг Северного моря) – тем не менее представляет интерес. География гаплогруппы **R-M467** наводит на предположение о миграции её носителей с территории современной Германии вдоль Северного моря к берегам Британских островов. Датировки для гаплогруппы **R-M467**, представленные на сайте www.yfull.com, указывают на её возникновение около 4,7 тыс. лет назад (в интервале 4,1–5,4 тыс. лет назад), а время экспансии (разделения на субветви) спустя тысячу лет – около 3,4 тыс. лет назад (в интервале от 2,8–4,0 тыс. лет назад). Сопоставляя ареал гаплогруппы **R-M467** на карте с указанными датировками возникновения и распространения, можно предположить две модели появления такой ситуации. Первая – когда миграционные процессы с территории центральных областей современной Германии времен бронзового века прошли вдоль берега Северного моря к Британским островам и оставили свой след в современном генофонде. Вторая – когда эта гаплогруппа могла быть распространена несколько шире на юг, запад и восток, но в результате прихода туда другого населения и без того невысокая частота гаплогруппы **R-M467** снизилась на юге до величин, почти не улавливаемых при текущих объемах выборки.

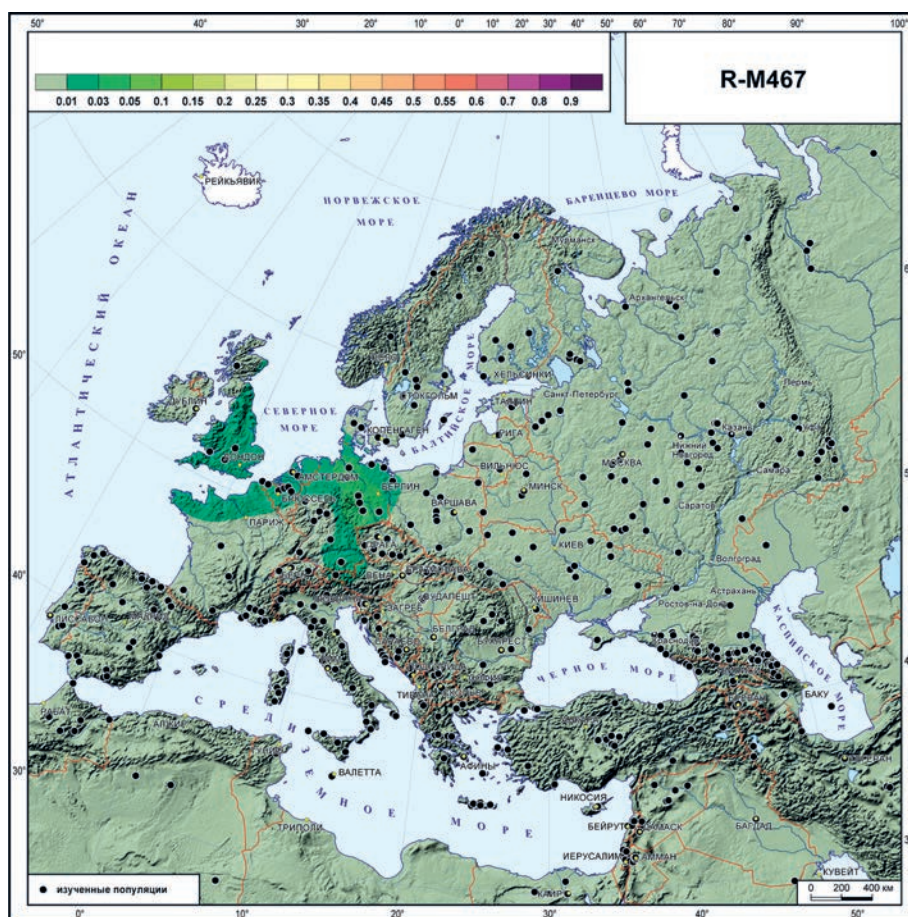


Рис. 2.13. География гаплогруппы R1b-M467 Y-хромосомы в Европе.

«ПИРЕНЕЙСКО-БРИТАНСКАЯ ВЕТВЬ» R1b-P312 В ЦЕЛОМ

Распространение гаплогруппы **R-P312** (рис. 2.14) интересно тем, что во многом отражает общие особенности распределения в населении Европы как своей предковой гаплогруппы **R-L265**, так и всей гаплогруппы **R1b** в целом. Самые высокие частоты гаплогруппы **R-P312** на самом западе Европы (Пиренейский полуостров, территория западной Франции, Британские острова) с постепенным снижением частоты к востоку и локальный всплеск в южном Приуралье – это почти те же особенности, которые были описаны выше при характеристике ареала гаплогруппы **R1b** в целом. Однако следование общему тренду совсем не значит полное сходство. Так, обширная область красно-фиолетовых оттенков (частоты 50–75%) на территории большей части Западной Европы, которую наблюдали ранее на картах гаплогрупп **R1b-L10** и **R-L265** (рис. 2.8 и рис. 2.11), на карте гаплогруппы **R-P312** (рис. 2.14) превращается в локальный всплеск частоты (до 75%) на севере Британских островов и область желто-красных оттенков (частота 30–50%) на территории Пиренейского полуострова и Франции. Южноуральский красно-фиолетовый всплеск общей частоты

R1b на карте гаплогруппы **R-P312** сокращается до небольшого «глазка» в точке для северных башкир, которую со всех сторон обтекают области низких частот (не более 3%).

Гаплогруппа **R-P312**, таким образом, охватывает значительную часть как разнообразия всей гаплогруппы **R-L265**, так и её ареала. Можно предполагать, что в таком случае внутри гаплогруппы **R-P312** тоже существует ряд локальных субветвей с более узкими ареалами. И это предположение оправдывается: на сегодня известно пять больших ветвей, которые в свою очередь подразделяются на несколько десятков более глубоких. Однако остановимся только на трех ветвях **R-P312**, распространение которых в Европе хорошо изучено: гаплогруппа **R-M167** и её субвариант **R-M153**, а также гаплогруппа **R-M222** (рис. 2.15-2.17).

«БАСКО-РОМАНСКАЯ» ГАПЛОГРУППА R1b-M167

Такое название для гаплогруппы **R-M167** условно: её география (рис. 2.15) охватывает население Пиренейского полуострова и Западной Франции, небольшую область на юге Германии и единичный «глазок» в Голландии, а еще область в Западном Причерноморье (румыны и болгары). Большинство

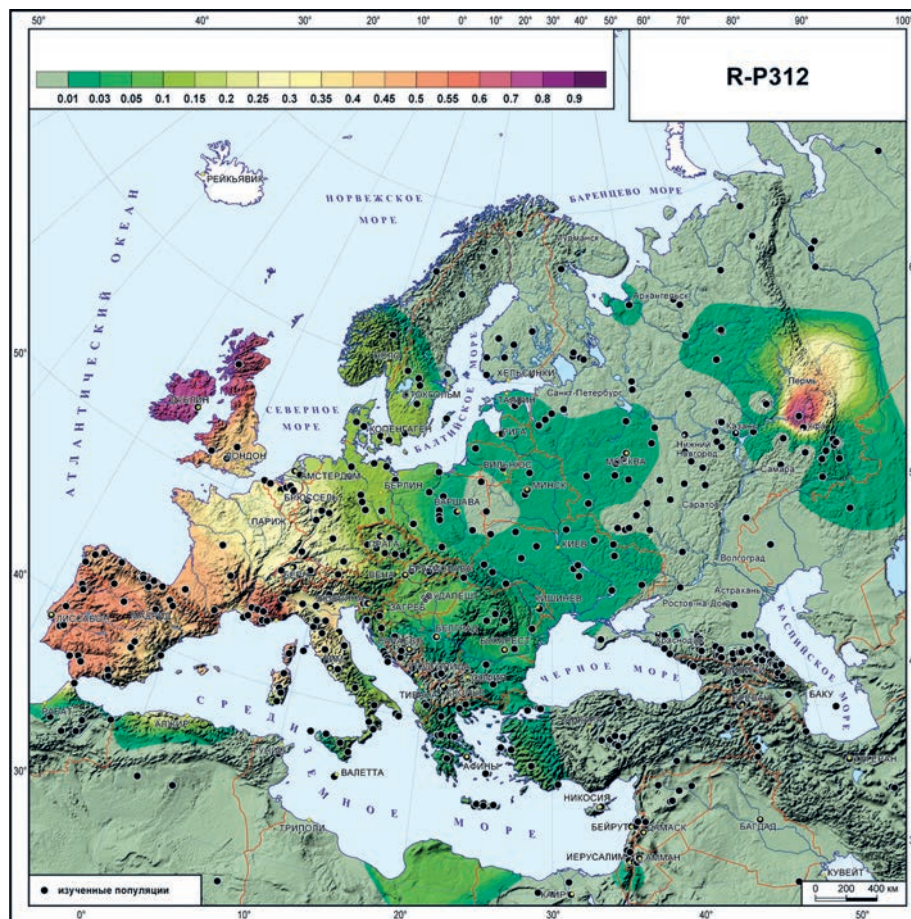


Рис. 2.14. География гаплогруппы R1b-P312 Y-хромосомы в Европе.

ареала гаплогруппы, таким образом, населяют народы, говорящие на языках романской (испанцы, португальцы, французы и румыны) группы индоевропейской языковой семьи и баскской языковой семьи. Впрочем, южные немцы и болгары, у которых тоже встречена гаплогруппа **R-M167**, говорят на языках иных групп индоевропейской семьи.

Область максимальных частот гаплогруппы **R-M167** (желтовато-зеленоватые тона на рис. 2.15, частота гаплогруппы от 6% до 25%) охватывает территорию Пиренейского полуострова и юго-западной Франции. Эта же область является единственной территорией распространения дочерней гаплогруппы **R-M167** – субварианта **R-M153** (рис. 2.16).

“БАСКСКАЯ” ВЕТВЬ R1b-M153

Как можно проследить по карте (рис. 2.16), гаплогруппа **R-M153** своей максимальной частоты (почти 16%) достигает у басков, в два раза реже встречается в населении испанских Пиренеев, и с очень низкими частотами (1–3%) охватывает оставшуюся часть своего ареала. Из-за такой географии гаплогруппу **R-M153** можно условно назвать «баскской» и предположить, что она появилась и распространялась преимущественно в популяци-

ях этого народа. Возраст возникновения этой гаплогруппы по данным www.yfull.com составляет около 2,5 тысяч лет (в пределах от 3500 до 1500 лет назад). Иными словами, эта гаплогруппа, вероятнее всего, появилась в популяциях Пиренейских гор на заре нашей эры и успела распространиться по обе стороны горного хребта (причем эффективнее – к югу от Пиренеев) в результате локальных миграций.

“БРИТАНСКАЯ ВЕТВЬ” R1b-M222

Удивительной, только не по широте ареала, а как раз по его компактности, можно считать гаплогруппу **R-M222** (рис. 2.17). Распространение этой гаплогруппы практически исключительно на Британских островах наводит на предположение о ее местном происхождении. Согласно оценкам www.yfull.com, экспансия гаплогруппы **R-M222** датируется всего лишь 1,8 тыс. лет назад (в интервале 1,3–2,4 тыс. лет назад). Иными словами, гаплогруппа появилась в популяциях Британских островов во 2 тыс. до н.э., а миграции, которые распространили её, происходили в начале нашей эры, то есть примерно в римское время.

Пример филогенетики гаплогруппы **R-L265** показывает, сколь подробная генетическая инфор-

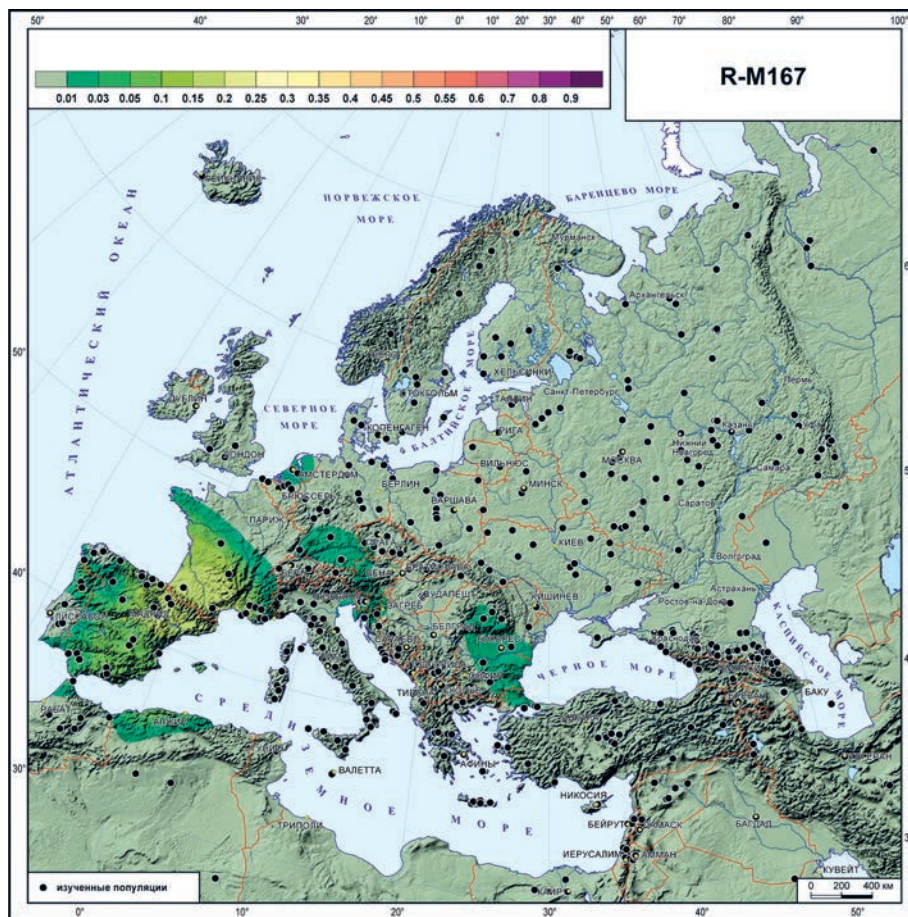


Рис. 2.15. География гаплогруппы R1b-M167 Y-хромосомы в Европе.

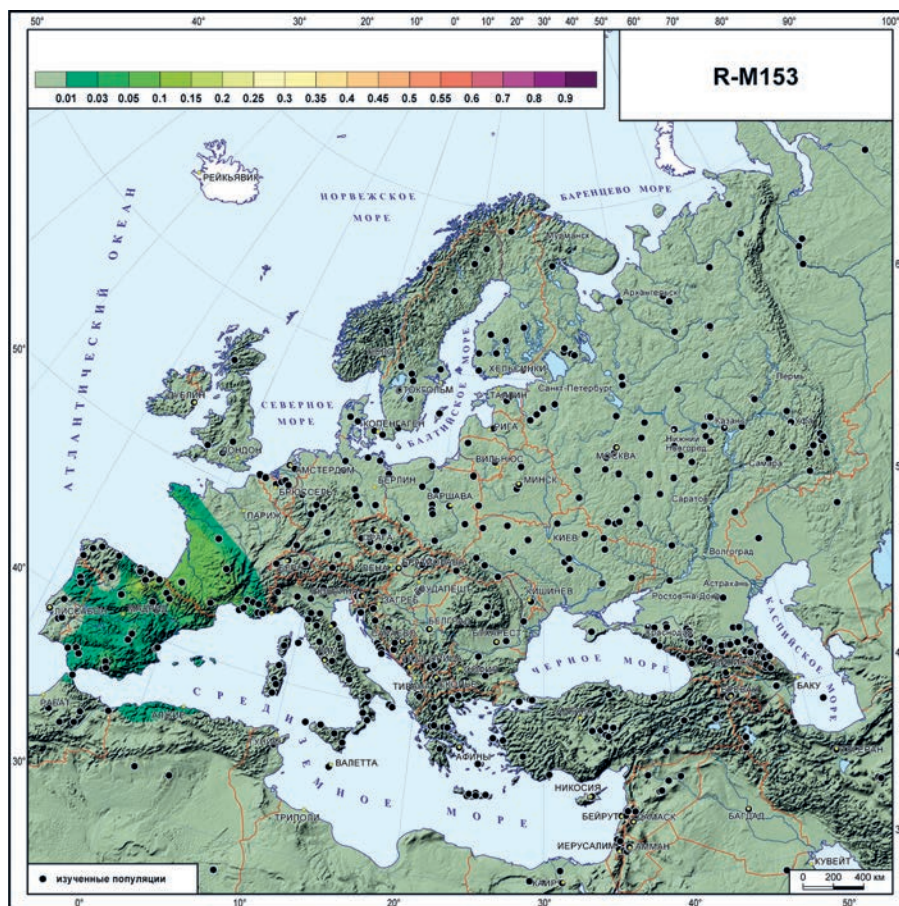


Рис. 2.16. География гаплогруппы R1b-M153 Y-хромосомы в Европе.

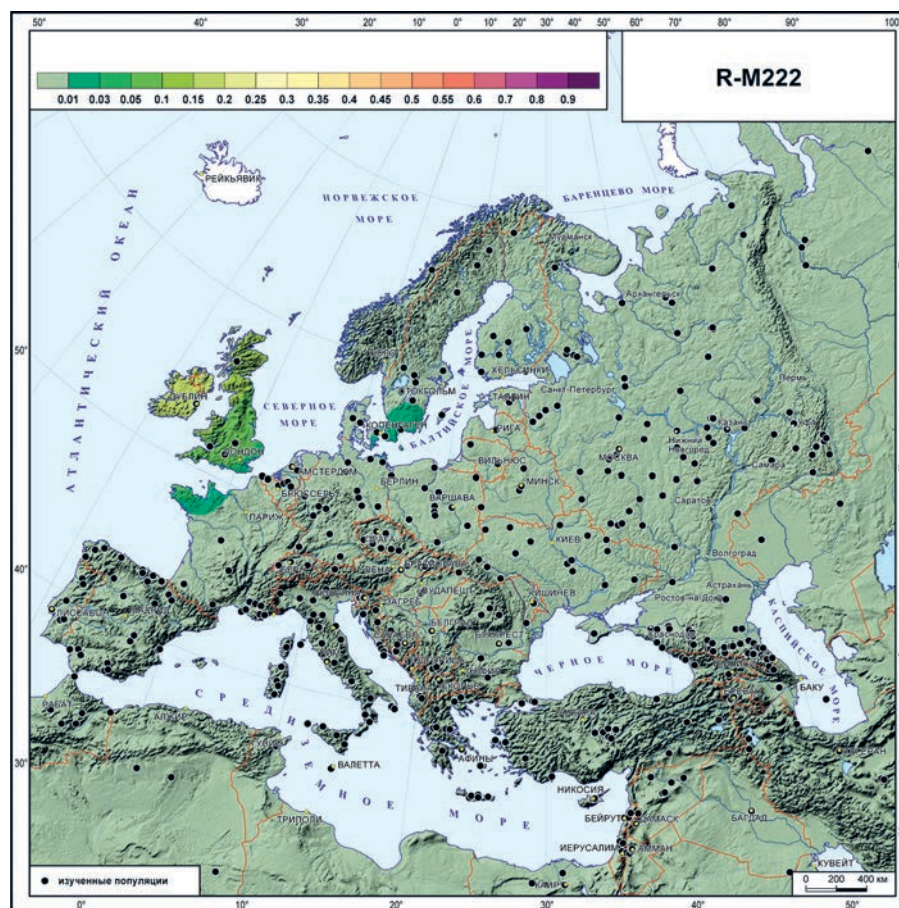


Рис. 2.17. География гаплогруппы R1b-M222 Y-хромосомы в Европе.

мация скрывается в пределах каждой гаплогруппы Y-хромосомы. Остается надеяться, что по мере накопления данных по полным геномам будут обнаружены информативные маркеры и в пределах других гаплогрупп.

“ИНДИЙСКАЯ” ГАПЛОГРУППА R2 (R-L266)

Гаплогруппа **R2 (R-L266)** происходит из того же корня, что и гаплогруппа **R1**, давшая начало гаплогруппам **R1a** и **R1b**. Карта распространения гаплогруппы **R2 (R-L266)** (рис. 2.18) показывает, что для народонаселения Европы в целом **R2** не характерна: она встречается с крайне

низкими частотами (менее 3%) лишь на севере Аппенинского полуострова и в Сардинии, у популяции турок вблизи пролива Дарданеллы, в центральной Анатолии и Закавказье. С чуть более высокой частотой 5% гаплогруппа **R2 (R-L266)** замечена в юго-восточной Турции, с частотой 6% – у калмыков. Такая скудная география гаплогруппы в пределах картографируемого ареала, пожалуй, неудивительна: ведь основной ареал **R2 (R-L266)** охватывает далекие территории Индийского полуострова и Центральной Азии, а с низкими частотами гаплогруппа доходит до Китая на востоке и до Юго-Западной Азии – на западе своего ареала.

ГАПЛОГРУППА I И ЕЕ СУБВАРИАНТЫ

Продолжая тему о высокой информативности выделения субтипов и анализа их географического распределения, напомним ставшую классической работу [Rootsi et al., 2004], в которой гаплогруппа **I-M170** была филогенетически подразделена на три субтипа. Оказалось, что эти субтипы имеют совершенно различную географическую приуроченность: гаплогруппа **I1-M253** (на картах – **I-L118**) приурочена к северной Европе, **I2a-P37** – к Балканскому региону, а **I2b-M223** (на картах – **I-L35**) ло-

кализована преимущественно в Северо-Западной Европе. Именно после этой работы и ряда ей подобных стало ясно, что обнаружение новых SNP маркеров, подразделяющих гаплогруппу на субветви, является магистральным путем повышения информативности анализа Y-хромосомы. Но долгое время движение по этому пути было медленным, и лишь после внедрения полного секвенирования Y-хромосомы новые маркеры стали накапливаться лавинообразно. Так, в течение долгого времени не уда-

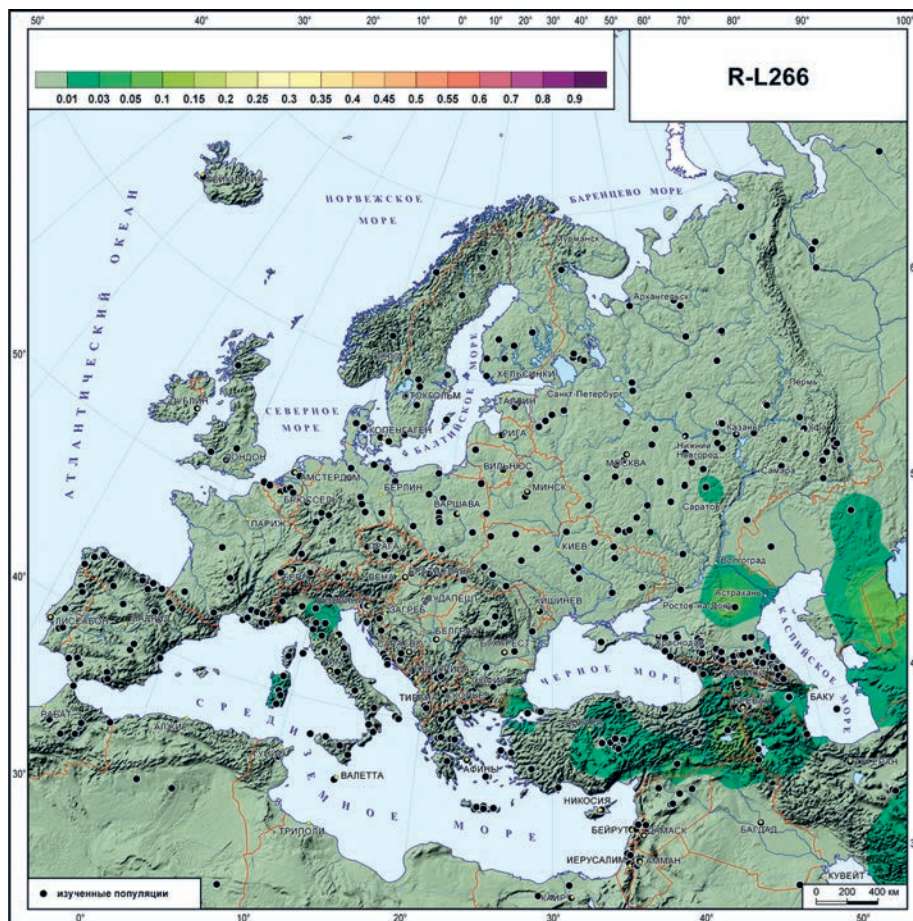


Рис. 2.18. География гаплогруппы R2-L266 Y-хромосомы в Европе.

валось обнаружить новые SNP маркеры в пределах гаплогруппы **I**, или же новые открываемые SNP маркеры не выделяли новых ветвей с четкой географией. Например, в работе [Underhill et al., 2008] были обнаружены новые маркеры и реорганизована топология ветвей гаплогруппы **I-M170**, но сохранились, пусть под новыми названиями, те же три основные ветви, а новые обнаруженные варианты были редкими и не характеризовались четкими географическими трендами.

«СКАНДИНАВСКАЯ» ГАПЛОГРУППА **I-L118**

Карта распространения «скандинавской» гаплогруппы **I-L118 (I-M523)**, рис. 2.19) показывает ее максимальные частоты не только в географической Скандинавии, но и на территориях, тесно связанных с ней исторически: Дании, Шотландии, западных районах Финляндии. Поэтому любопытна область умеренных частот у русских популяций к северу от Волги, выделяющаяся на фоне низких частот у большинства славянских популяций. Нельзя исключать, что это отражает исторические контакты с норманнскими популяциями. Хотя эти контакты были более интенсивны для северо-западных русских популяций (на пути «из варяг в греки»), но территории северо-запада были сравнительно густо населены, поэтому влияние «варя-

гов» могло оказаться более значимым для меньших по размеру популяций Заволжья.

«БАЛКАНСКАЯ» ГАПЛОГРУППА **I-P37**

Карта распространения «балканской» гаплогруппы **I-P37** (рис. 2.20) показывает ее высокие частоты также и в Италии. В целом, эта гаплогруппа распространена практически по всей Европе, кроме Скандинавии. В Восточной Европе плавный градиент убывания ее частоты с юго-запада (с Украины) к северу и востоку очень напоминает карту первой главной компоненты по классическим генетическим маркерам [Рычков и др., 2002]. Также стоит отметить достаточно выраженное совпадение ареала этой гаплогруппы с географическими границами Европы – при высокой частоте в Причерноморье гаплогруппа очень редка на Северном Кавказе и почти отсутствует в Закавказье, а при максимальных частотах на Балканах очень редка в соседней Малой Азии.

«ВАРЯГО-ГРЕЧЕСКАЯ» ГАПЛОГРУППА **I-L35**

Вдоль берегов Северного и Балтийского морей, и немного вглубь от них - так можно было бы описать географию гаплогруппы **I-L35 (I2b-M223)**, если бы еще не два дополнения на северо-востоке и юго-

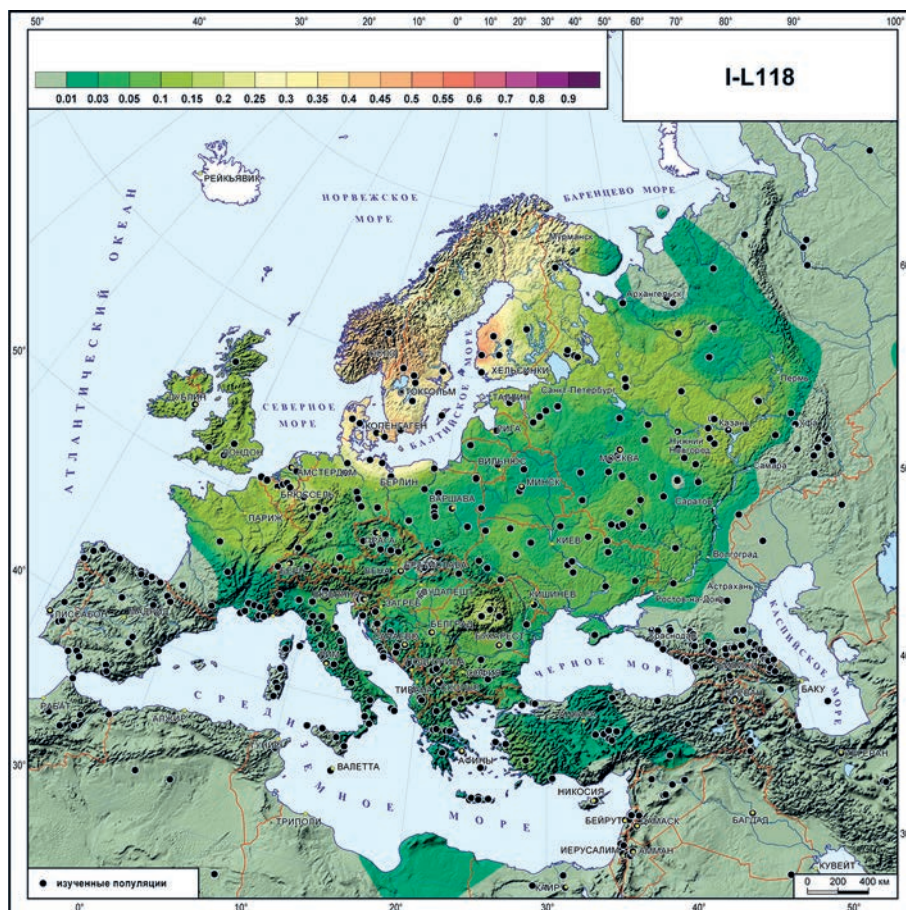


Рис. 2.19. География гаплогруппы **I-L118 (I1-M253)** Y-хромосомы в Европе.

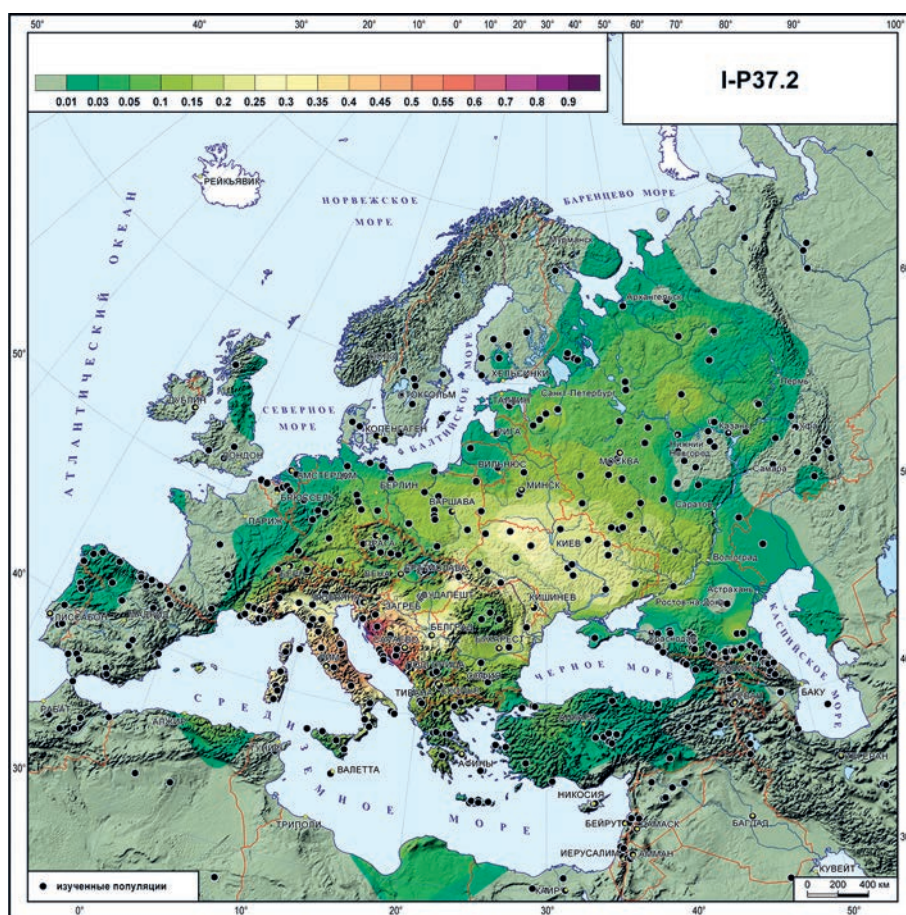


Рис. 2.20. География гаплогруппы I-P37 (I2a) Y-хромосомы в Европе.

востоке. Хотя максимум частоты гаплогруппы **I-L35** (17%) отмечен в Швеции (рис. 2.21), зона более редкой встречаемости гаплогруппы (3–7%) охватывает северную часть Центральной Европы, на юго-востоке отдельной областью очень низких частот (1–3%) охватывает Балканский полуостров и западную часть Малой Азии, а на северо-востоке проявляется у русских популяций Тверской и Костромской областей. Можно заметить, что ареал гаплогруппы **I-L35** – довольно умеренный и однородный по зна-

чениям частот гаплогруппы (на большинстве ареала частота варьирует от 2–3% до 7%, только пик в Швеции до 17%) – связывает несколько крайних областей Европы: Скандинавию, северные районы Восточной Европы и восток Южной Европы. Вполне возможно, что география гаплогруппы **I-L35** действительно отражает один из генетических следов торговых путей между севером и югом Европы. Не менее вероятно, что каждая из этих зон связана со своим субвариантом этой гаплогруппы.

ГАПЛОГРУППА N И ЕЕ СУБВАРИАНТЫ

Гаплогруппа **N** в пределах Европейского региона представлена двумя большими ветвями – гаплогруппами **N1c** и **N1b** (далее и на картах **N-M46** и **N-P43** соответственно). Хотя общие ареалы этих двух гаплогрупп в Европе пересекаются, зоны их максимальных частот разграничиваются на севере рекой Печорой, а южнее – Уральским хребтом: к западу от этой условной “границы” остаются зоны повышенной встречаемости гаплогруппы **N-M46**, а к востоку – зона максимальных частот **N-P43**. Рассмотрим географию этих двух “братских” ветвей гаплогруппы **N** чуть подробнее.

“СЕВЕРОЕВРОПЕЙСКАЯ” ЗАПАДНАЯ ГАПЛОГРУППА N-M46 (N1c)

Карта распространения гаплогруппы **N-M46** (рис. 2.22) показывает ее максимальные частоты у финнов. Ее частоты высоки и у других популяций северо-востока Европы. Хотя в целом эта гаплогруппа оказывается приуроченной к финно-угорским популяциям, но она столь же часта и у северных русских (что, по-видимому, объясняется асимилиацией дославянского населения, как будет описано в следующих главах). Наиболее любопытным

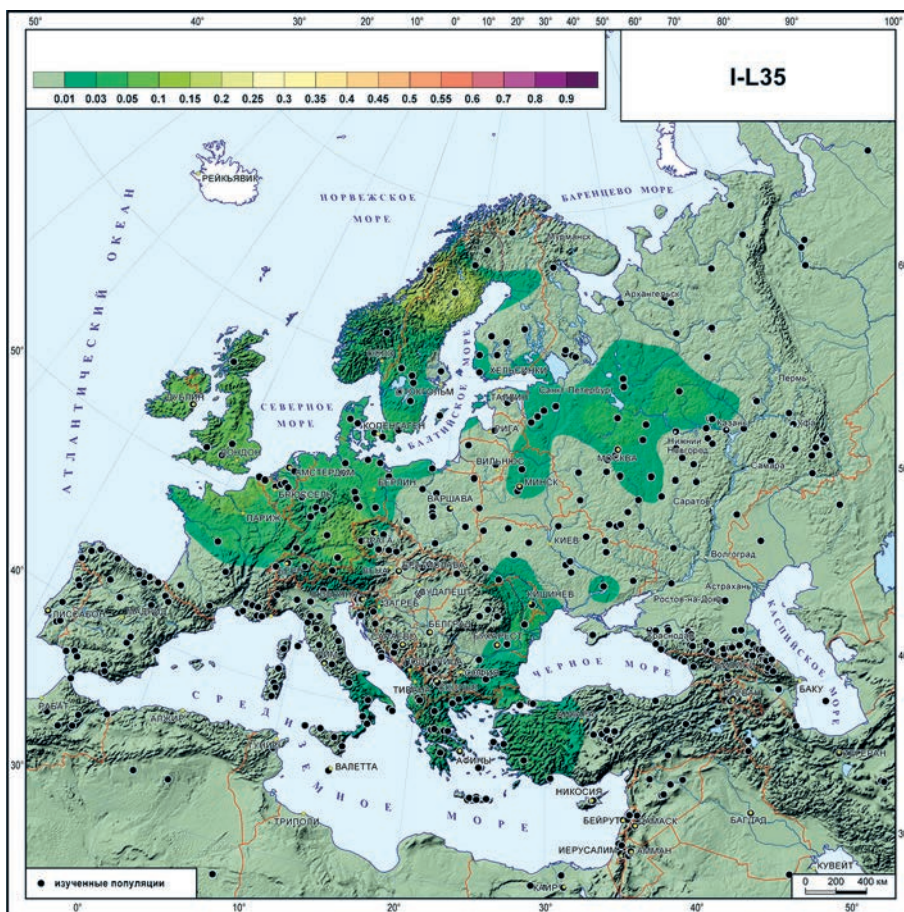


Рис. 2.21. География гаплогруппы I-L35 (I2b-M223) Y-хромосомы в Европе.

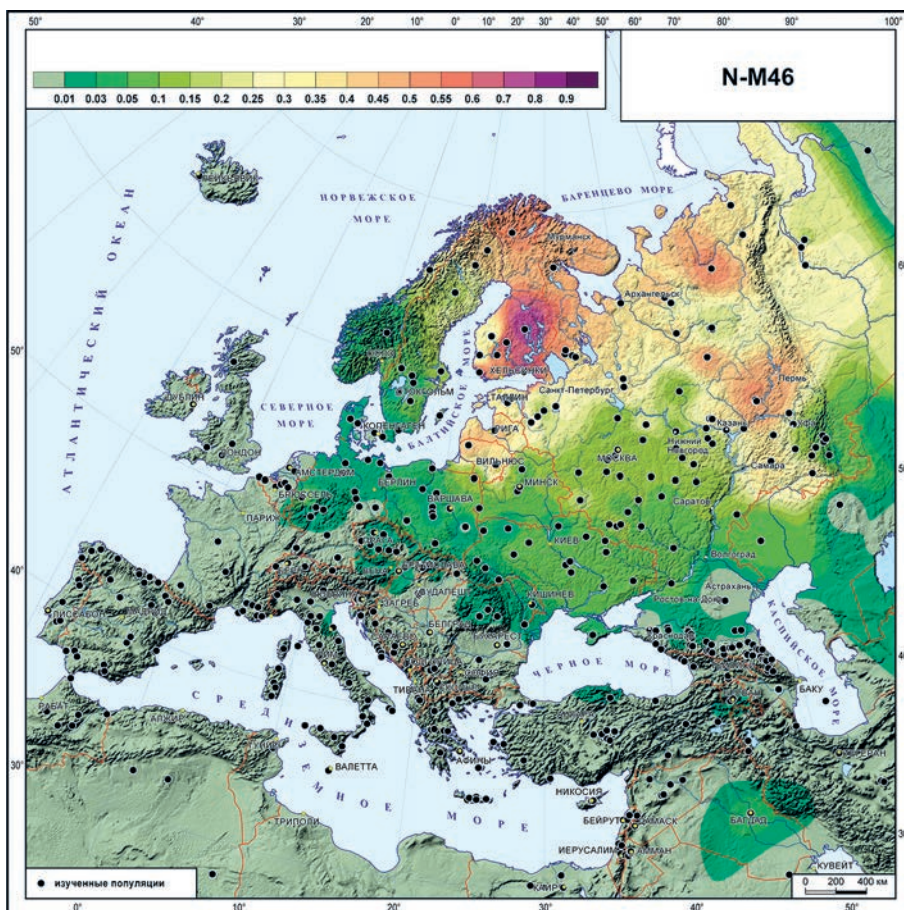


Рис. 2.22. География гаплогруппы N-M46 (N1c) Y-хромосомы в Европе.

представляется область высокой частоты этой гаплогруппы у балтов (латышей и литовцев). В данном случае генетические границы хорошо совпадают с лингвистическими: граница ареалов балтов и западных славян совпадает с резко выраженным «обрывом» частоты гаплогруппы. Похожая картина наблюдается и на границе с белорусами.

«СЕВЕРОУРАЛЬСКАЯ» ГАПЛОГРУППА N-P43

Карта распространения гаплогруппы N-P43 (рис. 2.23) выявляет также северный, но куда бо-

лее узкий ареал. Максимальные частоты этой гаплогруппы наблюдаются на крайнем северо-востоке Европы и плавно переходят в мировой максимум этой гаплогруппы в Западной Сибири. С умеренными частотами эта гаплогруппа распространена по всему Уралу и Приуралью. Заметен плавный градиент убывания частоты к западу и югу, так что у популяций Центральной России и нижней Волги эта гаплогруппа еще встречается (хотя и с минимальными частотами), а у других популяций Европы практически полностью отсутствует.

ГАПЛОГРУППА E И ЕЕ СУБВАРИАНТЫ

Большинство вышеописанных гаплогрупп в основном характеризовали условно северную половину Европейского континента. Рассматривая географию следующей гаплогруппы – E – мы переместим внимание на юг, к берегам Средиземного моря. В Европе встречаются четыре основные ветви гаплогруппы E: E-L142 и E-V22 (восходящие к общему корню – гаплогруппе E-M78), E-M81 и E-M123.

«ВОСТОЧНО-СРЕДИЗЕМНОМОРСКАЯ» ГАПЛОГРУППА E-L142

Карта гаплогруппы E-L142 (рис. 2.24) отчасти напоминает карту гаплогруппы I-P37 (рис. 2.20) – максимальные частоты обеих этих гаплогрупп приурочены к Балканскому полуострову. Также очень похожи их паттерны: убывание частоты к северу, западу и востоку. Основные различия между трен-

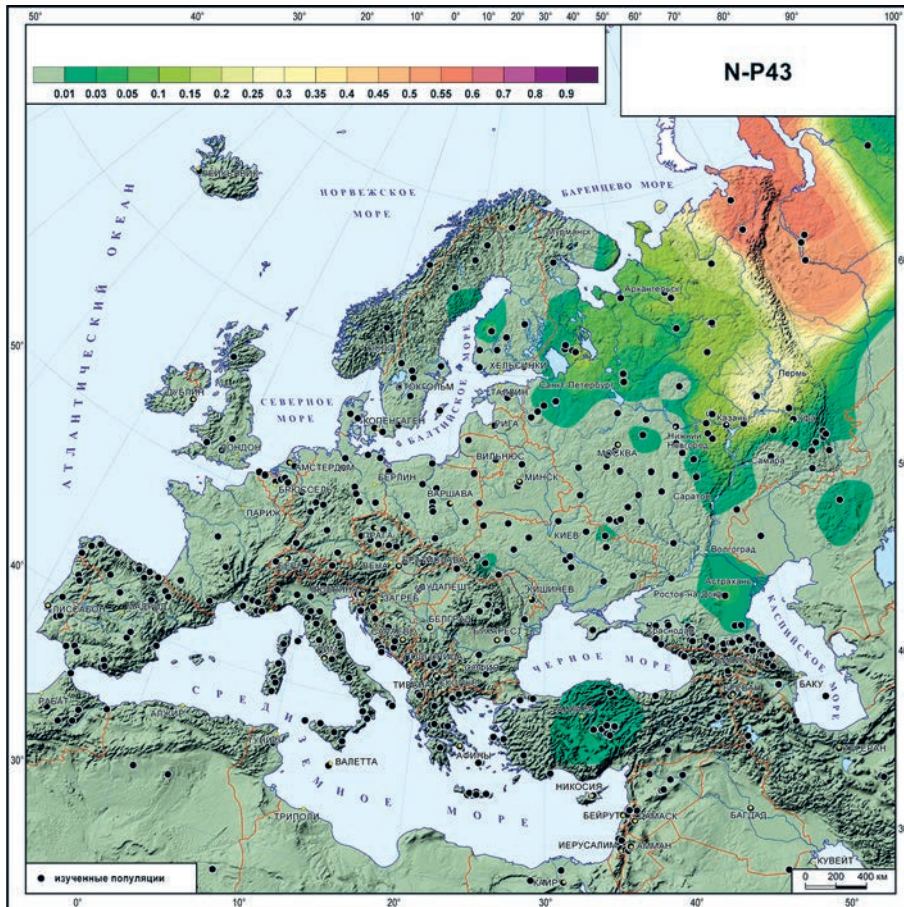


Рис. 2.23. География гаплогруппы N-P43 (N1b) Y-хромосомы в Европе.

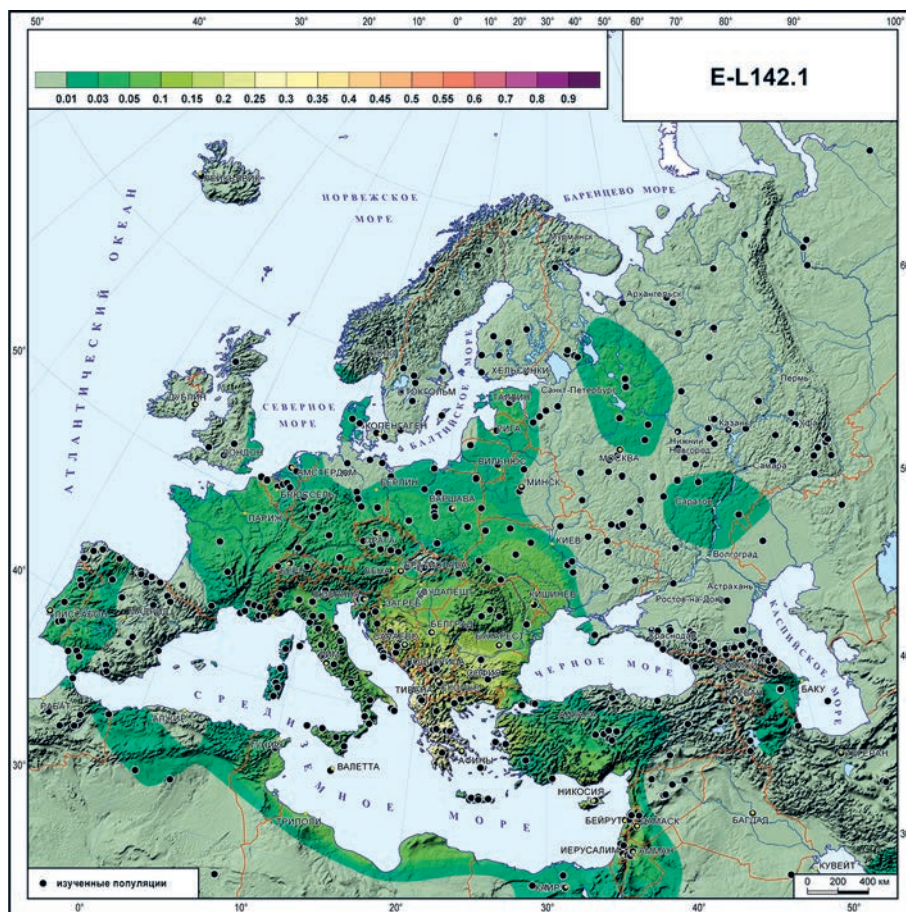


Рис. 2.24. География гаплогруппы E-L142 Y-хромосомы в Европе.

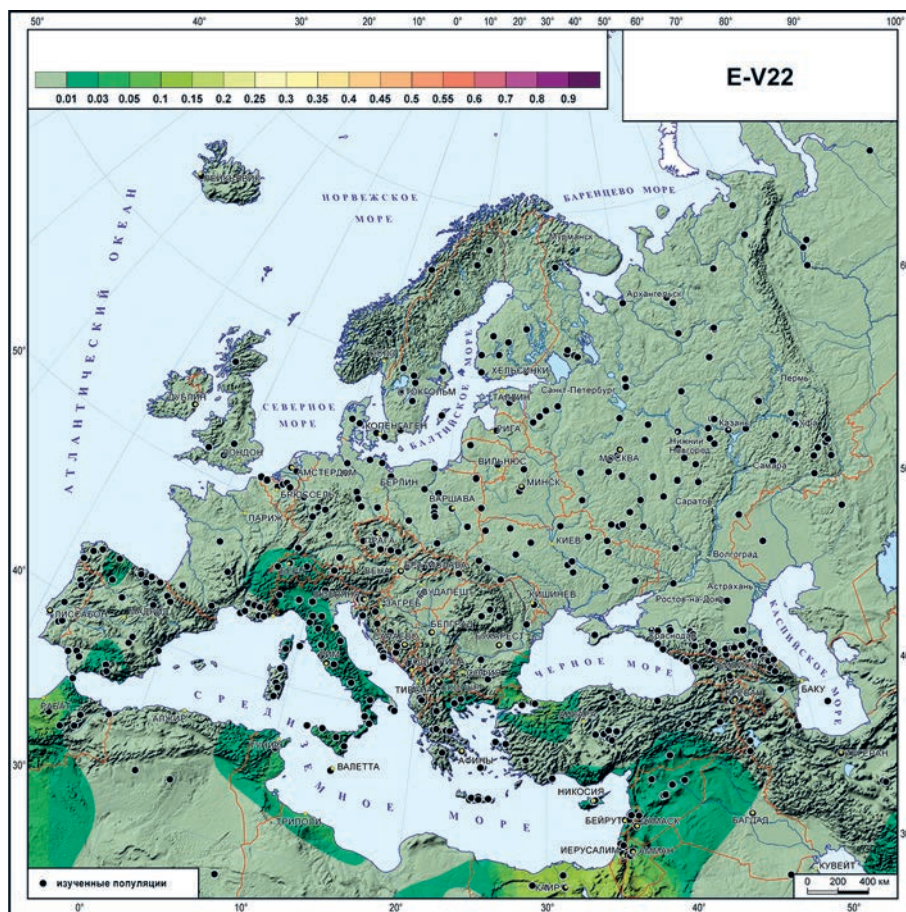


Рис. 2.25. География гаплогруппы E-V22 Y-хромосомы в Европе.

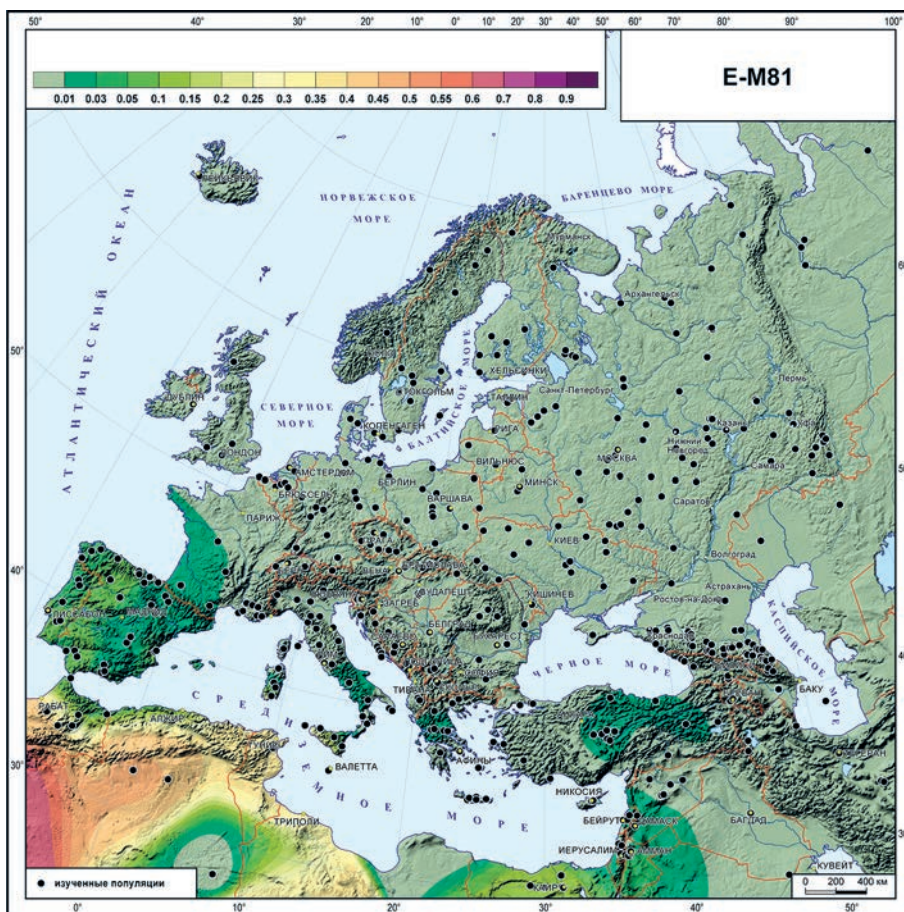


Рис. 2.26. География гаплогруппы E-M81 Y-хромосомы в Европе.

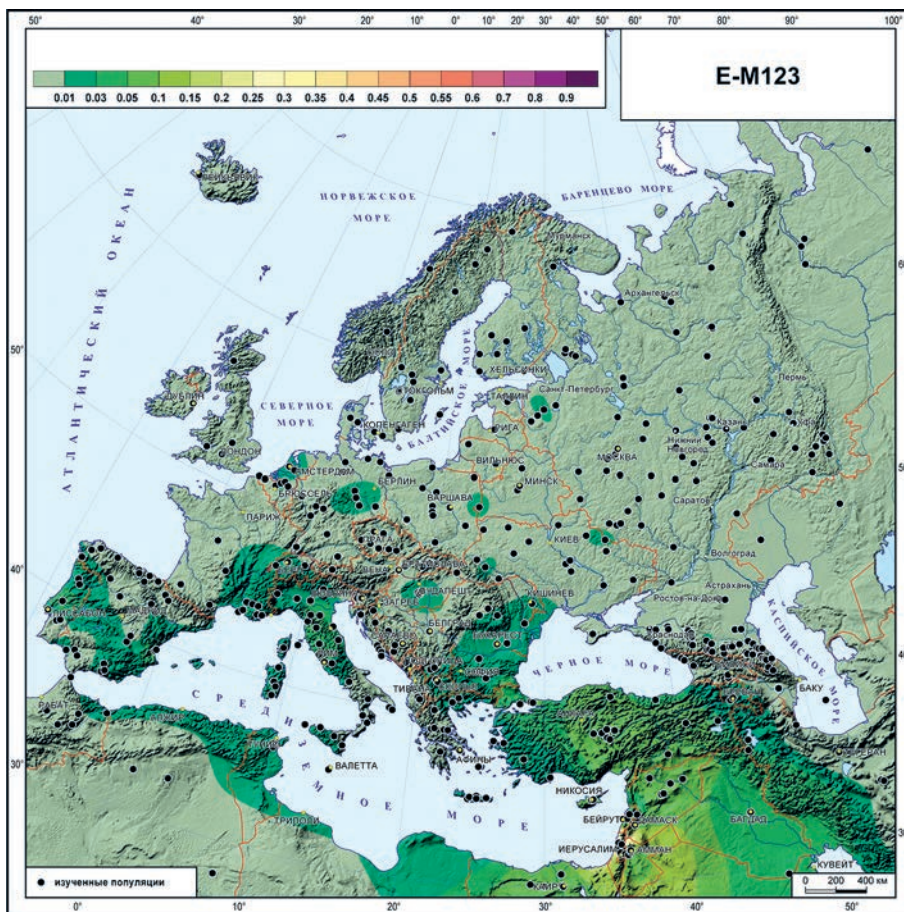


Рис. 2.27. География гаплогруппы E-M123 Y-хромосомы в Европе.

дами двух этих гаплогрупп состоят, во-первых, в меньшей частоте **E-L142** (на большинстве территорий она «уступает» **I-P37** на один интервал шкалы карты) и, во-вторых, в распространении **E-L142** не только в Европе, но и по всему Средиземноморью, включая Ближний Восток и северные окраины Африки. В пределах Балкан **I-P37 (I2a)** особенно часта у боснийцев и хорватов, а **E-L142** преобладает южнее – у сербов, албанцев и греков.

«ЕГИПЕТСКАЯ» ГАПЛОГРУППА E-V22

Два небольших «глазка» на севере и юге Пиренейского полуострова и Аппенинский полуостров – территории с частотой гаплогруппы **E-V22** до 5%, небольшая область на северо-западе Малой Азии (в окрестностях Стамбула) с частотой до 6%: на этом можно было бы закончить описание географии гаплогруппы **E-V22** в географических границах Европы (рис. 2.25). Однако география гаплогруппы **E-V22** в Европе в основном следует Средиземноморскому побережью, поэтому стоит проследить её распространение и южнее. Карта на рис. 2.25 показывает плавное повышение частоты гаплогруппы **E-V22** в двух противоположащих регионах: на юго-западе и юго-востоке Средиземноморья. В юго-западном Средиземноморье у берегов Гибралтарского пролива и вглубь Марокко частота гаплогруппы **E-V22** вырастает до 7%. В юго-восточном Средиземноморье повышение частоты гаплогруппы **E-V22** отмечается от Малой Азии на юг – через Левант вплоть до дельты Нила, где она достигает 14%, и ещё южнее вдоль Нила, где у египетских арабов оазиса Бахария она достигает мирового максимума 22%. Из этих двух неевропейских регионов распространения гаплогруппы **E-**

V22, вероятно, именно последний (с более высокими частотами и широким распространением) послужил основным источником распространения гаплогруппы **E-V22** в Европе.

САХАРСКАЯ ГАПЛОГРУППА E-M81

Если две предыдущих ветви гаплогруппы **E – E-L142** и **E-V22** – демонстрировали в основном распространение вдоль северных и восточных берегов Средиземного моря, иногда с углублением на север и восток континента, то география третьей ветви – гаплогруппы **E-M81** – больше следует южному Средиземноморью. Карта распространения гаплогруппы **E-M81** на рис. 2.26 хотя и показывает заметное присутствие гаплогруппы на Ближнем Востоке (частота до 20%), однако основная зона высоких частот находится в северо-западной Африке. Из этого максимума отмечено плавное снижение частоты в северном направлении через Гибралтарский пролив с умеренным распространением (2–10%) гаплогруппы **E-M81** на территории Пиренейского полуострова и юго-западной Франции, узкие области низких частот (не более 3%) на Аппенинском и Балканском полуостровах, а также северо-востоке Малой Азии.

«ПЕРЕДНЕАЗИАТСКАЯ» ГАПЛОГРУППА E-M123

Гаплогруппа **E-M123** достигает максимума в Передней Азии (20–25%), ближе к Европе (в Малой Азии) снижается до 15%. В самой же Европе она с низкими частотами распространена вдоль северных берегов Средиземного моря, и отдельными вкраплениями проявляется в Центральной Европе (рис. 2.27).

ГАПЛОГРУППА G И ЕЕ СУБВАРИАНТЫ

Карты распространения гаплогруппы двух ветвей гаплогруппы **G2a-P15** – гаплогрупп **G-P303** и **G-P16** – в целом похожи, и даже в общеевропейском масштабе карт видно, что обе ветви приурочены в первую очередь к Кавказу. Однако своих максимальных частот они достигают в разных частях Кавказского региона: один вариант **G2a** преобладает на крайнем северо-западе у причерноморских шапсугов, а другой вариант **G2a** (гаплогруппа **G-P16**) – на Центральном Кавказе у осетин-иронцев (рис. 2.28 и рис. 2.29). Это полностью подтверждается при отдельном изучении генофонда Кавказа [Balanovsky et al., 2011; глава 7 этой книги].

«ЗАПАДНОКАВКАЗСКАЯ» ГАПЛОГРУППА G-P303

Карта показывает, что гаплогруппа **G-P303** (рис. 2.28) также распространена на Ближнем Востоке, а в Европе встречается в основном в ее юго-

восточных областях. Зоны высокой частоты гаплогруппы **G-P303** тянутся вдоль восточного побережья Черного моря и приурочены к популяциям народов Западного Кавказа, в основном, абхазо-адыгской лингвистической группы. С меньшими частотами (до 10%) гаплогруппа **G-P303** распространена в Северном Причерноморье (на территории Крымского полуострова и в Приазовье), область низких частот (3–5%) тянется далее на север в зону украинской лесостепи. Локальное повышение частоты гаплогруппы **G-P303** до 13% также наблюдается в Поволжье у популяции мордвы мокша и окружается областью низких частот.

«ЦЕНТРАЛЬНОКАВКАЗСКАЯ» ГАПЛОГРУППА G-P16

География гаплогруппы **G-P16** (рис. 2.29) более узкая по сравнению с её «братской» ветвью **G-P303**. Она охватывает исключительно Кавказский

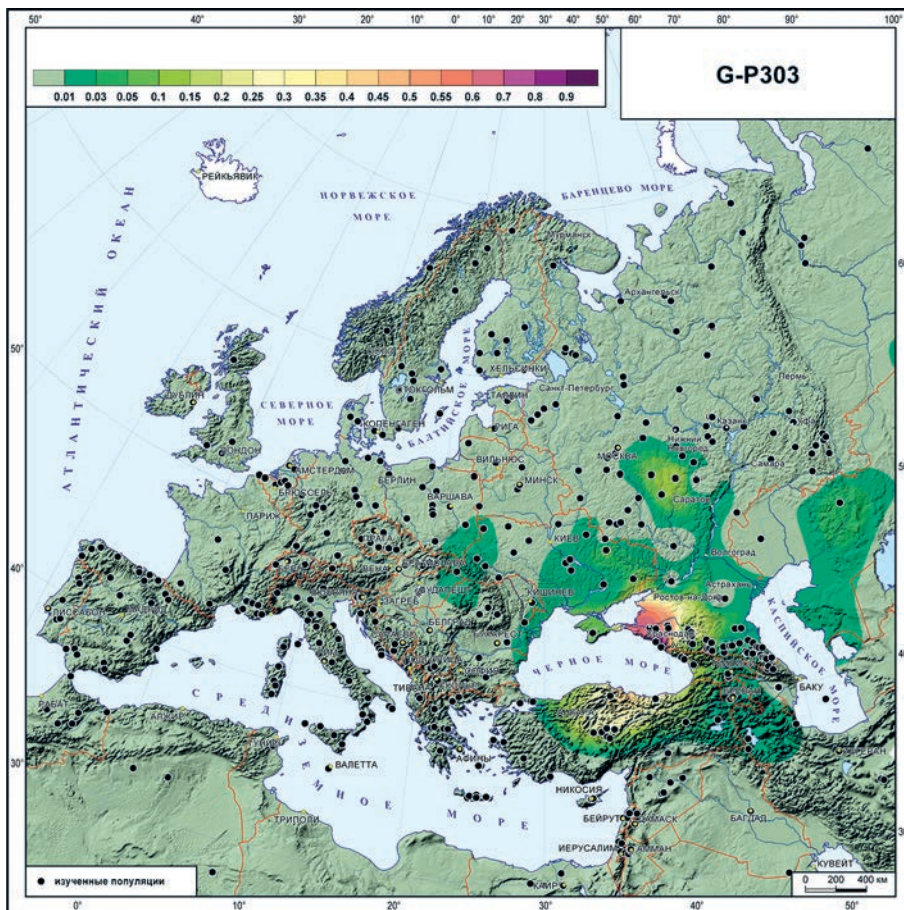


Рис. 2.28. География гаплогруппы G-P303 Y-хромосомы в Европе.

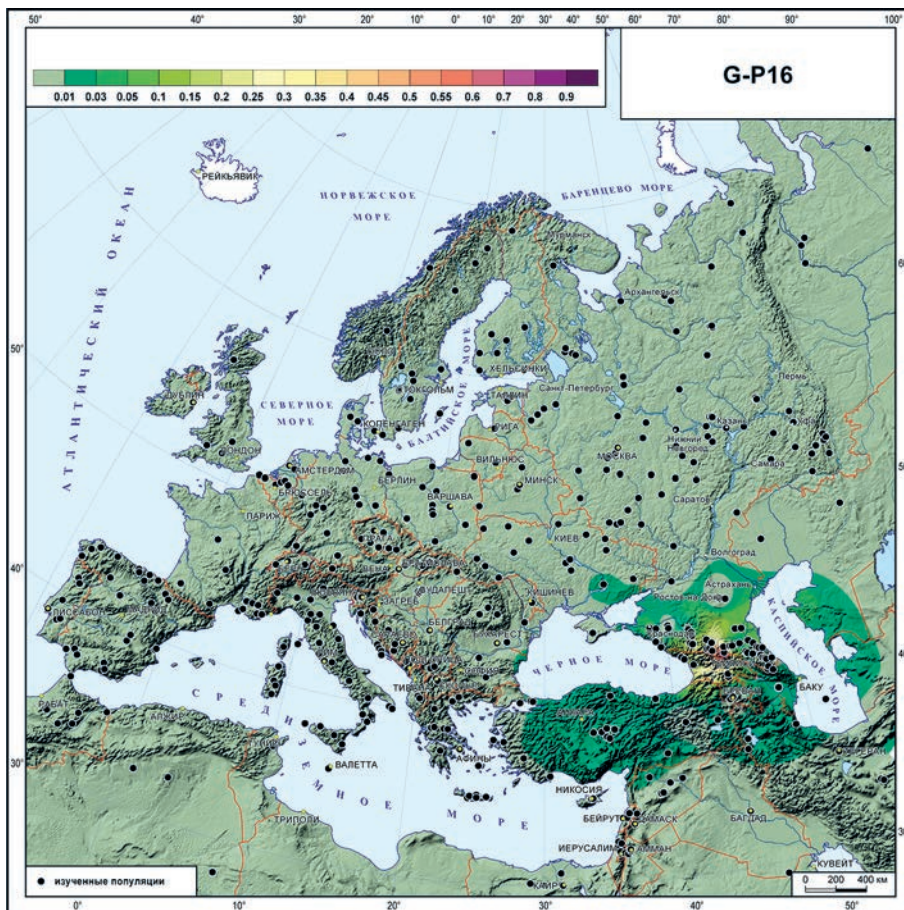


Рис. 2.29. География гаплогруппы G-P16 Y-хромосомы в Европе.

регион и Малую Азию. В пределах этого ареала частота гаплогруппы **G-P16** резко падает от максимальных значений у осетин (73% у осетин-иронцев и 56% у осетин-дигорцев) к значениям в диа-

пазоне 10–20% у народов Северо-Западного Кавказа (абхазов, черкесов, балкарцев и карачаевцев) и затем снижается до минимальной отметки у турок, армян и азербайджанцев.

ГАПЛОГРУППА J И ЕЕ СУБВАРИАНТЫ

Гаплогруппа **J** представлена двумя крупными ветвями: **J-L255 (J1)** и **J-L228 (J2)**, из которых в Европе большее распространение получила вторая. Хотя обе эти ветви по своему происхождению ближневосточные и отражают древние взаимосвязи населения Европы и жителей намного более южных регионов Евразии, имеет смысл рассмотреть отдельно географию и каждой из этих линий в целом, и каждой из внутренних ветвей более распространенной в Европе гаплогруппы **J-L228 (J2)**.

ГАПЛОГРУППА J-L255 (J1): ОТ ПЕРЕДНЕЙ АЗИИ К ПИКАМ ВОСТОЧНОГО КАВКАЗА

Распространение гаплогруппы **J-L255** (рис. 2.30) в картографируемом ареале охватывает преимущественно его юго-восточные области: Кавказ, Малую Азию, Северное Причерноморье, Балканский полуостров; небольшими зонами – юг Аппенинского полуострова и Франции, юго-запад Пиренейского по-

луострова. В пределах региона максимальных частот – Кавказа – гаплогруппа **J-L255** выделяет народы Восточного Кавказа, где на ее долю приходится львиная доля всего генофонда. Максимум гаплогруппа **J-L255** достигает в Дагестане (у кубачинцев, даргинцев, табасаранцев и аварцев ее частоты колеблются от 63% до 98%), и снижается до 44% у лезгин и 21% у чеченцев. На территории Малой Азии доля гаплогруппы **J-L255** составляет от 5% до 15%.

ГАПЛОГРУППА J-L228 (J2) В ЦЕЛОМ

Карта распространения гаплогруппы **J-L228 (J2)** рисует иную картину (рис. 2.31). В отличие от большинства рассмотренных выше гаплогрупп, **J-L228 (J2)** распространена с высокими частотами не в Европе, а на Ближнем Востоке и в Северной Африке. Конечно, ее ареал заходит и в соседние с этими регионами южные области Европы: Испанию (особенно в ее южные районы), Италию (так-

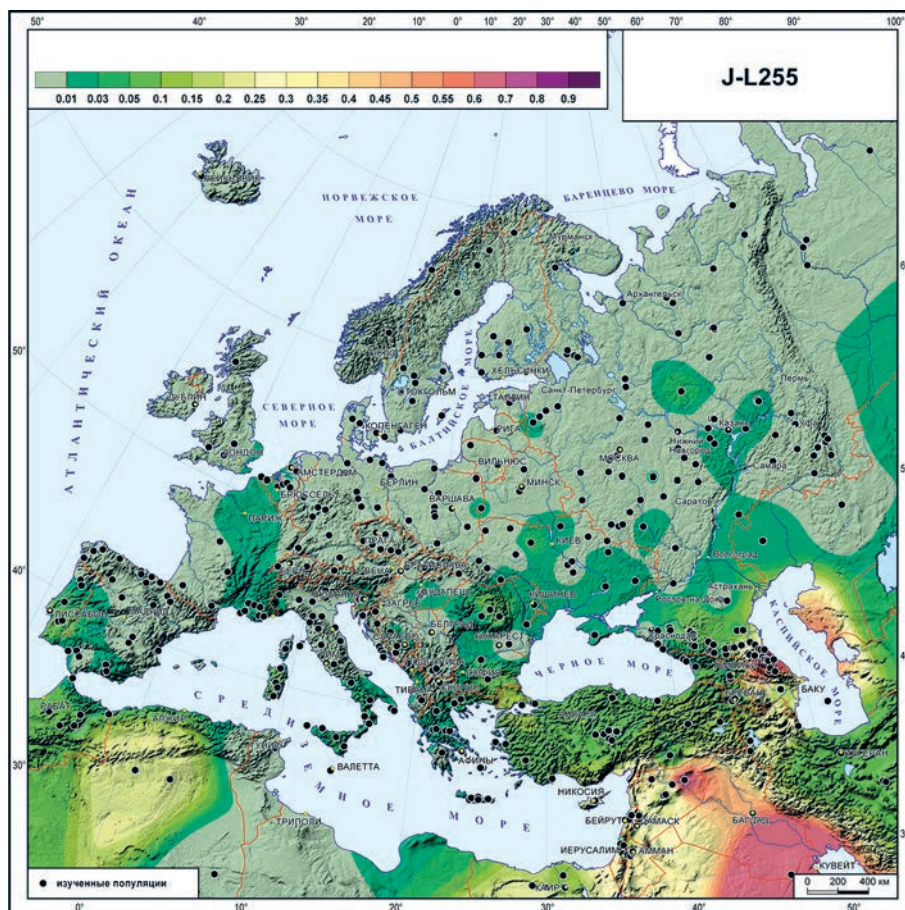


Рис. 2.30. География гаплогруппы J-L255 (J1) Y-хромосомы в Европе.

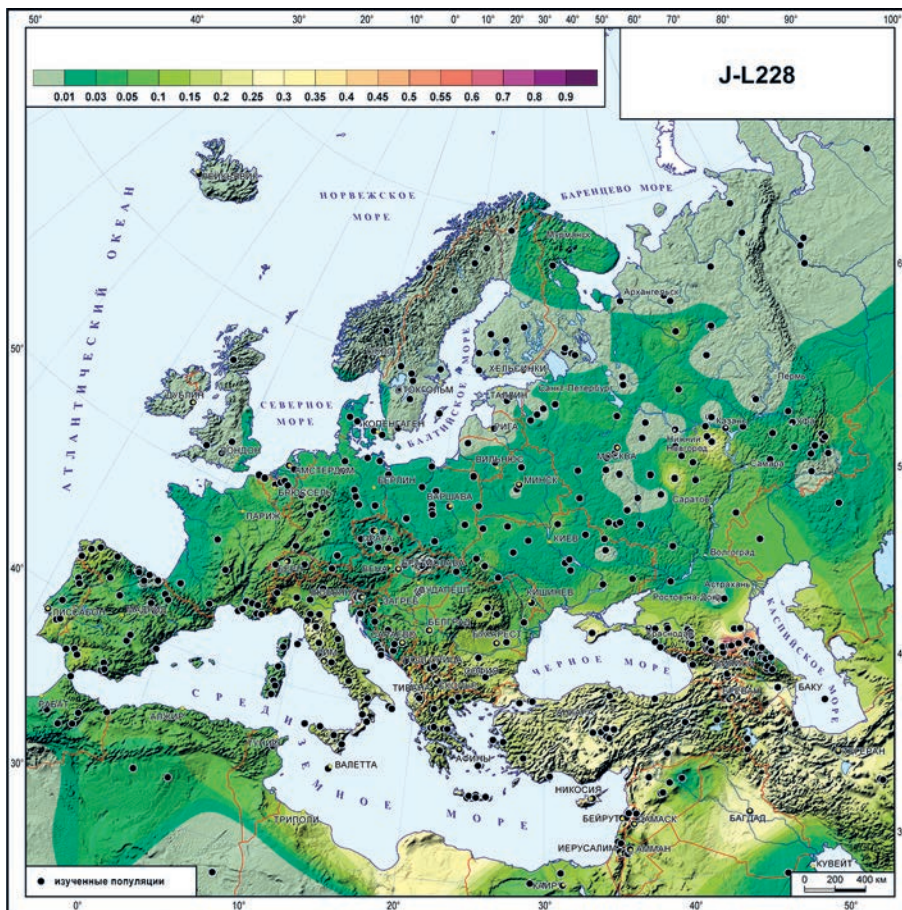


Рис. 2.31. География гаплогруппы J-L228 (J2) Y-хромосомы в Европе.

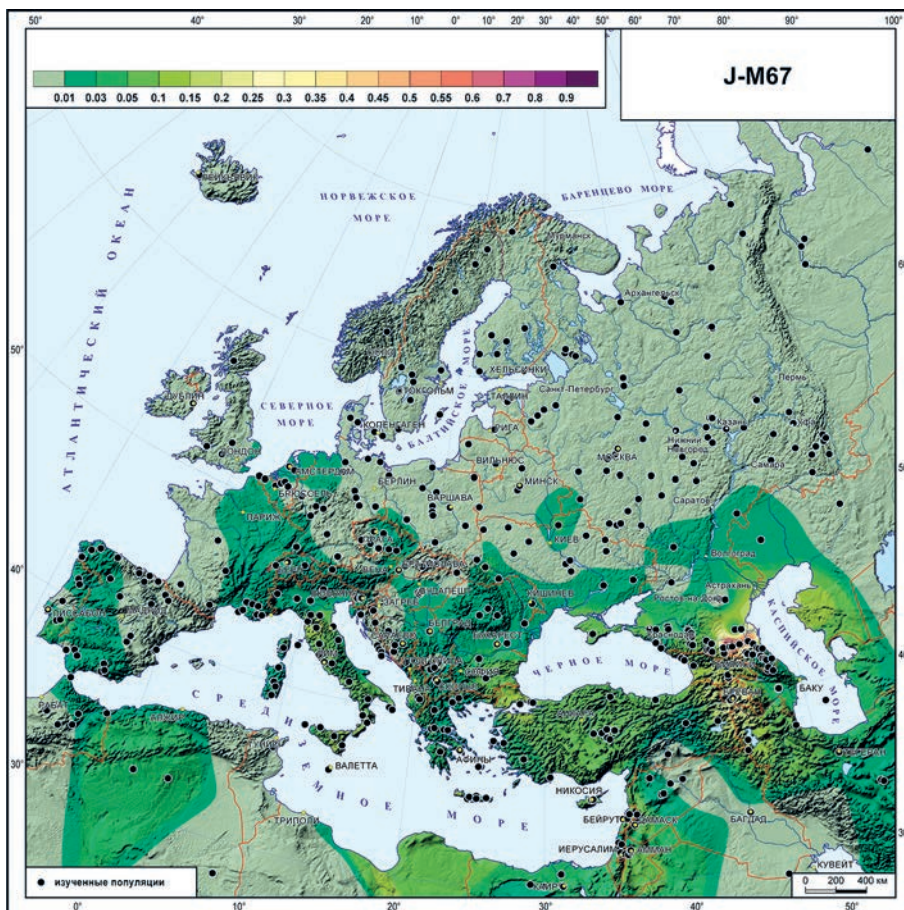


Рис. 2.32. География гаплогруппы J-M67 Y-хромосомы в Европе.

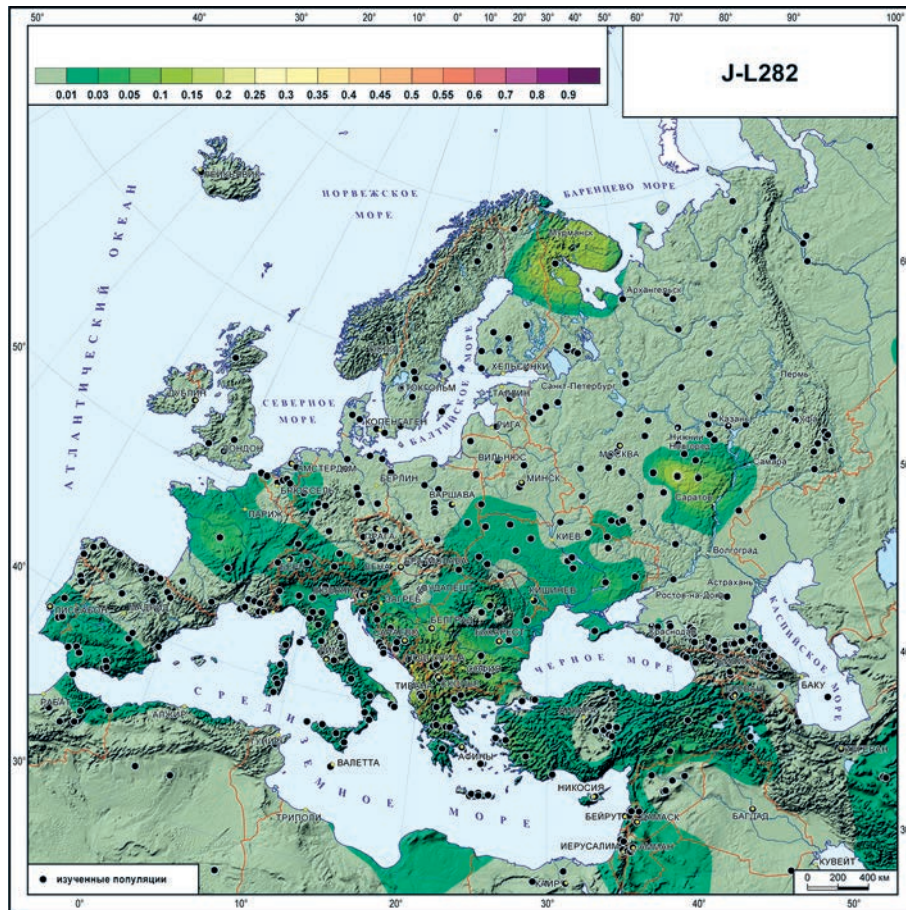


Рис. 2.33. География гаплогруппы J-L282 Y-хромосомы в Европе.

же особенно южную), юг Балканского полуострова. Но максимальных частот (красный цвет на карте) эта гаплогруппа достигает в некоторых популяциях Северного Кавказа. Как было показано в нашем исследовании [Balanovsky et al., 2011; глава 7 данной книги], а также в работе [Yunusbaev et al., 2012] кавказский максимум приурочен к популяциям нахской группы (чеченцам и ингушам).

ГАПЛОГРУППА J-L152 И ЕЕ СУБВАРИАНТ J-M67

География гаплогруппы **J-L152** на территории Европы довольно обширна: зона низких частот тянется с северо-запада Балканского полуострова через Центральную Европу и до Пиренейского полуострова. Зона повышенных частот (от 15%) охватывает Малую Азию и Кавказ. Практически повторяет эту закономерность, но несколько в более

узком географическом охвате и в целом с более низкими частотами, распространение ее «сыновней» ветви – гаплогруппы **J-M67** (рис. 2.32). Однако при этом гаплогруппа **J-M67** по своей частоте больше приурочена к Кавказу, составляя на территории Малой Азии не более 13% генофонда.

ГАПЛОГРУППА J-L282

Распространение гаплогруппы **J-L282** в Европе (рис. 2.33) заметно отличается от предыдущей ветви. На фоне обширной зоны низких частот выделяются три области, где доля гаплогруппы **J-L282** в генофонде повышена до 10–15%: север Фенноскандии (правда, этот максимум опирается лишь на данные по одной популяции), Поволжье (популяция мордвы мокша) и север Балканского полуострова (несколько популяций албанцев).

РЕДКИЕ ГАПЛОГРУППЫ L, Q, T

ГАПЛОГРУППА L-M11

Гаплогруппа **L-M11** (рис. 2.34) в Европе приурочена к юго-восточным регионам (за исключением единичного случая в Бельгии): Крымский полуостров, европейская часть Турции и Северный

Кавказ. При этом частота гаплогруппы **L-M11** плавно увеличивается в юго-восточном направлении, достигая 14% на территории Восточного Кавказа у чеченцев и 13% на северо-востоке Малой Азии у турок. Далее ее частота повышается на Ближнем и Среднем Востоке, достигая индийско-

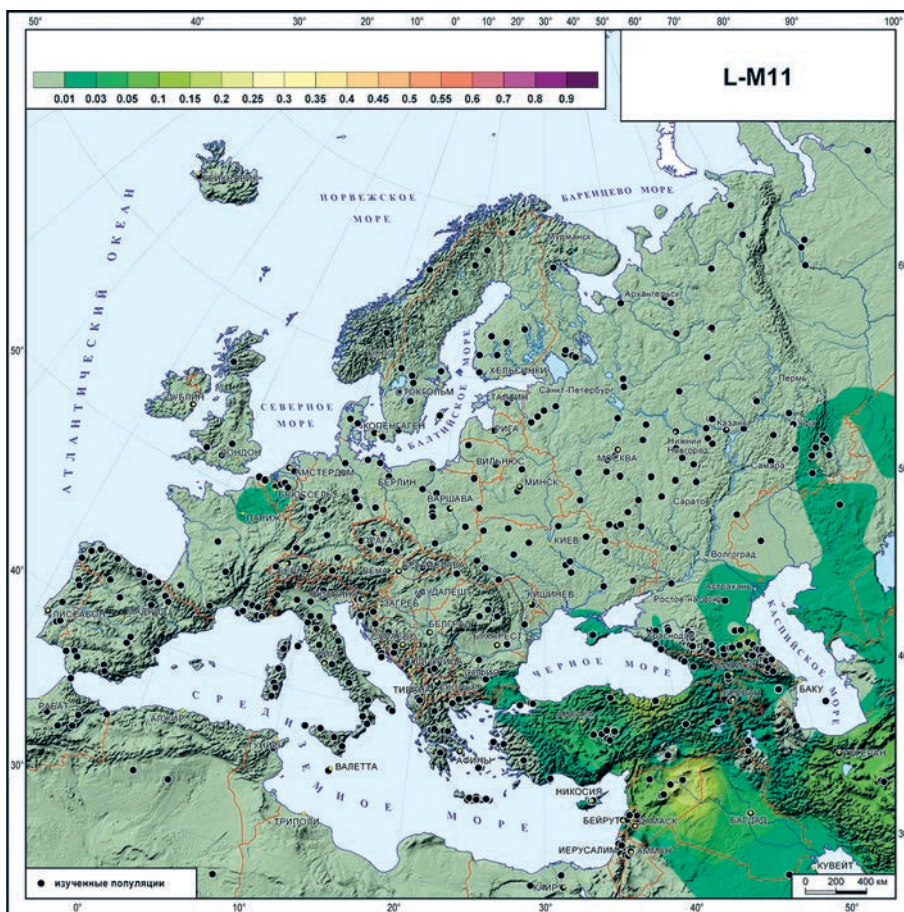


Рис. 2.34. География гаплогруппы L-M11 Y-хромосомы в Европе.

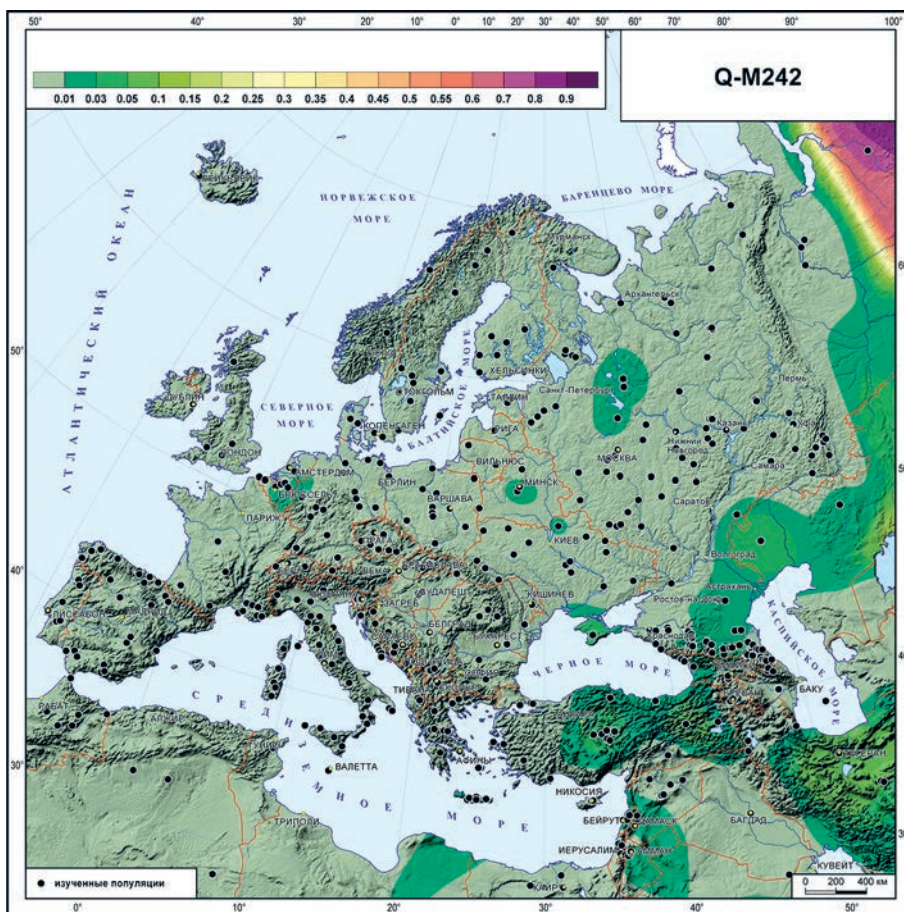


Рис. 2.35. География гаплогруппы Q-M242 Y-хромосомы в Европе.

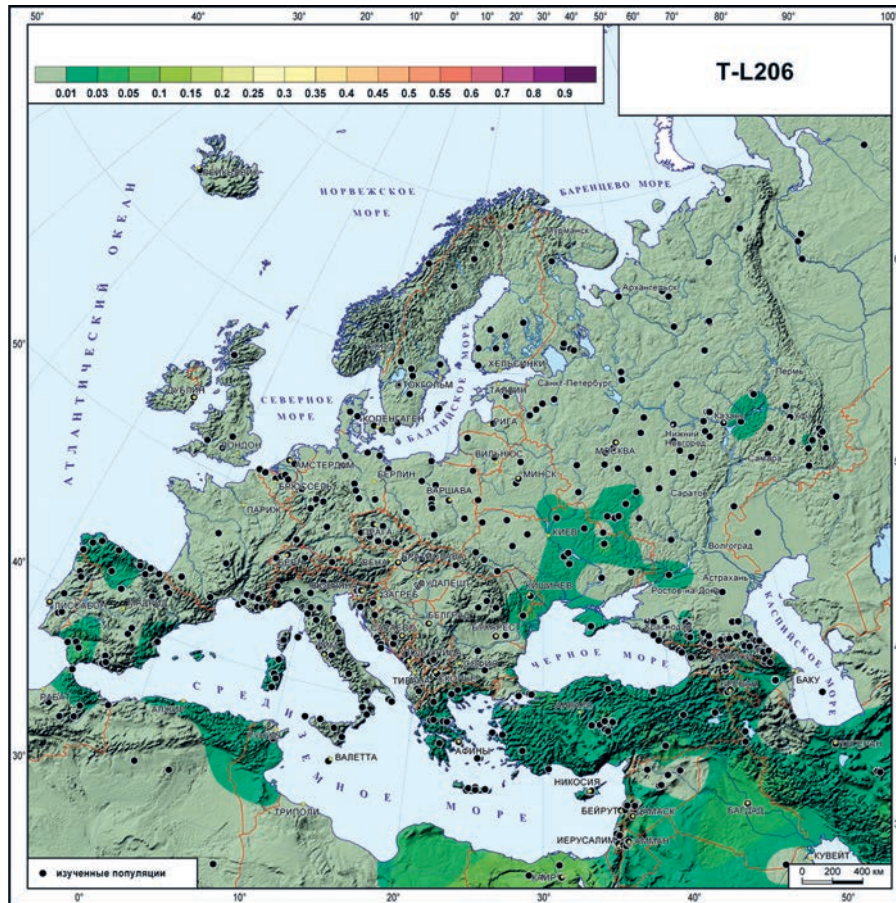


Рис. 2.36. География гаплогруппы T-L206 Y-хромосомы в Европе.

го максимума уже далеко за пределами картографируемого ареала.

ГАПЛОГРУППА Q-M242

Гаплогруппа **Q-M242** в Европе практически не встречается, за исключением нескольких точек с низкими частотами (до 3%) в Центральной и Восточной Европе, на Крымском полуострове и в восточной половине Малой Азии (рис. 2.35). Плавным коридором от северо-западного побережья Каспийского моря далее на восток с постепенным ростом частоты ареал гаплогруппы **Q-M242** переходит в Центральную Азию и Сибирь, где она достигает своего евразийского максимума.

ГАПЛОГРУППА T

Частота гаплогруппы **T-L206** в народонаселении Европы не превышает 5% (рис. 2.36). Ее ареал в Европе лоскутный и сводится на западе к трем узким областям на Пиренейском полуострове и в Сардинии, а на востоке – к зоне в Северном Причерноморье с переходом на северо-восток в украинскую лесостепь, единичными случаями в Поволжье у казанских татар и на Северном Кавказе у кубанских ногайцев. За пределами Европы частота гаплогруппы **T** столь же невелика, но зато ее ареал становится уже не лоскутным, а почти непрерывным, охватывая всю Малую Азию, Ближний Восток и северо-восток Африки.

Y-ГЕНОФОНД ЕВРОПЫ – ПРИНЦИП ПАЗЗЛОВ

Таким образом, проведенное исследование изменчивости Y-хромосомы в популяциях Европы, основанное на объединении обширных собственных данных с литературными, подтвердило и уточнило вывод о том, что основной чертой структурированности генофонда является четкое выделение географических зон, в каждой из которых преобладает своя гаплогруппа. Это убедительно де-

монстрируют карты распределения гаплогрупп (рис. 2.2–2.36).

Однако при размещении данных по каждой гаплогруппе на отдельной карте трудно отследить перекрывание – или же, напротив, не перекрывание – ареалов гаплогрупп. Поэтому мы совместили зоны распространения всех гаплогрупп на одной карте (рис. 2.37). На этом рисунке контур каждой из девяти-



Рис. 2.37. Структура генофонда Европы по Y-хромосоме: «принцип пазлов».

В каждой географической части Европы доминирует одна гаплогруппа, редко встречающаяся в остальных частях. Карта показывает, к каким частям Европы приурочена каждая из основных гаплогрупп. Карта основана на точных значениях частот. Цветом показаны зоны с частотой гаплогруппы выше порогового значения 0,35 (то есть на выделенных цветом территориях более трети генофонда относится к данной гаплогруппе).

ти основных европейских гаплогрупп строго соответствует территории, на которой частота этой гаплогруппы выше 35%, т.е. более трети генофонда. Можно видеть, что каждая гаплогруппа действительно занимает свой собственный ареал. Области, остающиеся белыми, невелики – это те территории, на которых разнообразие гаплогрупп высоко и ни одна гаплогруппа не достигает уровня 35%.

Полученная карта ярко демонстрирует «принцип пазлов» – высокую географическую специфичность гаплогрупп Y-хромосомы. Благодаря этому на каждой территории, хотя и присутствует много гаплогрупп, но доминируют лишь одна-две. Таким образом генофонд Европы оказывается составлен, как из пазлов, из ареалов преимущественного распространения разных гаплогрупп.

2.3. ГЕНЕТИЧЕСКИЕ ВЗАИМООТНОШЕНИЯ ПОПУЛЯЦИЙ: ТРИ УРОВНЯ

Проведя картографический анализ отдельных гаплогрупп Y-хромосомы и убедившись в четкой географической приуроченности большинства из

них, перейдем к обобщенному (по всем гаплогруппам совместно) статистическому анализу генофонда Европы.

МАССИВЫ ДАННЫХ ДЛЯ СТАТИСТИЧЕСКОГО АНАЛИЗА

Массив данных для статистического анализа, как и массив данных для карт в предыдущем разделе, сформирован на основе нашей базы данных Y-base (раздел 1.1). Как и для картографирования, в массив данных для статистического анализа отбирались только выборки достаточного объема (по

возможности более 70 образцов), а меньшие выборки или исключались, или – в пределах одного народа – объединялись друг с другом. Отличие статистического анализа от картографического состоит в том, что каждая популяция должна быть изучена по каждой гаплогруппе. Однако в большин-

стве статей, особенно не самых последних лет, использовались ограниченные наборы гаплогрупп. Поэтому, если требовать, чтобы каждая выборка была изучена по каждой гаплогруппе, то резко сократится или набор гаплогрупп (менее чем до 10) или набор популяций.

Чтобы избежать и того, и другого, данные по всем выборкам в пределах одного народа были усреднены и в анализе использовались эти среднеэтнические частоты. В таком случае, если по данному народу изучено 10 популяций (выборок), то достаточно, чтобы хотя бы одна из этих популяций была изучена по данной гаплогруппе, и тогда частота этой гаплогруппы в этой популяции будет присвоена всему народу. Если же гаплогруппа была изучена в нескольких региональных популяциях народа, то еще лучше – эти значения усредняются и тогда присвоенное народу значение частоты гаплогруппы получается более надежным. При таком подходе, ориентированном не на локальные популяции, а на народы (этнический уровень популяционной системы) число пустых ячеек в таблице частот гаплогрупп резко сокращается и удается охарактеризовать большинство народов Европы по большинству гаплогрупп.

Правда, какое-то число пустых ячеек все равно остается (для какого-либо народа вообще нет информации о частоте у него какой-либо гаплогруппы). В таких случаях сохраняется проблема выбора – исключать из анализа такой народ или исключать такую гаплогруппу. Анализ был проведен в обоих вариантах и были созданы два массива данных: массив частот 17 гаплогрупп у 62 народов (меньше гаплогрупп, но больше народов) и массив частот 40 гаплогрупп у 40 народов (больше гаплогрупп, но меньше народов). Это увеличение чис-

ла гаплогрупп достигнуто за счет подразделения трех основных гаплогрупп Европа (R1b, R1a и N1c) на их ветви.

При описанном анализе на этническом уровне мы избегаем и искажений частот гаплогрупп, и потери числа гаплогрупп или числа народов. Но кое-что все-таки теряется: при таком подходе из поля зрения выпадает изменчивость на внутриэтническом уровне, то есть никак не могут проявиться различия между популяциями, относящимися к одному народу. Чтобы изучить изменчивость и на этом – внутриэтническом – уровне, был создан третий массив данных. В него вошли данные только по тем народам, для которых имелись данные хотя бы о двух разных популяциях этого народа. Число гаплогрупп составило 15 (две гаплогруппы из набора 17 гаплогрупп были исключены из рассмотрения), число народов составило 20, а число популяций этих народов было большим – 120 популяций.

Наконец, поскольку ориентироваться в большом числе точек на графике бывает нелегко, был создан еще один упрощенный массив данных – региональный. Были выделены 15 регионов Европы, частоты гаплогрупп в которых были рассчитаны усреднением частот у народов, относящихся к этим регионам.

В итоге были получены массивы данных, характеризующие генофонд Европы на трех иерархических уровнях: региональном, этническом и субэтническом. Причем этнический уровень охарактеризован двумя массивами данных: по 17 и по 40 гаплогруппам. Таблица 2.4 суммирует важнейшую информацию о всех этих массивах данных. А таблица 2.5 показывает для каждого народа, в каких массивах он представлен.

Таблица 2.4. Массивы данных, сформированные для статистического анализа генофонда Европы по маркерам Y-хромосомы

Название массива данных	Число анализируемых популяций	Число гаплогрупп	Перечень гаплогрупп
Региональный	15 регионов Европы	17	C-M130, E-M35, G-M201, I-M253, I-P37.2, I-M223, I-M170(xM253,P37.2,M223), J-M304, N-LLY22g, R-M198(xM458), R-M458, R-M343, L-M11, Q-M242, R-M124, T-M70, other (A, D, E, F, H, I, K, O)
Этнический 17	62 народа Европы	17	Те же гаплогруппы
Этнический 40	40 народов Европы	40	E-M35, G-M201, I-M253, I-P37.2, I-M223, J-M304, C-M130, I-M170(xM253,P37.2,M223), L-M11, Q-M242, R-M124, T-M70, R-M198(xM458,M558,Z284,Z93), R-M458, R-M558, R-Z284, R-Z93, R-M343(xM73xM269), R-V88, R-M73, R-M269(xL23), R-L23(xM412), R-M412(xL11), R-L11(xU106xS116), R-U106(xU198), R-U198, R-S116(xM529xU152), R-M529(xM222), R-M222, R-U152, N-M178(xP298), N-L392(xCTS10760,Z1936,F4205,B202), N-Z1936, N-CTS10760, N-F4205, N-B202, N-P298, N-M2118, N-B211, other (A, D, E, F, H, I, K, O)
Субэтнический	120 популяций двадцати народов	15	C-M130, E-M35, G-M201, I-M253, I-P37.2, I-M223, J-M304, L-M11, N-LLY22g, Q-M242, R-M198, R-M343, R-M124, T-M70, other (A, D, E, F, H, I, K, O)

Таблица 2.5. Представленность народов Европы в анализируемых массивах данных

Народ	Представлен в массиве «Этнический 17»	Представлен в массиве «Этнический 40»	Представлен в массиве «Субэтнический» (в скобках – число популяций)	К какому региону относится в массиве «Региональный».
Абазины	+	+	-	Западный Кавказ
Абхазы	+	+	-	Закавказье
Аварцы	+	+	-	Восточный Кавказ
Адыги	+	+	-	Западный Кавказ
Албанцы	+	-	+ (3)	Балканский п-в
Англичане	+	+	-	Британские о-ва
Армяне	+	+	-	Анатолия
Балкарцы	+	+	-	Западный Кавказ
Баски	+	-	-	Пиренейский п-в
Башкиры	+	-	+ (5)	Степи Европы
Белорусы	+	+	+ (8)	Восточно-Европейская равнина
Бельгийцы	-	-	+ (5)	Западная Балтика
Болгары	+	-	+ (9)	Балканский п-в
Боснийцы	+	+	+ (2)	Балканский п-в
Венгры	+	+	-	Степи Европы
Галисийцы	+	-	-	Пиренейский п-в
Герцеговинцы	+	+	-	Балканский п-в
Греки	+	+	+ (5)	Балканский п-в
Грузины	+	-	-	Закавказье
Даргинцы	+	+	-	Восточный Кавказ
Датчане	+	+	-	Западная Балтика
Ингуши	+	+	-	Восточный Кавказ
Ирландцы	+	+	-	Британские о-ва
Испанцы	+	+	-	Пиренейский п-в
Итальянцы	+	+	-	Аппенинский п-в
Караногайцы	+	+	-	Степи Европы
Карачаевцы	+	+	-	Западный Кавказ
Карелы	+	+	-	Северо-Восточная Европа
Каталонцы	+	-	-	Пиренейский п-в
Коми	+	+	+ (2)	Северо-Восточная Европа
Кубачинцы	+	+	-	Восточный Кавказ
Кумыки	+	+	-	Восточный Кавказ
Латыши	+	-	-	Восточная Балтика
Литовцы	+	-	+ (6)	Восточная Балтика
Македонцы	+	+	+ (2)	Балканский п-в
Мокша и эрзя (мордва)	+	-	+ (2)	Правобережье Волги
Немцы	+	+	-	Западная Балтика
Ногайцы	+	-	-	Степи Европы
Осетины	+	+	-	Западный Кавказ
Поляки	+	+	+ (6)	Восточно-Европейская равнина
Португальцы	+	+	-	Пиренейский п-в
Румыны	+	+	-	Балканский п-в

Таблица 2.5. Представленность народов Европы в анализируемых массивах данных (окончание)

Народ	Представлен в массиве «Этнический 17»	Представлен в массиве «Этнический 40»	Представлен в массиве «Субэтнический» (в скобках – число популяций)	К какому региону относится в массиве «Региональный».
Русские	+	+	+ (20)	Восточно-Европейская равнина
Саамы	+	-	-	Северо-Восточная Европа
Сардинцы	+	-	-	Аппенинский п-в
Сербы	+	-	+ (4)	Балканский п-в
Словаки	+	-	-	Восточно-Европейская равнина
Словенцы	+	+	-	Балканский п-в
Сорбы	+	-	-	Западная Балтика
Татары казанские	+	+	-	Восточно-Европейская равнина
Татары крымские	+	-	-	Степи Европы
Турки	+	+	+ (6)	Анатолия
Украинцы	+	+	+ (13)	Восточно-Европейская равнина
Финны	+	-	+ (7)	Северо-Восточная Европа
Французы	+	-	-	Франция
Хорваты	+	+	+ (3)	Балканский п-в
Черкесы	+	-	-	Западный Кавказ
Чехи	+	+	+ (5)	Восточно-Европейская равнина
Чеченцы	+	-	-	Восточный Кавказ
Чуваши	+	+	-	Правобережье Волги
Шведы	+	+	+ (7)	Западная Балтика
Эстонцы	+	+	-	Восточная Балтика

РЕГИОНАЛЬНЫЙ УРОВЕНЬ

Рассмотрение на региональном уровне является предварительным – мы сначала посмотрим на Европу как бы из космоса, чтобы уловить скелет, основной паттерн в структуре ее генофонда, а потом уже перейдем к главному анализу – на этническом уровне. Генетические взаимоотношения регионов Европы представлены на рис. 2.38.

Прежде всего, хорошо видны отличия генофонда Кавказа: Западный Кавказ и Восточный Кавказ при взгляде «из космоса» генетически близки и друг к другу, и к народам Закавказья и Анатолии. (Последние два, строго говоря, не относятся к Европе, но включены в анализ для полноты картины). Зато все эти четыре региональных генофонда расположены далеко от генофондов собственно Европы.

Европейские же генофонды делятся на два кластера – Западной Европы и Восточной. В пределах Западной Европы генофонды Средиземноморья (Аппенинский и Пиренейский полуострова) находятся хотя и не близко к анатолийско-кавказскому кругу, но все же ближе к ним, чем генофонды

Франции и Британских островов. А эти генофонды, географически более северные, и в генетическом пространстве находятся выше («севернее») средиземноморских.

Мостом между западно-европейскими и восточно-европейскими генофондами служит население Балтики, причем Западная Балтика генетически ближе к Западной Европе, а Восточная Балтика – к Восточной Европе. Географически так и должно быть, поскольку «Западная Балтика» в нашем понимании (табл. 2.5) включает и Швецию, и Данию, и Германию, то есть значительную часть пограничья Западной и Восточной Европы. Правда, регион «Восточно-Европейская равнина» также граничит с Западной Европой. Но ее генофонд на графике и расположился очень близко к генофонду Восточной Балтики.

Положение восточноевропейских генофондов относительно друг друга также вторит географии: популяции Северо-Восточной Европы и в генетическом пространстве (рис. 2.38 А) располагаются справа и выше («северо-восточнее») от популяций

Восточно-европейской равнины. А популяции Правобережья Волги (табл. 2.5), которые географически южнее большинства популяций Восточно-Европейской равнины, и на генетическом графике располагаются ниже («южнее») ее (рис. 2.38 А). Популяции восточноевропейских степей, географически расположенные на самом юге Восточной Европы и примыкающие к Кавказу, и на генетическом графике оказываются «южнее» остальных восточно-европейских регионов, и оказываются единственным регионом Европы, генетически приближенным к Кавказу.

Итак, генетические взаимоотношения регионов Европы отражают соотношения их географического положения друг относительно друга. Учитывая, как много только исторически документированных миграций (и, вероятно, еще больше не документи-

рованных) происходило в Европе за все время истории ее населения и во всех мыслимых направлениях (что и сделало показатель генетических различий популяций Европы наименьшим среди таких показателей для всех регионов мира [Балановская, 1998]), неудивительно, что по географической близости или удаленности регионов друг от друга можно хорошо предсказывать, соответственно, обилие или немногочисленность миграций между ними и, соответственно, близость или удаленность их генофондов. На самом деле картина интереснее простого отражения географии в генофонде: хотя взаимное положение регионов следует географии, но вот генетические расстояния между регионами отнюдь не прямо пропорциональны географическим. Особенно хорошо это видно в удаленности кластеров Восточной и Западной Европы (географически боль-



Рис. 2.38. Структура генофонда Европы на региональном уровне.

График главных компонент, отражающий взаимоотношения генофондов 15 регионов Европы. Используются данные о частотах 17 гаплогрупп Y-хромосомы. Первая главная компонента описывает 29%, а вторая – 17% общей дисперсии.

шинство входящих в них регионов не настолько удалены друг от друга), а также в генетической удаленности Кавказа, хотя географически он близок восточноевропейским регионам.

Кроме общего правила согласованности генетического и географического положения, из рисунка 2.38 следуют и другие исключения. Наиболее яркое из них – совпадение генофонда Британских островов и Балкан: их генофонды не должны бы быть похожи. Они действительно не похожи, что хорошо видно по частотам гаплогрупп и их картам. И анализ главных компонент тоже хорошо уловил эти раз-

личия, но отразил их по третьей главной компоненте. По ней генофонды Британских островов и Балканского полуострова характеризуются противоположными, полярными значениями, но при проекции многомерного графика на плоскость первой и второй главных компонент эти два региона наложились друг на друга. Несмотря на это, рассмотрение первых двух главных компонент весьма информативно, поскольку вместе они описывают почти половину суммарной изменчивости, что редко когда достигается при анализе более 10 независимых признаков (в нашем случае 17 гаплогрупп).

ЭТНИЧЕСКИЙ УРОВЕНЬ: 17 ГАПЛОГРУПП, МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ

ЭТНИЧЕСКИЙ УРОВЕНЬ

Этнический уровень является основным – и наиболее устойчивым, и наиболее хорошо изученным – среди всех уровней иерархической структуры популяций. Это показано и экспериментально [Балановская, 1990] и интуитивно признается большинством исследователей, потому что любую изученную ими выборку называют прежде всего по названию народа, к которому она относится. (Правда, в Европе могут называть как народ, так и страну, но большинство стран Европы, за исключением разве что России и Великобритании, если и не мононациональны, то близки к этому – в отношении их коренного населения). Этнический уровень является основным и для нашего анализа структуры генофонда Европы, поэтому его мы проанализируем трижды: методом многомерного шкалирования на основе массива данных «Этнический 17» (табл. 2.4), методом главных компонент на основе того же массива данных и методов многомерного шкалирования на основе массива данных «Этнический 40» (табл. 2.4).

4 ЗОНЫ НА ГРАФИКЕ

Генетические взаимоотношения народов Европы при анализе методом многомерного шкалирования показаны на рис. 2.39. При взгляде на этот график бросается в глаза его сложность – почти все поле заполнено многочисленными популяциями и поэтому их группировка в кластеры неоднозначна. Но генетические взаимоотношения между почти всеми народами, существующими в Европе, и не должны быть простыми при их проецировании на двумерный график. Попробуем разобраться в полученной картине. Прежде всего поделим график на четыре зоны – в соответствии с географией Европы и прилегающих регионов. В левой нижней зоне окажутся генофонды Средиземноморья (в соответствии с их географическим положением на юго-западе Европы). В правой нижней зоне окажутся генофонды Передней Азии (включая и

Кавказ), что соответствует их географическому положению на юго-восток от Европы. В правой верхней зоне окажутся генофонды Восточной Европы, а в левой верхней – генофонды Западной Европы, опять-таки в соответствии с географией. То есть первый вывод, который мы можем сделать и который облегчит дальнейшее рассмотрение графика – это то, что согласованность положения регионов Европы в географическом и генетическом пространстве сохраняется (как и должно быть) и после того, как каждый регион представлен уже не одной точкой (рис. 2.38), а множеством точек входящих в этот регион народов (рис. 2.39). Рассмотрим поочередно каждую из этих четырех зон.

ЗОНА 1: СРЕДИЗЕМНОМОРЬЕ

Средиземноморье (южная часть Европы) географически представляет собой цепь из трех больших полуостровов и ряда больших и малых островов. Полуострова, перечисляя с запада на восток – Пиренейский, Аппенинский, Балканский. Далее к востоку находится полуостров Малая Азия (Анатолия), но он формально уже не относится к Европе. И на генетическом графике мы видим ту же цепь генофондов народов, населяющих эти полуострова: слева расположились популяции Пиренейского полуострова; правее (впрочем, без резкого перехода) расположился генофонд Аппенинского полуострова и соседнего с ним острова Сардиния; еще правее – генофонды многочисленных народов Балканского полуострова. Часть балканских генофондов уходит далеко за пределы зоны Средиземноморья, вклиниваясь в скопления этнических точек не только Восточной Европы (что соответствует отсутствию четкой географической границы между Балканами и Восточной Европой), но и в зону графика, занятую популяциями Западной Европы, что выглядит странно. Но в целом, хотя идеально-го соответствия географического и генетического положения внутри регионов и не прослеживается (чтобы его достичь, каталонцы должны были бы поменяться местами с галисийцами, испанцы – с

португальцами, а сардинцы – с итальянцами), то, что популяции одного региона кластеризуются совместно, а расположение этих регионов формирует такую же цепь, как и в географическом пространстве, показывает явный параллелизм между структурированностью генофонда Европы и ее географической структурированностью.

ЗОНА 2 – ПЕРЕДНЯЯ АЗИЯ

Этот обширный регион в нашем анализе представлен только двумя своими частями, пограничными с Европой – Анатолией и Кавказом. Генофонд Анатолии прекрасно вписывается в цепь полуостровов Средиземноморья, примыкая к генофондам народов Балкан, с которыми Анатолия соседствует и географически. Анатолия в анализе представлена не только турками, но и армянами (поскольку восточные области Анатолии вплоть до

геноцида армян в начале 20 века были населены преимущественно армянами). На графике оба этих народа Анатолии кластеризуются совместно, отражая, по всей видимости, общий древний анатолийский субстрат в их генофондах или даже прямую ассимиляцию части армян турками. Кавказ в анализируемых данных представлен подробно, и четко делится на кластеры Западного Кавказа, Восточного Кавказа и Закавказья. Как и на графике регионов (рис. 2.38), из всех европейских популяций Кавказ приближен только к популяциям степей Восточной Европы. Причем этот мостик переброшен через тюркоязычные народы Западного Кавказа (карачаевцев и балкарцев). Конечно, одного лишь положения этих народов на паневропейском графике недостаточно, чтобы делать выводы об их этногенезе (генофонду Кавказа посвящена отдельная глава 7), но все же можно выдвинуть гипотезу о наличии в их генофондах не только ассимилиро-



Рис. 2.39. Структура генофонда Европы на этническом уровне: многомерное шкалирование по данным о 17 гаплогруппах. Использованы данные о частотах 17 гаплогрупп Y-хромосомы у 63 народов Европы (массив данных «Этнический 17», табл. 2.4. и 2.5.). Для каждого народа цвет значков обозначает регион, а форма значка – лингвистическую группу, цвет надписей народов и регионов соответствует цвету значка, цвета и формы стандартизованы для всех рисунков серии (рис. 2.39. – 2.41.). Приведен график многомерного шкалирования, основанный на матрице генетических расстояний Нея. Уровень стресса составляет 0.19, алиенации – 0.20.

ванного тюрками автохтонного компонента Кавказа, но и компонента, пришедшем на Кавказ из степей вместе с носителями тюркских языков.

ЗОНА 3 – ЗАПАДНАЯ ЕВРОПА

Поскольку Средиземноморье мы рассматриваем отдельно, на долю Западной Европы осталось не так много народов. На графике присутствуют народы Британских островов, французы, немцы, датчане и шведы. (Еще несколько народов, например, голландцы и норвежцы, к сожалению, надежно не изучены по выбранной панели гаплогрупп). Популяции Британских островов, в четком соответствии с их географическим положением формируют островной кластер в самой левой («западной») части графика. Французы, географически принадлежащие частью к Средиземноморью, а частью граничащие с англичанами и немцами, на генетическом графике тоже находятся между двумя этими популяциями, но решительно приближаются к пиренейскому кластеру Средиземноморской зоны, отдаляясь от немцев. Немцы, напротив, не преподносят никаких сюрпризов, располагаясь на графике в точном соответствии с их географическим положением. К «северу» от них, опять-таки в соответствии с географией, располагаются датчане, а еще «северо-восточнее» (выше и правее на графике), как можно было бы ожидать – шведы. В целом, генетические взаимоотношения западноевропейских народов с немецкой пунктуальностью воспроизводят их географическое положение, и эту монотонную картину нарушает лишь пара неугомонных балканских народов – боснийцы и герцеговинцы – далеко отлетевшие на графике от остальных популяций Балкан в западно-европейскую зону и занявшие место, которое должно бы – по географии – принадлежать отсутствующим в нашем анализе норвежцам. (Это исключение уже упоминалось при описании зоны Средиземноморья, и к нему мы еще раз вернемся ниже).

ЗОНА 4 – ВОСТОЧНАЯ ЕВРОПА

Среди восточноевропейских народов наиболее многочисленной группой являются народы Восточно-Европейской равнины. Они занимают и самую большую территорию, значительно превышающую площадь остальных выделенных в нашем анализе географических регионов. Но на генетическом графике все они сбились в компактный кластер, отражая генетическую общность населения огромной территории – от Польши до Урала. Правда, внутри кластера, как мы видели и для кластеров Средиземноморья, географические закономерности не прослеживаются, например, татары расположились немного «западнее» русских. В эту группу популяций Восточно-Европейской равнины вошли,

прежде всего, западные и восточные славяне, а также татары. Об этой генетической общности мы еще будем подробно говорить в главе 6, посвященной балто-славянским народам. Пока же отметим единственное исключение – чехов, которые, в противовес их географической, лингвистической и исторической близости к словакам и полякам, на графике отлетели далеко на «запад» и расположились вблизи генофондов немцев. Второе исключение – северные русские, которые демонстрируют большие генетические отличия от прочих славян и близость к популяциям Восточной Балтики и Северо-Восточной Европы, но о своеобразии северных русских уже написано достаточно [Балановская, Балановский, 2007; Balanovsky et al., 2008; Балановская и др., 2011].

Следующий регион, географически расположенный к северу от Восточно-Европейской равнины – Восточная Балтика (в данном анализе соответствует обычному понятию Прибалтики и включает латышей, литовцев и эстонцев) – и на генетическом графике расположен к «северу» от популяций Восточно-Европейской равнины. Близкое положение этих популяций на генетическом графике соответствует небольшим географическим расстояниям между ними. Эти популяции Прибалтики служат мостом от популяций Восточно-Европейской равнины к популяциям Северо-Восточной Европы – опять-таки в соответствии с их географически промежуточным положением. Эти популяции северо-востока Европы (карелы, финны, саами и коми) не преподносят никаких сюрпризов, кроме разве что какой-то слишком уж выраженной близости своих генофондов, не соответствующей ни большим географическим расстояниям между ними, ни хорошо известному для них действию дрейфа генов, который должен был бы подальше разнести их генофонды друг относительно друга. Впрочем, при анализе Y-хромосомы слишком много зависит от уровня филогенетического разрешения отдельных гаплогрупп: поскольку во всех этих генофондах преобладает гаплогруппа N1c-M178, а в этом анализе она еще не подразделена на недавно обнаруженные субтипы [Karmin et al., 2015], то большим межпопуляционным различиям проявиться труднее. Следующий регион – Правобережье Волги – объединяет мордву-мокшу, мордву-эрзю и чувашей. Конечно, географически их ареалы расположены на Восточно-Европейской равнине, но их особое положение на графике вкупе с тесным географическим соседством друг с другом побудило меня рассматривать их как отдельный регион. Они и географически, и генетически располагаются между другими популяциями Восточно-Европейской равнины и популяциями степей Европы. А среди их соотношений друг с другом стоит отметить разве что большие генетические различия между мокшей и эрзей.

Последний из восточно-европейских регионов – степи Европы – выделен не столько по географическому, сколько по «историческому» признаку. Хотя степная зона Европы и представляет собой единую полосу от Урала до Среднедунайской низменности, но тогда в нее надо было бы включать и украинцев (которые обозначены мной как представители народов Восточно-Европейской равнины отчасти из-за их большого генетического сходства с другими славянскими популяциями этой равнины). Все же стоит обратить внимание, что большинство включенных в анализ степных народов Европы (караногайцы, ногайцы, крымские татары, венгры) формируют единую, хотя и вытянутую, зону на графике (только башкиры занимают совсем другое место на графике). Это может служить основанием для гипотезы о наличии общего «степного» пласта в их генофондах, которую, как и любую гипотезу, надо далее проверять.

Итак, планомерно пройдя по всему графику, мы каждый раз выявляли хорошее согласие между, с одной стороны, географическим положением регионов и народов друг относительно друга и, с другой стороны, их генетическими соотношениями. Из этого правила – «генетика следует за географией» – есть, конечно, и несколько исключений. Наиболее яркое из них – уже дважды упоминавшееся положение боснийцев и герцеговинцев вблизи западноевропейских популяций на графике. Но судя по самой матрице генетических расстояний, рассчитанных по частотам гаплогрупп, пять народов, наиболее генетически сходных с боснийцами, это герцеговинцы, румыны, сербы, хорваты, македонцы; а пять народов, наиболее сходных с герцего-

винцами, это боснийцы, румыны, хорваты, сербы, македонцы. То есть и боснийцам, и герцеговинцам прежде всего генетически близки другие народы Балкан, а вовсе не Западной Европы. Потому и на графике боснийцы и герцеговинцы должны бы занять место в балканском кластере, но при визуализации матрицы в виде двумерного графика были помещены далеко в стороне. В том-то и состоит основной недостаток метода многомерного шкалирования, что, не теряя информации (используются генетические расстояния, в расчете которых принимали участие все гаплогруппы), график шкалирования теряет в точности. Ведь многомерное пространство (сколько точек на графике, столько измерений) никогда не проецируется на двумерную плоскость идеально, и геометрические расстояния между точками воспроизводят генетические расстояния между популяциями лишь с наибольшей возможной степенью. Показатели стресса и аллиениции как раз и измеряют степень искажения, внесенную при визуализации на двумерном графике. В нашем случае они высоки (0.19 и 0.20), поэтому скорее удивительно, что популяций с искажением их положения так немного (явные случаи такого искажения только боснийцы и герцеговинцы), чем то, что популяции с искажениями вообще есть.

Поэтому попробуем проанализировать тот же самый массив данных другим методом – главных компонент, который, в отличие от многомерного шкалирования, не теряет в точности, хотя и теряет значительную часть информации: первые две компоненты описывают лишь часть общей изменчивости использованных для анализа гаплогрупп.

ЭТНИЧЕСКИЙ УРОВЕНЬ: 17 ГАПЛОГРУПП, ГЛАВНЫЕ КОМПОНЕНТЫ

Генетические взаимоотношения народов Европы, выявленные методом главных компонент, показаны на рис. 2.40. По идее, этот график должен бы совпасть с предыдущим (рис. 2.39), потому что исходные данные для обоих видов анализа идентичные (массив «Этнический 17», способ визуализации тоже идентичен (расположение точек-популяций на двумерном графике) и отличается только сам статистический метод визуализации – многомерное шкалирование на рис. 2.39 и главные компоненты на 2.40.

В первом приближении графики действительно похожи, поскольку на рис. 2.40. выделяются те же самые четыре зоны, соответствующие четырем географическим частям Европы и ближайших к ней частей Передней Азии: переднеазиатская зона в правой нижней части графика (юго-восток изучаемого региона), средиземноморская зона в левой нижней части графика («юго-запад»), западноевропейская зона в левой верхней части графика («се-

веро-запад») и восточноевропейская зона в правой верхней части графика («северо-восток»). Но положение регионов и тем более народов в пределах этих зон более хаотично, чем на рис. 2.39. Рассмотрим зоны поочередно.

ЗОНА 1 – СРЕДИЗЕМНОМОРЬЕ

Как и на предыдущем графике, можно выделить цепь региональных кластеров, соответствующую полуостровам Средиземноморья. Но народы Пиренейского и Аппенинского полуостровов перемешаны друг с другом, хотя их совместное облако почти не перекрывается с облаком народов Балканского полуострова. Также совершенно не наблюдается близости генофондов Балканского полуострова и Анатолии (рис. 2.40), которая столь явно проявляется и на географической карте, и при анализе тех же данных методом многомерного шкалирования (рис. 2.39).

ЗОНА 2 – ПЕРЕДНЯЯ АЗИЯ

ЗОНА 3 – ЗАПАДНАЯ ЕВРОПА

Народы Анатолии (турки и армяне), как и на предыдущем графике, в целом близки к генофондам Северного Кавказа и Закавказья, хотя парадоксально, что армяне занимают на графике обособленное положение, а турки попадают в само облако кавказских популяций. В пределах Кавказа географическая структурированность почти не проявляется: народы Восточного Кавказа, Западного Кавказа и Закавказья не только не формируют отделенных друг от друга кластеров, но даже не прослеживаются закономерности в их распределении внутри кавказского облака. Зато кардинальные отличия генофонда Северного Кавказа от остальных популяций, входящих в географическое понятие Европы, и на этом графике видны очень четко (рис. 2.40).

В этой зоне, в отличие от Кавказа, закономерности могут быть прослежены, хотя они и выражены очень слабо: французы снова приближены к генофондам Средиземноморья, а оба изученных народа Британских островов (англичане и ирландцы) расположены на графике вблизи друг друга и если их выделить в отдельный кластер, то он будет непосредственно примыкать, но все же не перекрываться с народами Западной Балтики – шведами, датчанами и немцами. Правда, британский кластер тесно граничит и с народами Балкан (что не имеет ни географического, ни исторического смысла), но, как и при анализе на уровне регионов, Балканы и Британские острова, почти совпадая по значениям первых двух главных компонент, принимают противоположные значения по третьей



Рис. 2.40. Структура генофонда Европы на этническом уровне: анализ главных компонент по данным о 17 гаплогруппах. Массив данных и обозначения идентичны рис. 2.39: использованы массив данных «Этнический 17» (табл. 2.4 и 2.5), цвет значков и надписей обозначает регион, а форма значка – лингвистическую группу, как и для всех рисунков серии (рис. 2.39 – 2.41). Приведен график главных компонент. Первая главная компонента описывает 20 %, а вторая – 13 % общей дисперсии.

главной компоненте. То есть различия их генофондов несомненны, и выявляются они в том числе и методом главных компонент, пусть и не на основном графике.

ЗОНА 4 – ВОСТОЧНАЯ ЕВРОПА

Как и на предыдущем графике, генофонды Восточно-Европейской равнины, в соответствии с географией, граничат и с облаком западно-европейских точек, и с балканскими генофондами. Как и на предыдущем графике, народы этого региона, несмотря на обширность Восточно-Европейской равнины, демонстрируют выраженное сходство друг с другом. Характерно, что только чехи опять смещены на «запад», демонстрируя такое генетическое сходство с западноевропейскими народами, которое не проявляют другие славянские и неславянские народы Восточно-Европейской равнины. Опять-таки, повторяя предыдущий график и географическую карту, на «северо-восток» от Восточно-Европейской равнины располагаются генофонды Восточной Балтики (Прибалтики), а еще далее – генофонды народов Северо-Восточной Европы (финны, саамы, карелы, коми). И северные русские опять приближены к этим генофондам и к балтам, тогда как южные и центральные русские располагаются в середине популяций Восточно-Европейской равнины, большинство которых составляют славянские народы. Завершая аналогию с предыдущим графиком, отметим, что народы Правобережья Волги снова расположены между остальными популяциями Восточно-Европейской равнины и популяциями степей Евразии, и в пределах этого региона мокша и эрзя снова генетически далеки друг от друга. Но вся эта согласованность с географией выражена куда слабее, чем на графике многомерного шкалирования (рис. 2.39) – четких кластеров нет и говорить можно лишь о тенденциях на графике, и в пределах каждого региона положение большинства народов хаотично. Пожалуй, единственное, в чем график главных компонент лучше отражает географическое положение (и, можно полагать, региональную структуру генофонда), чем график многомерного шкалирования, это положение башкир, которые на рис. 2.40 присоединились к другим популяциям степей Европы. Эти степные популяции вновь являются единственным мостиком, приближающим европейские генофонды к кавказским, и их приближенность друг к другу на графике снова позволяет задуматься о наличии у них общего «степного» компонента в генофонде. Хотя выражена эта тенденция куда слабее, чем на графике многомерного шкалирования – крымские татары отчего-то попали в середину кавказского облака, а венгры (что, напротив, вполне разумно) присоединились к балканским популяциям, удалившись от прочих степняков.

СРАВНЕНИЕ РЕЗУЛЬТАТОВ МНОГОМЕРНОГО ШКАЛИРОВАНИЯ И ГЛАВНЫХ КОМПОНЕНТ

Итак, два метода анализа одних и тех же данных (рис. 2.39 и 2.40) дали похожие, но отнюдь не идентичные результаты. Поэтому те черты, которые проявились на обоих графиках, можно считать наиболее устойчивыми и объективными. Перечислим их:

- положение народов на генетическом графике соответствует их географическому положению на географической карте (с точностью до положения региона Европы эта закономерность прослеживается очень хорошо, с точностью до положения отдельных народов прослеживается не всегда – это часто видно на графике шкалирования и реже на графике главных компонент);

- генофонд Северного Кавказа отличается от генофонда Европы и примыкает к генофондам Передней Азии;

- генофонд чехов, в отличие от других славянских народов, приближен к западноевропейским генофондам;

- генофонд северных русских является вторым исключением из общей тенденции большого сходства всех западно-славянских и восточно-славянских генофондов, потому что приближен к финно-угорским народам северо-востока Европы (и отчасти к балтским народам);

- изо всех популяций Европы только народы степной зоны приближены к генофонду Кавказа и, несмотря на большую протяженность степей (от Урала до Венгрии) демонстрируют некоторое генетическое сходство друг с другом;

- остальные народы Восточно-Европейской равнины, несмотря на обширность этого региона, генетически близки друг другу.

Основное отличие между двумя графиками – то, что на графике многомерного шкалирования (рис. 2.39) географические закономерности видны очень четко, а на графике главных компонент (рис. 2.40) они смазаны. Конечно, когда два разных метода показывают разные паттерны, в общем случае неизвестно, который из них более правильный. Но в данном случае предпочтение можно обоснованно отдать паттерну многомерного шкалирования. Ведь, во-первых, если бы график многомерного шкалирования отличался бы лишь в силу технических причин, то популяции никак не могли бы сами «догадаться» о своем географическом положении – ведь в анализ закладывались лишь частоты гаплогрупп, а география в исходных данных не присутствовала. Значит, высокая степень структурированности в генофонде присутствует объективно, но лучше выявляется методом многомерного шкалирования, чем главных компонент. Во-вторых, к тому же выводу приводит и рассмотрение отдельных регионов, в ча-

стности, Кавказа. Четкое деление его генофонда в соответствии с географическим делением на Западный и Восточный Кавказ несомненно (глава 7, а также [Balanovsky et al., 2011; Yunusbaev et al., 2011]) и хорошо выявляется при многомерном шкалировании (рис. 2.39), но не проявляется на графике главных компонент (рис. 2.40). Вообще, хотя теоретически эти два метода равноценны, на

практике, по моему опыту, многомерное шкалирование лучше выявляет реально существующую структуру генофонда (не обязательно эта структура географическая, но все равно шкалирование ее обычно выявляет лучше). Потому-то в работах нашего коллектива куда чаще используются графики многомерного шкалирования, чем главных компонент.

ЭТНИЧЕСКИЙ УРОВЕНЬ: 40 ГАПЛОГРУПП, МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ

В этом виде анализа использовано большое число гаплогрупп, хотя при этом и пришлось исключить из анализа народы, по которым не было столь подробных данных. Основное отличие массивов данных «Этнический 17» и «Этнический 40» – то, что основные западно-европейские гаплог-

руппы **R1b**, **R1a** и **N1c** подразделены на субтипы. (Снова и снова повторим, что, хотя этих субтипов сейчас известно очень много, для изучения структуры генофонда и истории популяций полезны лишь те их них, для которых уже проведен скрининг их распространенности в популяциях и оп-



Рис. 2.41. Структура генофонда Европы на этническом уровне: многомерное шкалирование по данным о 40 гаплогруппах. Используются данные о частотах 40 гаплогрупп Y-хромосомы у 40 народов Европы (массив данных «Этнический 40», табл. 2.4 и 2.5). Как и на других рисунках серии (рис. 2.39 – 2.41) для каждого народа цвет значка и надписей обозначает регион, а форма значка – лингвистическую группу. Приведен график многомерного шкалирования, основанный на матрице генетических расстояний Нея. Уровень стресса составляет 0.16, алиенации – 0.17.

ределены частоты их встречаемости хотя бы в двух-трех десятках популяций Европы.) Поэтому график шкалирования по 40 гаплогруппам потенциально может точнее выявить особенности генетических взаимоотношений народов Европы.

Рассматривая этот график (рис 2.41) мы снова видим те же 4 зоны (Средиземноморье, Передняя Азия, Западная Европа, Восточная Европа), а в пределах зон обнаруживаем те же, географические в своей основе, соотношения между генофондами регионов Европы. Чтобы не повторять в третий раз описание тех же закономерностей, обратим внимание лишь на две главные черты.

Первая особенность этого графика – то, что генофонд Балканского кластера переехал из средиземноморской зоны в восточноевропейскую. Это означает, что после разделения основных гаплогрупп на субтипы (что приближает к реальной структуре линий Y-хромосомы) генофонд Балкан проявляет большее сходство с генофондом Восточной Европы, чем Средиземноморья или Западной Европы. Исключением стали лишь греки, генетически приблизившиеся к своим соседям по Средиземноморью – народам Аппенинского полуострова и Анатолии.

Вторая черта – то, что подразделение в каком-либо регионе его мажорной гаплогруппы на субтипы не приводит автоматически к резкому возрастанию генетических отличий народов этого региона друг от друга. Так, несмотря на подразделение гаплогруппы **R1b**, англичане, датчане, ирландцы и шведы остаются на графике вблизи друг от друга, хотя и более четко отделяются от народов Средиземноморья (испанцев, португальцев, итальян-

цев) и опять примкнувших к этому кругу французов. Точно так же, несмотря на подразделение гаплогруппы **R1a**, народы Восточно-Европейской равнины остаются довольно компактным кластером, а подразделение гаплогруппы **N1c** заметно не увеличивает различия между карелами и коми. Можно думать, что такая стабильность положения народов в генетическом пространстве отчасти задается прочими гаплогруппами (кроме мажорной), данные по которым не меняются, а отчасти определяется тем, что даже при подразделении гаплогруппы на субтипы, ареалы субтипов значительно перекрываются если не по зонам максимумов, то по зонам их умеренных частот.

В целом, увеличение числа гаплогрупп с 17 до 40, при неизбежном уменьшении числа рассматриваемых народов, не увеличило информативность графика. Поэтому, если из серии рассмотренных графиков (рис. 2.39 – 2.41) выбирать какой-то один, я бы однозначно остановился на многомерном шкалировании по 17 гаплогруппам (рис. 2.39). Впрочем, вдумчивым читателям, которые будут использовать материалы этой главы для собственных исследований или собственных выводов, не стоит ограничиваться каким-то одним графиком – исследования будут куда надежнее, а выводы обоснованнее при учете и различных видов статистического анализа, и карт распространения гаплогрупп, и роли лингвистического фактора, о котором скоро пойдет речь. Не стоит забывать и о том, что каждый народ, представленный на графике одной точкой, в действительности состоит из многих локальных популяций, которые могут значительно отличаться друг от друга.

СУБЭТНИЧЕСКИЙ УРОВЕНЬ: ЭТНИЧЕСКИЕ ОБЛАКА

ИДЕЯ АНАЛИЗА

Поэтому рассмотрим теперь изменчивость генофонда Европы на внутриэтническом уровне. В этот вид анализа включены те европейские народы, для которых имелись данные хотя бы по двум популяциям. Таких народов набралось двадцать, число изученных популяций для разных народов варьирует от 2 до 20 (особенно подробно изучены восточнославянские народы), а суммарное число популяций – 120 (табл. 2.4 и 2.5). Этот анализ, хотя и ограничен по числу народов, показывает структуру генофонда Европы настолько детально, насколько позволяет внушительный современный объем данных, накопленный за 15 лет интенсивных исследований. Ведь в этом анализе мы имеем дело уже не с усредненными этническими частотами, а со всем многообразием конкретных популяций, населяющих различные уголки Европы.

При рассмотрении этих графиков мы не будем уже специально обращать внимание на взаимоот-

ношения народов друг с другом – об этом выше сказано уже достаточно, да 20 народов и маловато для детальных выводов. Вместо этого мы сосредоточимся на вопросе о том, насколько сходны популяции одного народа друг с другом и насколько они перекрываются с популяциями других народов. То есть речь пойдет о том, насколько важен этнический уровень в структурировании генофонда Европы, то есть насколько принадлежность популяций к одному народу задает сходство их генофондов и, как следствие, насколько обоснован был использованный выше прием усреднения данных по локальным популяциям в среднеэтнические характеристики.

Ранее этот вопрос об уровне внутриэтнической изменчивости в Европе по Y-хромосоме специально не рассматривался (за исключением моей собственной работы [Balanovsky et al., 2008]). Даже исследования, в которых какой-либо народ Европы изучен не в одной, а в нескольких региональных популяциях, немногочисленны. Среди таких

исследований следует особо выделить работу [Kauyser et al., 2005], в которой изучены популяции двух народов – немцев и поляков. Правда, авторы не анализируют и тем более не сравнивают степень внутренней гетерогенности этих народов, сконцентрировавшись на выявлении генетической границы между ними. Ряд популяций поляков (и их субэтнической группы кашубов) изучен и в статье [Rebala et al., 2012]. В работе [Cinnioplou et al., 2004] изучены паттерны географической изменчивости гаплогрупп Y-хромосомы в популяциях Турции. В работе Б.А. Мальярчука [Malyarchuk et al., 2004] исследована изменчивость в пределах русских популяций (главным образом центральных и южных). В исследованиях [Kasperaviciute et al., 2004; Karlsson et al., 2006; Luca et al., 2007; Lappalainen et al., 2008; Mirabal et al., 2009; Лобов, 2009; Larmuseau et al., 2010; Karachanak et al., 2013; Kushnarevich et al., 2013] изучены популяции, соответственно, литовцев, шведов, чехов, финнов, коми, башкир, бельгийцев, болгар, белорусов. В наших собственных исследованиях были очень подробно изучены русские, украинские и белорусские популяции [Balanovsky et al., 2008; Kushnarevich et al., 2015], а также популяции мордвы [наши неопубликованные данные]. В данном анализе, кроме перечисленных источников, использовались и данные разных авторов по финнам, грекам, албанцам, сербам, хорватам, боснийцам, македонцам (для этих народов, хотя и не проводилось специальных исследований ряда популяций, но данные по разным популяциям одного и того же народа появлялись в разных статьях, и попали в нашу базу данных Y-base).

Анализ генофонда Европы на субэтническом уровне я проводил дважды – в 2007 году (опубликовано в [Balanovsky et al., 2008]), и в 2015 году (впервые публикуется в этой книге). Два этих анализа методически однотипны, но различаются объемом данных – за 8 лет значительно увеличилось и число народов, по которым есть данные о нескольких субпопуляциях (9 народов в 2007 году и 20 народов в 2015 году) и число гаплогрупп, по которым есть данные обо всех субпопуляциях всех этих народов (7 гаплогрупп в 2007 году и 15 гаплогрупп в 2015 году). Тем интереснее сравнить полученные закономерности.

АНАЛИЗ 2007 ГОДА (9 НАРОДОВ, 7 ГАПЛОГРУПП)

Результаты анализа 2007 года представлены на рис. 2.42. Этот график демонстрирует удивительно четкую картину: почти во всех случаях популяции, относящиеся к одному народу, кластеризуются совместно. Причем образующиеся «этнические облака» (скопления точек, соответствующих популяциям одного народа) почти не перекрываются друг с другом. Этот результат имеет принципиаль-

ное значение, поскольку он указывает на четкие различия между генофондами народов Европы по маркерам Y-хромосомы.

Два своеобразных исключения касаются славянских групп. Во-первых, северные русские популяции не присоединяются к южным и центральным, а образуют отдельный кластер, находящийся между популяциями шведов, финнов и немцев. Во-вторых, остальные популяции западных и восточных славян – поляков, южных русских и украинцев – формируют не три отдельных, а единое облако, демонстрируя общность своих генофондов. Здесь опять проявляется ведущая роль лингвистического, а не географического фактора: славянские популяции, занимающие половину территории Европы, оказываются не различными (как должно бы быть в случае доминирования географического фактора), а генетически сходными, в согласии с лингвистическими данными о близком родстве восточнославянских и западнославянских языков. Хотя эти данные сами по себе еще не отвергают вывод Россера с соавторами [Rosser et al., 2000] о географической (а не лингвистической) структурированности генофонда Европы, эти результаты указывают, что в отдельных регионах Европы географически различные, но лингвистически сходные популяции оказываются сходными и генетически, указывая на важность и лингвистического фактора в формировании структуры генофонда.

Из полученных результатов можно сделать два вывода. Первый вывод методологический: Y-хромосома является надежной генетической системой с высокой разрешающей способностью. В противном случае генетические различия между этносами не были бы обнаружены или же разные популяции одного народа вели бы себя по-разному, занимая случайные места на графике. Второй вывод – феноменологический: в структуре генофонда проявляется не только географическая, но и лингвистическая структурированность. (В данном случае этнический фактор совпадает с лингвистическим, поскольку этническая принадлежность популяции определяется в основном по языку). К анализу роли географии и лингвистики в структурировании славянских генофондов мы еще вернемся в главе 6.

АНАЛИЗ 2015 ГОДА (20 НАРОДОВ, 15 ГАПЛОГРУПП)

В анализ 2015 года удалось включить в два раза больше и гаплогрупп, и народов, и их субпопуляций. Генетические взаимоотношения этих ста двадцати европейских популяций показаны на рис. 2.43. Популяций так много, что даже современная версия программы Statistica не выводит их на график одновременно; поэтому на рис. 2.43 А выведены популяции всех народов, но восточнославянские народы представлены лишь некоторыми из изучен-

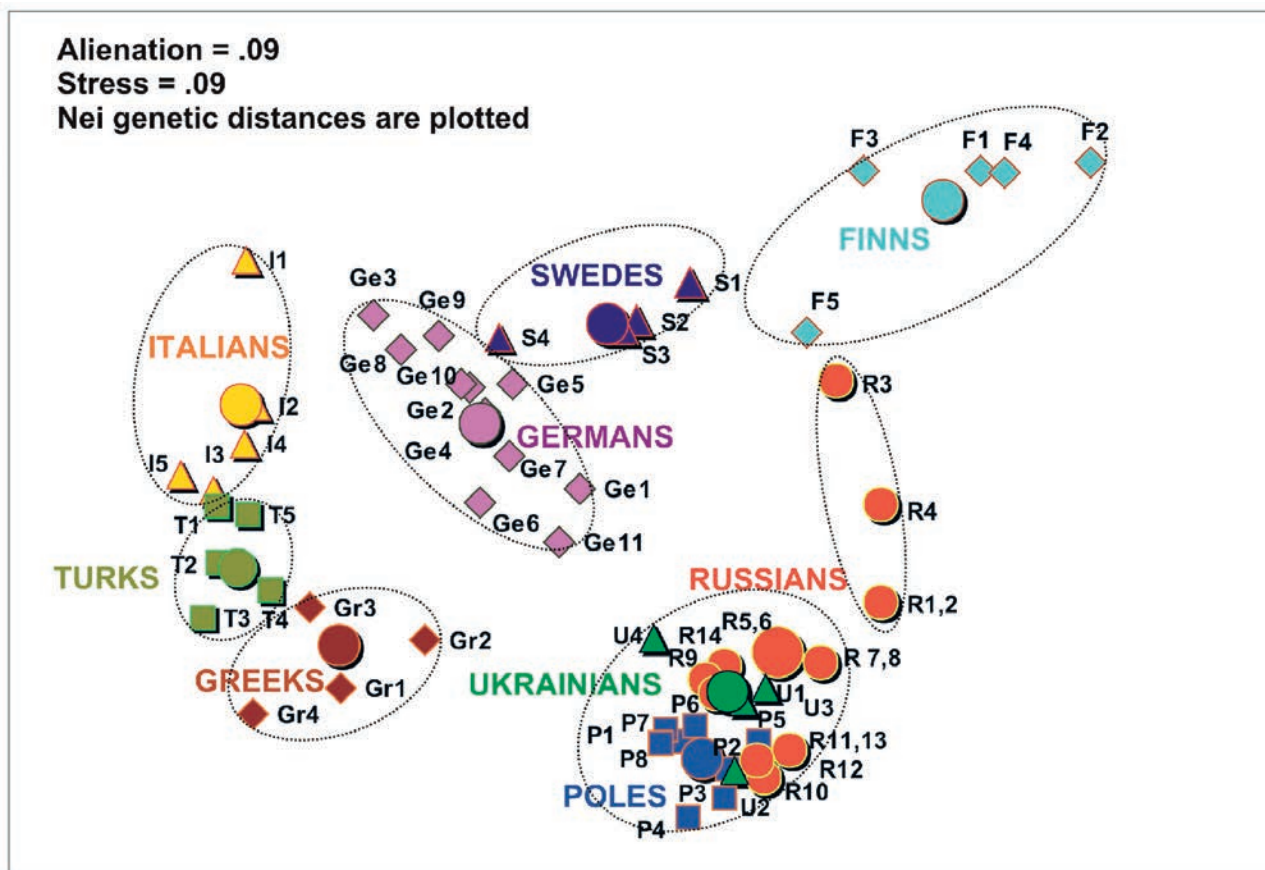


Рис. 2.42. Структура генофонда Европы на субэтническом уровне (анализ 2007 года, воспроизводится по [Balanovsky et al., 2008]). Приведен график многомерного шкалирования. Популяции, относящиеся к одному народу, показаны однотипными значками одного цвета.

ных популяций, а на рис. 2.43 Б выведены лишь народы центрального кластера, проявившегося на рис. 2.43. А, но восточнославянские популяции выведены уже все.

Несмотря на новые данные, закономерность осталась старой – наличие четко выраженных этнических облаков. Действительно, разные региональные популяции одного народа группируются совместно, и эти этнические облака или отделены от этнических облаков других народов или граничат с ними, но за редким исключением не перекрываются. Рассматривая сначала график 2.43 А, видим отграниченные друг от друга этнические облака финнов, шведов, бельгийцев и граничащие друг с другом облака греков, турок и албанцев. При этом облака албанцев и турок четкие и довольно компактные, а греческие популяции генетически очень разнообразны – греки Малой Азии и Крита близки к туркам (что неудивительно, поскольку турки Малой Азии должны были в значительной степени ассимилировать греческое население, особенно на побережье).

Переходя к популяциям остальных народов (рис 2.43 Б) снова видим ту же закономерность: для коми, литовцев, поляков (и кашубов), чехов, украинцев, боснийцев, хорватов и сербов характерны четкие этнические облака: многочисленные по-

пуляции каждого из этих народов сбиваются друг с другом в кучки. С остальными народами – мордвой, русскими, белорусами, македонцами и болгарами картина чуть сложнее. Впрочем, с болгарами и македонцами особенность только в том, что облако болгар оказывается довольно рыхлым, показывая большие межпопуляционные различия, но с другими облаками оно почти не перекрывается – лишь немного с облаком македонцев. Особенность мордвы в том, что популяции мокши и эрзи не кластеризуются друг с другом, хотя размах генетических различий между ними на этом графике оказывается не так уж и велик. Особенность русских, проявлявшаяся и на предыдущем графике (рис. 2.42) – это генетическая обособленность северных популяций, которые на этом графике (рис. 2.43) частично перекрываются с этническим облаком литовцев. С белорусами еще интереснее – они четко делятся на южных (Полесье) и северных (вся остальная часть Белоруссии, то есть центр и север); южные белорусы не формируют своего облака, а попадают в облако украинцев, аналогично северные белорусы попадают в самую середину облака русских популяций. Таким образом, увеличив число гаплогрупп и число изученных локальных популяций, удалось увидеть внутреннюю структуру поляко-русско-украинской генетической

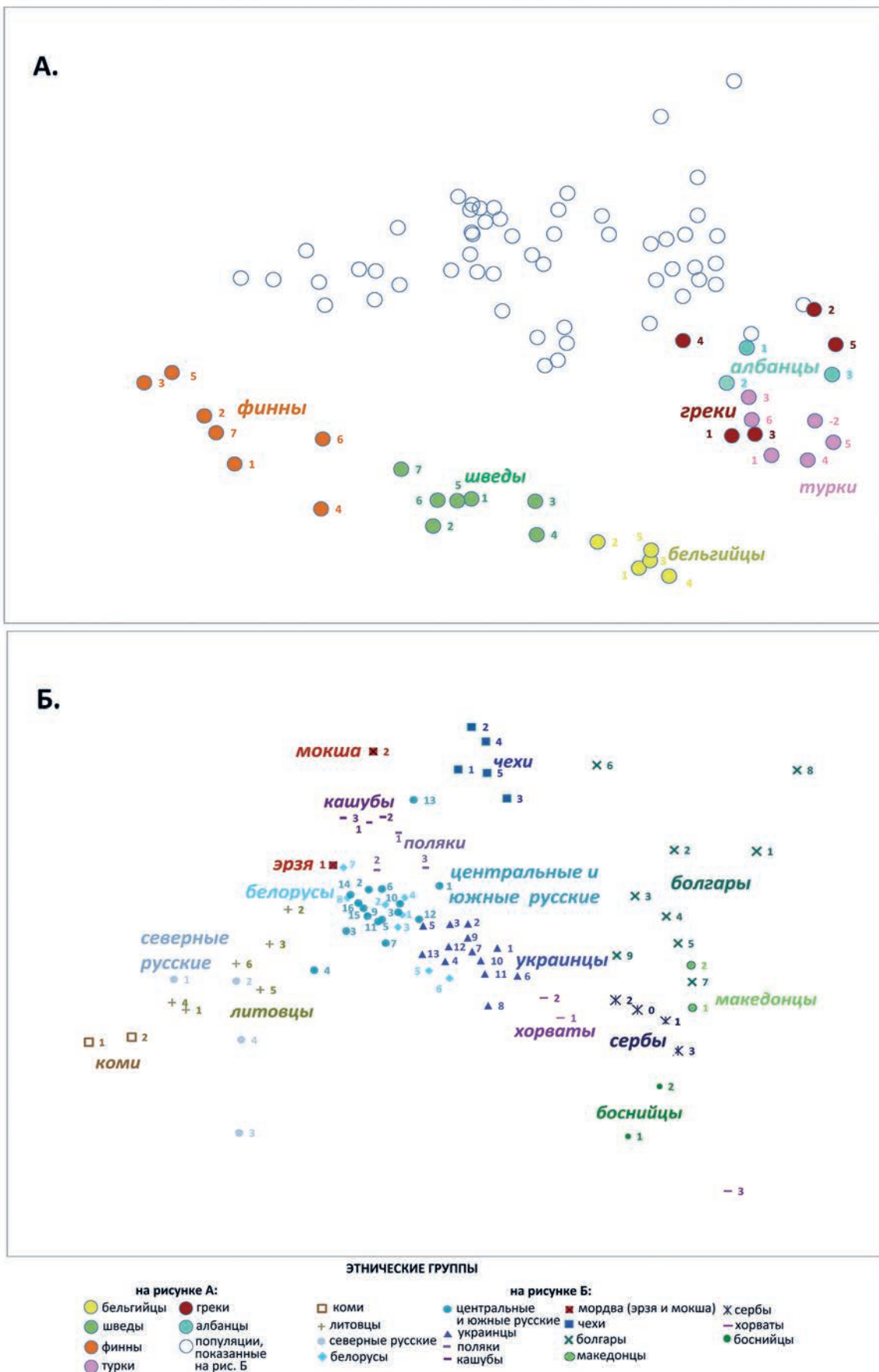


Рис. 2.43. Структура генофонда Европы на субэтническом уровне.

Использованы данные о частотах 15 гаплогрупп Y-хромосомы (массив данных «Субэтнический», табл. 2.4. и 2.5.). В связи с большим генетическим своеобразием башкир их популяции не выведены на график, соответственно, проанализированы 115 популяций 19 народов.

Цвет (а на рис. Б – и форма) значка специфичны для каждого народа. Эти цвета не соответствуют цветам на предыдущих графиках (рис. 2.39 – 2.41) поскольку на них цвет означал регион Европы, а здесь цвет обозначает народ.

А. График многомерного шкалирования, основанный на матрице генетических расстояний Нея и включающий все популяции, кроме некоторых восточнославянских. Уровень стресса составляет 0.12, алиенации – тоже 0.12.

Б. График многомерного шкалирования, основанный на матрице генетических расстояний Нея и включающий все популяции всех народов, показанных выколотыми кружками на рис. А. На этот график выведены все изученные восточнославянские популяции. Уровень стресса составляет 0.10, алиенации – тоже 0.10.

Конкретные популяции в пределах каждого народа обозначены цифрами.

На рисунке А:

Финны : 1 – Хяме, 2 – Северной Остроботнии, 3 – Северной Савонии, 4 – Сатакунты, 5 – Южной Карелии, 6 – Южной Остроботнии, 7 – Юго-Западной Финляндии.

Шведы: 1 – Блекинге и Кристианстада, 2 – Готланда, 3 – Эстергётланда и Йёнчёпинга, 4 – Скараборга, 5 – Упсалы, 6 – Вэрмланда, 7 – Вестерботтена.

Бельгийцы: 1 – Антверпена, 2 – Кампина, 3 – Мехелена, 4 – Северного Брабанта, 5 – Фландрии.

Албанцы: 1 – Албании, 2 – Македонии, 3 – Косово.

Греки: 1 – Малой Азии, 2 – Фессалии, 3 – Крита, 4 – Македонии, 5 – Пелопоннеса.

Турки: 1 – центральные, 2 – восточные, 3 – Стамбула, 4 – северо-восточные, 5 – северо-западные, 6 – юго-восточные.

На рисунке Б:

Коми: 1 – ижемские, 2 – прилужские.

Мордва: 1 – эрзя, 2 – мокша.

Литовцы: 1 – аукшайты восточные, 2 – аукшайты южные, 3 – аукшайты западные, 4 – жемайты северные, 5 – жемайты южные, 6 – жемайты западные.

Белорусы: 1 – юго-западные (Брестская обл.), 2 – северные (Витебская обл.), 3 – центральные, 4 – восточные, 5 – юго-восточные, 6 – юго-западные, 7 – западные, 8 – северные.

Северные русские: 1 – Архангельская обл., Лешуковский р-н, 2 – Архангельская обл., Пинежский р-н, 3 – Архангельская обл., Красноборский и Ленский р-ны, 4 – Вологодская обл.

Центральные и южные русские: 1 – Костромская обл., Мантуровский и Межевской р-ны, 2 – Тверская обл., Кашинский р-н, 3 – Псковская обл., Порховский и Дедовичский р-ны, 4 – Псковская обл., Островский р-н, 5 – Смоленская обл., Рославльский и Ершичский р-ны, 6 – Орловская обл., Ливненский р-н, 7 – Курская обл., Пристенский р-н, 8 – Воронежская обл., Репьевский р-н, 9 – Белгородская обл., 10 – донские казаки (нижний Дон), 11 – донские казаки (верхний Дон), 12 – кубанские казаки, 13 – терские казаки, 14 – Рязанская обл., Сапожковский р-н, 15 – Рязанская обл., Сараевский р-н, 16 – Рязанская обл. (из разных пунктов).

Украинцы: 1 – западные (Закарпатская обл.), 2 – западные (Ивано-Франковская обл.), 3 – западные (Львовская обл.), 4 – центральные (Черкасская обл.), 5 – западные (Хмельницкая обл.), 6 – западные (Черновецкая обл.), 7 – западные (Ровненская обл.), 8 – северные (Житомирская обл.), 9 – северные (Черниговская обл.), 10 – северные (Сумская обл.), 11 – восточные (Белгородская обл.), 12 – Белгородской обл., 13 – запорожские казаки (Запорожская обл.).

Поляки: 1 – Косіеуіе, 2 – курпы, 3 – из Вроцлава.

Кашубы: 1 – центральные, 2 – северные, 3 – южные.

Чехи: 1 – Брно, 2 – Йиндржихув-Градец, 3 – Клатови, 4 – Писек, 5 – Тршебич.

Болгары: 1 – Бургас, 2 – Ловеч, 3 – Монтана (Болгария), 4 – Пловдив, 5 – Софийская обл., 6 – Хасково, 7 – Варна, 8 – Разград, 9 – София.

Македонцы: 1 – из Скопье, 2 – из разных регионов Македонии.

Сербы: 1 – из Приштины (Косово), 2 – из разных регионов Сербии, 3 – из Боснии.

Хорваты: 1 – из Сплита, 2 – из разных регионов материковой Хорватии, 3 – из Боснии.

Боснийцы: 1 – центр и юг Боснии и Герцеговины (Зеница, Мостар, Широки-Бриег), 2 – из разных регионов Боснии и Герцеговины.

общности, которая была гомогенной при анализе 2007 года, основанном на меньшем объеме данных: теперь популяции каждого народа занимают свою собственную зону в пределах этой общности. Учитывая, что украинское и русское облака, хотя не перекрываются друг с другом, но тесно соприка-

саются, генетические взаимоотношения популяций восточных славян на графике многомерного шкалирования можно описать как генетическую общность, состоящую из русской и украинской половин, причем белорусские популяции распределены почти поровну между обеими половинами.

2.4. ГЕОГРАФИЧЕСКАЯ И ЛИНГВИСТИЧЕСКАЯ СТРУКТУРИРОВАННОСТЬ – ЗАКОНОМЕРНОСТИ

Из всего предыдущего рассмотрения структуры европейского генофонда по Y-хромосоме очевиден высокий уровень межпопуляционных различий и четкая кластеризация популяций. Возникает естественный вопрос, каков основной принцип этой структуры, какими факторами она сформирована? Рассматривая те или иные конкретные закономерности, мы уже не раз видели кластеризацию по географическому принципу. Но видели и кластеризацию по принципу лингвистического сходства, а также – близкую по смыслу к лингвистической – кластеризацию в этнические облака. Поэтому в этом разделе описывается межпопуляционную изменчивость основных лингвистических групп Европы и анализируется роль географического и лингвистического факторов. Это вечный вопрос для генофонда любого региона мира – попробуем решить его на текущем уровне знаний применительно к генофонду Европы и к данным по Y-хромосоме.

ВНУТРИЭТНИЧЕСКАЯ ДИФФЕРЕНЦИАЦИЯ

Этот вид анализа является дополнением графика многомерного шкалирования на субэтническом уровне, которым оканчивался предыдущий раздел этой главы. Но график основной упор делает на соотношениях разных популяций друг с другом (в том числе становится видно, существуют или не существуют этнические облака для каждого народа). А о степени сходства разных популяций одного народа друг с другом по графику можно судить лишь по размеру этнического облака. Расчет же межпопуляционной изменчивости для популяций каждого народа, напротив, ничего не говорит о том, как генофонды разных народов соотносятся друг с другом. Но зато он дает количественный и точный ответ на вопрос, насколько сильно разные популяции одного народа отличаются друг от друга генетически, то есть оценивает параметр, который мы называем «этническая гетерогенность».

Такое исследование я тоже проводил неоднократно. В таблице 2.6 приведены результаты 2007 года для Европы (для Сибири – 2010 года), а в таблице 2.7 – результаты 2015 года по расширенному массиву данных. В таблице 2.6, кроме изменчивости популяций внутри отдельных народов (внутриэтническая дифференциация), рассчитана также изменчивость среднеэтнических частот разных народов (межэтническая дифференциация).

Для максимальной объективности, анализ 2007–2010 годов был проведен тремя методами, с использованием трех различных показателей межпопуляционной изменчивости: G_{ST} , AMOVA (оцен-

ки F_{ST}) и средние попарные генетические расстояния. Таблица 2.6 показывает, что в пределах Европы наибольшая гетерогенность характерна для финнов, хорватов, русских и итальянцев; шведы и немцы демонстрируют умеренную гетерогенность; другие народы (греки, турки, поляки, украинцы) характеризуются сниженной гетерогенностью. Высокая этническая гетерогенность турок объясняется высоким филогенетическим разрешением, с которым изучены эти популяции (26 гаплогрупп). Если эти гаплогруппы объединить до уровня, с которым изучены другие народы (8 гаплогрупп), то средние расстояния между популяциями турок падают с 16,5 до 3,5, а G_{ST} оказывается менее чувствительным к набору гаплогрупп, снижаясь с 1,3 до 0,8.

Массив данных 2015 года включает большее число народов, и он выровнен по числу гаплогрупп, что увеличивает сопоставимость результатов по разным народам (табл. 2.7). Закономерности похожи – финны, хорваты и русские вновь в числе лидеров (средние расстояния 0,12–0,15), хотя для русских это определяется дифференциацией на северных и южных, а различия внутри каждой из этих групп не так уж велики. Этим лидерам лишь немного уступают шведы (итальянцы и немцы в таблице 2015 года отсутствуют, поскольку по их региональным популяциям нет надежных данных по панели 15 гаплогрупп). Для турок, поляков и украинцев новыми данными подтверждается сниженная этническая гетерогенность (0,02–0,06). Из числа народов, проанализированных и в 2007, и в 2015 годах, различаются только результаты по грекам: им по новым данным оказывается свойственна не сниженная, а напротив, повышенная этническая гетерогенность; хотя это отчасти объясняется иным набором проанализированных популяций, а отчасти большим числом гаплогрупп, все же такие различия между старым и новым анализами заставляют быть осторожными в оценке этнической гетерогенности греков и подождать появления дополнительных данных по этому народу. Что касается народов, появившихся только в анализе 2015 года, то заслуживает упоминания неожиданно высокая этническая гетерогенность болгар (это мы видели и на графике шкалирования). А максимальная этническая гетерогенность, далеко оставляющая за собой гетерогенность иных народов Европы, оказывается свойственной башкирам. В разных местах этой книги по самым разным данным мы наталкиваемся на этот результат, который не является неожиданностью ни для антропологов (поскольку башкиры весьма гетерогенны в расовом отношении), ни для этнографов (поскольку башкиры состоят из многих родов, контрастных по своему

происхождению). Остальные народы, добавившиеся в анализ 2015 года (белорусы, бельгийцы, поляки-кашубы, литовцы, сербы, чехи) характеризуются сниженной гетерогенностью (средние расстояния 0,02-0,05) и лишь албанцы более гетерогенны (0,08).

В среднем, по результатам 2007 года, гетерогенность типичного европейского народа (его внутриэтническая дифференциация) составляет (табл. 2.6) около 0.03: $G_{ST}=0.029$, $F_{ST}=0.027$. Различия уже не внутри одного народа, а между народами (межэтническая дифференциация) оказываются намного выше: $G_{ST}=0.149$, $F_{ST}=0.166$ (табл. 2.6). Пятикратное превышение внутриэтнической изменчивости над межэтнической является количественным выражением того, что подразделение европейских популяций по Y-хромосоме в соответствии с их этнической принадлежностью высоко информативно и объясняет львиную долю всей межпопуляционной изменчивости в генофонде Европы.

Аналогичный анализ этнической гетерогенности проведен нами для народов Южной Сибири (табл. 2.6). Суммарные межпопуляционные различия в этом регионе (0,20) оказались практически такими же, как и для популяций Европы (0,19). Но основная часть изменчивости приходится на изменчивость внутри этнотерриториальных объединений

(0,15), а на различия между ними остается только 0,06. Такой результат объясняется тем, что этнотерриториальные объединения, выделяемые традиционно (хакасы, алтайцы), включают ряд малых народностей, генофонды которых значительно различаются между собой [Балаганская и др., 2011].

ДИФФЕРЕНЦИАЦИЯ ВНУТРИ ЛИНГВИСТИЧЕСКИХ ГРУПП

Рассмотрим теперь закономерности межпопуляционной изменчивости на более высоком иерархическом уровне – не отдельных народов, а их групп. Народы можно классифицировать по-разному, но основной является лингвистическая классификация их языков. Конечно, для оценки роли не только лингвистического, но и географического фактора было бы хорошо классифицировать народы и по регионам их проживания, и затем сравнить генетическую дифференциацию при лингвистической и географической классификациях. Такой анализ мы провели для Кавказа (глава 7), где географическая классификация достаточно очевидна. Не то для Европы – в ее пределах регионы можно выделять по-разному, и результаты слишком зависели бы от моего субъективного взгляда. Лингвистическая же классификация достаточно устойчива, почти общепризнана и

Таблица 2.6. Закономерности изменчивости Y-хромосомы на внутриэтническом и межэтническом уровнях популяционной иерархии (данные 2007 и 2010 года)

	N_{POP}	N	N_{HG}	Средние генетические расстояния	G_{ST} ($G_{ST} \cdot 10^2$)	AMOVA ($F_{ST} \cdot 10^2$)
ЕВРОПА						
Финны	5	107	6	0.13	8.2	8.1
Хорваты	5	100	8	0.18	8.0	7.4
Русские	10	123	8	0.14	5.2	4.5
Итальянцы	5	105	8	0.13	3.9	5.0
Шведы	4	76	12	0.13	2.7	1.4
Немцы	11	110	9	0.07	2.1	1.5
Греки	4	91	8	0.10	1.4	1.1
Турки	5	87	26(8)	0.17 (0.04)	1.3(0.8)	1.0
Поляки	8	114	7	0.02	1.1	0.3
Украинцы	4	102	10	0.02	0.9	0.4
Межэтнический уровень	10		6		14.9	16.6
Внутриэтнический уровень	6		6		2.9	2.7
Внутрипопуляционный уровень	61		6		82.2	80.7
СИБИРЬ						
Хакасы	4	62	30	0.85	18.2	30.3
Северные алтайцы	3	67	30	0.80	15.1	20.0
Шорцы	2	70	30	0.09	4.9	4.7
Южные алтайцы	2	104	30	0.07	3.9	3.2
Киргизы	2	123	30	0.15	2.9	5.1
Межэтнический уровень	5		30		-	5.5
Внутриэтнический уровень	3		30		-	14.7
Внутрипопуляционный уровень	13		30		-	79.8

Примечания: N_{POP} – число региональных популяций, изученных для данного народа; N – средний размер выборки региональных популяций; N_{HG} – число основных гаплогрупп, включенных в анализ; этносы расположены в пределах региона по убыванию значений G_{ST} .

Таблица 2.7. Закономерности изменчивости Y-хромосомы на внутриэтническом уровне (данные 2015 года)

Этнос	Средние генетические расстояния	N_{POP}	N	N_{HG}
Башкиры	0,425	5	86	15
Греки	0,197	5	195	15
Финны	0,155	7	82	15
Хорваты	0,146	3	89	15
Болгары	0,136	9	82	15
Русские	0,129	20	104	15
<i>Русские северные</i>	<i>0,065</i>	<i>4</i>	<i>95</i>	<i>15</i>
<i>Русские центральные и южные</i>	<i>0,047</i>	<i>16</i>	<i>106</i>	<i>15</i>
Шведы	0,110	7	53	15
Албанцы	0,082	3	110	15
Турки	0,064	6	69	15
Литовцы	0,046	6	50	15
Белорусы	0,040	8	96	15
Чехи	0,037	5	51	15
Украинцы	0,032	13	92	15
Кашубы	0,030	3	68	15
Сербы	0,020	4	100	15
Бельгийцы	0,017	5	95	15
Поляки	0,016	3	234	15

Примечания: Для получения более устойчивых оценок в анализ включены народы из массива «Субэтнический», представленные как минимум тремя популяциями, соответственно, не включены народы, представленные только двумя – боснийцы, коми, македонцы, мордва (мокша и эрзя). N_{POP} – число региональных популяций, изученных для данного народа; N – средний размер выборки региональных популяций; N_{HG} – число гаплогрупп, включенных в анализ;

уж в любом случае от меня не зависит. Поэтому в таблицах 2.8 и 2.9 приведена генетическая гетерогенность основных лингвистических групп Европы, рассчитанная как межпопуляционная изменчивость между всеми народами, относящимися к этой группе. (Отметим, что этот анализ, таким образом, основывался на среднеэтнических, а не популяционных частотах: частоты гаплогрупп во всех популяциях одного народа были усреднены, такая процедура проведена для всех народов, и сравнивались полученные среднеэтнические частоты).

При анализе по панели 17 гаплогрупп (табл. 2.8) наиболее гетерогенными оказываются славянская, тюркская и финно-угорская группы (средние генетические расстояния между народами в пределах каждой из этих групп составляют около 0,5). Этот результат по финно-уграм и тюркам хорошо перекликается с аналогичным анализом по мтДНК, о котором речь пойдет в следующей главе 3, а причины гетерогенности славян разбираются в главе 6. Романская группа уже заметно менее гетерогенна (0,3), германская еще менее (0,1), а балтская, абхазо-адыгская и нахско-дагестанская почти гомогенны (0,01–0,05). Надо, впрочем, оговориться, что эти выводы зависят от набора гаплогрупп – например, доминирующие на Северном Кавказе гаплогруппы в этом наборе гаплогрупп (оптими-

зированном под остальную Европу) представлены только гаплогруппами **G** и **J** безо всякого их подразделения, тогда как с учетом их подразделения на субтипы гомогенность обеих кавказских лингвистических групп значительно бы выросла.

В наборе 40 гаплогрупп кавказские варианты опять не подразделены (это повлекло бы за собой сокращение числа анализируемых популяций по остальной Европе, не изученных по этим субвариантам), поэтому дифференциация абхазо-адыгской и нахско-дагестанской групп остается минимальной. По балтской группе по этой панели данных на момент написания книги недостаточно, гетерогенность германской группы несколько возрастает (0,2), славянской группы даже немного снижается (0,4), зато гетерогенность тюркской группы увеличивается в полтора раза (до 0,7), финно-угорской – в два (до 1,0), а романской – в три (до 0,9). Хотя очевидна большая зависимость результатов как от числа гаплогрупп, так и от степени их подразделения, все же отметим, что при наилучшем возможном на данный момент наборе данных наиболее гетерогенной является не германская группа (хотя лучше всего подразделена доминирующая у нее гаплогруппа **R1b**), а финно-угорская группа – по всей видимости, за счет преобладающих у нее процессов дрейфа генов вследствие изоляции.

Таблица 2.8. Гетерогенность лингвистических групп Европы по панели 17 гаплогрупп

Лингвистическая группа	Число популяций	Средние генетические расстояния	Расстояния после коррекции*	Исключенные обособленные популяции
Абхазо-адыгская	4	0,048	-	-
Балтская	2	0,009	-	-
Германская	3	0,095	-	-
Нахско-дагестанская	5	0,009	-	-
Славянская	15	0,505	0,381	Северные русские и сорбы: ген. расстояния в 1,7 и 1,9 раза больше средней
Романская	8	0,312	0,214	Румыны: ген. расстояния в 2,5 раза больше средней
Тюркская	10	0,471	-	-
Финно-угорская	8	0,470	-	-

*В некоторых случаях одна-две генетически обособленные популяции за счет больших генетических расстояний до остальных популяций завышали уровень гетерогенности всей группы. Поэтому была проведена коррекция: повторный расчет при исключении из расчета тех популяций, среднее расстояние до которых значительно превышало аналогичный показатель для остальных популяций данной группы.

Таблица 2.9. Гетерогенность лингвистических групп Европы по панели 40 гаплогрупп

Лингвистическая группа	Число популяций	Средние генетические расстояния	Расстояния после коррекции	Исключенные обособленные популяции
Абхазо-адыгская	3	0,058	-	-
Германская	3	0,230	-	-
Нахско-дагестанская	4	0,012	-	-
Славянская	10	0,444	0,374	Чехи: ген. расстояния в 1,6 раза больше средней
Романская	4	0,902	0,586	Румыны: ген. расстояния в 1,4 раза больше средней
Тюркская	7	0,751	0,594	Казанские татары: ген. расстояния в 1,4 раза больше средней
Финно-угорская	4	0,968	-	-

РОЛЬ ЛИНГВИСТИЧЕСКОГО И ГЕОГРАФИЧЕСКОГО ФАКТОРОВ

Напомню, что первое обобщающее исследование Y-хромосомного генофонда Европы [Rosser et al., 2000] называлось «Изменчивость Y-хромосомы в Европе является клинальной и задается, прежде всего, географическим, а не лингвистическим фактором». Этот вывод был основан на полученной Россером с соавторами достоверной корреляции матрицы генетических расстояний между популяциями Европы с матрицей географических расстояний между теми же популяциями (коэффициент корреляции 0.4), при недостоверной корреляции

между генетическими и лингвистическими расстояниями [Rosser et al., 2000].

Поэтому наш коллектив тоже провел анализ корреляции между матрицами генетических, географических и лингвистических расстояний, включив в него и многочисленные новые данные, опубликованные после выхода в свет упомянутой статьи [Rosser et al., 2000]. Анализ проводился по точным лексико-статистическим данным и только в пределах индоевропейских народов (составляющих большинство в Европе), поскольку расстояния между разными лингвистическими семьями не поддаются точному измерению.

Таблица 2.10. Роль лингвистического и географического факторов в структурировании Y-хромосомного генофонда Европы (первый вариант – 17 гаплогрупп, 26 народов)

Вид корреляции	Параметры корреляции	Коэффициент корреляции
Парная корреляция	Генетика и лингвистика	0,45
Парная корреляция	Генетика и география	0,56
Частная корреляция	Генетика и лингвистика (география остается постоянной)	0,30
Частная корреляция	Генетика и география (лингвистика остается постоянной)	0,46

Таблица 2.11. Роль лингвистического и географического факторов в структурировании Y-хромосомного генофонда Европы (второй вариант – 40 гаплогрупп, 18 народов)

Вид корреляции	Параметры корреляции	Коэффициент корреляции
Парная корреляция	Генетика и лингвистика	0,54
Парная корреляция	Генетика и география	0,74
Частная корреляция	Генетика и лингвистика (география остается постоянной)	0,42
Частная корреляция	Генетика и география (лингвистика остается постоянной)	0,69

При использовании массива данных по 17 гаплогруппам (табл. 2.10) корреляция генетических расстояний с географическими (0,56) лишь немного больше корреляции с лингвистикой (0,45). При использовании частных корреляций величины уменьшаются, но разрыв становится рельефнее: 0,46 с географией и только 0,30 с лингвистикой.

При использовании массива данных по 40 гаплогруппам (подразделяются **R1a**, **R1b**, **N1c** – табл.2.4) приходится исключить из анализа некоторые народы, не изученные по этой панели. Но это полностью оправдывается тем, что резко возрастают коэффициенты корреляции и с географией, и с лингвистикой. То есть подразделение гаплогрупп не только лучше выявляет их географическую приуроченность на картах, но и использование этих новых субгаплогрупп позволяет лучше увидеть характер структурированности европейского генофонда, его связь и с геогра-

фическим, и с лингвистическим факторами. Парная корреляция генетических расстояний (табл. 2.11) с географическими составляет 0,74, а с лингвистическими 0,54. При анализе частных корреляций связь с лингвистикой хотя и остается существенной (0,42), но значительно уступает связи с географией: 0,69. Отметим, что полученный нами коэффициент частной корреляции между матрицами генетических и географических расстояний превышает 0,7, что в полтора с лишним раза больше величины 0,4, полученной ранее Россером с соавторами. Это убедительно доказывает, что по мере накопления данных по изменчивости Y-хромосомы географическая структурированность генофонда Европы становится все более и более явной. В целом же, оба фактора – и география, и лингвистика – играют каждый свою роль, совместно формируя генетический ландшафт Европы.

ГЛАВА 3.

СТРУКТУРА ГЕНОФОНДА ЕВРОПЫ В ЗЕРКАЛЕ МИТОХОНДРИАЛЬНОЙ ДНК

Уже около двадцати лет митохондриальная ДНК является одной из наиболее востребованных генетических систем в популяционно-генетических исследованиях. Вместе с Y-хромосомой мтДНК образовала своеобразный дуэт, и эти две нерекombинирующие генетические системы с однородительским типом наследования стали наиболее популярными маркерами в исследовании популяций человека. Даже сейчас, после внедрения широкогеномных маркеров и полногеномного секвенирования, число выходящих статей по изменчивости мтДНК и Y-хромосомы в несколько раз превышает число полногеномных исследований, хотя и уступает им по импакт-фактору журналов. В целом можно уверенно констатировать, что появление широкогеномных маркеров добавило третью систему, но не заменило первые две.

Поскольку митохондриальная ДНК и Y-хромосома настолько параллельны друг другу – и по методам анализа, и по объему накопленных данных – в книге анализ этих двух систем проведен по возможности аналогично, чтобы было легче сопоставить результаты: и общие закономерности, выявляемые обеими системами, и частные факты, обнаруженные только по одной из них. Поэтому план изложения исследований митохондриального генофонда Европы такой же, как и для Y-хромосомного генофонда. Сначала мы кратко рассмотрим итоги предшествующих исследований. Затем проанализируем генетические взаимоотношения популяций, причем, как и для Y-хромосомы, рассмотрим закономерности в трех приближениях. (Правда, поскольку межпопуляционная вариабельность мтДНК в Европе куда меньше, то ее изменчивость на внутризтническом уровне, где она совсем мала, не рассматривается). Потом изучим роль географического и лингвистического факторов в структурировании генофонда, и рассчитаем генетическую гетерогенность основных лингвистических групп населения Европы.

Однотипность плана анализа мтДНК и Y-хромосомы придется нарушить лишь в двух частных пунктах: для Y-хромосомы мы не стали анализировать гаплотипическое разнообразие, а для мтДНК не будем рассматривать карты отдельных гаплогрупп. Это вызвано тем, что внутривидовое гаплотипическое разнообразие Y-хромосомы, оцениваемое по наиболее популярной панели из 17 STR маркеров, практически во всех популяциях Европы близко к максимуму, и потому различия между популяциями невелики. Что же касается карт отдельных гаплогрупп мтДНК, то для них географические закономерности (например, более высокая частота гаплогруппы **H** на севере Европы) известны, но маловыразительны. Причина – уже упоминавшаяся гомогенность Европы по частотам гаплогрупп мтДНК. Поэтому геногеография гаплогрупп мтДНК в Европе специально не рассматривается (в отличие от анализа Y-хромосомы, четким геногеографическим паттернам отдельных гаплогрупп которой посвящен раздел 2.2).

Отметим, что в этой главе анализируется панель данных по гипервариабельному сегменту мтДНК и SNP маркерам кодирующей части, а данные полного секвенирования мтДНК не рассматриваются. Дело в том, что геногеография оперирует выборками, а не индивидуальными образцами. И если полное секвенирование нескольких десятков образцов какой-либо гаплогруппы выявляет новые субгаплогруппы, эта информация касается лишь молекулярной, а не популяционной составляющей анализа мтДНК: мы знаем, что эти субгаплогруппы существуют, но мы не знаем, с какой частотой они встречаются в каких популяциях. Поэтому и по Y-хромосоме (глава 2), и по мтДНК (данная глава) анализируются лишь те гаплогруппы, по которым уже накопились данные об их частотах в разных популяциях, а новейшие данные о гаплогруппах, выявленных при полногеномном секвенировании, рассматриваются отдельно, в одном из разделов главы 8.

3.1. ПРЕДШЕСТВОВАВШИЕ ИССЛЕДОВАНИЯ ИЗМЕНЧИВОСТИ мтДНК В ЕВРОПЕ

Авторитет митохондриальной ДНК (мтДНК) в популяционной генетике человека утвердился после яркой иллюстрации африканских корней митохондриальных генофондов населения всех континентов. Это явилось чрезвычайно убедительным аргументом в пользу моноцентристской концепции происхождения человека современного вида и африканской прародины. Меньший географический масштаб, но ничуть не меньший резонанс, имело и крупнейшее исследование митохондриального генофонда Европы [Richards et al., 2000]. В популяционной генетике человека Европа всегда являлась наиболее изученным континентом, поскольку многочисленные европейские лаборатории концентрируются в первую очередь на изучении населения своих собственных стран, и лишь во вторую – населения других частей света.

НЕОЛИТ ИЛИ ПАЛЕОЛИТ?

Митохондриальная ДНК не стала исключением, и, начиная с 90^х годов XX века, стали стремительно накапливаться данные о полиморфизме мтДНК европейских популяций. Эти исследования привели к результатам, которые можно рассматривать как смену «неолитической» парадигмы в трактовке формирования генофонда Европы на «палеолитическую». В 1970^х-90^х годах преобладала гипотеза формирования европейского генофонда в результате миграции популяций неолитических земледельцев из Передней Азии в Европу [Ammerman, Cavalli-Sforza, 1984]. Но на рубеже XXI века большинство специалистов склонилось на сторону гипотезы неолитизации путем заимствования неолитической культуры без миграции самих земледельцев, что предполагает значительно более древнее (палеолитическое) время формирования европейского генофонда [Richards et al., 1996, 2000]. Тем самым была почти отвергнута знаменитая теория «демической диффузии» [Ammerman, Cavalli-Sforza, 1984; Cavalli-Sforza et al., 1994], возводящая (по данным об аутосомных классических маркерах) формирование основных черт генофонда Европы ко времени расселения неолитических земледельцев, пришедших с Ближнего Востока и ассимилировавших аборигенные палеолитические популяции охотников-собирателей. Взамен была реабилитирована теория «культурной диффузии», рассматривающая неолитизацию Европы как распространение земледелия путем культурных заимствований без миграции самого населения. Правда, в последующие годы однозначность такого вывода несколько поколебалась, о чем мы будем говорить в главе 8.

Ключевым аргументом в работе [Richards et al., 2000] состоял в выявлении характерных для Европы гаплогрупп и кластеров гаплотипов мтДНК и расчете их возраста: древность большинства из них далеко превышала датировки неолита и даже мезолита, уводя в верхний палеолит. Итак, благодаря исследованию [Richards et al., 2000] не только (временно) возобладала теория палеолитического времени формирования генофонда Европы с лишь незначительным вливанием ближневосточного генетического компонента в неолите, но и изучение митохондриальной ДНК приобрело исключительный авторитет в решении проблем истории формирования генофондов.

ГОМОГЕННОСТЬ МИТОХОНДРИАЛЬНОГО ГЕНОФОНДА ЕВРОПЫ

Однако основным успехом исследования Richards с соавторами [Richards et al., 2000] заключался в сравнении европейского митохондриального генофонда с генофондом внешнего по отношению к Европе региона Ближнего Востока, а вопрос о внутренней структуре европейского генофонда не ставился. Во многом это было связано с устоявшимся к тому времени взглядом на европейский митохондриальный генофонд как на исключительно гомогенный, однородный по всему европейскому ареалу: исследования показали, что в самых разных частях Европы распространен один и тот же набор гаплогрупп, причем с очень близкими частотами [Simoni et al., 2000].

Последующее накопление данных лишь отчасти внесло коррективы в такую трактовку: было показано, что при увеличении суммарной выборки (по всем изученным европейским популяциям) до 3113 человек и при объединении данных по крупным регионам Европы удастся выявить некоторую степень структурированности по географическому принципу: популяции Средиземноморья, Центральной и Северной Европы входят в разные кластеры на графике генетических взаимоотношений населения этих регионов [Richards et al., 2002]. И все же вывод о гомогенности митохондриального генофонда Европы и очень слабых различиях региональных популяций остался в силе и до сих пор часто цитируется.

РЕГИОНАЛЬНЫЕ ИССЛЕДОВАНИЯ

Возможно, именно сложившейся уверенностью в бесперспективности поиска внутренней структуры в европейском митохондриальном генофонде (а также интенсивной обобщающей работой в предшествовавший период, 1995–2002 гг.) объяс-

няется отсутствие новых попыток такого обобщающего анализа за десять с лишним лет, прошедших после работы [Richards et al., 2002].

Вместо этого большинство исследователей сосредоточились на получении данных по отдельным странам и популяциям Европы, и было опубликовано более ста подобных работ. Перечислим лишь важнейшие из них, опубликованные в период 1995–2011 годов и опирающиеся на выборки более, чем 150 образцов мтДНК [Sajantila et al., 1995; Dupuy et al., 1996; Richards et al., 1996, 2000; Opdal et al., 1998; Pfeiffer et al., 1999; Dimo-Simonin et al., 2000; Helgason et al., 2001; Larruga et al., 2001; Meinila et al., 2001; Nasidze et al., 2001; Bermisheva et al., 2002; Malyarchuk et al., 2002, 2004, 2006, 2010; Gonzalez et al., 2003; Maca-Meyer et al., 2003; Poetsch et al., 2003; Dubut et al., 2004; McEvoy et al., 2004; Pereira et al., 2004; Babalini et al., 2005; Goodacre et al., 2005; Alvarez et al., 2006; Bosch et al., 2006; Falchi et al., 2006; Picornell et al., 2006; Tonks et al., 2006; Achilli et al., 2007; Grzybowski et al., 2007; Brandstaetter et al., 2007, 2008; Hedman et al., 2007; Irwin et al., 2007, 2008; Richard et al., 2007; Tetzlaff et al., 2007; Zimmermann et al., 2007; Lappalainen et al., 2008; Lehocky et al., 2008; Martinez et al., 2008; Turchi et al., 2008; Alvarez-Iglesias et al., 2009; Santos et al., 2003; Alvarez et al., 2010; Garcia et al., 2010; Mikkelsen et al., 2010; Pereira et al., 2010; Tillmar et al., 2010; Morozova et al., 2011; Karachanak et al., 2011].

ИССЛЕДОВАНИЯ В РОССИИ

Реальность бума митохондриальных работ иллюстрирует тот факт, что только в России эти исследования были развернуты целым рядом коллективов: Института биологических проблем Севера в Магадане [Derenko et al., 2003, 2007; Malyarchuk et al., 2001, 2002, 2003, 2004, 2006, 2008, 2010, 2012], Института цитологии и генетики в Новосибирске [Гольцова и др., 2005; Derbeneva et al., 2002a,b; Сукерник и др., 2010; Starikovskaya et al., 1998, 2005; Rubinstein et al., 2008; Volodko et al., 2008], Уфимского научного центра в Уфе [Бермишева и др., 2002; 2004], Якутского научного центра в Якутске [Fedorova et al., 2003, 2013], Южного федерального университета в Ростове-на-Дону [Kornienko et al., 2004], Института общей генетики РАН [Orekhov et al., 1999; Орехов, 2002; Наумова и др., 2008; Bulayeva et al., 2003; Morozova et al., 2011], Института молекулярной генетики РАН [Belyaeva et al., 2003], Медико-генетического научного центра РАМН в Москве [Балановский и др., 2011; Pshenichnov et al., 2013] и другими научными коллективами страны. Среди них особого упоминания заслуживают работы Б.А. Мальярчука, который из всех российских исследователей опубликовал наибольшее

число работ по изменчивости мтДНК в Европе [Malyarchuk et al., 2002, 2003, 2004, 2006, 2008, 2010, 2012], во многих из которых проводится изучение различных аспектов митохондриального генофонда как популяций Сибири, так и восточных, западных и южных славян. Огромную роль в проведении исследований российских коллективов сыграл Эстонский биоцентр и его директор Р. Виллемс. В годы, когда секвенирование в России было технически и/или финансово недоступно, молодые специалисты (включая и автора этой книги) из практически всех перечисленных институтов России, так же, как и специалисты из многих других стран Европы, толпами съезжались в Эстонский биоцентр со своими образцами, проводя там и технический этап секвенирования, и обучаясь методам анализа митохондриальных данных.

ЦЕЛЬ – АНАЛИЗ В МАСШТАБЕ ВСЕЙ ЕВРОПЫ

Но, несмотря на такое обилие сильных региональных работ, обобщающие исследования в масштабе всей Европы за последнее десятилетие мне неизвестны. Свою роль в этом сыграли и объективные трудности составления сколько-нибудь полной сводки о полиморфизме мтДНК в Европе в условиях растущей лавины информации по этой тематике, публикуемой во множестве генетических, судебно-медицинских и даже археологических изданий. В связи с этим создававшиеся общедоступные базы данных по мтДНК, аккумулирующие эту информацию, либо переключились более на молекулярную, чем на популяционную сторону изменчивости мтДНК (например, популярные интернет-ресурсы www.mitomap.org и www.phylotree.org), либо рост базы данных прекратился или замедлился после нескольких обновлений (база данных HVRbase, достигшая объема 13 873 образцов мтДНК, но затем угасшая, EMPOP с текущим объемом 34 717 образцов). Из целого ряда предпринятых попыток создания баз данных о популяционном полиморфизме мтДНК, насколько нам известно, лишь база данных, развиваемая нашим коллективом («MURKA»), выдержала целый ряд последовательных обновлений и к настоящему времени содержит сведения о 132 600 образцов мтДНК в 2 101 популяциях мира (табл. 1.3). Применительно к генофонду Европы и соседних территорий в базе данных содержатся 31 935 образцов мтДНК, что более чем в 10 раз превышает объем данных, использованный в последнем обобщающем исследовании полиморфизма мтДНК в Европе [Richards et al., 2002].

Именно последовательная работа по созданию максимально полной базы данных по изменчивости мтДНК («MURKA», раздел 1.2.) является первой предпосылкой для новой попытки обобщающего анализа европейского митохондриального генофонда, предпринятой в данной монографии.

Второй предпосылкой стало наше исследование полиморфизма Y-хромосомы в Европе, тоже основанное на обширной базе данных и включившее картографический и статистический анализ этих данных (это исследование было начато в работе [Balanovsky et al., 2008] и завершено в главе 2 данной книги). Естественным продолжением этой ра-

боты стало проведение аналогичного анализа данных по митохондриальной ДНК: было решено предпринять попытку выявить основную структуру европейского генофонда по мтДНК и сравнить основные черты генофонда Европы, реконструируемые благодаря дуэту обеих информативных «однородительских» систем.

3.2. ГЕНЕТИЧЕСКИЕ ВЗАИМООТНОШЕНИЯ ПОПУЛЯЦИЙ: ТРИ МАСШТАБА

ФОРМИРОВАНИЕ МАССИВА ДАННЫХ

Для выявления генетических взаимоотношений популяций Европы и смежных регионов использован ставший традиционным метод многомерного шкалирования. Однако особенностью проведенного анализа является большое число тщательно отобранных популяций.

Первым критерием отбора был размер выборки. В базе данных MURKA по Европе имелись данные по 336 популяциям, но многие из них были изучены по небольшой выборке. Для обеспечения репрезентативности мы установили минимальный объем выборки $N=70$. При выборках меньшего объема по возможности объединялись соседние популяции одного народа так, чтобы в сумме составить выборку более 70 образцов. Если такой возможности не было, малые выборки исключались из анализа.

Вторым критерием отбора была надежность определения гаплогрупп. Мы стремились использовать только те выборки из нашей базы данных, в которых принадлежность образцов к данной гаплогруппе была подтверждена анализом SNP маркеров кодирующей части мтДНК (как правило, методом ПДРФ-анализа). Если для данного этноса или региона такие выборки отсутствовали, в анализ включались выборки, где ПДРФ данных не было, но гаплогруппы можно было надежно (с высокой вероятностью) определить по ГВС1. Пороговым уровнем было наличие не более 5% образцов, для которых надежное определение гаплогруппы оказывалось невозможно. Выборки, в которых доля таких образцов превышала 5%, не включались в анализ.

Третьим критерием было наличие четкой географической привязки в описании выборки. Поэтому, если для выборки было известно ее происхождение лишь с точностью до страны, но по этой стране имелись и выборки с региональной привязкой внутри страны, первую выборку приходилось исключать.

В итоге были сформированы 118 выборок, характеризующих все части Европы и смежных регионов, причем каждая представлена не менее чем 70 образцами. Средний объем выборок составляет $N=200$, суммарно 23 500 образцов. Таким образом, даже после того как две трети изученных к насто-

ящему времени европейских популяций были отсеяны или объединены, суммарный объем анализируемых образцов оказался все-таки на порядок больше, чем в предыдущих обобщающих исследованиях европейского митохондриального генофонда [Richards et al., 2000, 2002]. Но еще более важно для получения надежных результатов то, что мы опираемся на большой объем выборки из популяции ($N=200$). Карта на рис. 3.1 показывает географическое положение популяций Европы, изученных по полиморфизму мтДНК.

Следующей задачей было формирование набора гаплогрупп. Из двух критериев полиморфизма, принятых в популяционной генетике (однопроцентный и пятипроцентный) мы применили «мягкий». Поэтому рассматривали только гаплогруппы, средняя частота которых в европейских популяциях составляет хотя бы один процент. Для этого из большого числа гаплогрупп часть была объединена (с учетом их филогенетического родства). Например, были объединены **V** и **preV**, также объединялись разные варианты **U2** и т.д. Часть редких гаплогрупп, отдаленных от других гаплогрупп филогенетически или с отдаленным от Европы ареалом их преимущественного распространения, была объединена в сборную группу «другие» (**other**), причем средняя частота этой сборной парагруппы составила только 2%. В итоге в анализ были включены следующие 35 гаплогрупп: **A, B, C, D, F, G, Z, L, M1, preHV, H, HV, J, T*, T1, T2, I, N1a, N1b, R, K, V, W, X, U1a, U1b, U2, U3, U4, U5a, U5b, U6, U7, U8, other**.

В результате была сформирована таблица из 118 строк-популяций и 35 столбцов-гаплогрупп, которая и послужила основой для всех видов анализа, проведенных на уровне частот гаплогрупп мтДНК. Данные на уровне отдельных гаплотипов используются только для анализа внутривнутрипопуляционного гаплотипического разнообразия.

Для рассмотрения основных закономерностей европейского генофонда анализ был проведен на этническом уровне. Для этого данные по популяциям одного народа были усреднены и анализировались частоты для получившихся среднеэтнических популяций. Число анализируемых популяций при этом сократилось до 62 народов. Они включили следующие этнические популяции Европы и прилегающих

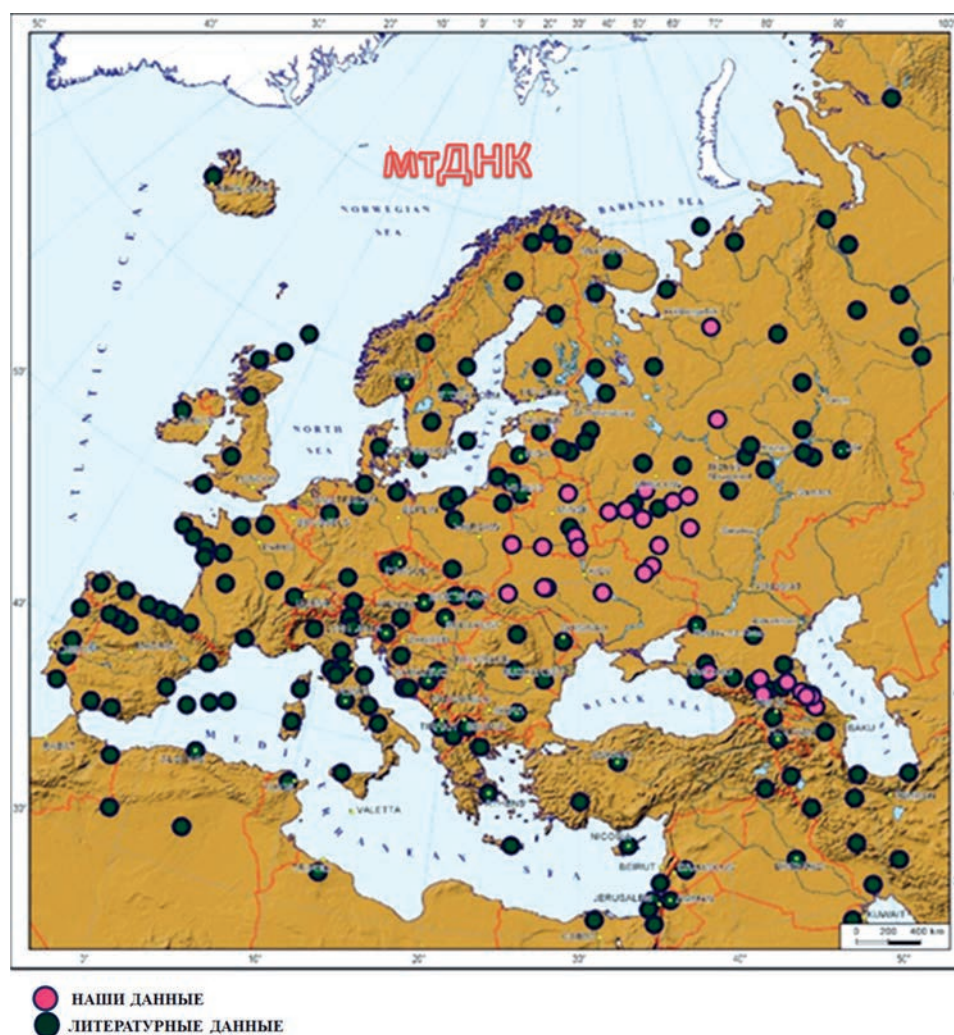


Рис. 3.1. Популяции Европы и смежных регионов, изученные по мтДНК.

регионов Северной Африки, Ближнего Востока и Западной Сибири (приводятся в соответствии с англоязычными обозначениями базы данных MURKA): Lithuanians, Estonians, Finns, Karelians, Komi, Mari, Mordvinians, Saami, Udmurts, Czech, Poles, Byelorussians, Russians, Slovaks, Ukrainians, Bashkirs, Chuvash, Tatars, Albanians, Greeks, Aromuns, Romanians, Slovenians, Bosnians, Bulgarians, Croatians, Hungarians, Basques, Irish, Scottish, Austrians, English, Germans, Icelanders, Norway, Swedes, Swiss, French, Italians, Portugals, Sardinians, Sicilians, Spaniards, Turkey, Armenians, Georgians, Kabardinians, Ossets, Azeris, Nogays, Kurds, Iranians, Syrians, Iraq, Jordanians, Palestinians, Saudi Arabia, Morocco, Kazakhs, Mansi, Nenets, Nganasans.

АНАЛИЗ ПОПУЛЯЦИЙ ЕВРОПЫ И БЛИЖНЕГО ВОСТОКА

Результаты анализа митохондриальных генофондов этих 62 этнических популяций представлены на рис. 3.2. На рисунке 3.2А показаны генетические взаимоотношения этих популяций (их

положение в генетическом пространстве), а на рисунке 3.2Б – локализация тех же популяций на карте (их положение в географическом пространстве).

В генетическом пространстве четко выделяются четыре кластера. Первый кластер включил только саамов, что не удивительно, учитывая их генетическое своеобразие [Cavalli-Sforza et al., 1994], в том числе показанное и по мтДНК [Tambets et al., 2004]. Вторым кластером включили популяции восточных рубежей Европы, у которых повышена частота восточноевразийских гаплогрупп. Третий кластер включил популяции Передней Азии и Кавказа. Наконец, все остальные популяции с основной территории Европы (от Волги до Пиренейского полуострова) помещаются в четвертый «пан-европейский» кластер. Его небольшие размеры на генетическом графике (при огромных размерах на географической карте) свидетельствуют о низкой межпопуляционной изменчивости внутри этого кластера. Эти результаты подтверждают генетическую гомогенность Европы (кроме Приуралья) по мтДНК и указывают на своеобразие генофондов Приуралья и Ближнего Востока.

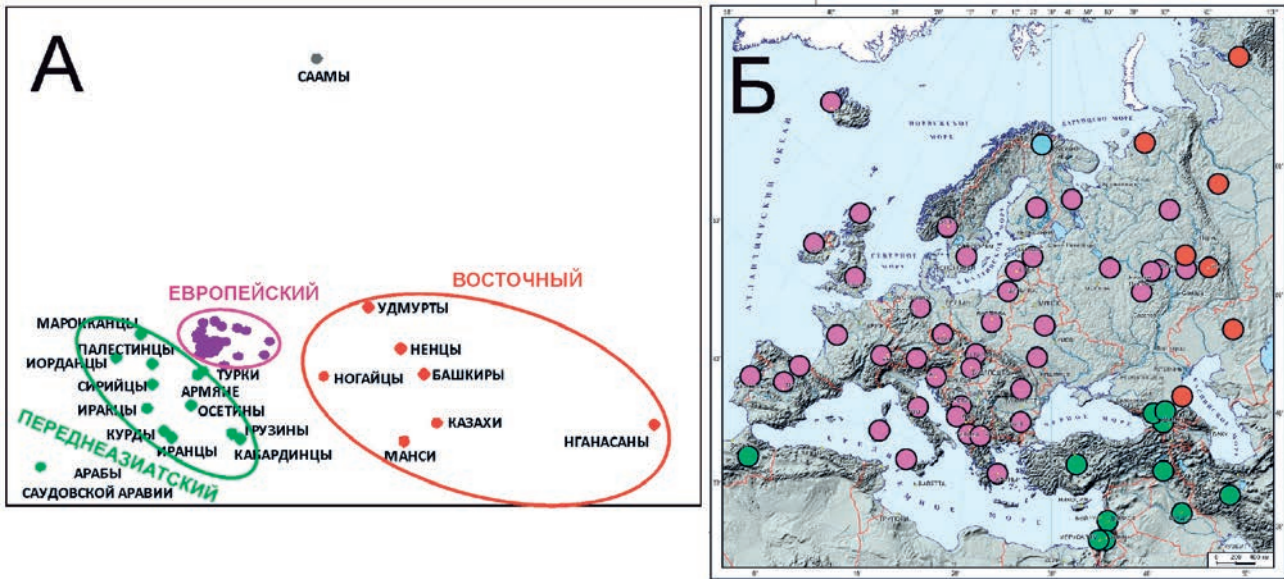


Рис. 3.2. Генетические взаимоотношения европейских и ближневосточных популяций по данным об изменчивости мтДНК.

А. График многомерного шкалирования (стресс=0,085). Популяции разных кластеров показаны разным цветом.
 Б. Географическое положение изученных популяций. Цвет каждой популяции такой же, как на графике А.

Результаты, представленные на рис. 3.2А, таким образом, подтверждают на новом витке исследований и резко возросшем объеме информации два явления: гомогенности митохондриального генофонда Европы и эффективности выделения двух основных вариантов митохондриального генофонда Евразии – западноевразийских и восточноевразийских гаплогрупп, поскольку именно наличие восточноевразийских гаплогрупп обуславливает своеобразие Приуралья.

Но попробуем узнать о Европе новое, – и для этого рассмотрим этот европейский кластер как бы под микроскопом, удалив из анализа внеевропейские популяции и «генетических чужаков» Европы: пограничную полосу красных кружков восточных рубежей Европы и зеленые популяции Ближнего Востока. То есть рассмотрим только одну Европу – только популяции, обозначенные фиолетовым цветом на рис. 3.2А и 3.2Б.

АНАЛИЗ ТОЛЬКО ЕВРОПЕЙСКИХ ПОПУЛЯЦИЙ

При анализе только популяций, вошедших в «европейский» (фиолетовый) кластер на рис. 3.2, получаем новый график и новую карту (рис. 3.3). График снова показывает положение популяций в генетическом пространстве и вновь популяции, попавшие в один кластер, обозначены одним цветом (3.3 А), а карта (3.3 Б) отображает географическое положение тех же популяций с теми же цветовыми обозначениями. Мы обнаруживаем новый генетический кластер со значительным разнообразием популяций внутри его (желтый цвет). Обратим внимание, что эти популяции (желтого цве-

та) занимают большую область только в генетическом пространстве (3.3А). На географической же карте (3.3Б) желтые кружки близки друг к другу, занимая не столь уж большую область на северо-востоке Европы. Таким образом, высокое генетическое разнообразие (большие генетические различия между популяциями, проявляющиеся в большом размере генетического кластера) уместается на географически небольшой территории северо-востока Европы. Рассмотрение того, какие популяции входят в этот кластер, позволяет условно назвать его «финно-угорским». Действительно, в него входят финноязычные народы (коми, мордва, мари, финны, карелы). К этому кластеру приближаются также тюркоязычные чуваша и татары, для которых по антропологическим данным предполагается значительный финно-угорский субстрат. Отметим, что другие тюркские народы Европы (например, башкиры и ногайцы) в этот кластер не входят.

Интерпретация высокого межпопуляционного разнообразия в «финно-угорском» кластере будет дана ниже, при рассмотрении гаплотипического разнообразия мтДНК. Пока же отметим, что, хотя этот кластер популяций географически располагается на востоке Европы вблизи границы с Азией, его генетическое своеобразие нельзя объяснить азиатскими влияниями. Ведь миграции из Азии, отразившиеся в высокой частоте восточноевразийских гаплогрупп, уже сформировали кластер, проявившийся при предыдущем генетическом масштабе (рис. 3.2 А), в котором финноязычные популяции были генетически неотличимы от остальной Европы. Мы видели, что лишь при повыше-

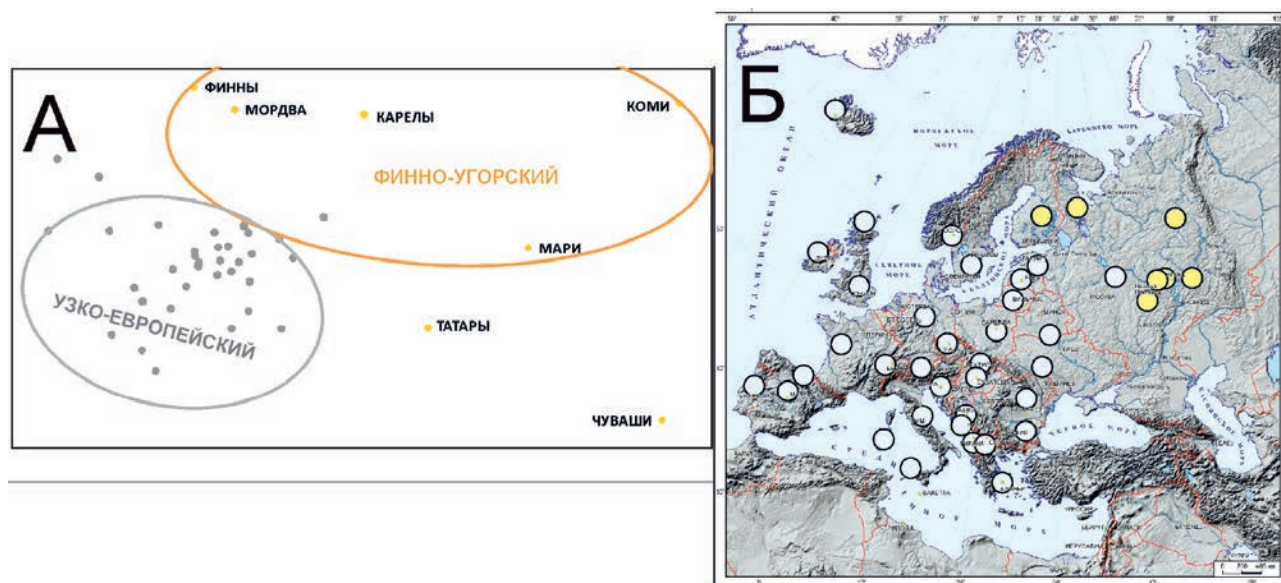


Рис. 3.3. Генетические взаимоотношения популяций, вошедших в «пан-европейский» кластер на рис. 3.2.
 А. График многомерного шкалирования (стресс=0,138). Популяции разных кластеров показаны разным цветом.
 Б. Географическое положение изученных популяций. Цвет каждой популяции такой же, как на графике А.

нии разрешающей способности анализа и переходе от генетического масштаба рис. 3.2 к генетическому масштабу рис. 3.3. (при рассмотрении изменчивости генофонда только самой Европы, без восточноевразийского влияния) проявилось своеобразие финноязычных народов. Таким образом, финноязычные народы (с примкнувшими к ним чувашами и татарами) следует признать по митохондриальному генофонду европейцами, но четкое разделение генофонда Европы на финноязычную и остальную Европу оказывается новым и важным результатом нашего анализа.

Отметим, что все остальные популяции Европы снова оказались генетически почти неразличимы, опять формируя гомогенный кластер, который теперь можно назвать «узко-европейским». Это снова демонстрирует генетическую гомогенность Европы – но на этот раз уже не всего европейского генофонда, а только «не-финноязычной» его части.

АНАЛИЗ ЕВРОПЕЙСКОГО ГЕНОФОНДА БЕЗ ФИННОЯЗЫЧНЫХ ПОПУЛЯЦИЙ

В попытке выявить, присутствует ли все же генетическая структурированность в этом «узко-европейском» кластере, мы вновь повторили тот же прием смены генетического масштаба. Для этого исключили из рассмотрения генетически отличающиеся популяции («финно-угорского» кластера) и проанализировали остальные популяции Европы (только «узко-европейского» кластера).

Результаты этого анализа показаны на рис. 3.4. Только при этом, третьем по счету, «увеличении» нашего популяционного микроскопа, мы, наконец, обнаруживаем различия между этносами внутри

Европы. Хотя изученные популяции занимают почти все пространство графика и четкие кластеры поэтому не образуются, все же по сходству популяций в генетическом пространстве можно выделить несколько групп популяций. Более строго их было бы назвать «зонами», поскольку термин «кластер» предполагает отграниченность от соседних групп, а в нашем случае одна зона тесно соседствует с другой. Но для сохранения единства терминологии мы будем называть выделенные группы генетически сходных популяций по-прежнему «кластерами».

В один кластер (рис. 3.4 А) вошли испанцы, португальцы, французы, сардинцы, итальянцы. Все это народы романской языковой группы индоевропейской лингвистической семьи. В другой кластер вошли латыши и литовцы, то есть народы балтской языковой группы. В третий – соседний – кластер попали русские, словенцы, поляки, чехи, словаки, боснийцы, хорваты, украинцы, болгары. Все это – народы славянской языковой группы. Народы германской языковой группы – немцы, шведы, норвежцы, швейцарцы, австрийцы, англичане, исландцы – тоже образуют свой собственный кластер. Учитывая, что население Шотландии лишь исторически недавно сменило кельтский язык на язык германской группы, их генетическое сходство с ирландцами обрисовывает общность митохондриального генофонда и кельтской группы тоже. Албанцы – единственные представители албанской языковой группы – располагаются примерно в центре графика. Баски – единственный представленный на этом графике народ, который не относится к индоевропейской семье, соответственно занимает обособленное положение.

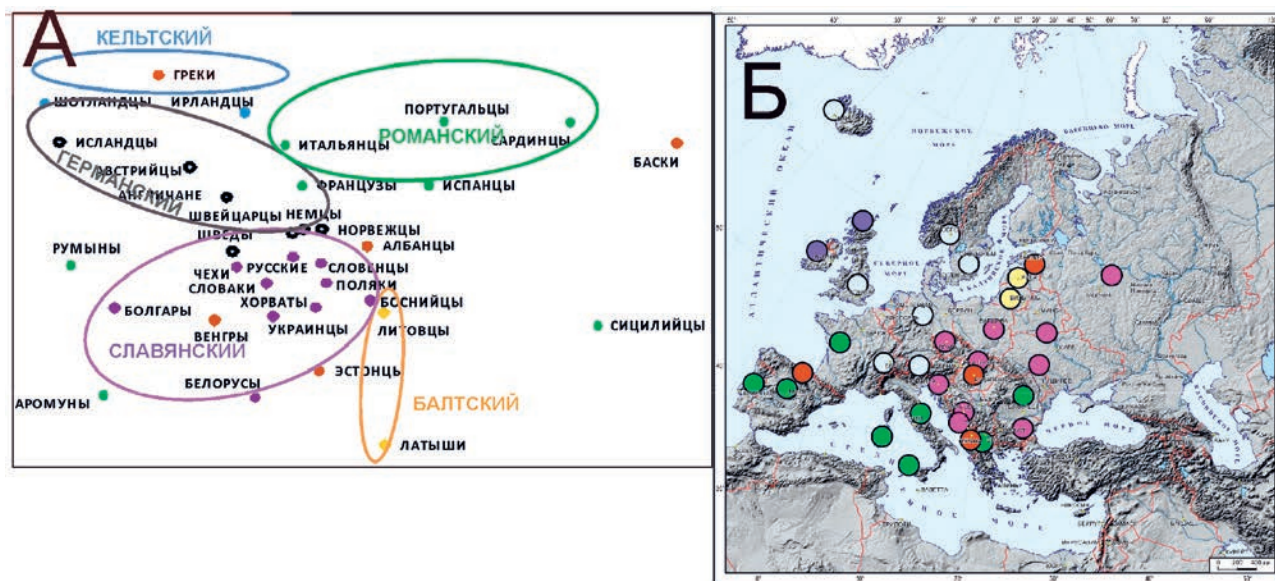


Рис. 3.4. Генетические взаимоотношения популяций, вошедших в «узко-европейский» кластер на рис. 3.3. А. График многомерного шкалирования (стресс=0,19). Популяции разных кластеров показаны разным цветом. Б. Географическое положение изученных популяций. Цвет каждой популяции такой же, как на графике А.

Такой впервые полученный результат позволяет нам сделать два важных вывода.

1) **Европа не гомогенна по мтДНК: в ее пределах есть генетические кластеры.**

2) **Эти кластеры образованы народами, сходными по языку, то есть митохондриальный генофонд Европы упорядочен по лингвистическому принципу.**

Эти выводы подтверждаются и тем, что при предыдущем генетическом масштабе формирование кластеров также следовало лингвистическому принципу, поскольку отделились финно-язычные народы (уральская лингвистическая семья), а остались в кластере только народы, говорящие на языках индоевропейской лингвистической семьи, и баски – осколок мира, предшествовавшего распространению индоевропейских языков.

Из общей закономерности, выявленной в ходе анализа митохондриального генофонда Европы (популяции одной языковой группы характеризуются сходными митохондриальными генофондами, рис. 3.3 и 3.4), обнаруживаются всего лишь три исключения. Они заданы наложением географического фактора на лингвистический. Так, романоязычные румыны и аромуны, географически оторванные от большинства романских наро-

дов, и генетически располагаются особняком. Эстонцы, единственные из финно-угров, расположились между славянским и балтским кластером – то есть между своими географическими соседями. Итак, на графике можно увидеть как роль сходства языков, так и важность фактора географического соседства. Поэтому необходимо обратиться к решению вопроса: так что же все-таки определяет генетические дистанции между популяциями Европы – лингвистика или география? Этот вопрос рассматривается в следующем разделе.

Отметим, что успех проведенного анализа, по сути, впервые выявившего структурированность европейского генофонда по мтДНК, обусловлен, на наш взгляд, прежде всего обширностью созданной базы данных и использованием в анализе лишь больших выборок. Если бы в анализ были включены выборки малого объема, то статистический «шум» не позволил бы разглядеть закономерности в структуре генофонда. Вторым фактором успешности проведенного анализа явился примененный метод последовательного «увеличения» разрешающей способности через последовательное выявление и исключение из анализа генетически удаленных популяций.

3.3. ГЕОГРАФИЧЕСКАЯ И ЛИНГВИСТИЧЕСКАЯ СТРУКТУРИРОВАННОСТЬ – ТЕНДЕНЦИИ

Вопрос о том, что больше определяет сходство генофондов – географическое соседство или лингвистическое родство популяций – имеет давнюю историю в популяционной генетике. Это старый

вопрос, и чтобы ответить на него на новом витке развития генетики уже недостаточно анализировать графики генетических взаимоотношений. Тут требуются другие методы анализа, позволяющие

по формальным статистическим критериям сравнить значимость этих двух факторов.

Классическим и математически строгим методом ответа на этот вопрос является тест Мантеля, оценивающий тесноту статистической связи между матрицами сходства популяций по разным системам признаков. Для применения этого теста были сначала получены матрицы, оценивающие сходство каждой пары популяций по трем признакам: генетическим, лингвистическим и географическим.

Генетическая матрица представляла собой обычную матрицу генетических расстояний между всеми изученными народами.

Лингвистическая матрица расстояний представляла собой инвертированную матрицу лингвистического сходства между языками, на которых говорят эти народы. Матрица лингвистического сходства между индоевропейскими народами Европы была получена в ходе многолетних исследований московской школы глоттохронологии (проект «Вавилонская башня», www.staring.rinet.ru) и любезно предоставлена член-корр. РАН А.В. Дыбо. Поскольку расстояния между языками, относящимися к разным лингвистическим семьям, трудно оценить количественно, в анализ были включены только популяции индоевропейской языковой семьи. Впрочем, к этой семье относится подавляющее большинство народов Европы. Для сопоставимости генетических и лингвистических данных, при расчете генетических расстояний популяции были объединены в соответствии с их языком, выделяемым лингвистами. Хотя в большинстве случаев это выделение совпадает с обычными представлениями генетиков о народах Европы, имеется и несколько различий. Так, популяции южной Франции были проанализированы отдельно от большинства французских популяций и сопоставлены провансальскому языку, популяции Уэльса и Корнуолла сопоставлены, соответственно, с валлийским и корнуэльским языками, и ряд других подобных примеров.

Географическая матрица представляла собой географические расстояния (в километрах) между центрами этнических ареалов. Для большей сопоставимости с генетической матрицей географические координаты этих центров получены усредне-

нием географических координат генетически изученных популяций, относящихся к данному народу (языку).

Эти три матрицы послужили входной информацией для теста Мантеля, результаты которого представлены в табл. 3.1. В верхних двух строках таблицы показаны «прямые» оценки тесноты связи генетической матрицы с двумя остальными матрицами. Оказалось, что генетическая матрица примерно в равной степени и достоверно связана и с лингвистической (коэффициент корреляции $r=0.31$), и с географической ($r=0.37$). Однако достаточно взглянуть на лингвистическую карту Европы, чтобы убедиться, что лингвистика и сама связана с географией (теснота этой связи в нашем анализе составила $r=0.39$). Поэтому в двух нижних строках таблицы 3.1 теснота той же связи двух факторов оценена при исключении влияния третьего фактора (технически для такого исключения применяется коэффициент частной корреляции.) При этом все связи стали слабее, не достигая уровня достоверности, но опять можно отметить, что уровень связи генетики с лингвистикой лишь немногим уступает уровню ее связи с географией.

Таким образом, все три параметра (генофонд, язык, география) оказываются столь тесно переплетены друг с другом, что не удастся выявить лидирующую связку между членами этой триады. Однако можно констатировать, что при формальном статистическом анализе роль лингвистического фактора в структурировании генофонда Европы оказывается почти столь же значимой, что и роль географического фактора.

РАЗНООБРАЗИЕ ГЕНОФОНДОВ ЛИНГВИСТИЧЕСКИХ ГРУПП ЕВРОПЫ

Как и в разделе 2.1. по изменчивости Y-хромосомы, кроме анализа скоррелированности генетических и лингвистических расстояний, был проведен и анализ генетического разнообразия лингвистических групп Европы (табл. 3.2). При этом оценивались не различия между генофондами лингвистических групп, а различия между популяциями в пределах одной группы, то есть генетическая гетерогенность каждой группы. Оказалось, что

Таблица 3.1. Роль географического соседства и лингвистического родства в структурировании митохондриального генофонда Европы (тест Мантеля)

Матрицы расстояний	Коэффициент корреляции
Генетика и лингвистика*	0.31
Генетика и география*	0.37
Генетика и лингвистика (без влияния географии)**	0.19
Генетика и география (без влияния лингвистики)**	0.29

Обозначения: *коэффициент парной корреляции; **коэффициент частной корреляции. Указаны корреляции между матрицами генетических, лингвистических и географических расстояний.

Таблица 3.2. Гетерогенность генофондов лингвистических групп Европы по мтДНК

Лингвистическая группа	Число популяций	Средние генетические расстояния	Расстояния после коррекции*	Исключенные обособленные популяции
Германская	14	0.013	0.013	-
Кельтская	7	0.039	0.039	-
Славянская	27	0.042	0.034	Хорваты о-ва Крк, белорусы Бобруйска
Романская	27	0.048	0.044	Кантабрия
Тюркская	5	0.240	0.240	-
Финно-угорская	9	0.420	0.070	Саамы, удмурты

*В некоторых случаях одна-две генетически обособленные популяции (например, островные популяции хорватов) за счет больших генетических расстояний до остальных популяций завышали уровень гетерогенности всей группы. Поэтому была проведена коррекция: исключение из расчета тех популяций, среднее расстояние до которых более чем в два раза превосходит аналогичный показатель для остальных популяций группы.

митохондриальные генофонды всех групп индоевропейской семьи характеризуются умеренной гетерогенностью (от 0.1 до 0.4), тогда как гетерогенность финно-угорской группы уральской языковой семьи в два раза выше. Это согласуется с обособленностью их генофонда и дисперсным расположением финно-угорских популяций при многомерном шкалировании (рис. 3.3 А).

Второе место по уровню гетерогенности (межпопуляционного разнообразия) занимает тюркская группа (табл. 3.2, рис. 3.5). Это легко понять, учитывая предполагаемую для многих из них (турки, чуваша и др.) смену автохтонного языка на тюркский без значительных изменений генофонда.

ФИННО-УГОРСКИЕ ГЕНОФОНДЫ: ВЫСОКОЕ РАЗНООБРАЗИЕ МЕЖДУ ПОПУЛЯЦИЯМИ И НИЗКОЕ – ВНУТРИ ПОПУЛЯЦИЙ

При анализе генетических взаимоотношений популяций Европы мы выявили обособленность финно-угорских и родственных им популяций Восточной Европы. Анализ генетической гетерогенности разных лингвистических групп Европы только что тоже указал нам на особенность финно-угорских групп. В следующем разделе мы увидим, что и по показателю внутривнутрипопуляционного (гаплотипического) разнообразия финно-угорские народы своеобразны: среди двух обнаруженных в Европе зон сниженного гаплотипического разнообразия (рис. 1.20) «пиренейская зона» является второсте-

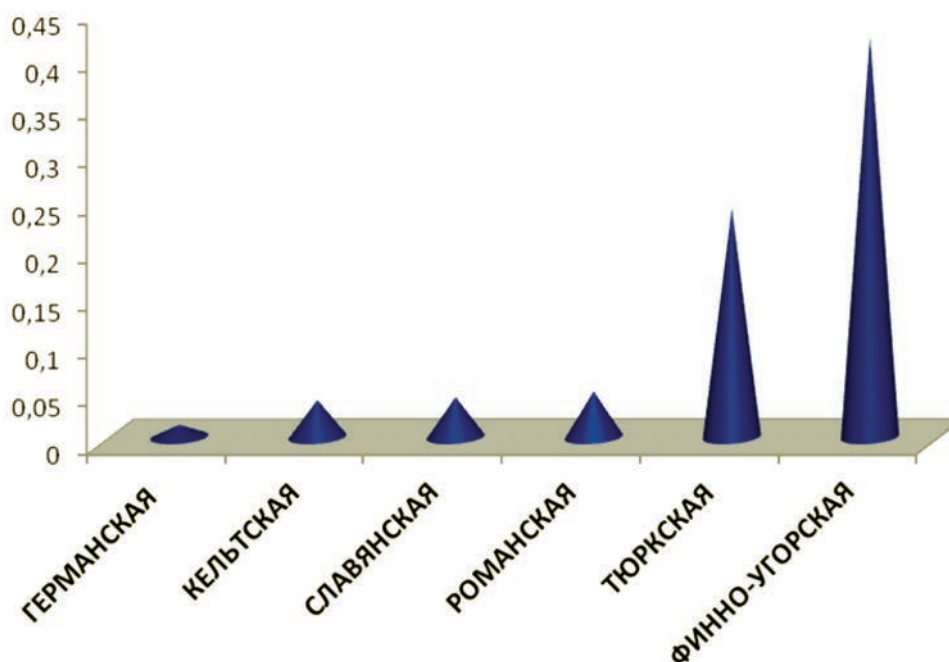


Рис. 3.5. Сравнение лингвистических групп Европы по гетерогенности их митохондриального генофонда. Показаны значения из таблицы 3.2.

пенной, а основной и по территории, и по масштабу снижения является зона сниженного разнообразия на северо-востоке Европы (рис. 1.20, рис. 3.6), в которую входят главным образом финно-угорские народы.

Таким образом, сниженное внутрипопуляционное разнообразие (рис. 3.6) сочетается у финно-угорских популяций с повышенным межпопуляционным разнообразием (рис. 3.5). Обе эти черты (сниженное внутрипопуляционное разнообразие и повышенное межпопуляционное) являются двумя сторонами одной медали: дрейфа генов. Действительно, хорошо известный в популяционной генетике «эффект Валунда» как раз объясняет снижение внутрипопуляционного разнообразия в подразделенной популяции. При этом, чем выше степень подразделенности, тем больше и межпопуляцион-

ные различия. Можно полагать, что у финно-угорских народов северо-востока Европы сниженный (в силу эколого-демографических причин) эффективный размер популяций и повышенная (по сравнению с другими европейскими народами) изоляция популяций друг от друга и сформировала обе особенности, выделяющие финно-угорский генофонд среди всех европейских генофондов: большие различия между популяциями и сниженное гаплотипическое разнообразие внутри них. Эта закономерность, таким образом, скорее географическая, чем лингвистическая: она была бы характерна для населения северо-восточных областей Европы, на каком бы языке они ни говорили, но так сложилось, что в этом ареале исторически были распространены преимущественно финно-угорские языки.

3.4. ГАПЛОТИПИЧЕСКОЕ РАЗНООБРАЗИЕ И МОНИТОРИНГ ГЕНОФОНДА

ГАПЛОТИПИЧЕСКОЕ РАЗНООБРАЗИЕ мтДНК

ГАПЛОГРУППЫ И ГАПЛОТИПЫ мтДНК

В предыдущем разделе была выявлена структурированность европейского митохондриального генофонда в соответствии с лингвистическим фактором. Не менее важной (и более легко выявляемой) является упорядоченность географическая (что следует и из табл. 3.1). Поскольку география гаплогрупп мтДНК в Европе (по крайней мере, на обычном уровне филогенетического разрешения) оказывается малоинформативной, то географическая изменчивость митохондриального генофонда изучена не на уровне гаплогрупп, а на уровне отдельных гаплотипов мтДНК. Конечно, рассмотрение географии каждого из нескольких тысяч гаплотипов не было бы разумным. В главе 6 рассматриваются географически скоррелированные группы гаплотипов, а здесь рассмотрим географию обобщающего показателя, суммирующего изменчивость всех гаплотипов мтДНК – гаплотипического разнообразия мтДНК [Балановский и др., 2011].

Для расчета величины гаплотипического разнообразия использована информация нашей базы данных «MURKA» о встречаемости каждого обнаруженного гаплотипа мтДНК в популяциях Восточной Европы и смежных территорий. Значительная часть этой информации представляет собой наши неопубликованные данные, но рассчитанные уровни гаплотипического разнообразия опубликованы в работе [Балановский и др., 2011]. Информация по другим популяциям, показанным на карте, введена в базу данных из опубликованных работ других авторов.

РАСЧЕТ ГАПЛОТИПИЧЕСКОГО РАЗНООБРАЗИЯ

Простейшей характеристикой митохондриального генофонда является частота в популяции той или иной гаплогруппы (или гаплотипа) мтДНК. Вторым показателем является гаплотипическое разнообразие, рассчитываемое как дополнение до единицы суммы квадратов частот всех гаплотипов, встречающихся в популяции. Если в популяции обнаружен только один гаплотип (т.е. все обследованные члены популяции одинаковы по гаплотипу мтДНК) – показатель гаплотипического разнообразия равен нулю; если все обследованные различаются по мтДНК (ни один из гаплотипов не встречен дважды) – показатель гаплотипического разнообразия составляет единицу. Можно видеть, что расчет гаплотипического разнообразия идентичен расчету ожидаемой гетерозиготности для аутомсомных маркеров: сумма квадратов частот соответствует суммарной частоте гомозигот, а дополнение частоты гомозигот до единицы является частотой гетерозигот, то есть гетерозиготностью популяции. Конечно, термин «гетерозиготность» применим лишь к диплоидным генетическим системам, тогда как митохондриальная ДНК гаплоидна. Поэтому можно считать, что гаплотипическое разнообразие для гаплоидных систем является аналогом гетерозиготности.

КАРТА ГАПЛОТИПИЧЕСКОГО РАЗНООБРАЗИЯ

Рассчитанные показатели гаплотипического разнообразия были картографированы. Созданная карта представлена на рис. 3.6.

Карта показывает закономерное убывание уровня разнообразия к северу: от 0.99 у южных популяций Причерноморья до 0.95 в северных областях Восточной Европы и ниже 0.9 у саамов на берегах Северного Ледовитого океана. Такая отрицательная корреляция с географической широтой местности нередко обнаруживается для отдельных генов, но впервые показана для разнообразия гаплотипов.

Для популяционной генетики сниженное внутривнутрипопуляционное разнообразие является индикатором действия дрейфа генов, особенно сильно-

го в небольших по размеру и значительно изолированных популяциях севера Восточной Европы. Такое объяснение является классическим и, по всей видимости, верным. Но можно сделать следующий шаг и задаться вопросом: а почему население северных областей существует в форме популяций небольшой численности? И почему интенсивность миграций между ними слабее, чем на юге?



Рис. 3.6. Карта гаплотипического разнообразия мтДНК в Восточной Европе.

ЭКОЛОГИЧЕСКАЯ ТРАКТОВКА

Возможный ответ на этот вопрос лежит в эколого-географических условиях обитания популяции. Если южные области представляют высокопродуктивную среду обитания для популяций человека (черноземная полоса), то более северные области относятся к лесной зоне, уже не столь благоприятной для сельского хозяйства, а значит, способной снабжать продуктами питания меньшую по численности популяцию. Еще далее к северу, в полосе лесотундры и тундры, может существовать только малочисленное население. А с учетом больших пространств эти популяции оказываются изолированными друг от друга самим географическим расстоянием, а также зачастую и меньшей «проходимостью» территории. Все эти факторы имели большое значение в прошлые столетия и тысячелетия, когда и происходило формирование генофондов севера Восточной Европы. Но даже и сейчас сравнение, например, густоты и качества дорожной сети северных и южных областей Европейской России делает понятным сохранение большей изолированности друг от друга именно северных популяций.

Эти общие рассуждения находят четкое подтверждение в прямых генетико-демографических данных. Так, проведенное исследование индексов эндогамии одной из южных областей (Белгородской) показало небольшие изменения параметров миграций, вызванные административным переподчинением районов [Чурносов и др., 2008], и в целом выявило небольшой уровень различий между генофондами разных районов Белгородской области. В то же время имеющиеся генетические данные по северным русским популяциям (Архангельской области) показывают их существенные генетические различия [Balanovsky et al., 2008; Балановская и др., 2011]. Большая генетическая дифференциация северных, чем южных, русских по-

пуляций показана и прямым расчетом (табл. 2.7), и эту закономерность можно распространить и на популяции других народов.

Другим аргументом, подтверждающим правомочность эколого-географической трактовки факта убывания гаплотипического разнообразия к северу, является его сходство с картой плотности населения (карту плотности населения можно найти, например, здесь <http://map-geo.ru/667943.html>). Оба показателя (и плотность населения, и гаплотипическое разнообразие) принимают максимальные значения в южных, промежуточные – в центральных, и минимальные – в северных областях Восточной Европы.

Таким образом, вырисовывается единая и закономерная картина взаимосвязанных изменений от юга к северу сразу нескольких параметров: географической широты (возрастает), продуктивности среды (убывает), плотности населения (убывает), изоляции субпопуляций (возрастает), эффективного размера популяции (убывает), дрейфа генов (возрастает), гаплотипического разнообразия мтДНК (убывает). Эти параметры перечислены в порядке предполагаемой причинно-следственной цепи.

В МАСШТАБЕ ЕВРОПЫ

Карта гаплотипического разнообразия была построена также и в масштабе всей Европы, а не только Восточной (эта карта приводилась на рис. 1.20 в качестве примера для сравнения возможностей картографических пакетов GGMAG и GeneGeo). В масштабе Европы выявляется и вторая область сниженного разнообразия – Пиренейский полуостров, главным образом у басков. Это также согласуется с предложенной экологической трактовкой. Хотя корреляция с широтной зональностью в этой области карты не прослеживается, но очевидна связь с высотной зональностью: снижение гаплотипического разнообразия в Пиренеях.

ЭКОЛОГО-ГЕНЕТИЧЕСКИЙ МОНИТОРИНГ

ТЕОРИИ

Предложенное «экологическое» объяснение изменчивости гаплотипического разнообразия в Европе позволяет несколько отклониться от центральной темы книги и рассуждать о роли экологических факторов в структуре генофонда человека – и вообще о возможностях отслеживания изменений в генофонде (мониторинге генофонда). Это позволит нам оглянуться на важный этап развития отечественной популяционной генетики в 70^е и 80^е годы. Сравнивая с нынешними работами, нельзя не заметить, что современный этап отличается стремительным ростом объема экспериментальных данных – важных, но частных результатов. Рабо-

там же Ю.П. Алтухова, Ю.Г. Рычкова и других видных популяционных генетиков того периода была свойственна любовь к обобщениям, поиску закономерностей. Создаваемые в результате концепции были мостом между теоретическими конструкциями и практическими исследованиями разнообразия генофондов.

Одним из таких обобщений, активно развиваемых Ю.П. Алтуховым, было представление об уровне гетерозиготности как об интегральном показателе, связанном и с продолжительностью жизни индивида [Алтухов, 1996, 1999], и с состоянием генофонда популяции: «Как убыль гетерозиготности, так и ее чрезмерное нарастание одинаково неблагоприятны для нормального функционирова-

ния популяции» [Динамика популяционных генофондов, 2004]. При этом исторически сформировавшееся соотношение гетерозиготности (внутрипопуляционной изменчивости) и межпопуляционной изменчивости рассматривалось им как важнейшая видоспецифичная «константа». Отклонение от нее в любую сторону свидетельствует о «неблагополучном» состоянии генофонда [Динамика популяционных генофондов, 2004]. В этом его работы теснейшим образом [Алтухов, Рычков, 1970] перекрываются с работами другого классика отечественной популяционной генетики – профессора Ю.Г. Рычкова и его школы, в особенности с их исследованиями распределения межпопуляционной изменчивости [Рычков, Ящук, 1980; Балановская, Рычков, 1990 а, б] и с работами в области эколого-генетического мониторинга [Рычков, Балановская, 1996].

Отдавая дань памяти этим ведущим ученым, а также желая хотя бы отчасти вернуть отечественной популяционной генетике ее теоретическую составляющую, рассмотрим представленные выше новые данные об изменчивости митохондриальной ДНК, пользуясь концепциями Ю.П. Алтухова, Ю.Г. Рычкова и их коллег о важнейшей роли гетерозиготности и о подходах к эколого-генетическому мониторингу. Для этого проанализируем связь географической изменчивости гаплотипического разнообразия (аналог гетерозиготности для митохондриальной ДНК) как с экологическими параметрами, так и с отягощенностью населения наследственной патологией (генетическим грузом).

ГЕН ИЛИ ГЕНОФОНД?

Понятие эколого-генетического мониторинга [Алтухов, 1995; Рычков, Балановская, 1996] основывается на представлении о генофонде как о «жизненном ресурсе» популяции, то есть о совокупности наследственности (определяющей ее облик) и изменчивости (задающей пределы ее пластичности, необходимой для адаптации к меняющимся условиям). Такое представление о генофонде восходит к родоначальнику геногеографии А.С. Серебровскому и получило дальнейшее развитие в работах отечественной геногеографической школы, возглавлявшейся Ю.Г. Рычковым. Похожие работы ведутся и в мировой науке: достаточно напомнить о бурно развивающейся сейчас фармакогенетике, обнаруживающей различия популяций в их реакции на лекарства. В более общем плане такие исследования являются областью экологической генетики [Спицын, 2008], изучающей те механизмы адаптации популяций к конкретной среде обитания, которые отражаются в повышении частот генов, наиболее благоприятных в данных условиях. Важно отметить, что в большинстве исследований акцент делается на конкретном *гене*, играю-

щем ту или иную роль в адаптации его носителя к условиям естественной или искусственной среды. Особенностью же отечественной популяционно-генетической традиции является акцент на *генофонде*, то есть на изменениях общих параметров распределения в популяции всей совокупности генов. Примером являются работы Ю.П. Алтухова как о связи средней (по многим генам) гетерозиготности и продолжительности жизни, так и о существовании оптимального уровня межпопуляционных различий [Алтухов, 1996, 1999; Динамика популяционных генофондов, 2004].

И в то же время, несмотря на такое множество близких по тематике исследований, нам неизвестны работы, в которых бы рассматривались проблемы адаптации народонаселения не по отдельным генам, а по генофонду в целом, причем в аспекте мониторинга, то есть отслеживания этих изменений во времени. В данном разделе я кратко суммирую итоги своих размышлений и исследований в этом направлении.

СХЕМА ЭКОЛОГО-ГЕНЕТИЧЕСКОГО МОНИТОРИНГА

Чтобы легче ориентироваться в комплексе сложных и взаимосвязанных понятий, предлагается следующая схема эколого-генетического мониторинга.

1. Генетические изменения, вызванные средой:

1А – изменения по отдельным генам (адаптация);

1Б – изменения генофонда (в силу экологических причин).

2. Отслеживание генетических изменений:

2А – вызванных средой (эколого-географический фактор);

2Б – вызванных миграциями (исторический фактор).

Проанализируем под этим углом зрения данные по изменчивости митохондриальной ДНК в народонаселении Восточной Европы. При этом наибольшее внимание уделим разделу 2А (отслеживанию экологически обусловленных изменений генофонда).

ИЗМЕНЕНИЯ ПО ОТДЕЛЬНЫМ ГЕНАМ (АДАПТАЦИЯ)

Адаптивно значимые изменения частот отдельных генов интенсивно изучаются многими исследователями на многочисленных примерах конкретных генов [Спицын, 2008; Borinskaya et al., 2009; Козлов и др., 1998 и многие другие]. Подавляющее большинство этих генов относится к аутосомным, поскольку именно они проявляются в фенотипе организма, подвергающегося действию отбора. Изучение митохондриальной ДНК практически не имеет отношения к этому разделу, поскольку

в ней находится слишком мало генов. Хотя отдельные примеры предположительного действия естественного отбора на гаплотипы и гаплогруппы мтДНК известны, все же они являются слишком частными, чтобы придавать мтДНК значимую роль во всем процессе адаптации популяций к окружающей среде. Кроме того, все эти гены находятся в кодирующей части мтДНК, а наиболее информативным для популяционных исследований является контрольный регион, считающийся селективно-нейтральным (безразличным для естественного отбора).

ИЗМЕНЕНИЯ ГЕНОФОНДА (В СИЛУ ЭКОЛОГИЧЕСКИХ ПРИЧИН)

Зато митохондриальная ДНК оказывается информативной для изучения изменений не отдельных генов, а генофондов в целом. Важно отметить, что такие изменения вовсе не обязательно должны быть «адаптацией», то есть улучшать существование популяции. Например, популяция, оказавшаяся в экологических условиях Севера, из-за меньшей продуктивности среды распадается на ряд изолированных субпопуляций, между которыми, с одной стороны, нарастают различия, а с другой стороны, члены одной субпопуляции становятся в черед поколений все более генетически сходными (вследствие случайного инбридинга). Это изменение не является адаптацией, поскольку оно не улучшает приспособленность ни отдельных членов, ни популяции в целом, но оно является неизбежным в тех экологических условиях, в которые попала такая популяция. Примером такой популяции является генофонд саамов [Tambets et al., 2004; Балановский, 2008] и северных русских популяций [Балановская, Балановский, 2007], и подобные же явления были обнаружены Ю.П. Алтуховым с коллегами для ряда популяций лососевых рыб [Динамика популяционных генофондов, 2004].

Другим примером действия экологических условий не на отдельные гены, а на генофонд в целом, является неолитическая революция: переход к земледелию вызвал резкий рост населения и резкое повышение его локальной плотности (сосредоточенность в поселениях). Это повлекло за собой изменения в структуре генофонда этих популяций – ослабело действие дрейфа генов, то есть стабилизировались частоты генетических вариантов в поколениях, сократились межпопуляционные генетические различия. Предполагается, что это же повлекло за собой и изменения в генофонде человечества в целом, поскольку те генетические варианты, которые случайно оказались свойственными популяциям, перешедшим на производящее хозяйство, по мере экспоненциального роста численности этих популяций могли стать преобладающими и в генофонде всего человечества.

Третьим примером, взятым из работ Ю.П. Алтухова, являются генетические последствия интенсивного промыслового лова рыбы. Использование сетей со стандартным размером ячеи приводит к преимущественному выживанию более мелких особей, которые (в силу связи гетерозиготности с размерами тела) оказываются в среднем более гетерозиготными [Динамика популяционных генофондов, 2004]. Тем самым экологический фактор (появление в ареале популяции многочисленных хищников с определенным способом ловли жертвы – рыболовецких судов) приводит к изменению одного из важнейших параметров генофонда – средней гетерозиготности.

Все эти примеры иллюстрируют то общее положение, что экологические условия, в которых находится популяция, оказывают закономерное воздействие не только на частоты отдельных генов, но и на общие параметры генофонда этой популяции.

В нашей работе исследуется вопрос о связи митохондриального генофонда Восточной Европы и особенностей географической среды. Высокая скорость мутаций мтДНК приводит к тому, что в подавляющем большинстве популяций мира гаплотипическое разнообразие очень высоко (более 97%). Но тем интереснее оказывается обнаружение популяций даже с небольшим снижением этого показателя. Как описывалось выше, обнаруженное снижение гаплотипического разнообразия на севере мы объясняем изменчивостью цепочки эколого-демографо-генетических явлений: географической широты (возрастает), продуктивности среды (убывает), плотности населения (убывает), эффективного размера популяции (убывает), изоляции субпопуляций (возрастает), дрейфа генов (возрастает), гаплотипического разнообразия мтДНК (убывает). Данное исследование может считаться примером комплексного изучения изменений в генофонде, проведенного с позиций эколого-генетического мониторинга.

Таким образом, мтДНК оказалась весьма эффективным инструментом для отслеживания изменений в географической структуре генофонда. При этом данные по мтДНК в целом согласуются и с данными по другим системам генетических маркеров (сходное убывание гетерозиготности к северу и востоку было показано по совокупности классических маркеров [Рычков и др., 1999]). Однако данные по мтДНК (одна генетическая система) в силу большого разнообразия вариантов и в 4 раза сниженного для этой системы эффективного размера популяции (для аутомных маркеров эффективный размер той же самой популяции в 4 раза больше) выявляют этот тренд в целом даже лучше, чем совокупность тридцати трех аутомных генетических маркеров. К тому же популярность мтДНК в популяционно-генетических исследова-

ниях, а также в работе криминалистической экспертизы обуславливает накопление огромного количества данных по множеству популяций. Это является дополнительным аргументом в пользу эффективности и целесообразности использования мтДНК для изучения общих параметров генофонда, в том числе и в эколого-генетическом аспекте.

ОТСЛЕЖИВАНИЕ ГЕНЕТИЧЕСКИХ ИЗМЕНЕНИЙ, ВЫЗВАННЫХ СРЕДОЙ. ГЕНЕТИЧЕСКИЙ ГРУЗ

Важным аспектом является добавление в ту же цепь еще одного важнейшего показателя: заболеваемости населения. Работами научной школы под руководством академика РАМН Е.К. Гинтера убедительно показана взаимосвязь между грузом наследственной патологии (следствие) и генетической структурой популяции, в частности уровнем случайного инбридинга (причина) [Гинтер, 1993; Зинченко и др., 2000; Наследственные болезни в популяциях человека, 2002]. Поскольку случайный инбридинг прямо связан с действием дрейфа генов, который, как мы видели, резко уменьшает гаплотипическое разнообразие, то возможна и тесная связь между уровнем гаплотипического разнообразия мтДНК и величиной сегрегационной составляющей груза наследственных болезней.

Для проверки этого предположения нами были сопоставлены величины этих двух показателей в одних и тех же популяциях (табл. 3.3). К настоящему времени методом тотального скрининга на-

селения определены величины генетического груза для 11 популяций России [Зинченко, Гинтер, 2008]. Из них для семи популяций имеются и данные об изменчивости мтДНК. Также мы включили в анализ и восьмую популяцию (Брянской области), взяв в качестве генетической пары наши данные по самому югу Смоленской области (районы, пограничные с Брянской областью и исторически с ней связанные).

Рассчитанные коэффициенты корреляции гаплотипического разнообразия с разными показателями генетического груза варьировали от 0.1 до 0.4. При этом наибольшая корреляция (коэффициент корреляции Пирсона, равный 0.4) получен с аутосомно-рецессивной патологией. Это указывает на правильность избранного подхода, поскольку наиболее прямая связь с инбридингом (и с генетическим дрейфом) ожидается как раз для рецессивных генов (выход которых в гомозиготу в результате инбредного брака обуславливает проявление наследственного заболевания).

Сама величина корреляции ($r=0.4$) является значимой, хотя и не слишком высокой. Нужно иметь в виду, что эта корреляция является *нижней* оценкой тесноты реально существующей связи, поскольку реальный уровень наследственной патологии трудно поддается количественному определению: интенсивность скрининга, неизбежно различная в разных областях, приводит к разному проценту выявляемости реально существующих в популяции больных с наследственной патологией.

Таблица 3.3. Уровни генетического груза и гаплотипического разнообразия мтДНК

Популяция	Аутосомно-доминантная патология	Аутосомно-рецессивная патология	X-сцепленная патология	Генетический груз суммарно	Гаплотипическое разнообразие мтДНК
Башкиры	2.73	1.53	0.68	4.94	0.984
Удмурты	3.47	1.56	0.96	5.99	0.953
Чуваши	1.82	1.25	0.47	3.54	0.967
Марийцы	2.23	1.43	0.33	3.99	0.922
Русские (Брянская обл.)	1.01	0.81	0.39	2.21	0.981
Русские (Костромская обл.)	1.25	0.9	0.3	2.45	0.958
Русские (Ростовская обл.)	2	1.34	0.44	3.78	0.976
Русские (Краснодарский кр.)	1.01	0.64	0.35	2	0.980
Русские (Кировская обл.)	1.56	1.28	0.36	3.2	
Русские (Тверская обл.)	0.96	0.72	0.62	2.3	
Адыгейцы	1.07	1.41	0.5	2.98	

Примечание: данные по генетической отягощенности взяты из работы [Зинченко, Гинтер, 2008]. В исследованиях, обобщенных в этой работе, генетический груз определялся методом тотального скрининга населения в выбранных регионах, при этом число обнаруженных семей с наследственной патологией приводится в формате «число случаев на 100 тысяч человек».

При статистических сравнениях это приводит к повышенному уровню шума, снижающему величину коэффициента корреляции.

Более того, корреляции такого уровня традиционно рассматриваются как информативные при сравнениях связи между признаками различной природы. Например, вывод о зависимости распределения гаплогрупп Y-хромосомы в Европе от географических факторов [Rosser et al., 2000] основывался на величине корреляции между генетическими и географическими расстояниями $r=0.3$; а общепризнанное в настоящее время положение об информативности фамилий в качестве квазигенетического маркера было первоначально обосновано А.А. Ревазовым с соавторами при такой же величине коэффициента корреляции $r=0.3$ [Ревазов и др., 1986].

Резюмируя, можно сказать, что гаплотипическое разнообразие митохондриальной ДНК оказывается весьма информативным для изучения связи генетической структуры популяций и груза наследственной патологии. Поэтому данные по мтДНК (совместно с другими генетическими и квазигенетическими маркерами) могут использоваться для прогноза груза наследственной патологии для тех популяций, по которым нет прямых генетико-эпидемиологических данных.

ОТСЛЕЖИВАНИЕ ГЕНЕТИЧЕСКИХ ИЗМЕНЕНИЙ, ВЫЗВАННЫХ ИСТОРИЧЕСКИМ ФАКТОРОМ

Последний аспект эколого-генетического мониторинга – это отслеживание изменений в генофонде, вызванных миграциями населения. Не имея тесной связи с экологической составляющей (поскольку миграции человека объясняются все же чаще историческими, а не экологическими причинами), фактор миграций имеет ключевое значение для мониторинга генофонда. Действительно, массовые миграции населения, усиливающиеся с каждым столетием, становятся решающим фактором, изменяющим структуру генофондов.

Отметим попутно, что рассмотрение фактора миграций имеет значение и для эколого-генетических исследований, поскольку каждое наблюдаемое изменение генофонда может быть вызвано как адаптацией (естественным отбором), так и миграциями. Поэтому прежде чем делать вывод об экологической обусловленности наблюдаемых различий, необходимо исключить действие других факторов микроэволюции, и в первую очередь миграций.

Рассмотрим информативность мтДНК для отслеживания действия миграций на генофонд – сначала для прошедших, а затем и для будущих этапов микроэволюции генофонда.

Высокая информативность для отслеживания миграций, имевших место в прошлом, доказывается всей совокупностью сотен научных публика-

ций, реконструировавших основные этапы и пути расселения человека по планете. Что касается возможности отслеживания по мтДНК итогов миграций, происходящих в настоящее время и ожидаемых в ближайшем будущем, то мы предлагаем подразделять их на два типа. Первый тип – миграции русских, украинцев и белорусов внутри их исторического ареала. Митохондриальные генофонды всех этих популяций очень похожи [Балановский и др., 2010], и вряд ли мы сможем обнаружить потоки миграций между этими генетически сходными этносами восточных славян. Однако мы можем оценивать общую интенсивность миграций по *средним* генетическим расстояниям между восточнославянскими популяциями. Эта величина составляет 0.049 между всей совокупностью восточнославянских популяций; 0.043 между популяциями белорусов; 0.053 между популяциями русских; 0.029 между популяциями украинцев. При резком возрастании интенсивности миграций между популяциями, величины средних генетических расстояний будут убывать. Таким образом, при повторных исследованиях генофондов мы имеем возможность сравнивать получаемые величины средних генетических расстояний с наблюдаемыми нами для коренного (в трех поколениях) населения, проводя тем самым генетический мониторинг общей интенсивности миграций, снижающих подразделенность восточнославянского генофонда. Для этого, правда, будет необходимо провести повторное исследование строго по стандартам первого, включая формирование выборок, использованные маркеры мтДНК (ГВС1 + отдельные маркеры кодирующей части) и набор выделяемых гаплогрупп. Необходимо при этом учитывать, что сравнительно небольшие величины этих различий делают такой подход пригодным лишь для отслеживания принципиально нового уровня межпопуляционных миграций, тогда как небольшие изменения, по-видимому, не будут достигать порога статистической значимости.

Второй тип миграций – это включение в восточнославянский генофонд населения из географически удаленных и генетически контрастных регионов. Характер распределения гаплогрупп мтДНК у народов мира позволяет четко отслеживать миграции из восточных частей Евразии, в частности Сибири, Китая, Юго-Восточной Азии. Прослеживание же миграций из Юго-Западной Азии возможно только по частотам характерных «ближневосточных» гаплогрупп, таких как J и U7. Что же касается миграций с Кавказа и Урала, то значительное сходство митохондриальных генофондов этих регионов с восточнославянским делает такую оценку в настоящее время проблематичной. Анализ мтДНК на основе ее полного секвенирования позволяет более детально дифференцировать потоки миграций. Тем не менее, можно

полагать, что для этих целей наиболее информативным будет анализ Y-хромосомы, варианты которой обладают значительно большей географической специфичностью, чем варианты мтДНК.

Заключая, можно констатировать, что из четырех аспектов эколого-генетического мониторинга митохондриальная ДНК весьма информативна для двух: 1) для изучения связи экологи-

ческих условий и структуры генофонда в целом; 2) для мониторинга таких изменений. При этом митохондриальная ДНК мало информативна для изучения адаптации по отдельным генам и обладает средней информативностью в четвертом аспекте – при отслеживании изменений, происходящих в генофонде по причине миграций населения.

ГЛАВА 4. СТРУКТУРА ГЕНОФОНДА ЕВРОПЫ В ЗЕРКАЛЕ ПОЛНОГЕНОМНЫХ МАРКЕРОВ

4.1. ОСОБЕННОСТИ ШИРОКОГЕНОМНОГО АНАЛИЗА

Специалист подобен флюсу – полнота его односторонняя.

Козьма Прутков

«ПОЛНОГЕНОМНЫЕ» и «ШИРОКОГЕНОМНЫЕ»

Быстрый прогресс технологий генетического анализа обеспечил возможность подробной генетической характеристики каждого образца - по всей длине его генома. Термин «полногеномные» используется в двух смыслах, различающихся степенью охвата генома.

В первом случае под «полногеномными» данными понимаются чипы высокой плотности покрытия генома, состоящие из множества (от нескольких сот тысяч до нескольких миллионов) SNP-маркеров. Эти панели маркеров точнее было бы называть широкогеномными (калька с англоязычного термина *genome-wide*), и именно так я буду их называть в этой книге, хотя в русскоязычной литературе их обычно называют «полногеномные».

Это нередко вызывает путаницу, потому что во втором случае под тем же термином «полногеномный анализ» имеется в виду именно расшифровка полной последовательности всего генома. Правда, и эта полнота не абсолютна – в геноме есть регионы, состоящие из часто повторяющихся последовательностей, секвенирование которых из-за этого крайне затруднено. Этим последовательностям особенно много на Y-хромосоме – они составляют более половины ее длины. Поэтому «полное» секвенирование Y-хромосомы чаще всего включает 10–15 млн. п.н., то есть примерно четверть ее длины, но и этого оказывается вполне достаточно для точных филогенетических реконструкций.

В связи с очень высокой стоимостью полного секвенирования геномов большинство «полногеномных» популяционных исследований (в том числе представленных в нашем исследовании) выполнены по «широкогеномным» данным. Но и действительно полногеномный анализ, пусть и в меньшей степени, представлен в мировой литературе и в этой книге – например, секвенирование полных древних геномов (глава 8) и результаты секвенирования Y-хромосомы (глава 7).

ПОЛНОГЕНОМНЫЕ = АУТОСОМНЫЕ?

В связи с особой важностью Y-хромосомы для популяционных работ надо пояснить, что и широкогеномные панели, и полногеномные данные, конечно, включают данные и по этой хромосоме. Но поскольку методы анализа рекомбинирующих и нерекомбинирующих систем различны, то в популяционных исследованиях широкогеномных данных информация по Y-хромосоме, а также по мтДНК и X-хромосоме, исключается из анализа. Анализируются, таким образом, лишь аутосомные маркеры, поэтому анализ широкогеномных данных является фактически синонимом анализа подробных аутосомных данных.

Нужно лишь сделать важную оговорку. Большинство широкогеномных данных разрабатывались для целей медицинской генетики и поэтому включили мало маркеров по Y-хромосоме, да и те попались в основном филогенетически мало информативные. Но две широкогеномные панели – панель Human Origin (разработанная под руководством David Reich из Гарвардского университета) и панель GenoChip (разработанная в проекте Genographic 2.0) – специально предназначались для популяционных исследований и включили множество филогенетически информативных маркеров Y-хромосомы. Поэтому при использовании этих широкогеномных панелей данные по Y-хромосоме не игнорируются. Но и в этом случае они рассматриваются отдельно, а в основной анализ включаются лишь аутосомные маркеры.

Эта «аутосомность» широкогеномного анализа касается и полногеномного. Хотя при полном секвенировании добываются также данные по Y-хромосоме и мтДНК, они так же рассматриваются отдельно, а в основные виды анализа включаются только маркеры аутосомных хромосом.

МНОГО МАРКЕРОВ, НО МАЛО ОБРАЗЦОВ

Высокая стоимость широкогеномного и, особенно, полногеномного генотипирования резко ограни-

чивает число изучаемых образцов. Если классическими для популяционной генетики являются выборки в 70-100 образцов, а при анализе Y-хромосомы и мтДНК их иногда старались даже увеличивать (что для гаплоидных систем действительно необходимо), то выборки при широкогеномном анализе составляют обычно 10-20 образцов, а полные геномы и вовсе чаще всего одиночны. Считается, что малый объем выборки отчасти компенсируется большой подробностью генотипирования каждого образца. Но, во-первых, только отчасти, а во-вторых, это только так считается. Что действительно помогает получать надежные результаты при столь малых выборках, это то, что единицей анализа при широкогеномных исследованиях выступает не популяция, а отдельный образец. Соответственно, если популяция представлена десятком-другим образцов, и почти все они ведут себя одинаково (например, входят в один кластер на графике главных компонент и характеризуются почти одинаковым соотношением предковых компонентов на графике ADMIXTURE) – это является сильным аргументом воспроизводимости и надежности полученного результата даже при малом объеме выборки.

Но малый объем выборок представляет еще одну опасность. Ведь если в выборке в 15 образцов из популяции А многие образцы случайно окажутся от людей, в своих родословных имеющих предков из популяции Б, то для всей популяции А будет сделан вывод о ее большом генетическом сходстве с популяцией Б – и этот вывод будет неверным. Иными словами, поскольку по данным о малой выборке – 10–15 человек – зачастую судят о генофонде многомиллионного народа, решающее значение имеет, в какой степени эти 10 человек репрезентативны для всего генофонда. Это накладывает огромную ответственность на тех, кто собирает популяционные выборки и отбирает из них образцы для широкогеномного анализа. Конечно, в популяционной генетике всегда анализируются выборки, а выводы делаются о популяциях. Но при широкогеномных исследованиях контраст между выборкой и всей популяцией может быть разительным. Итак, при широкогеномных исследованиях принципиальное значение имеет качество формирования выборок.

РАЗНЫЕ ПАНЕЛИ

Широкогеномное генотипирование технически проводится на двух платформах, предлагаемых компаниями Illumina и Affimetrix. Для генотипирования нужно иметь и оборудование от одной из этих компаний, и наборы для генотипирования, соответствующие оборудованию, и поэтому разрабатываемые теми же производителями.

Illumina за несколько лет своей работы последовательно предлагала около десятка основных

наборов – для панелей маркеров, включающих от 300 тысяч до 2 миллионов SNP-маркеров, равномерно покрывающих весь геном. Кроме того, компания предлагает разработку наборов на заказ, включающих любые маркеры, нужные пользователю, но объем заказа должен быть достаточно большим (1000 и более образцов). Основные панели были разработаны для нужд медицинской генетики, ведь для картирования генов болезней как раз нужно иметь равномерное покрытие генома. Но эти же панели оказались высокоинформативны и для популяционной генетики. Из панелей, разработанных к настоящему времени на заказ, для целей популяционной генетики предназначена лишь одна – панель GenoChip, разработанная биоинформатиком Eran Elhaik для проекта Genographic 2.0. Автор этой книги и сам участвует в проекте Genographic 2.0. И потому, зная принцип тщательного отбора маркеров – чтобы в их число были включены характеристические маркеры для каждой популяции мира, изученной ко времени создания набора, – может рекомендовать читателю не судить о качестве набора лишь по числу маркеров (которое кажется не слишком большим по современным меркам: около 130 тысяч аутосомных и около 13 тысяч Y-хромосомных маркеров).

Affimetrix также предлагает ряд панелей, но в целом они использовались в популяционной генетике меньше, чем Illumina. Однако самое первое, ключевое широкогеномное исследование популяций Европы [Novembre et al., 2008] выполнено на одной из первых панелей Affimetrix. И набор Human Origin, широко используемый в статьях команды David Reich, занявшей в последние пару лет лидирующие позиции, тоже генотипируется на платформе Affimetrix.

АНАЛИЗ РОССИЙСКИХ ПОПУЛЯЦИЙ

Еще с конца 90х годов ряд российских коллективов в тесном сотрудничестве друг с другом выполнял работы по изучению восточно-европейских популяций по панелям из 3–10 аутосомных маркеров. Были охвачены почти все народы Восточной Европы и Урала, а число маркеров было для того времени достаточно большим: использовались в основном микросателлиты и минисателлиты (Popova et al., 2001; Султанаева и др., 2001; Хуснутдинова и др., 2003; Verbenko et al., 2003ab; Stepanov et al., 2011; Ахметова и др., 2006), хотя применялись и маркеры других типов (Shabrova et al., 2004; Yunusbayev et al., 2006; Соловьева и др., 2010). Основная часть этих исследований была обобщена в монографии (Лимборская и др., 2002) и моей диссертации (Балановский, 2002).

Накопленный опыт исследований отдельных аутосомных маркеров позволил российским коллективам (в сотрудничестве с рядом зарубежных,

предоставлявших технические возможности генотипирования) включиться и в исследования полногеномных маркеров, когда они вошли в арсенал популяционной генетики. К настоящему времени народонаселение России и сопредельных стран, в основном благодаря работам Эстонского биоцентра и его многочисленных россий-

ских коллег, а также ряда других российских и зарубежных лабораторий, изучено на уровне, не уступающем большинству других регионов мира. По крайней мере, большинство народов России представлены хотя бы одной выборкой, генотипированной по широкогеномным маркерам.

4.2. ГЕНЕТИЧЕСКИЕ ВЗАИМООТНОШЕНИЯ ПОПУЛЯЦИЙ: ТРИ ВАРИАЦИИ НА ЕВРОПЕЙСКУЮ ТЕМУ

Все то же солнце ходит надо мной,
Но и оно не блещет новизной!

Шекспир. Сонет 76.

ВАРИАЦИЯ ПЕРВАЯ: «ГЕОГРАФИЧЕСКАЯ» КЛАСТЕРИЗАЦИЯ

Первой статьей, подробно охарактеризовавшей генофонд Европы по широкогеномным маркерам, стала работа международного коллектива, опубликованная в 2008 году в Nature [Novembre et al., 2008]. В этой статье популяции, представляющие большинство европейских стран, были охарактеризованы по панели Affimetrix из 500 тысяч SNP-маркеров. Основной результат показан на графике главных компонент (рис. 4.1).

Как говорилось, большое число SNP-маркеров позволяет надежно определять генетическое положение не только популяционной выборки, но и каждого индивидуального образца. Тем удивительнее, что даже индивиды, несмотря на заведомо большой размах межиндивидуальных различий внутри популяции, четко кластеризуются по стране происхождения. Более того, получившаяся «генетическая карта» в значительной степени воспроизводит географическую. Этот результат был вынесен в название статьи – «Гены отражают географию». Но, положив руку на сердце, этот результат не нов. Он ранее неоднократно намечался в популяционных исследованиях и школы Кавалли-Сфорца, и отечественной школы геногеографии.

Другое дело, что географическая кластеризация никогда еще не проявлялась с такой рельефностью и убедительностью. Действительно, почти все ирландские образцы образуют «этническое облако», примыкающее к облаку уроженцев Великобритании – точно так же как примыкают друг к другу два этих острова. Сразу за английским облаком – опять следуя географии – располагается облако французских образцов. А от французского облака отделено небольшим промежутком (хочется сказать – отделено Пиренеями) облако испанцев, почти слившихся – на одном полуострове – с португальцами. Полноту сходства с географической картой дополняет то, что если провести прямую ли-

нию от англичан к испанцам, то французы окажутся справа от нее и на географической, и на генетической карте, а само это пространство на генетической карте пусто – как и на географической оно занято Бискайским заливом.

Генетическая карта отлично воспроизводит географическую и в Средиземноморье – видны генетические облака, соответствующие Аппенинскому и Балканскому полуостровам, и промежутки между ними, соответствующие Балеарскому и Адриатическому морям. Правда, в Восточной Европе такого полного отражения географии уже не наблюдается. Там положение этнических облаков друг относительно друга хотя обычно и соответствует географическому положению народов, но пропорции генетической карты уже явно отличаются от географической.

Напомним, что очень близкий результат получен в нашем исследовании и по данным об Y-хромосоме (рис. 2.40): точно так же вместе группировались популяции Британских островов, близко друг к другу были популяции Пиренейского полуострова (аналогично Аппенинского и Балканского), соответствие географической и генетической карт прослеживалось и в Восточной Европе. Обнаружив один и тот же результат в двух обобщающих исследованиях, разных и по использованным системам, и по методам, и по изученным популяциям, нужно признать, что генофонд Европы действительно структурирован прежде всего по географическому принципу!

ВАРИАЦИЯ ВТОРАЯ: ОБОСОБЛЕННОСТЬ СЕВЕРО-ВОСТОКА

Последующие широкогеномные исследования генофонда Европы основывались, главным образом, на различных панелях с разным, все увеличивающимся, числом SNP-маркеров компании Illumina [Nelis et al., 2009; Yunusbaev et al., 2012], хотя использовались и панели Affimetrix [Xing et al., 2009]. Эти работы подтвердили вывод о при-

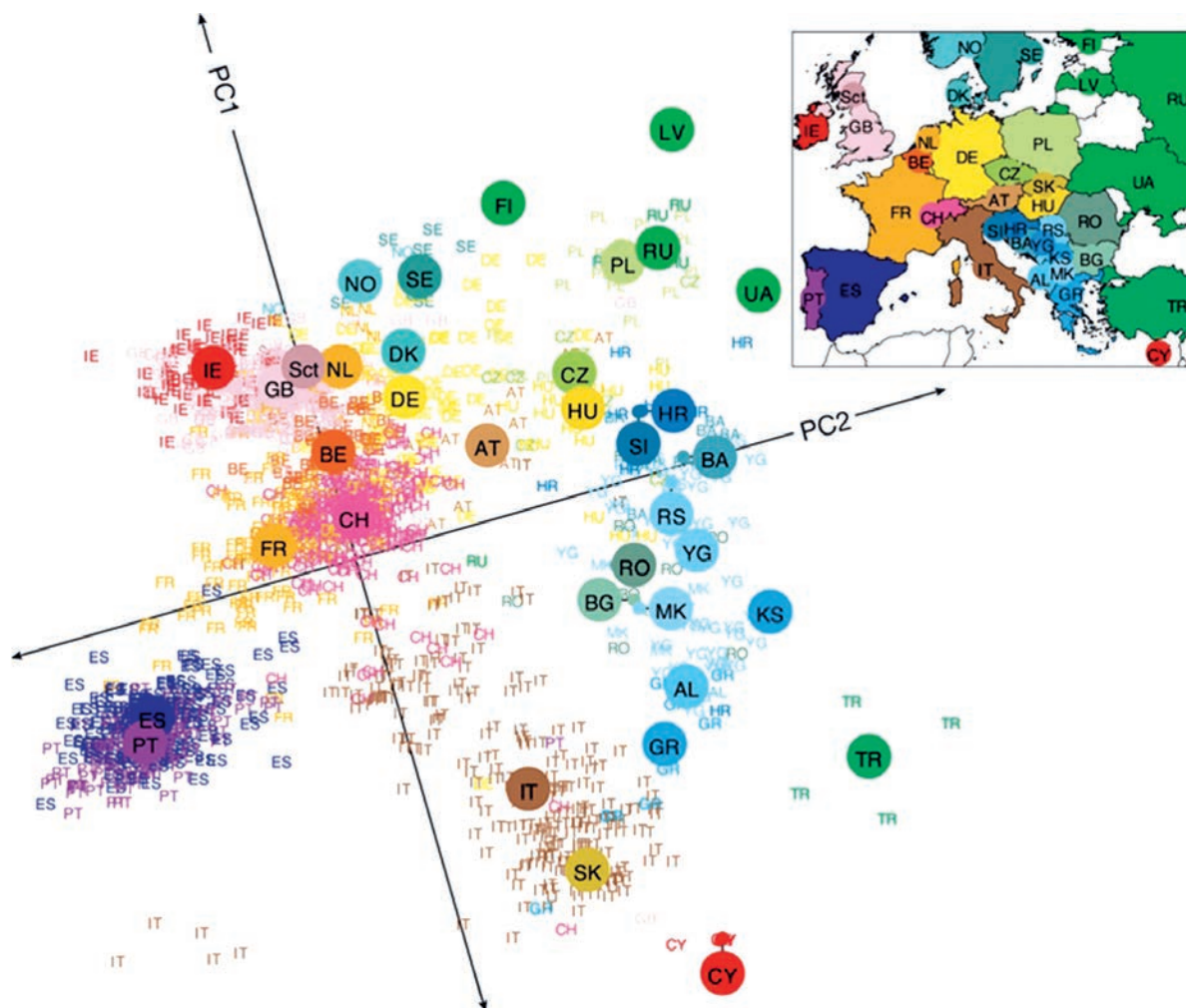


Рис. 4.1. Генетические взаимоотношения популяций Европы по данным о широкогеномных маркерах панели Affymetrix. График главных компонент, приводится по [Novembre et al., 2008].

мате географического фактора в структурировании аутосомного генофонда Европы по широкогеномным данным. Так, вышедшая уже в следующем году статья [Nelis et al., 2009] тоже констатировала кластеризацию по географическому принципу. Более того, роль географии была показана не только на уровне крупных популяций (уровня стран), но и внутри столь небольшой по площади страны, как Эстония (группы индивидов, выделенные на графике по критерию их генетического сходства, соответствуют их происхождению из различных районов Эстонии).

Другой важный вывод этой работы состоит в том, что популяции северо-востока Европы (представленные в этой статье главным образом южными и северными финнами, а также эстонцами) оказываются одним из трех генетических «полюсов» Европы. Степень выраженности этих полюсов такова, что все европейские популяции вместе формируют как бы трехлучевую звезду, с лучами, вытянутыми от общеевропейского центра к трем «полюсам» (рис. 4.2). В работе [Novembre et al., 2008] северо-восток был почти не представлен, – види-

мо, поэтому в ней не выявилась «трехлучевая» структура. А когда в работе [Nelis et al., 2009] добавились генетические контрастные финны (куда более контрастные, чем можно было бы прогнозировать из их географического положения), все остальное генетическое разнообразие Европы сплюснилось в одну линию. Правда, и в статье [Nelis et al., 2009] очень многие народы Европы отсутствовали, поэтому и трехлучевая картина не является окончательной.

Отметим, что вывод [Nelis et al., 2009] о резко выраженном своеобразии генофонда финнов по большому счету тоже не нов – не новее вывода о географической кластеризации [Novembre et al., 2008]. Это своеобразие финнов хорошо известно в популяционной генетике и объясняется сильным действием дрейфа генов на редко заселенных территориях северо-востока Европы. У медицинских генетиков есть даже термин «финские болезни» – те редкие наследственные варианты, которые за счет дрейфа генов достигли заметной частоты у финнов, а в других популяциях Европы практически отсутствуют. Финны во всех этих случаях явля-

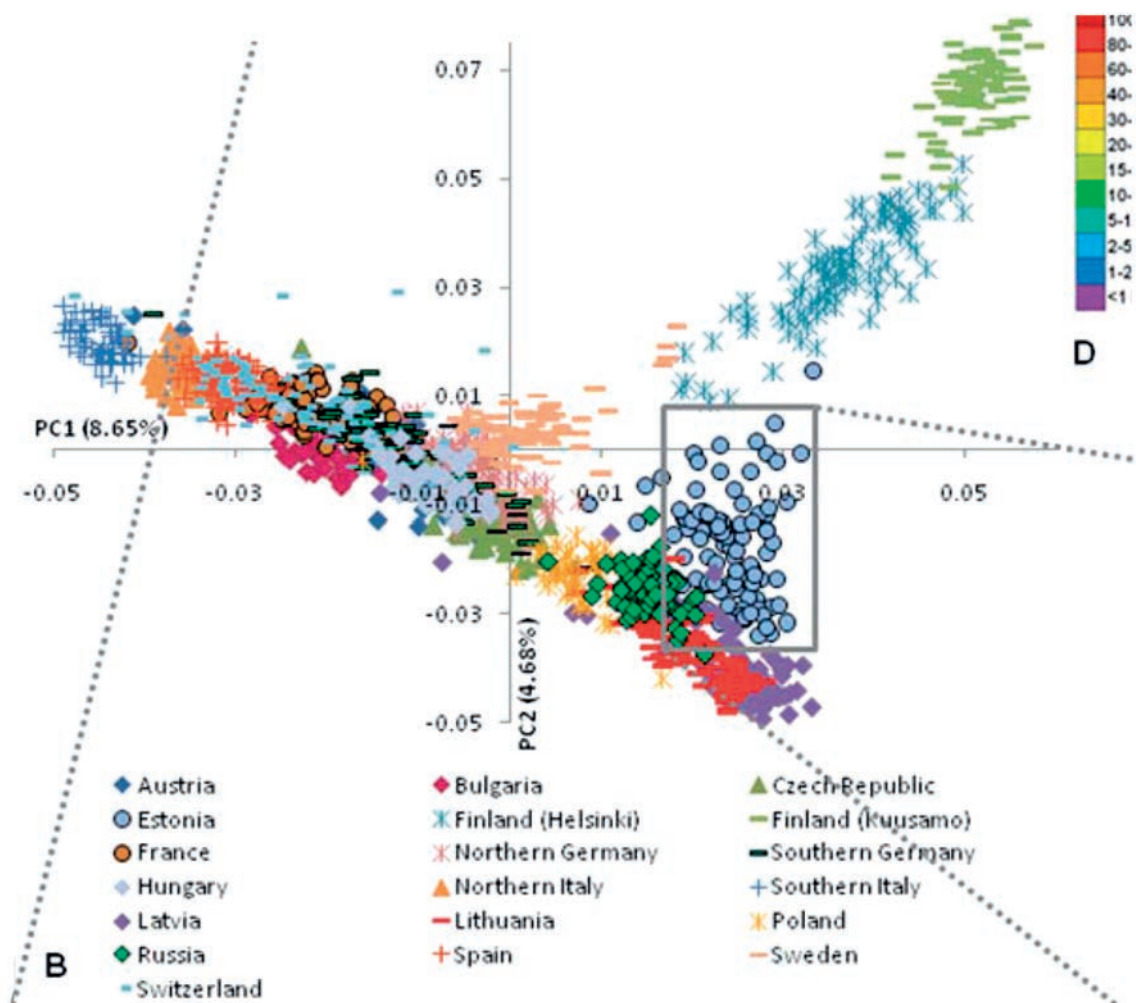


Рис. 4.2 Генетические взаимоотношения популяций Европы по данным о широкогеномных маркерах панели Illumina. График главных компонент [Nelis et al., 2009].

ются наиболее подробно изученными представителями всего населения северо-востока Европы (от Балтики до Урала), и выводы о финнах можно, хотя и с определенной осторожностью, относить ко всему этому региону.

В статье [Nelis et al., 2009] проведен также анализ генетических границ. Для этого использован не геногеографический подход через локальное межпопуляционное разнообразие (описанный в главе 8), а традиционный подход совмещения данных о генетических расстояниях между популяциями и их географических координатах, реализованный в программе *Barriér* (использованной и нами в главе 7). Любопытно, что генетические границы (барьеры для потока генов) выявлены как между этническими облаками, так и внутри некоторых народов (например, между южными и северными финнами, между южными и северными итальянцами).

Вообще эта работа ничем не уступает по своему уровню исследованию [Novembre et al., 2008], но, появившись чуть позже, она была опубликована и в менее престижном журнале (*PLoS One*), и реже цитируется.

Последующие работы не изменили сколь-либо существенно выводов этих двух основных работ, и были посвящены отдельным регионам Европы [Filipova et al., 2012; Karafet et al., 2015; Khrunin et al., 2013 и другие]. В частности, в статье [Khrunin et al., 2013] представлены данные по 4 русским популяциям, а также финно-угорским популяциям севера Европейской части России и показано, что население крайнего северо-востока Европы (представленного в статье популяцией коми) может быть еще одним полюсом генетического разнообразия европейского генофонда.

ВАРИАЦИЯ ТРЕТЬЯ: МОСТ К БЛИЖНЕМУ ВОСТОКУ

Следующий виток исследований связан с использованием двух методических приемов: специальных «исторических» панелей маркеров и совмещением данных по современным и древним популяциям.

Если первые полногеномные популяционно-генетические исследования опирались на чипы высокой плотности, разработанные для целей медицин-

ской генетики, то в последние годы все шире используются «прицельные» панели - широкогеномные панели маркеров, специально подобранные для целей выявления истории популяций. Наиболее подробные данные о современных популяциях опубликованы к настоящему времени по набору Human Origins в серии статей под руководством David Reich, и к рассмотрению их результатов мы сейчас и переходим. Что касается совмещения на одном графике данных по современным и древним образцам, то этот прием широко используется в тех же статьях, но мы его рассмотрим в главе 9, специально посвященной исследованиям древней ДНК.

График главных компонент, полученный по набору Human Origin, представлен на рисунке 4.3.

В этом исследовании были подробно изучены не только европейские, но и ближневосточные популяции. На графике прекрасно выделяются эти два региона, популяции которых протянулись двумя параллельными линиями. Генетическими полюсами Европы являются: сардинцы, а также народы Иберийского полуострова (первый полюс), и (второй полюс) народы Восточно-Европейской равнины. То есть опять паттерн в точности следует географической оси, вдоль которой с запада на восток вытянута Европа. Генетические полюса Ближнего Востока, рассматриваемого в самом широком смысле, включая и смежные регионы, – это Северная Африка и Кавказ. Обращает на себя внимание, что изменчивость генофонда Ближне-

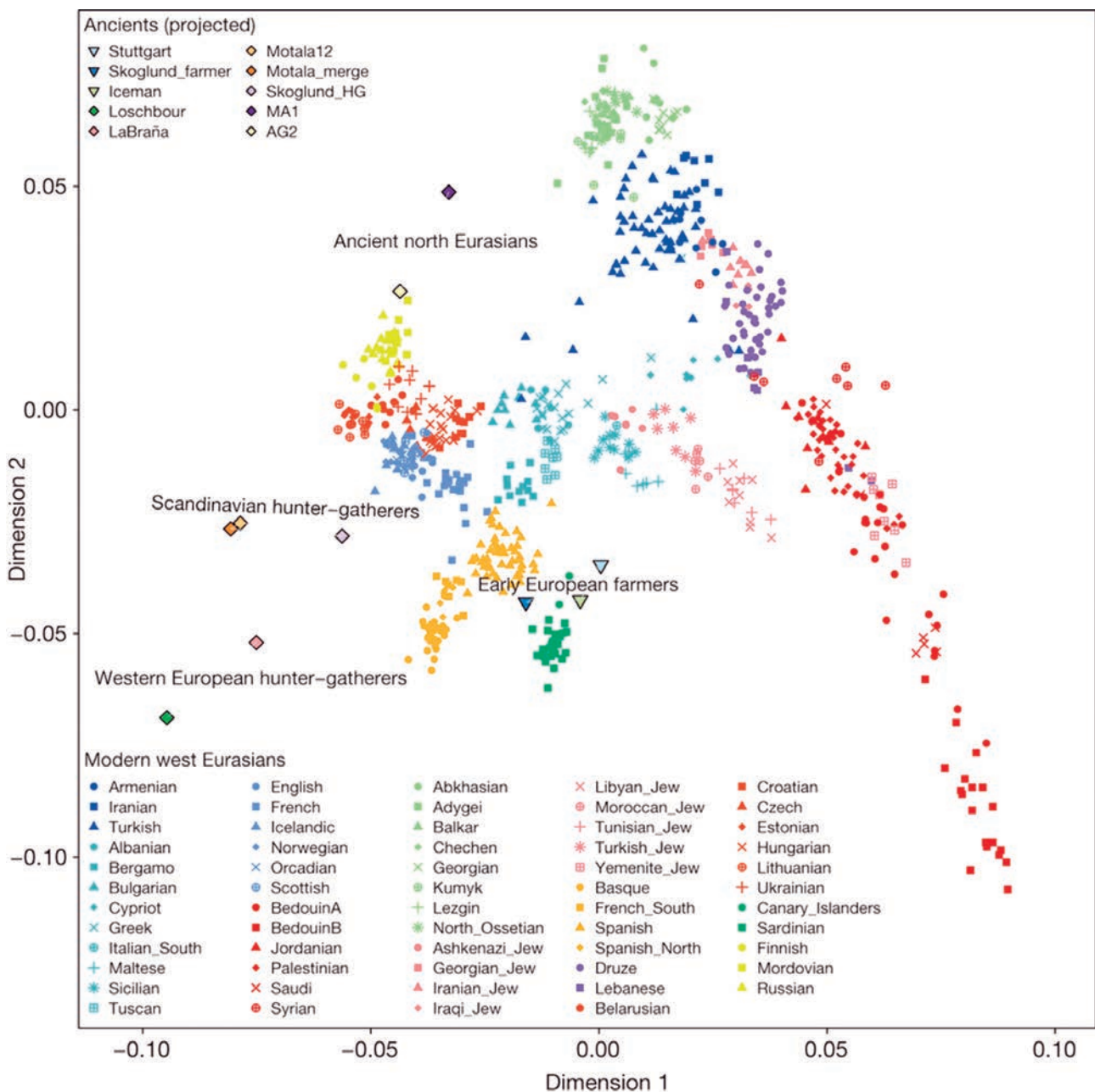


Рис. 4.3. Генетические взаимоотношения популяций Европы и Ближнего Востока по данным о широкогеномных маркерах (панель Human Origin – Affimetric). График главных компонент [Lazaridis et al., 2014].

го Востока (и Северной Африки) выше, чем Европы, хотя по площади территории эти регионы сопоставимы.

Два больших облака на графике – европейских и ближневосточных популяций – четко отделены друг от друга, как и географически они разделены Средиземным и Черным морями. Мостом между ними служат популяции двух народов, населяющих острова Средиземноморья – греков и сицилийцев. Острова Эгейского моря и Сицилия действительно являлись путями многих исторически документированных миграций между южным и северным побережьями Средиземного моря. В этом плане скорее неожиданно, почему население вдоль третьего пути (Гибралтарского) по этим данным разнится между собой (например, популяции испанцев и запада Северной Африки генетически удалены друг от друга).

Еще одним мостом являются популяции евреев (из Европы, Ближнего Востока и Северной Африки), многие из которых выведены на этот график. Возможно, в ходе своей сложной истории миграций они включили в себя как европейские, так и ближневосточные компоненты. Лишь евреи Грузии, Ирана и Йемена генетически почти неотличимы от основного населения этих стран – вероятно, за счет преобладания в их генофонде компонента принявших иудаизм местных уроженцев над компонентом, принесенным в ходе миграции евреев из их исторического ареала.

Любопытно положение сардинцев, находящихся в генетическом пространстве близко к основной массе европейских популяций, но не присоединяющихся к ней. Действительно, население Сар-

динии, как считается, сохранило генофонд неолитической миграционной волны в Европу, которая во всех других местах значительно смешалась с донеолитическим населением или приняла в себя позднейшие миграционные волны. Подтверждение этому мы увидим в главе 9, где с современным генофондом Сардинии почти совпадают геномы первых неолитических популяций Европы, реконструированные по древней ДНК.

Итак, результаты этого третьего крупного исследования, хотя и добавляют ряд важных штрихов – в основном про взаимодействие Европы и Ближнего Востока – тоже являются вариацией на тему географической кластеризации европейских генофондов.

Дальнейшее уточнение структуры европейского генофонда с помощью широкогеномных данных может идти двумя путями. Первый путь – это увеличение числа изученных популяций, ведь в каждом из трех описанных исследований карта изученных популяций зияет многочисленными и обширными дырами. Второй путь – это расширение спектра использованных методов анализа данных, ведь до сих пор мы рассматривали лишь графики главных компонент. Оба этих пути реализованы в данной книге в региональном масштабе – для балто-славянских народов Европы (глава 6). Но прежде чем перейти к анализу славянских генофондов, рассмотрим еще одни результаты по широкогеномным маркерам в глобальном масштабе. Речь идет о применении метода ADMIXTURE, который, наряду с главными компонентами, уже стал традиционным при обработке широкогеномных данных.

4.3. АНАЛИЗ ПРЕДКОВЫХ КОМПОНЕНТОВ (ADMIXTURE)

Я не сам по себе – я лишь тень отдаленного предка.
Постоянно мне кажется – с ним я один человек.
Я на дереве общем всего лишь случайная ветка,
От которой появится новый зеленый побег.

Миллионы за мною и передо мной миллионы.
Жар свирепой пустыни, зола от костров на снегу.
Как измученный зек, из невидимой этой колонны
Я и рад убежать бы, да только, увы, не могу.

А. Городницкий

ПРИНЦИП МЕТОДА

Чтобы сравнивать генофонды по широкогеномным данным, часто используется анализ «предковых компонентов» (не путать с «главными компонентами»!). Этот метод реализован в программе STRUCTURE и в ее усовершенствованном варианте – программе ADMIXTURE. Метод требует только двух вещей: широкогеномных данных по большому числу популяций и задания числа гипотетических предковых популяций, из которых

сформировались все эти современные популяции. Тогда программа вычисляет, каковы должны быть частоты каждого аллеля в каждой из предковых популяций, и рассчитывает для каждой современной популяции доли этих предковых популяций в ее генофонде. Эти разные предковые компоненты принято отображать разными цветами, и каждая популяция характеризуется своим спектром этих цветов. Понятно, что такая модель условна – в реальности вряд ли современные генофонды сформировались в результате смешения фиксированно-

го числа предковых популяций. Но выявляемые предковые компоненты, как мы сейчас увидим, имеют реальный смысл.

ВЫБОР ЧИСЛА КОМПОНЕНТОВ (k)

Основной проблемой является выбор числа предковых компонентов, который должен задать пользователь. Это число принято обозначать **k**. Хотя и существуют приемы определения оптимального **k**, они не всегда срабатывают, и выбор обычно остается на усмотрение исследователя. Ситуацию облегчает то, что выделяемые компоненты весьма устойчивы – при анализе самых разных массивов данных часто выявляются одни и те же предковые компоненты. Так, на низких **k** выделяются континент-специфичные генофонды, а при их увеличении выделяются давно известные из антропологических и геногеографических работ ареалы сходных популяций внутри континентов.

АНАЛИЗ В МИРОВОМ МАСШТАБЕ: ГЕНЕТИЧЕСКАЯ РЕАЛЬНОСТЬ РАС

Наиболее репрезентативно исследование [Pagani et al., submitted], основанное на секвенировании полных геномов, и поэтому набор использованных SNP-маркеров не смещен в сторону европейских популяций, как это свойственно большинству широкогеномных панелей. В этой работе к тому же проанализировано большое число популяций из всех основных регионов мира.

В результате при **k=2** один компонент включил все популяции Африки и Западной Евразии, а другой – все популяции Восточной Евразии и Америки.

При увеличении до **k=3** западно-евразийские и африканские популяции разделяются на отдельные компоненты.

При **k=4** выделяется отдельный компонент, доминирующий в популяциях американских индейцев, но также составляющий около половины генофонда и в Сибири (остальная часть сибирского генофонда приходится на восточно-евразийский и в меньшей мере на западно-евразийский компоненты).

Но уже при **k=5** население Сибири приобретает свой собственный компонент, вбирающий основную часть сибирского генофонда.

При **k=6** население Сахула (Австралия + Папуа Новая Гвинея, правда, представленное в этом исследовании лишь папуасскими популяциями), приобретает свой собственный компонент. До этого оно было частью восточноевразийского компонента, а после отделения сибирского – частью оставшегося восточноазиатского компонента.

Эта картина удивительно схожа (чтобы не сказать совпадает) с традиционной расовой классификацией. Ведь при **k=3** мы, по сути, видим при-

вычное со школы деление на три большие расы – негроидов, европеоидов и монголоидов. А если физическим антропологам сказать, что требуется разделить все антропологическое многообразие на 4 группы, тогда большинство из них выделит американоидную расу из состава монголоидной. Небольшое отличие от антропологических классификаций состоит в том, что австралоидная раса приобретает свой собственный компонент лишь начиная с **k=6**, тогда как в традиционных расовых классификациях австралоидная раса имеет чуть более высокий иерархический статус. Поразительно и печально, что при таком полном подтверждении традиционных расовых классификаций новейшими генетическими данными, до сих пор широко распространено мнение, что генетика «доказала» отсутствие рас. Впрочем, об этом уже давно писалось [Раса: миф или реальность, 1998].

Все эти расы, выделившиеся при **k=6** (негроидная, европеоидная, америндская, северомонголоидная, южномонголоидная, австралоидная), продолжают существовать и при более высоких **k**, лишь при каждом его повышении какая-то из них дробится на два компонента.

Так, при **k=7** из европеоидного (западноевразийского) компонента выделяется южноазиатско-ближневосточный, который при **k=8** разделяется на ближневосточный и южноазиатский (индийский), при **k=9** выделяется компонент северо-восточной Сибири, при **k=10** начинает дробиться негроидная раса и так далее.

АНАЛИЗ В ЕВРОПЕЙСКОМ МАСШТАБЕ

Но вернемся к генофонду Европы. В ходе анализа генофонда балто-славянских народов (глава 6) был проведен анализ ADMIXTURE, включивший не только балтов и славян, но и остальные изученные по широкогеномным маркерам популяции Европы, и, в несколько меньшей мере, популяции других регионов мира. Это позволяет нам воспользоваться этими результатами для анализа в европейском масштабе.

Приведем карты распространения предковых компонентов, полученных при контрастных величинах **k**, равных 6 и 18. Первый вариант близок к разумному минимуму, поскольку делит все генофонды Старого Света на шесть основных доменов: африканский домен, два домена в северной половине Евразии (европейский и сибирский), и три домена в южной половине Евразии (средиземноморский, южноазиатский и восточноазиатский). Это те регионы мира, которые выделяются и географически, и известны из предшествовавших геногеографических работ, что подтверждает осмысленность результатов, получаемых методом ADMIXTURE. Второй вариант (**k=18**) близок к максимуму, поскольку при таких высоких **k** многие популяции приобретают

«собственные» компоненты, и картина становится чрезвычайно подробной.

Чтобы различать предковые компоненты, выделенные при $k=6$ и $k=18$, будем далее обозначать их в первом случае ($k=6$) как 6.1, 6.2, ... 6.5, 6.6; а во втором случае ($k=18$) как 18.1, 18.2, ... 18.17, 18.18.

РАСПРОСТРАНЕНИЕ ПРЕДКОВЫХ КОМПОНЕНТОВ ПРИ $k=6$

На рисунке 4.4 картографировано распространение по Европе и смежным регионам всех шести предковых компонентов, выделяемых при $k=6$. Поскольку они выделены в масштабе мира, максимумы только двух из них приходится на изучаемый регион – это североевропейский и средиземноморский компоненты. Геногеографические карты всех шести основных компонентов, в разной мере создающих генофонд Европы, позволяют объективно оценить и степень влияния, и регионы их влияния (рис. 4.4. А-Е).

Североевропейский компонент (6.1, рис. 4.4.А) охватывает всю Европу, но максимума достигает на территориях вокруг Балтийского моря, где составляет более 90% генофонда. Паттерн его распространения поразительно сходен – почти совпадает – с распространением светлой пигментации волос и глаз. Это ведущий антропологический признак выделения североевропеоидной малой расы, и этот генетический компонент, по всей видимости, совпадает с антропологическим компонентом, давно уже выделенным в физической антропологии.

Средиземноморский компонент (6.2, рис. 4.4.Б) составляет более трети генофонда на европейском побережье и более половины генофонда – на азиатском и африканском побережьях Средиземного моря.

В Европе доля североевропейского компонента снижается к югу, а доля средиземноморского навстречу ему – к северу. Но в целом североевропейский компонент преобладает по всей Европе: даже на южных полуостровах Европы он составляет большую долю генофонда, чем средиземноморский предковый компонент. Центр тяжести средиземноморского компонента приходится на Юго-Западную Азию. Его можно сопоставить с южноевропеоидной (средиземноморской) малой расой. Основное отличие от нее кроется за пределами картографированного ареала – ареал средиземноморской малой расы включает и Индостан, тогда как генетический предковый компонент в Индостане свой, равноправный средиземноморскому.

Впрочем, и средиземноморский компонент заходит в Индостан, и индийский (южноазиатский) компонент заходит в Средиземноморье.

Южноазиатский компонент (6.3, рис. 4.4.В) в пределах Европы наиболее выражен на Северном

Кавказе, где составляет около трети генофонда. Примерно так же он выражен и по всей Юго-Западной Азии. Заметную долю генофонда (10-15%) он составляет на Балканском и Аппенинском полуостровах и в южных частях Центральной Европы, но далее к северу не распространяется, и вокруг Балтийского моря почти полностью отсутствует.

Сибирский компонент (6.4, рис. 4.4.Г). Примерно с той же степенью влияния, как и южноазиатский компонент, в Европе представлен и сибирский. На отдельных территориях (на южном Урале) он составляет более четверти генофонда, но чуть западнее у народов Поволжья частота его уже снижается, далее к западу на основной части Восточно-Европейской равнины он составляет совсем небольшую часть генофонда, а еще западнее – в Центральной и Западной Европе – практически отсутствует.

Что касается двух оставшихся компонентов мирового генофонда, то их доля в европейском генофонде ничтожна.

Восточноазиатский компонент (6.5, рис. 4.4.Е) составляет от 1 до 5% процентов у отдельных народов Урала, Поволжья и Восточного Кавказа, но отсутствует на всей остальной территории Европы.

Африканский компонент (6.6, рис. 4.4.Д) сходным образом составляет пару процентов генофонда Пиренейского полуострова, и отсутствует на остальной территории Европы.

РАСПРОСТРАНЕНИЕ ПРЕДКОВЫХ КОМПОНЕНТОВ ПРИ $k=18$

Хотя в нашем анализе генофондов мира методом ADMIXTURE Европа и представлена куда большим числом популяций, чем другие части света, однако из 18 предковых компонентов 10 являются африканскими или азиатскими. На территории Европы они либо вообще не встречаются, либо ограничиваются ее самыми восточными или самыми южными популяциями, но даже в них частоты этих компонентов лишь немного превышают 1%. Из остальных 8 компонентов основными в Европе являются западноевропейский и восточноевропейский компоненты; чуть меньшее значение имеют средиземноморский и кавказский; поволжский и южноуральский выражены локально, хотя и очень ярко; а южноиндийский и сибирский компоненты, напротив, покрывают почти всю Восточную Европу, но составляют лишь немногим более одного процента ее генофонда. Рассмотрим эти восемь карт подробнее.

Четыре основных предковых компонента Европы (**18.1–18.4**), выделяемые при значениях $k=18$, представлены на рис. 4.5. А-Г.

Западноевропейский компонент (18.1, рис. 4.5.А) составляет более половины генофонда на

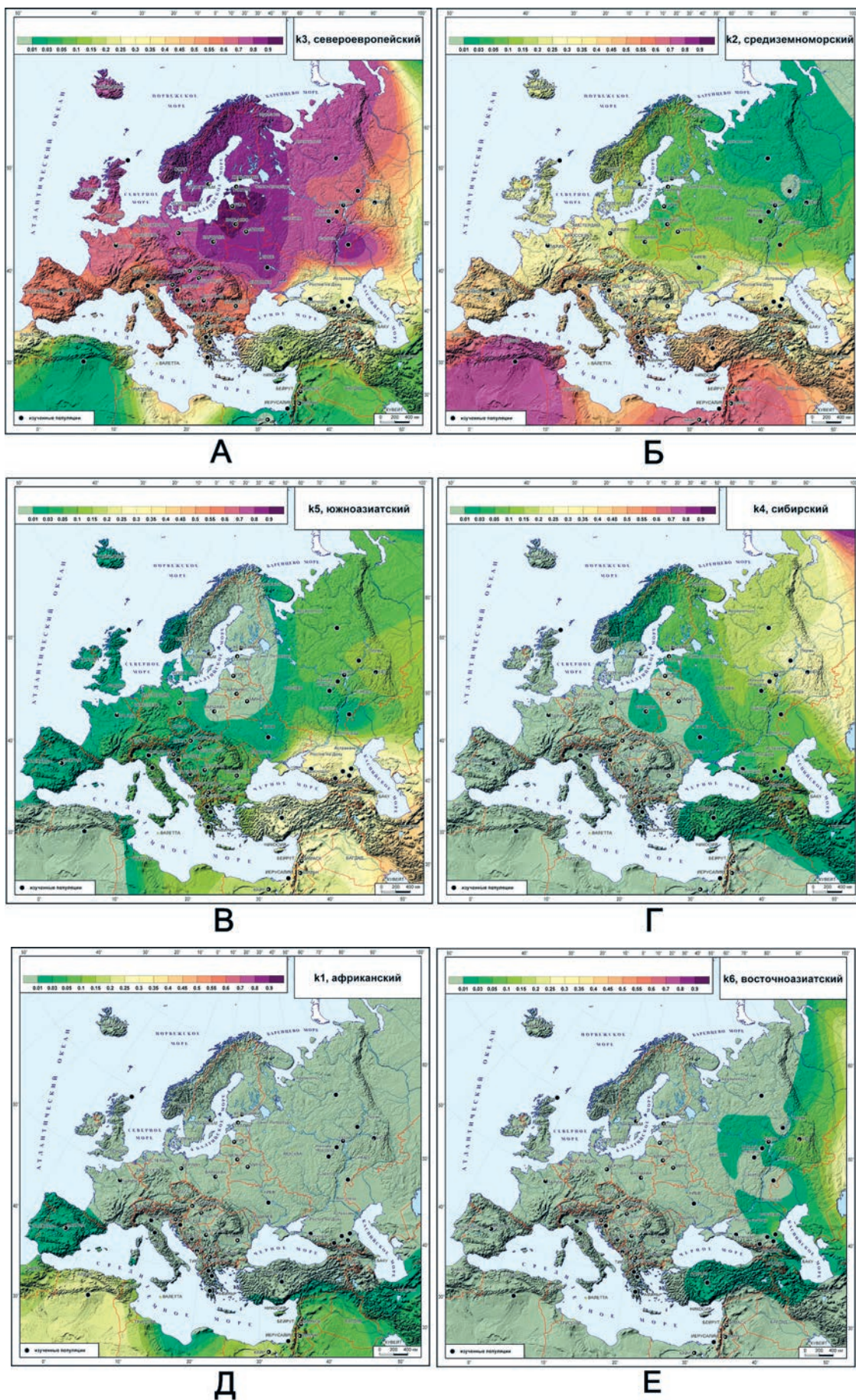
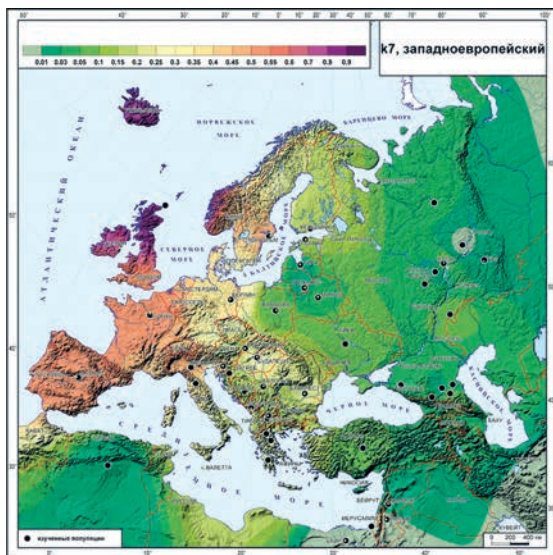
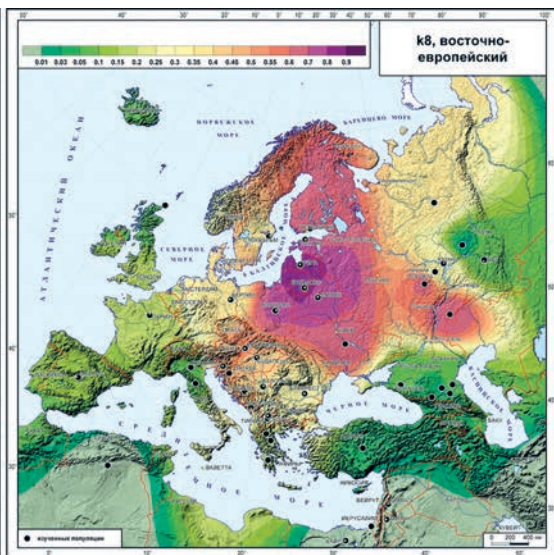


Рис. 4.4. Распространение предковых компонентов ADMIXTURE при k=6.

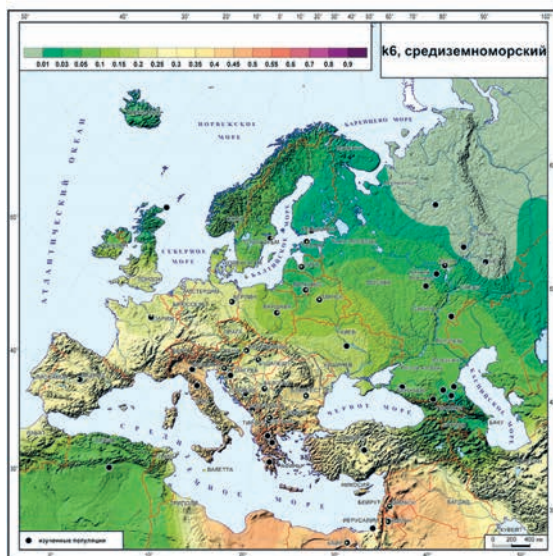
А. Североевропейский компонент (6.1); Б. Средиземноморский компонент (6.2); В. Южноазиатский компонент (6.3); Г. Сибирский компонент (6.4); Д. Африканский компонент (6.6); Е. Восточноазиатский компонент (6.5).



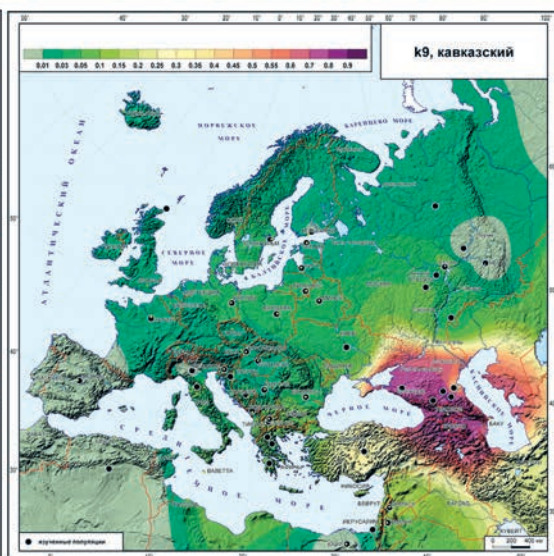
А



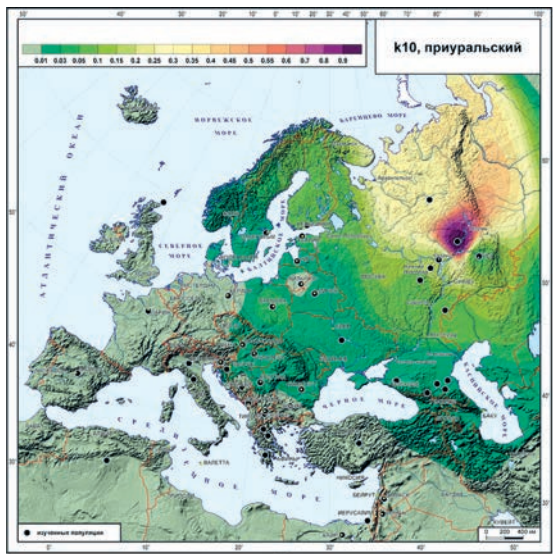
Б



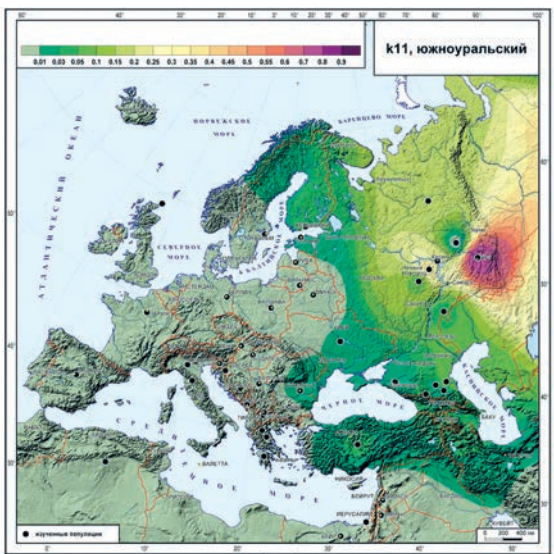
В



Г



Д



Е

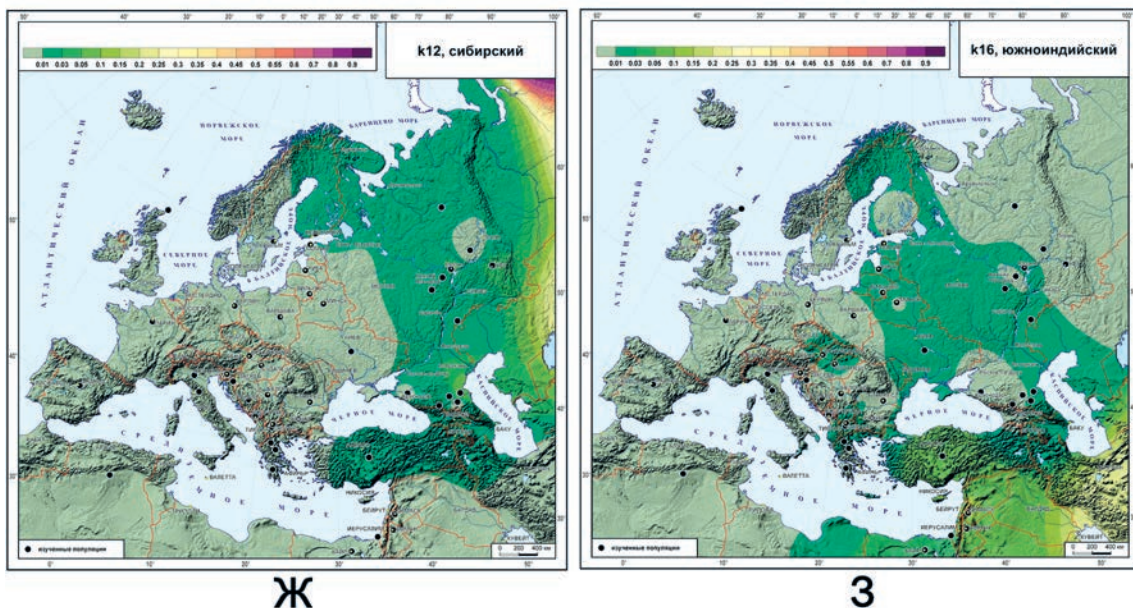


Рис. 4.5. Распространение предковых компонентов ADMIXTURE при $k=18$.

А. Западноевропейский компонент (18.1); Б. Восточноевропейский компонент (18.2); В. Средиземноморский компонент (18.3); Г. Кавказский компонент (18.4); Д. Приуральский компонент (18.5); Е. Южноуральский компонент (18.6); Ж. Сибирский компонент (18.7); З. Южноиндийский компонент (18.8).

Иберийском полуострове, во Франции, на Британских островах и юге Скандинавии. Перепрыгивая моря, он объединяет популяции всех побережий западной трети Европы.

Восточноевропейский компонент (18.2, рис. 4.5.Б) доминирует у популяций восточной трети Европы. На всем пространстве огромной Восточноевропейской равнины – от Черного моря на юге до Балтийского и Белого на севере, от Карпат на западе до Волги на востоке – он составляет даже не половину, а две трети генофонда, и чуть менее выражен и далее на восток – до Урала.

Средняя треть Европы – Центральная Европа и Балканы – примерно в равной мере характеризуется западноевропейским и восточноевропейским компонентами, причем на Балканах они оба уступают лидерство средиземноморскому компоненту.

Средиземноморский компонент (18.3, рис. 4.5.В). Максимум этого предкового компонента находится за пределами формальных границ Европы – в Леванте и Египте. Но это и показывает всю условность границ между выделяемыми для удобства частями света. Ведь население всех побережий Средиземноморья, и особенно восточной его части, издавна обменивалось миграциями. Скорее мир Средиземноморья можно было противопоставить миру более северных областей Европы, чем противопоставлять друг другу европейскую и африканско-азиатскую части Средиземноморья. Об этом и напоминает нам география средиземноморского компонента, составляющего более трети генофонда населения Балканского и Аппенинского

полуостровов (особенно Греции и Италии) и чуть менее трети – генофонда остальной части южной Европы, включая и Карпаты, и Альпы, и южную половину Франции.

Кавказский компонент (18.4, рис. 4.5.Г). Четвертый и последний из основных предковых компонентов Европы, напротив, распространен только в одной части Европы – на Северном Кавказе. Но зато он составляет более трех четвертей генофонда северокавказских популяций. Впрочем, генетически Кавказ является частью Передней Азии, а не Европы (как мы увидим в главе 7). Поэтому и предковый кавказский компонент на Переднюю Азию еще распространяется, а в Европе выражен очень слабо: его доля варьирует от 1% до 5% генофонда, причем популяции Восточной Европы ближе к верхней границе этого интервала, а Западной Европы – к нижней.

Этими четырьмя компонентами описывается почти весь генофонд Европы – за исключением Волго-Уральского региона. В этой восточной окраине Европы обнаруживается большое генетическое разнообразие, проявившееся сразу в двух предковых компонентах.

Приуральский компонент (18.5, рис. 4.5.Д). Один из них, названный нами приуральским, превышает 90% у удмуртов и составляет около трети генофонда коми.

Южноуральский компонент (18.6, рис. 4.5.Е). Другой – южноуральский предковый компонент – доминирует у башкир и составляет около трети генофондов татар и чувашей.

Но оба этих компонента распространены, хотя и с очень разными частотами, у всех народов Волго-Уральского региона. На остальной территории Восточной Европы встречаются с крайне низкими частотами, а в Западной Европе отсутствуют. Наличие этих двух специфических уральских компонентов указывает на большое генетическое разнообразие и своеобразие, скрытое в самых восточных областях Европы – вывод, совпадающий с результатами анализа митохондриальной ДНК (рис. 3.5) и с результатами анализа карт генетических расстояний (глава 5).

Для полноты картины упомянем еще два компонента, которые, хотя и с низкими частотами, все же обнаружены в генофонде Европы.

Сибирский компонент (18.7, рис. 4.5.Ж) встречается, как и можно ожидать, у народов на самом востоке Европы (с частотой около 2%).

Южноиндийский компонент (18.8, рис. 4.5.З), распространенный также по всей Передней Азии, заходит, пусть и с самыми низкими частотами, на Восточно-Европейскую равнину, но отсутствует в Западной Европе.

ГЛАВА 5.

ПАНОРАМА НАРОДОВ НА ФОНЕ ЕВРОПЫ

Лицом к лицу
Лица не увидеть.
Большое видится на расстоянии.
Сергей Есенин

Мы рассмотрели отражение лика генофонда Европы в трех зеркалах – Y-хромосомы (глава 2), митохондриальной ДНК (глава 3) и аутосомного генома (глава 4). Однако даже такое трехмерное отображение будет все же незаконченным, если мы от Европы в целом не обратимся еще и к лицам отдельных народов – к генетическим связям того или иного народа Европы со всем остальным европейским миром. Такое рассмотрение позволяет не только увидеть место того или иного этнического генофонда среди его ближних и дальних соседей. Оно дает больше – увидеть, как именно из отдельных пазлов складывается общая картина европейского генофонда. Может быть, это позволит разглядеть и пути истории в сложении этой общей картины. Для этой цели наиболее полезно зеркало Y-хромосомы: его информативность сопоставима с информативностью широкогеномных аутосомных панелей (как показано, например, в следующей главе 6), а палитра изученных популяций – несравнимо богаче.

Генетический портрет отдельных народов на фоне всего европейского генофонда лучше всего обрисуют карты генетических расстояний. Они показывают как генофонд данного народа вписывается в общую панораму народов Европы. Основываясь на всей совокупности гаплогрупп, карты генетических расстояния показывают для данного народа насколько он своеобразен, с кем он схож, от кого отличается, насколько далеко простираются его генетические связи с другими народами Европы и близлежащих регионов.

Карты генетических расстояний создаются так. Сначала строится серия карт – для каждой гаплогруппы своя карта. Как подробно описано в главе 1 (раздел 1.4), каждая карта представляет собой числовую матрицу – очень густую сетку, равномерно покрывающую все картографируемое пространство. В каждом из множества узлов сетки (на приводимых картах – почти 200 тысяч узлов сетки покрывают картографируемую территорию) записана частота данной гаплогруппы в данной географической точке. Затем выбирается интересующая нас группа популяций (она называется реперной)

– скажем, поляки – от которой и будут рассчитываться генетические расстояния до каждого узла сетки (включая и ареал самих поляков). Берутся средние частоты гаплогрупп у поляков. А для каждой точки Европы известны частоты в ней каждой гаплогруппы (ведь мы уже построили карту для каждой гаплогруппы). Этих данных достаточно, чтобы вычислить среднее генетическое расстояние от частот гаплогрупп у поляков до частот гаплогрупп в каждой точке Европы. Эти генетические расстояния и наносятся на карту.

Затем берем в качестве реперной популяции, например, сербов – и повторяем все те же действия с картами. И получаем карту генетического ландшафта, показывающего степень сходства Y-хромосомного генофонда сербов с Y-хромосомным генофондом каждой популяции Европы. И так для любой выбранной нами популяции – этноса или же субэтноса.

Однако, что делать с тем, что разные популяции изучены по разным наборам гаплогрупп? Конечно, при построении геногеографических карт интерполированные значения рассчитываются для каждой точки карты даже если опорных точек (непосредственно изученных популяций) немного. Но если мы хотим – при построении карт генетических расстояний – наиболее точно охарактеризовать генофонд всех популяций, то панель гаплогрупп начинает сокращаться как шагреновая кожа. Нашим коллективом используется обширная панель SNP-маркеров (44 основных, 32 дополнительных гаплогруппы, а также еще 32 «новейших» гаплогруппы, как описано в разделе 1.3), и большинство популяций Восточной Европы нами изучены по этой широкой панели. Но чтобы равномерно представить на картах генетических расстояний все уголки Европы, эту панель, к большому сожалению, на данном этапе изученности генофонда пришлось сократить до восьми основных европейских гаплогрупп – **E1b-M35, G-M201, I1-M253, I2a-P37, J-M304, N1c-M178, R1a-M198, R1b-M269**.

Дальнейшие исследования и массовый скрининг популяций Европы по субветвям этих гаплогрупп, открываемым благодаря полному секвени-

рованию Y-хромосомы, позволит постепенно уточнять эти карты. Читая любую карту, надо помнить, что эта модель создана для объема информации, имеющейся на данном временном срезе: ограничен и массив популяций, и панель гаплогрупп. Поэтому важно обращать внимание не на детали, а на наиболее общие и устойчивые структуры генетического ландшафта.

Карты генетических расстояний могут быть построены для всех народов Европы. В данной монографии мы приведем не все, но многие – 36 карт генетических расстояний от 36 этносов и субэтносов Европы, наиболее важных для остальных глав книги. Эти 36 генетических ландшафтов объединены в шесть серий:

Серия 1: народы Северо-Восточной Европы (карелы и вепсы, эстонцы, коми ижорские, коми прилузские, литовцы, латыши, северные русские, финны);

Серия 2: восточные и западные славяне (русские центральные и южные, украинцы, белорусы, белорусы Полесья, поляки, кашубы, словаки, чехи, сорбы);

Серия 3: неславянские народы Восточной Европы (башкиры, казанские татары, мишари, чуваша, мокша и эрзя);

Серия 4: на севере Балкан (молдаване, румыны, гагаузы, венгры, словенцы);

Серия 5: южные славяне (македонцы, сербы, хорваты, боснийцы, Герцеговина);

Серия 6: обрамление Европы (албанцы, шведы, ногайцы).

5.1. НАРОДЫ СЕВЕРО-ВОСТОЧНОЙ ЕВРОПЫ (СЕРИЯ I)

Эта серия объединяет восемь карт генетических расстояний – не только от генофондов этносов (карел и вепсов, эстонцев, латышей, литовцев и финнов), но и от отдельных субэтнических групп (коми ижемских, коми прилузских, Русского Севера). Почти все эти карты объединяет не только географический регион, но и сходство генетического ландшафта. При этом лингвистическая принадлежность этих народов поражает своей пестротой. Здесь есть и западные финноязычные народы (прибалтийско-финская ветвь финно-угорских языков) – карелы, эстонцы, финны; и восточные финноязычные коми (пермская ветвь финно-угорских языков); и славяне – северные русские; и балты – латыши и литовцы. И, тем не менее, их генофонды во многом схожи. Чтобы убедиться в этом, рассмотрим всю серию карт – восемь карт генетических расстояний от каждого из восьми реперных генофондов (рис. 5.2-5.9). А для того, чтобы видеть отличия каждой из восьми карт от обобщенного генетического ландшафта Северо-Восточной Европы, приведем среднюю карту генетических расстояний (рис. 5.1). Такой обобщенный ландшафт получен в результате простых арифметических действий с матрицами карт: суммированием всех восьми карт (для каждой точки карты были суммированы значения восьми карт в этой точке) и делением полученной суммарной карты на восемь.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ КАРЕЛ И ВЕПСОВ (рис. 5.2)

Основной ареал генофондов, сходных с карелами и вепсами (при расчете «реперных» частот Y-гаплогрупп наряду с данными о карелах и вепсах учтена также небольшая выборка ижорцев), очерчен географически четко (рис. 5.2). Наиболее

генетически близкие популяции (то есть генетические расстояния до них от карел и вепсов наименьшие), окрашены интенсивно зелеными тонами. Это генетические расстояния в интервале $0 < d < 0.05$. Такие небольшие генетические расстояния обнаруживаются для популяций стран Прибалтики, Русского Севера и частично Приуралья. Несколько генетически более отдаленные популяции, окрашенные в светло-зеленые и желтые тона (генетические расстояния до них от карел и вепсов лежат в интервалах $0.05 < d < 0.10$), ненамного расширяют этот ареал. На севере он охватывает Белое море. На юге его граница удивительно четко проходит по левобережью Волги. На юго-западе эта граница проходит между Литвой и Беларусью. На юго-востоке она охватывает Татарстан и вдоль правобережья Волги опускается до Самары, на востоке ее ограничивает Кама, и далее граница следует почти строго на север по «камскому меридиану».

Важное отличие карты генетического ландшафта карел и вепсов от других карт данной серии мы обнаруживаем не на востоке, а на северо-западе. Здесь область генетического сходства с карелами и вепсами не знает административных границ и пронизывает «желтым» коридором популяций, генетически все же схожих с карелами и вепсами ($0.05 < d < 0.10$), все страны Фенноскандии – Финляндию, Швецию и Норвегию.

Стоит отметить и то, что группа «оранжевых» интервалов (генетические расстояния от карел и вепсов $d \approx 0.2$), показывающая популяции генетически уже далекие, но все же еще не совсем чуждые генофонду карел и вепсов, охватывает значительную часть Фенноскандии, Восточной и Центральной Европы. При этом Западная и Южная Европа, Предкавказье, Прикаспийский регион и даже Зауралье генетически максимально далеки от ге-

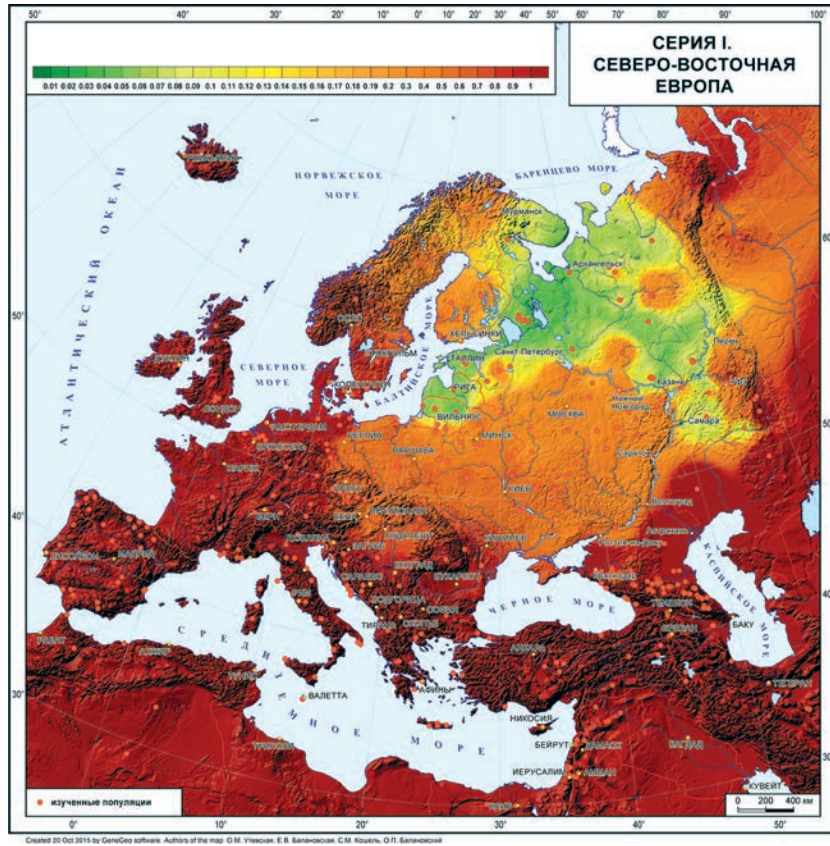


Рис. 5.1. Карта обобщенного генетического ландшафта Северо-Восточной Европы по гаплогруппам Y-хромосомы. (Усредненная карта генетических расстояний от вепсов и карел, коми ижемских, коми прилузских, латышей, литовцев, русских северных популяций, финнов, эстонцев).

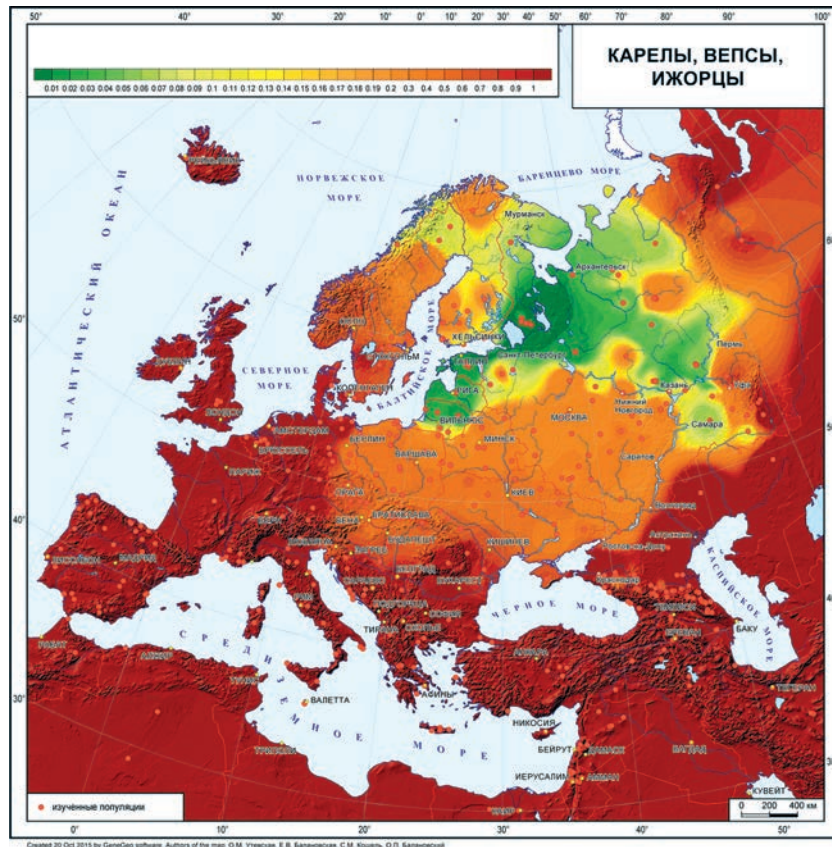


Рис. 5.2. Карта генетических расстояний от карел и вепсов (генетический ландшафт по гаплогруппам Y-хромосомы).

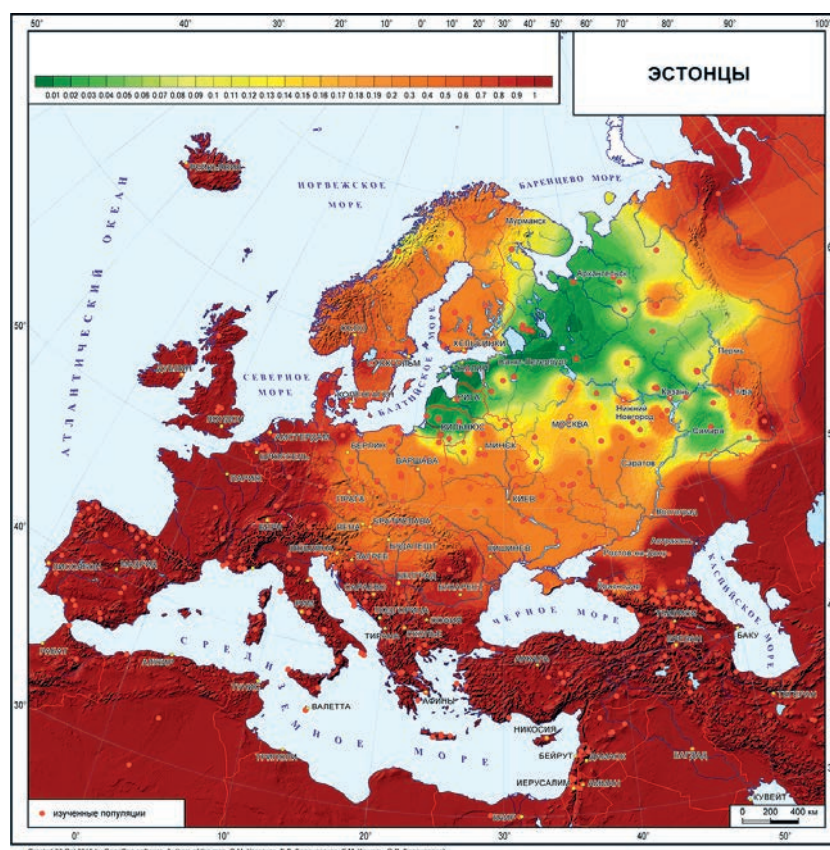


Рис. 5.3. Карта генетических расстояний от эстонцев (генетический ландшафт по гаплогруппам Y-хромосомы).

нофондов карел и вепсов (интенсивно красные тона).

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ЭСТОНЦЕВ (рис. 5.3)

Переходя к карте генетических расстояний от эстонцев (рис. 5.3), мы видим те же общие закономерности, что и на карте расстояний от карел и вепсов (рис. 5.2). Однако область генетически наиболее близких популяций, окрашенных интенсивно зелеными тонами (наименьшие генетические расстояния от эстонцев в интервале $0 < d < 0.05$), расширяется и смещается восточнее. Генетически более далекие популяции, окрашенные в светло-зеленые и желтые тона (генетические расстояния от эстонцев в интервалах $0.05 < d < 0.10$) охватывают еще больший ареал, местами заходя южнее и на правобережье Волги. В то же время северо-запад Европы – популяции Фенноскандии – генетически дальше от эстонцев, чем от карел и вепсов. Однако генетические ландшафты обеих карт (рис. 5.2 и 5.3) очень схожи, указывая на сходство генофондов всех трех народов.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ КОМИ-ЗЫРЯН (рис. 5.4 и 5.5)

Популяции коми традиционно подразделяют на две этнические общности – коми-зыряне и коми-

пермяки, хотя у них общий этноним, да и граница между их диалектами не совпадает с административной. Более южную общность составляют коми-пермяки, проживающие ныне в Пермском крае. Более северную общность – коми-зыряне, проживающие в Республике Коми (именно их часто называют просто коми). Истоки формирования коми относят ко II тыс. до н.э. в районе впадения в Волгу Оки и Камы. В ходе следующих веков общий ареал коми неуклонно расширялся на север, а под давлением новгородской колонизации смещался на восток. Коми расселялись по долинам крупных рек, ассимилируя различные группы более древнего населения – как прибалтийских финнов, так и другие уралоязычные группы, на что указывают данные топонимики.

Среди коми-зырян выделяют девять этнографических групп. Одна из них – коми-ижемцы (рис. 5.4), проживающие компактно в Ижемском районе на севере Республики Коми (в бассейне среднего течения Печоры и ее притоков) и занимающиеся – в отличие от других групп коми – кочевым оленеводством, которое они переняли от ненцев. Сложение этнографической группы коми-ижемцев относят к концу XVI века – в ее формировании приняли участие не только разные группы коми (вымские, удорские) и северные русские, но и ненцы. Основная масса коми-ижемцев относится к беломорскому антропологическому типу.

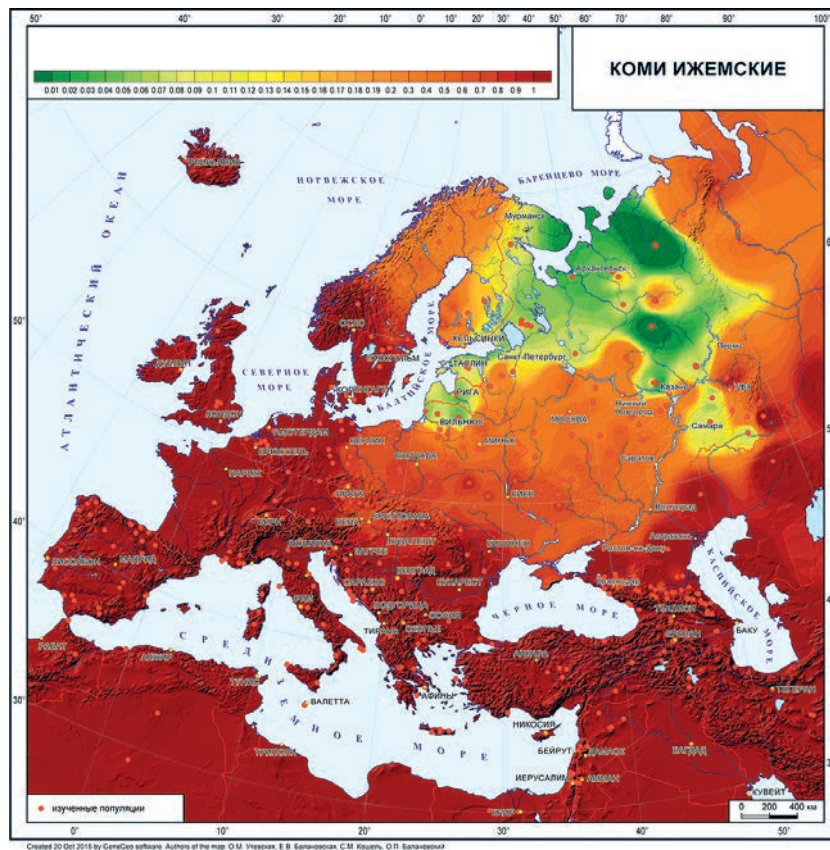


Рис. 5.4. Карта генетических расстояний от коми ижемских (генетический ландшафт по гаплогруппам Y-хромосомы).

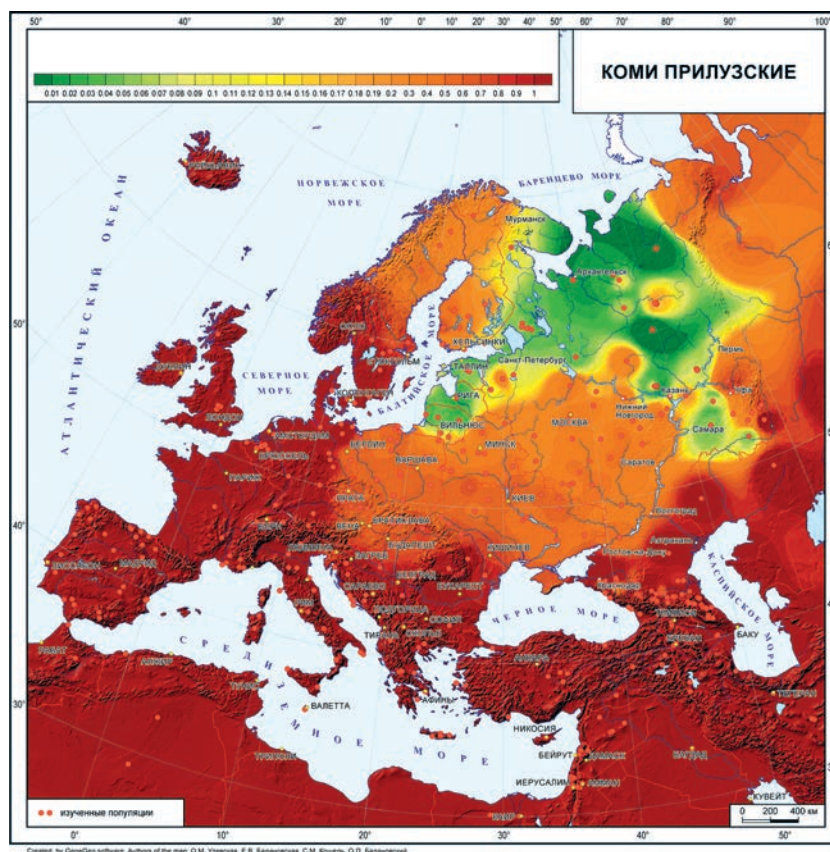


Рис. 5.5. Карта генетических расстояний от коми прилузских (генетический ландшафт по гаплогруппам Y-хромосомы).

Другая этнографическая группа – прилузских коми (рис. 5.5) – проживает на другом – южном – конце ареала коми-зырян: на самом юге Республики Коми в бассейне Лузы и в верховье Летки, где граничит уже с Пермским краем и Кировской областью.

Однако вопреки и географии, и хозяйственно-культурному типу, и адаптивному типу (ижемских коми относят к арктическому адаптивному типу), карты генетических расстояний от обеих этнографических групп коми-зырян удивительно схожи. Выделяется темно-зеленая область минимальных расстояний (наибольшего сходства) между обеими группами коми. Их разделяет популяция русских Красноборского района Архангельской области, резко отличающаяся (оранжевые тона) от них, как впрочем, и от основного массива северных русских популяций (рис. 5.8). Со всеми другими северными русскими популяциями коми обнаруживают наибольшее сходство (наиболее яркие зеленые тона на карте). Это особенно ярко видно на карте генетических расстояний от прилузских коми (рис. 5.5), отличающихся от генофонда своих южных географических соседей и генетически явно тяготеющих к северным, пусть и географически более далеким, генофондам.

Однако не будем забывать, что такая генетическая близость самой южной и северной групп коми может указывать на сохранение единства лишь Y-хромосомного генофонда коми: возможно, из иноэтничных групп брались, главным образом, жены, а приток мужских Y-хромосом из них был

невелик. Возможность «гендерной асимметрии браков» всегда необходимо учитывать, когда мы анализируем лишь одну из однородительских генетических систем – или Y-хромосому, или мтДНК.

За этим исключением – смещения наименьших генетических расстояний (ярко-зеленых) на восток и север – область генофондов, генетически близких к коми, окрашенная в светло-зеленые и желтые тона, очень схожа с ландшафтом, обнаруженным выше у карел, вепсов и эстонцев. Это невольно заставляет вспомнить работы палеоантропологов [Хартанович, 1991], указывавших, что по краниологическим данным коми-зыряне тяготеют к карелам, а не к коми-пермякам. Однако лишь будущее детальное изучение генофондов всего разнообразия популяций коми-зырян и коми-пермяков (а также своеобразных по языку коми-язьвинцев) позволит точно определить степень их генетического сходства как друг с другом, так и с другими народами Северо-Восточной Европы и Приуралья.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДАМИ ЛАТЫШЕЙ И ЛИТОВЦЕВ (рис. 5.6 и 5.7)

На выше рассмотренных четырех картах (рис. 5.2–5.5) «реперными» генофондами, от которых рассчитывались генетические расстояния, выступали популяции финно-угорских народов. Теперь перед нами карты генетических расстояний от двух

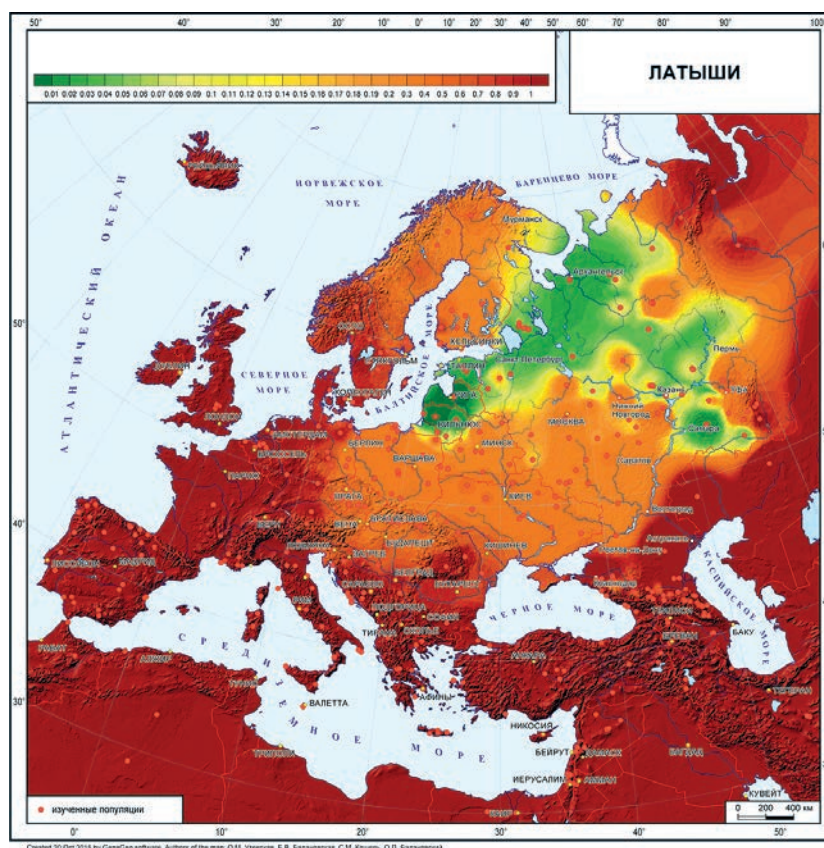


Рис. 5.6. Карта генетических расстояний от латышей (генетический ландшафт по гаплогруппам Y-хромосомы).

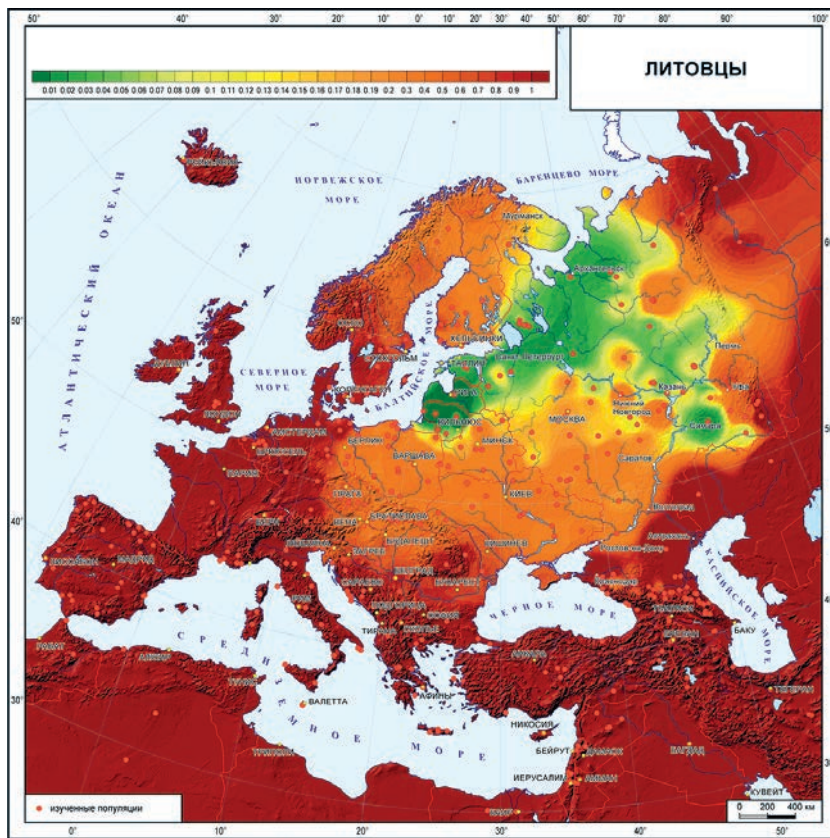


Рис. 5.7. Карта генетических расстояний от литовцев (генетический ландшафт по гаплогруппам Y-хромосомы).

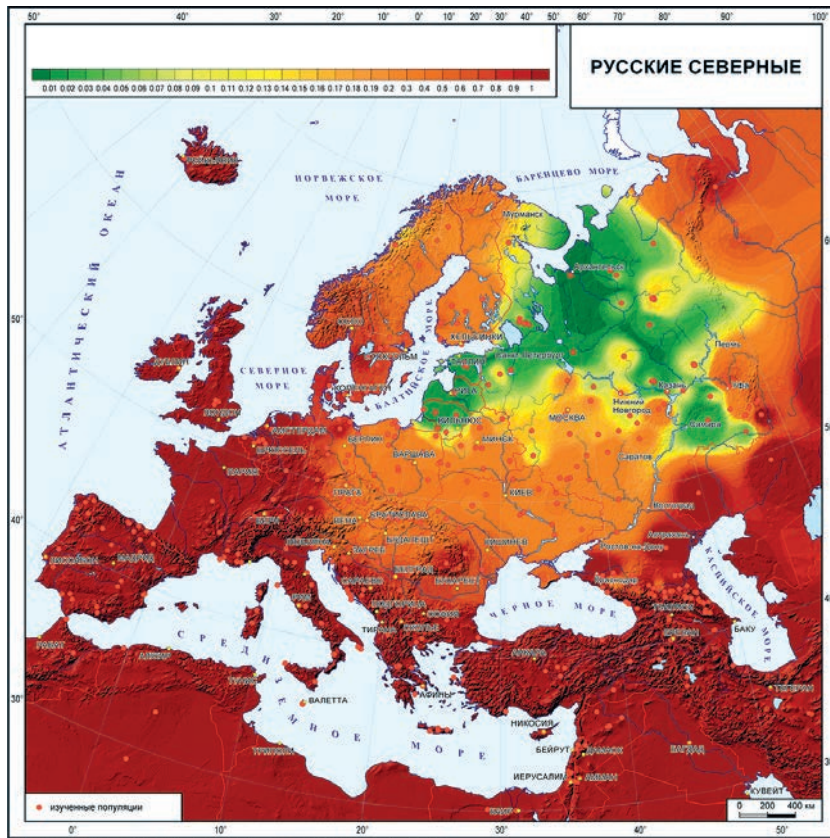


Рис. 5.8. Карта генетических расстояний от северных русских (генетический ландшафт по гаплогруппам Y-хромосомы).

балтоязычных народов – латышей (рис. 5.6) и литовцев (рис. 5.7). Лингвистически они относятся уже не к уральской семье, а к индоевропейской. Однако вопреки столь огромным лингвистическим различиям мы вновь видим все тот же генетический ландшафт, не требующий даже дополнительного описания. Ближе всего он к генетическому ландшафту соседних эстонцев (рис. 5.3). Отличие этих двух ландшафтов лишь в том, что область популяций, генетически близких к балтским народам, на северо-западе и северо-востоке максимально сужается, оставаясь широкой на юге и тем самым приближаясь к форме треугольника.

Предполагается, что носители балтских языков ранее были распространены в намного более широком ареале – от северо-востока современной Польши до верховьев Волги, бассейна Оки, среднего Днепра и Припяти. Поэтому совпадение генетических ландшафтов карел, вепсов, коми, эстонцев и латышей позволяет ставить вопрос о причинах такого совпадения. Налицо смена языковой принадлежности (или балтов, или финно-угров, или и тех, и других) при сохранении некоего общего древнего генофонда. Может быть, существовал некий более древний генофонд, о лингвистической принадлежности которого у нас нет даже гипотез, и именно он стал тем генетическим субстратом, который до сих пор задает ландшафт самых разных генофондов Северо-Восточной Европы?

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ СЕВЕРНЫХ РУССКИХ (рис. 5.8)

Эти сомнения и раздумья еще больше усиливает карта генетических расстояний от северных русских (рис. 5.8): генофонд наследников Новгородской Руси полностью повторяет все вышеописанные закономерности. Генетическое своеобразие северных русских популяций твердо установлено [Balanovsky et al., 2008; Балановская и др., 2011; Kushniarevich et al., 2015]. Но стало привычным клише связывать это своеобразие лишь с финно-угорским субстратом. Поэтому обратим внимание, что карта генетических расстояний от северных русских все же более схожа с генетическими ландшафтами балтов – латышей и литовцев, а не финноязычных народов. Это указывает, что будущие исследования палеодНК мезолитического и неолитического населения могут внести коррективы в ставшую привычной интерпретацию генетического своеобразия Русского Севера просто лишь как наследия генофонда финноязычного населения. Возможно, у нас будет возможность увидеть связь генофонда Русского Севера с балтами, которые в свою очередь унаследовали генофонд древнейшего населения перигляциальной зоны Восточной Европы.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ФИННОВ (рис. 5.9)

С этим призывом к осторожности интерпретаций согласуется и самая своеобразная карта данной серии – генетических расстояний от «самого финноязычного» народа – то есть от самих финнов (рис. 5.9). Их генетический ландшафт не похож ни на один из рассмотренных: мы вообще не видим никакого сходства с рассмотренными генофондами Северо-Восточной Европы. Область сходных значений уместается в Фенноскандии, да и то занимая лишь ее половинку: как самая восточная окраина Фенноскандии, так и огромная юго-западная часть Норвегии и Швеции оказались генетически далеки от генофонда финнов. И лишь очертания оранжевой области генетически далеких популяций (но все же не максимально удаленных генофондов) повторяют очертания зоны сходства на остальных картах этой серии.

Столь выраженное своеобразие генетического ландшафта финнов находится в противоречии с их близким лингвистическим родством с другими народами прибалтийско-финской группы (сформировавшейся исторически недавно – в I тысячелетии до н.э.) и географическим положением – финны естественным образом входят именно в регион Северо-Восточной Европы.

Традиционно считается, что своеобразие генофонда финнов (выражающееся, в том числе и в наличии особого «финского» спектра наследственных болезней) связано с тем, что популяция прошла через демографическое «бутылочное горлышко», приведшее к мощным эффектам дрейфа генов. Финны как бы оказались на периферии и финно-угорского, и скандинавского миров. Напомню, что у Андерсена при поисках чертогов Снежной королевы лапландка отсылает Герду на самый край света – к финке. Дальше уже некуда.

ЧТО ДАЛЬШЕ?

Итак, выявлен стойкий генетический ландшафт, характерный для большинства народов Северо-Восточной Европы. Но эти народы не объединяет ни принадлежность к общей языковой группе, ни принадлежность к общему региону (финны, несомненно, принадлежат к тому же региону, но их карта иная). Тогда что же их объединяет? Сохранение («консервация») генофонда древнейшего населения перигляциальной зоны Восточной Европы? Искушение выдвинуть такую гипотезу велико. Ведь даже если мы исключим из обобщенной карты генетического ландшафта генетически резко отличающихся (отдрейфовавших?) финнов, и построим карту заново по серии из семи карт (рис. 5.10) – мы получим все тот же устойчивый генетический ландшафт Северо-Восточной Евро-

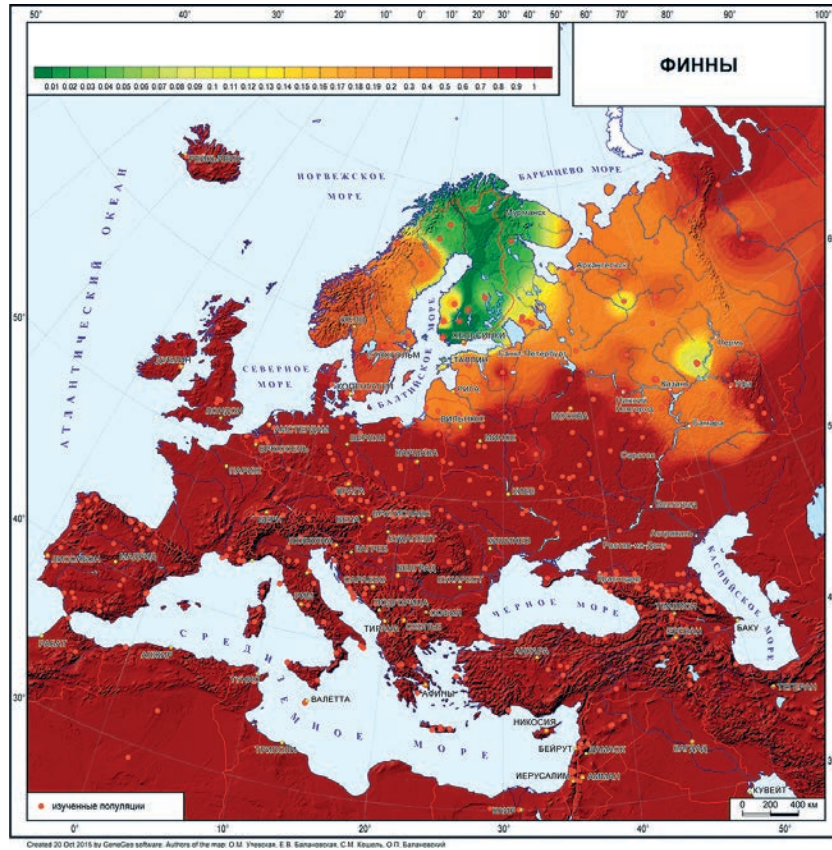


Рис. 5.9. Карта генетических расстояний от финнов (генетический ландшафт по гаплогруппам Y-хромосомы).

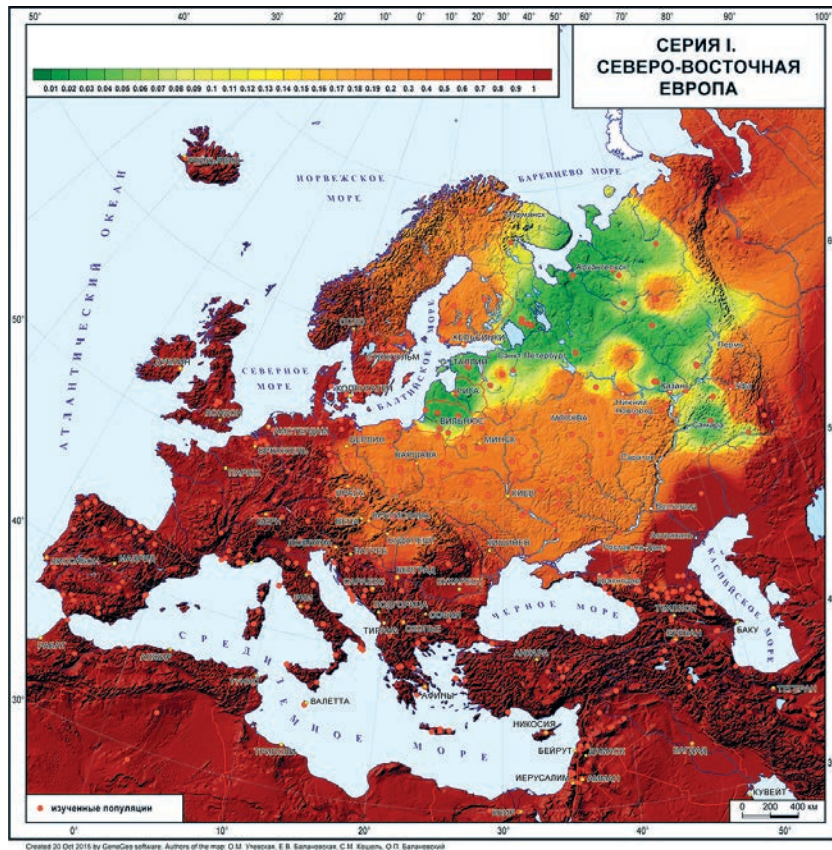


Рис. 5.10. Карта обобщенного генетического ландшафта Северо-Восточной Европы по гаплогруппам Y-хромосомы (без финнов) (Построена как средняя по семи картам генетических расстояний: от вепсов и карел, коми ижемских, коми прилузских, латышей, литовцев, северных русских, эстонцев).

пы (что и на рис. 5.1), только окрашенный в еще более яркие тона малых генетических различий. Именно его можно считать типичным, стандартным, «эталонным» генетическим ландшафтом коренного населения Северо-Восточной Европы.

Любой, даже поверхностно знакомый с географией, сразу скажет: эти карты объединяет гаплогруппа **N1c-M178**. Да, именно ее высокие частоты характерны для всех рассмотренных генофондов, и ареал этих высоких частот формирует выгнутую к северу дугу от Прибалтики до Приуралья. Но ее частота особенно велика у финнов (более половины генофонда), и своеобразие генетического ландшафта финнов объясняется во многом именно повышением частоты этой гаплогруппы. У других народов севера Восточной Европы частоты **N1c-M178** более умеренные. Но не будем забывать, что карты построены не по одной **N1c-M178**, а по данным обо всей совокупности основных европейских гаплогрупп, частоты которых значительно разнятся в пределах Северо-Восточной

Европы. Поэтому выявленные зоны сходства и их особенности определяется не одной лишь гаплогруппой **N1c**, а всем Y-хромосомным генофондом.

Но все же роль этой североевразийской гаплогруппы особенно велика. Поэтому ее углубленное изучение позволит продолжить историю, рассказанную в этом разделе. Продолжения ждать уже недолго: полногеномное исследование Y-хромосомы уже позволили выделить в евразийском ареале гаплогруппы **N1c**, по крайней мере, восемь географически приуроченных ветвей, по которым уже проведен скрининг целого ряда популяций Евразии. Как только число и спектр популяций Европы, для которых определены частоты новых ветвей гаплогруппы **N1c**, достигнет достоверного уровня для создания полноценных карт генетических расстояний, мы обновим эту серию карт, включив в спектр анализируемых гаплогрупп карты новых ветвей **N1c** и тогда, надеюсь, сумеем выделить различные миграционные потоки в генетическом ландшафте Северо-Восточной Европы.

5.2. ВОСТОЧНЫЕ И ЗАПАДНЫЕ СЛАВЯНЕ (СЕРИЯ II)

Следующая серия карт хотя и объединяет народы по принципу лингвистической общности – восточные (белорусы, русские, украинцы) и западные славяне (кашубы, поляки, словаки, сорбы, чехи), но также следует и географии, охватывая обширный географический ареал, но теперь уже не в Северо-Восточной, а в Восточной и Центральной Европе. Как и в предыдущей серии, мы не следуем строго этническому уровню при отборе «реперных точек», от которых строятся карты генетических расстояний, но, верные геногеографической логике, включаем и «субэтнические» карты – южных/центральных русских и белорусов Полесья.

Наглядный результат применения метода карт расстояний – то, что эта серия девяти карт, как и предыдущая, выявила общий паттерн геногеографического ландшафта. Этот паттерн можно назвать «западно-восточнославянским» – карта обобщенного генетического ландшафта западных и восточных славян приведена на рис. 5.11. Как и обобщенная карта первой серии, она получена в результате простых арифметических действий с матрицами карт: усреднением для каждой точки карты значений всех девяти карт в этой точке.

Единственное исключение из общего паттерна – чехи – лишь оттеняет правило, как и в предыдущей серии, было одно исключение – финны. Конечно же, каждая карта имеет и свои особенности. Однако генетическое, лингвистическое и географическое единство западных и восточных славян демонстрируется этой серией карт ярко и убедительно.

Во избежание недоразумений и поверхностных трактовок напомним международное правило популяционной генетики – принцип «трех поколений». В каждом этническом ареале исследуются только те индивиды, все бабушки и деды которых относились к этому этносу и родились в данной популяции. Казалось бы, для изучения Y-хромосомных линий достаточно знаний происхождения только деда по отцовской линии. Но показано, что учет всей родословной на три поколения позволяет охватить наиболее устойчивую часть каждого генофонда.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ЮЖНЫХ И ЦЕНТРАЛЬНЫХ РУССКИХ ПОПУЛЯЦИЙ (рис. 5.12)

Для краткости анализируемый «не-северный» массив русских популяций будем условно именовать «южнорусским» генофондом, помня, что это название соответствует не привычному административному делению на округа и зоны, а дихотомическому делению русского генофонда на «север» и «не-север». При этом «не север» включает и центр, и восток, и запад исторического ареала русского народа. Принципы выделения этого ареала даны в книге «Русский генофонд на Русской равнине» [Балановская, Балановский, 2007].

Карта (рис. 5.12) сразу обнаруживает две самые яркие особенности генетического ландшафта «южнорусских» популяций.

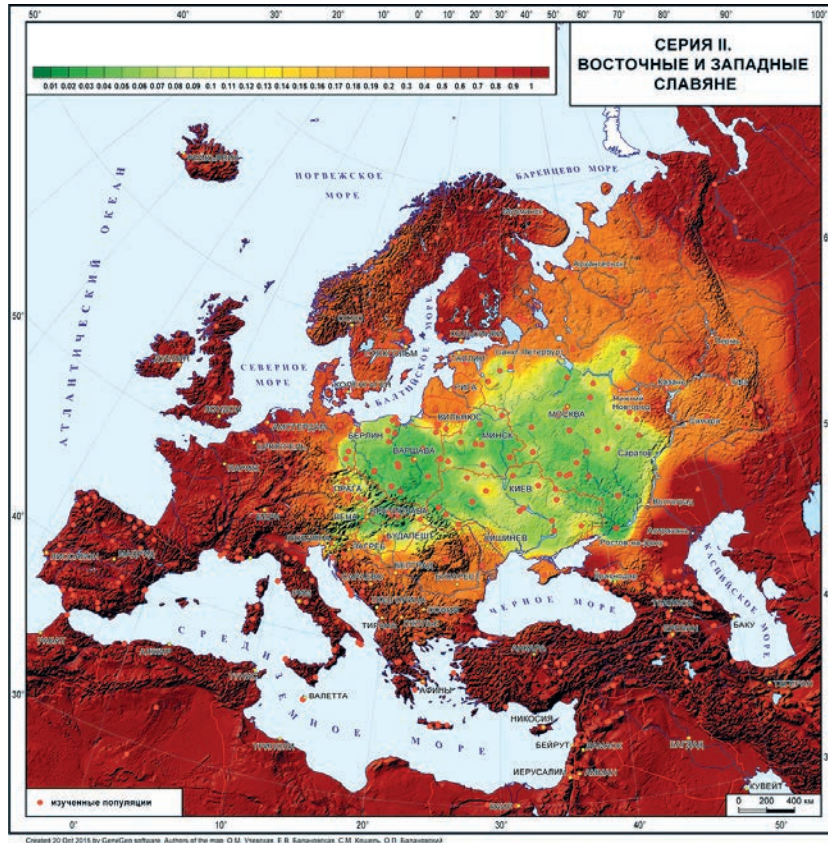


Рис. 5.11. Карта обобщенного генетического ландшафта восточных и западных славян по гаплогруппам Y-хромосомы. (Построена как средняя по девяти картам генетических расстояний: от белорусов, белорусов Полесья, кашубов, поляков, русских «южных», словаков, сорбов, украинцев, чехов).

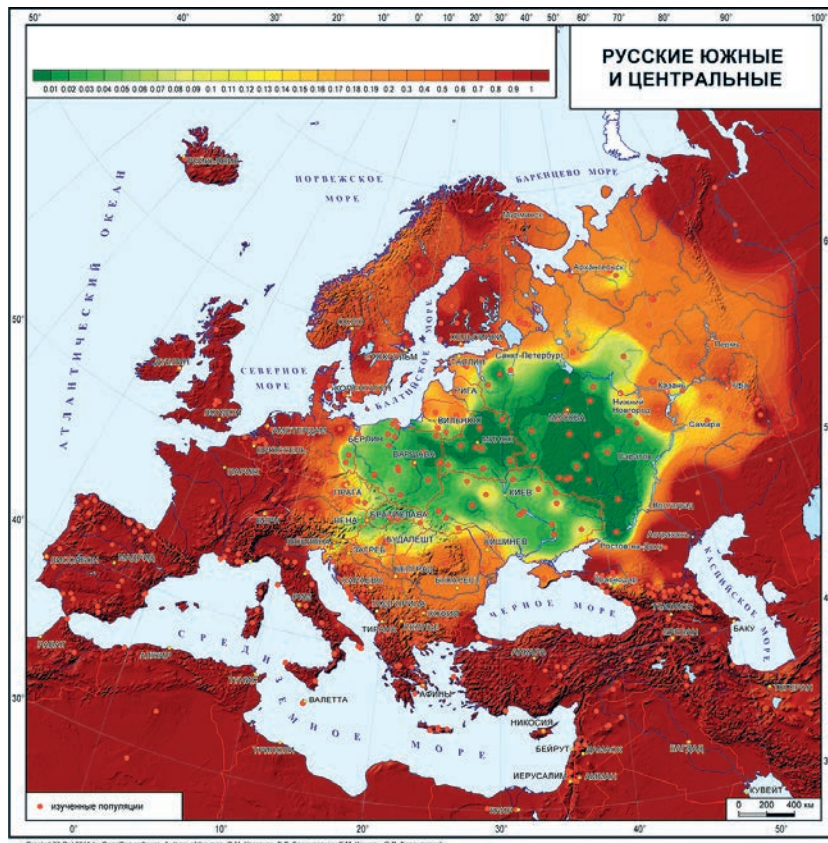


Рис. 5.12. Карта генетических расстояний от южных и центральных русских популяций (генетический ландшафт по гаплогруппам Y-хромосомы).

Во-первых, он является почти полной альтернативой генетическому ландшафту Русского Севера (рис. 5.8): ареал популяций, сходных с «южно-русским» генофондом, иной. «Северный» и «южный» русские ландшафты почти не перекрываются. Лишь небольшой «рукав» вдоль верхней Волги является как бы переходной зоной между ними и отчасти принадлежит обоим половинкам русского генофонда.

Во-вторых, (рис. 5.12) ареал популяций, генетически сходных с «южно-русским» генофондом, практически равномерно охватывает значительную часть не только Восточной, но и Центральной Европы. Приводит в изумление, что популяции, наиболее генетически близкие к «южно-русским» (минимальные расстояния в интервалах $0 < d < 0.05$, окрашенные зелеными тонами), охватывают и Беларусь, и Украину, и Польшу, и восточную Германию, и Словакию. Если попытаться все же ранжировать степень сходства, то практически не отличимыми от русского генофонда (интенсивно темно-зеленые тона) оказываются популяции Беларуси и восточной Польши, а генофонды Западной Польши, Украины и Словакии окажутся чуть дальше. Но, конечно же, при интерпретации крайне важно учитывать масштаб генетических различий: по сравнению со всеми другими окружающими генофондами Европы, окрашенными в оранжевые тона больших генетических расстояний, эти различия

между разными западно- и восточнославянскими генофондами несущественны.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ БЕЛОРУСОВ (рис. 5.13 и 5.14)

Отличить генетический ландшафт белорусского генофонда (рис. 5.13) от «южно-русского» (рис. 5.12) практически невозможно: они почти идентичны. Можно лишь заметить, что зеленые области генетического сходства с белорусами еще более равномерно покрывают практически весь ареал «западно-восточнославянского» ландшафта – и Украину, и основную часть исторического русского ареала, и Польшу, и восточную Германию, и Словакию.

Это заставляет выдвинуть неожиданную гипотезу, что именно эти два генофонда – белорусский и русский – наиболее полно воспроизводят общие черты единства Y-хромосомных генофондов западных и восточных славян. На других картах «западно-восточнославянской» серии мы увидим те же основные очертания ареала генетического сходства, но области наибольшего сходства, окрашенные насыщенно-зелеными тонами, наиболее обширны и наиболее равномерно покрывают этот ареал только на этих двух картах генетических ландшафтов – русском и белорусском.

При этом если мы рассмотрим часть белорусского генофонда – популяции Полесья (рис. 5.14),

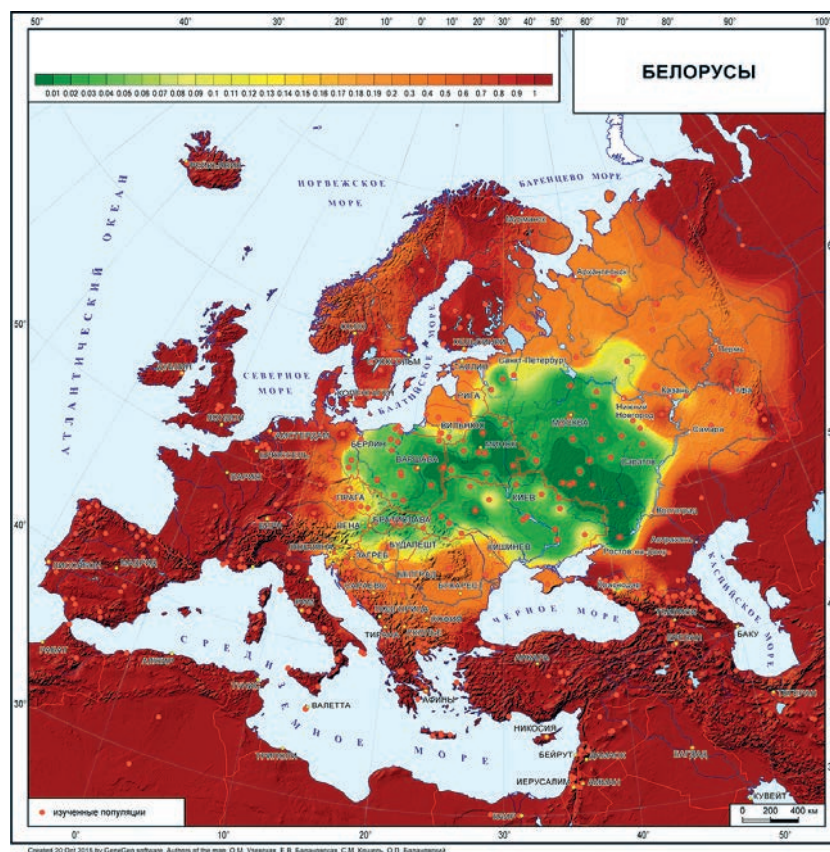


Рис. 5.13. Карта генетических расстояний от белорусов (генетический ландшафт по гаплогруппам Y-хромосомы).

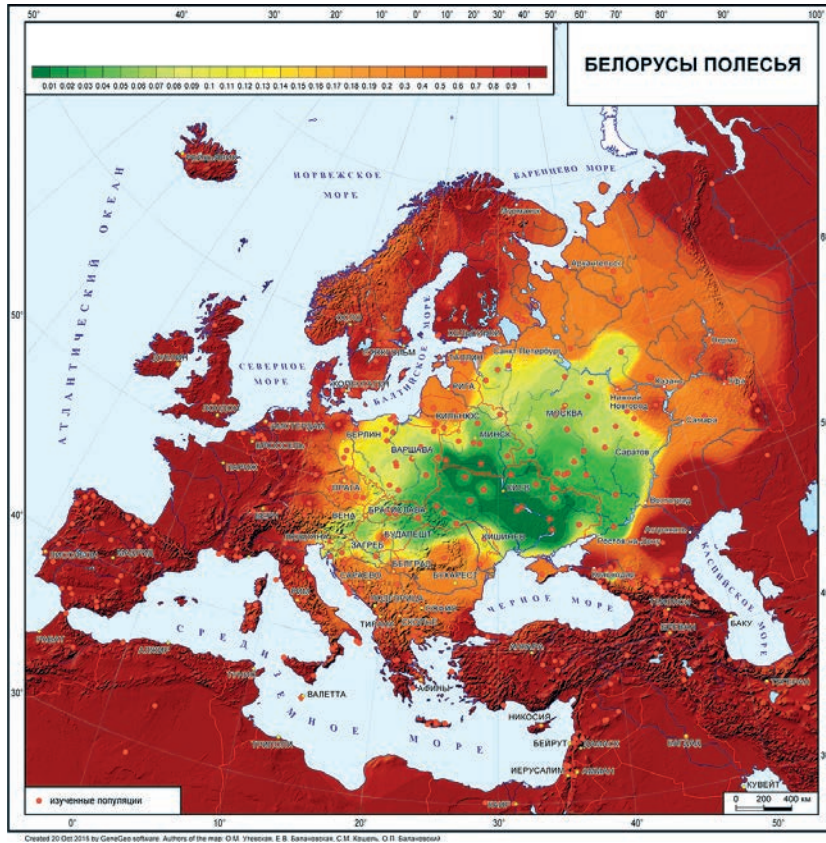


Рис. 5.14. Карта генетических расстояний от белорусов Полесья (генетический ландшафт по гаплогруппам Y-хромосомы).

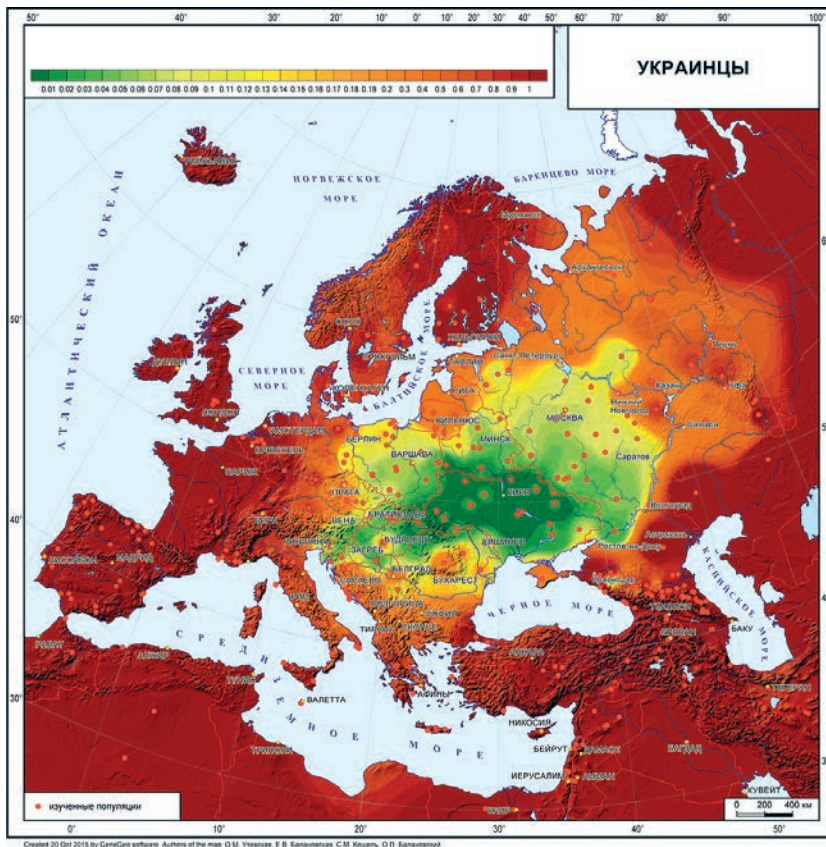


Рис. 5.15. Карта генетических расстояний от украинцев (генетический ландшафт по гаплогруппам Y-хромосомы).

то увидим, что их генофонд сохраняет историческое сходство с популяциями Украины, поскольку паттерн этой карты сходен с паттерном карты генетических расстояний от украинцев (рис. 5.15). Однако вновь напомним о генетическом масштабе – все остальные популяции Беларуси и юго-западной части “южнорусского” ареала, также окрашенные в зеленые тона, лишь чуть-чуть больше отличаются от населения Полесья, чем генофонд украинцев.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ УКРАИНЦЕВ (рис. 5.15)

Не удивительно, что на карте генетических расстояний от украинцев (рис. 5.15), мы обнаруживаем тот же самый общий «западно-восточнославянский» ландшафт. Отличие от других восточнославянских ландшафтов лишь в том, что темно-зеленые области минимальных расстояний концентрируются теперь главным образом на Украине, в Словакии и в белорусском Полесье. Это различие между генетическими ландшафтами украинцев, белорусов и русских позволяет предположить, что Y-хромосомный генофонд украинцев представляет собой более локальный вариант генетической общности западных и восточных славян.

Из характерных деталей генетического ландшафта украинцев важно отметить, что, несмотря на то, что для Украины имеются выборки для всех историко-этнографических областей, в масштабе Европы мы не видим генетических различий между ними: даже генофонд западной Украины неотличим от других ее частей.

Другая характерная деталь генетического ландшафта украинцев – юго-западный зеленый рукав сходных популяций, мало заметный на карте расстояний от русских, здесь более явно тянется через Румынию и Хорватию к Адриатике.

Во избежание каких-либо односторонних трактовок, напомним еще раз и о том, что на каждой территории в выборку, до которой рассчитаны генетические расстояния, включались только индивиды, обе бабушки и деда которых относили себя к коренному народу этой территории. Поэтому население Крыма, в соответствии с правилами популяционной генетики, представлено только наиболее коренным населением – крымскими татарами. Они были изучены содружеством украинских и российских генетиков при заинтересованной поддержке разных организаций крымских татар задолго до того, как Крым стал объектом мирового внимания. Ни русские, ни украинские популяции – в полном соответствии с правилами геногеографии – в анализируемых генетических данных по Крыму не фигурируют. Поэтому неудивительно, что территория Крыма, представленная крымскими

татами, обнаруживает отличия от генофондов украинцев, а обширные территории, представленные украинцами, словаками, поляками, белорусами, южными и центральными русскими обнаруживают высокое генетическое сходство с украинцами.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДАМИ ПОЛЯКОВ (рис. 5.16) И КАШУБОВ (рис. 5.17)

Ландшафт генетических расстояний от поляков вновь полностью вторит общему «западно-восточнославянскому» ландшафту, хотя и имеет три характерные черты.

Во-первых, области, наиболее воспроизводящие «среднестатистический» генофонд поляков (окрашенные в насыщенные темно-зеленые тона) слегка сдвигаются на запад: они более характерны для популяций западной половины Польши, соседних популяций Словакии и наиболее восточных областей Германии, чем для восточных окраин Польши.

Во-вторых, карта опровергает шаблонное представление об особенном сходстве генофондов Польши и Украины.

В-третьих, генофонд поляков обнаруживает яркое сходство с огромным массивом белорусских и русских популяций – это сходство больше, чем с генофондами географических соседей – чехов и украинцев.

Генофонд кашубов (рис. 5.17) еще ярче проявляет все эти три черты, несмотря на относительно небольшую численность: кашубов в Польше около 300 тыс. человек, в то время как поляков – в сто раз больше (36 млн.). Кашубы – это субэтническая группа поляков, компактно проживающая на севере Польши. Как предполагают, она восходит к древнеславянскому племени поморян. Ее современный генофонд обнаруживает максимальное генетическое сходство с популяциями вдоль южного побережья Балтийского моря (в восточной Германии, центре и на востоке Польши). И в то же время сохраняет память о генетическом родстве с поляками, словаками и обширным массивом белорусских и географически далеких русских популяций.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ СЛОВАКОВ (рис. 5.18)

В отличие от трех предыдущих карт (от генофондов украинцев, поляков и кашубов), генетические расстояния от словаков равномерно покрывают зеленым покрывалом всю зону «западно-восточнославянского» сходства, не делая различий между соседними украинцами и далекими белорусскими и русскими популяциями. В этом они обнаруживают неожиданное сходство с генетическим ландшафтом белорусов и русских (рис. 5.13 и 5.12).

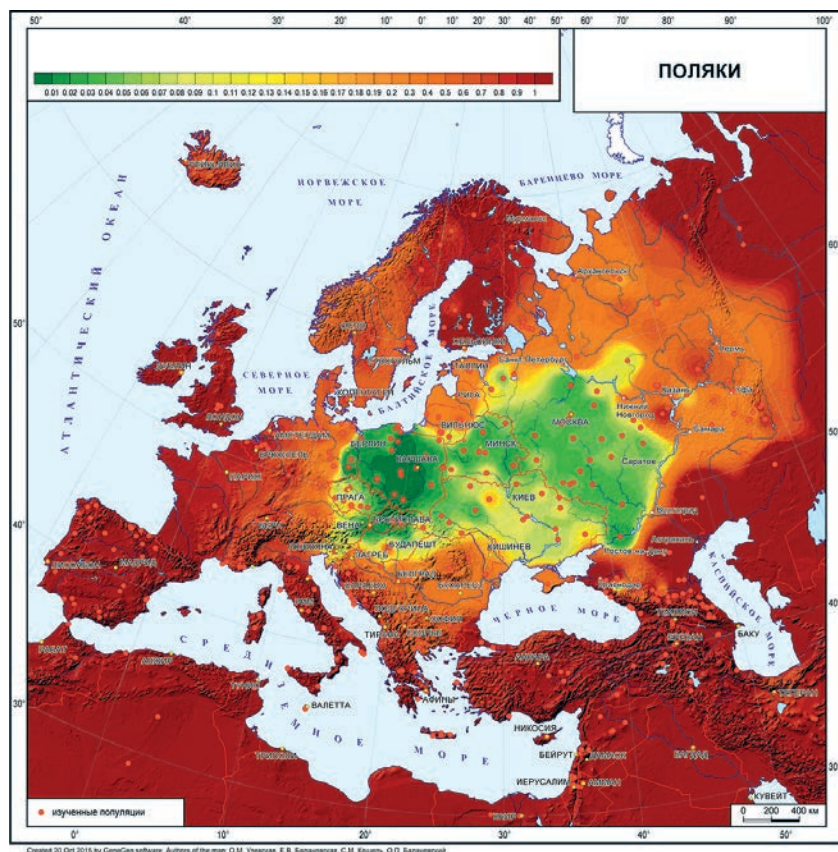


Рис. 5.16. Карта генетических расстояний от поляков (генетический ландшафт по гаплогруппам Y-хромосомы).

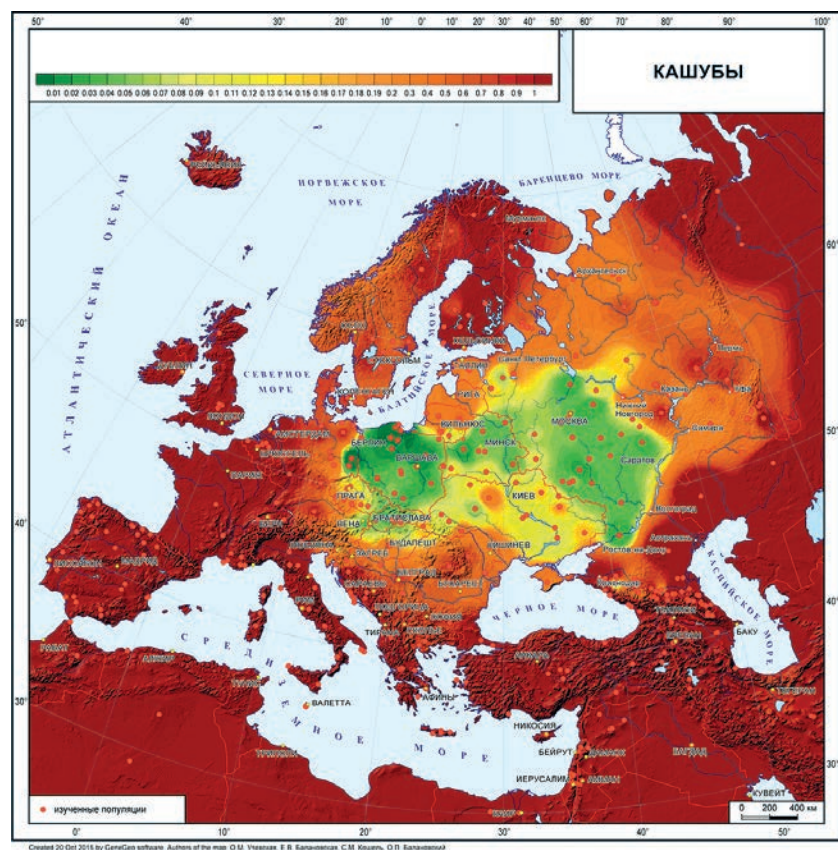


Рис. 5.17. Карта генетических расстояний от кашубов (генетический ландшафт по гаплогруппам Y-хромосомы).

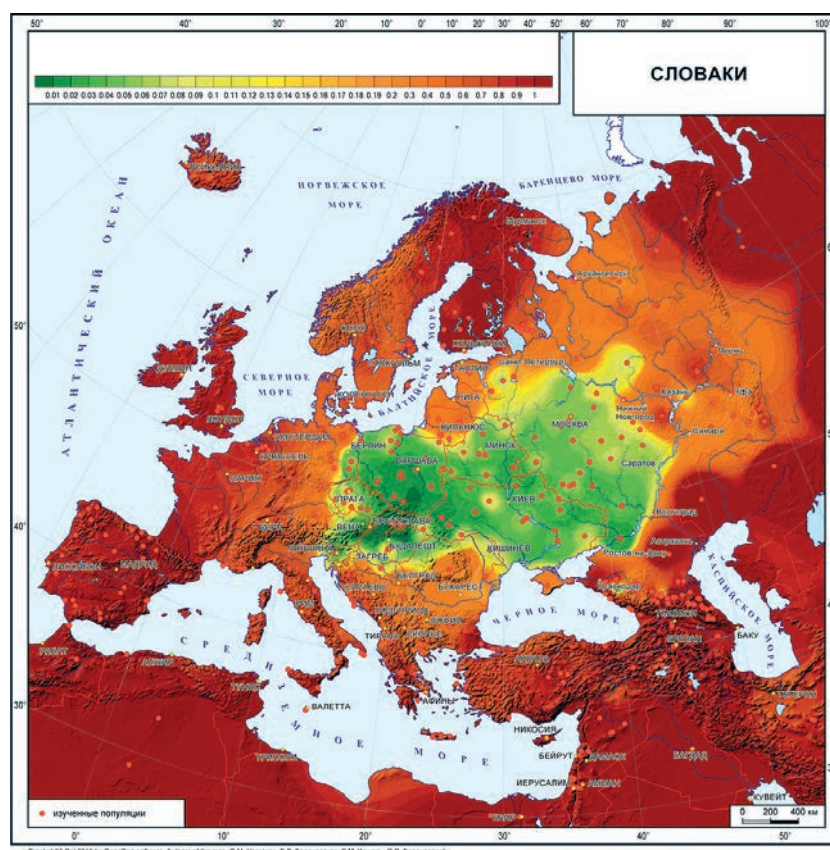


Рис. 5.18. Карта генетических расстояний от словаков (генетический ландшафт по гаплогруппам Y-хромосомы).

Среди особых черт ландшафта генофонда словаков – продвижение на юг зоны генетически сходных с ними значений. В нее попали Молдова, самый север Румынии, вся Венгрия и часть Хорватии, так что к Адриатике тянется уже не узкий коридор зеленых значений, а широкий тракт.

Наиболее близки генофонду словаков близлежащие области Польши и Венгрии – как ни удивительно, но они оказались генетически ближе к словакам, чем генофонд чехов.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ЧЕХОВ (рис. 5.19)

Генофонд чехов – единственное исключение из «западно-восточнославянского» паттерна среди всех западных и восточных славян. Их генетический ландшафт (рис. 5.19) убедительно говорит об особости чешского генофонда. Область наибольшего генетического сходства находится только в самой Чехии, а светло-зеленые тона умеренного сходства лишь небольшими отдельными языками заходят в соседние области Польши, Словакии, Словении, Венгрии, Австрии и Германии.

Особость генофонда чехов подтверждается и диаграммами многомерного шкалирования (главы 2 и 6), причем не только по Y-хромосоме, но и по широкогеномным аутосомным панелям. Можно предполагать, конечно, что поскольку в течение многих

веков история чехов была переплетена с историей германских государств, то чехи могли быть онемечены – или наоборот, немцы могли ассимилировать и включить в свой генофонд те славянские племена, которые генетически родственны чехам. Но этой гипотезе противоречит, например, то, что генетический ландшафт сорбов (рис. 5.20) – маленького анклава славян среди немцев – совсем иной, чем у чехов, и вновь воспроизводит общий «западно-восточнославянский» генетический ландшафт. Конечно, теоретически можно возвести особость генофонда чехов к наследии более древних времен (кельтов, обитавших в этом ареале в железном веке, или к народам эпохи бронзы). Но здесь свое слово смогут сказать только будущие исследования древней ДНК, если они будут проведены по обширным выборкам и с подробным покрытием территории Европы.

ГЕНЕТИЧЕСКИЙ ЛАНДШАФТ «СЕВЕРНЫХ СЛАВЯН»

О, Запад есть Запад, Восток есть Восток,
не встретиться им никогда,
Пока будут Небо с Землей таковы,
какими их Бог создал.
Р. Киплинг (перевод К. Филатова)

Переходя от карты к карте этой серии, мы убеждаемся, что реально существует общий генофонд

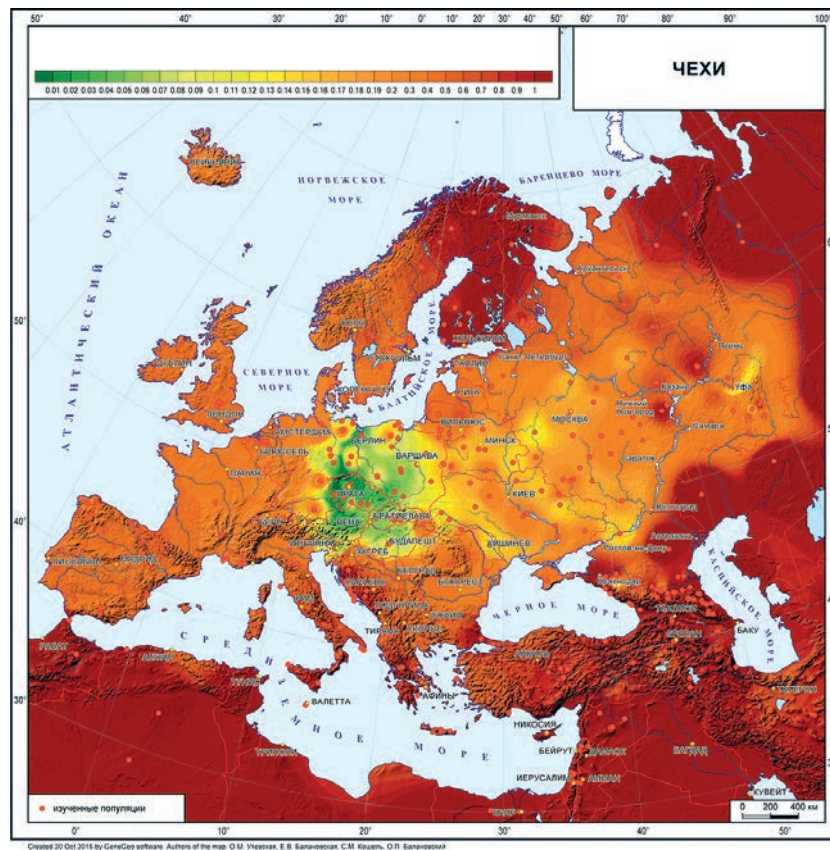


Рис. 5.19. Карта генетических расстояний от чехов (генетический ландшафт по гаплогруппам Y-хромосомы).

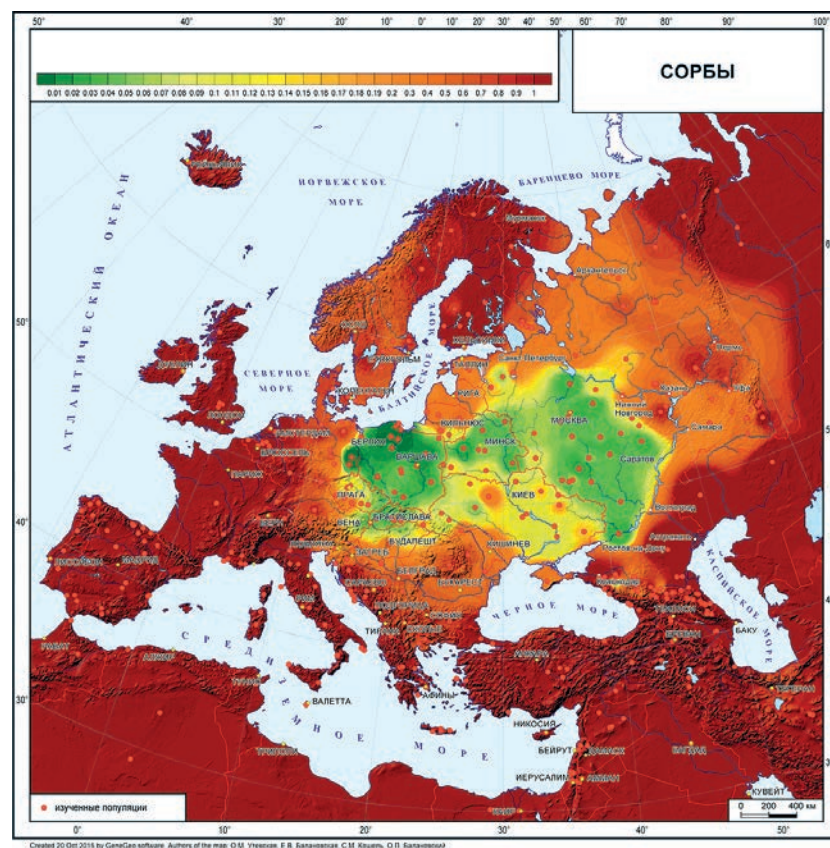


Рис. 5.20. Карта генетических расстояний от сорбов (генетический ландшафт по гаплогруппам Y-хромосомы).

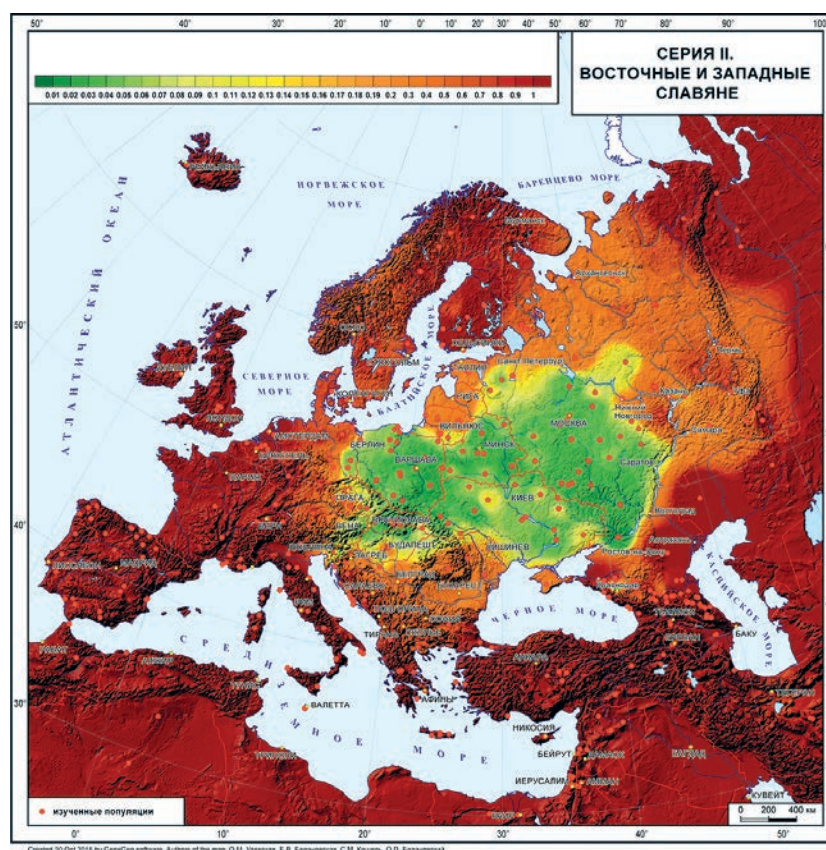


Рис. 5.21. Карта обобщенного генетического ландшафта «северных славян» по гаплогруппам Y-хромосомы (Построена как средняя по картам генетических расстояний от белорусов, кашубов, поляков, словаков, сорбов, украинцев, южных и центральных русских).

западных и восточных славян с четко очерченным ареалом его распространения. Однако термин «западно-восточнославянский» генетический ландшафт не просто неуклюж и не изящен, он еще и вызывает недоумение – нас уже давно приучили к мысли, что «не встретиться им никогда». Тем более что серия V – карт генетических расстояний от южных славян – демонстрирует, что и южные славяне также обладают своим общим генетическим ландшафтом. Это позволяет в дополнение к трехчленному лингвистическому делению (южные, западные и восточные) ввести двучленное генетическое деление – генофонд южных славян и генофонд северных славян. Конечно, задачу осложняет то, «самые-самые северные славяне» – население Русского Севера – в этот ландшафт не входят,

а образуют вместе со своими географическими соседями свой собственный (см. раздел 5.1.) генетический ландшафт Северо-Восточной Европы. Но если считать, что исключения лишь подтверждают правила, то мы можем создать карту генетического ландшафта «северных славян» (рис. 5.21). В нем исключены исключения – он построен по картам генетических расстояний от всех популяций северных славян (западных и восточных по лингвистическому делению), кроме популяций Русского Севера и чехов. Благодаря этому мы получаем наиболее устойчивый генетический ландшафт «северных славян», который можно считать типичным, стандартным, «эталонным» генетическим ландшафтом основного массива популяций западных и восточных славян.

5.3. НЕСЛАВЯНСКИЕ НАРОДЫ ВОСТОЧНОЙ ЕВРОПЫ (СЕРИЯ III)

Но прежде, чем перейти к южным славянам и другим балканским народам, следует завершить обзор народов Восточной Европы. Не имея возможности анализировать все этносы, рассмотрим шесть карт генетических расстояний от народов Волго-Уральского региона. При этом ограничимся попу-

ляциями лишь четырех соседствующих друг с другом республик, пересекающих этот регион единой полосой с востока на запад: Башкортостан, Татарстан, Чувашия и Мордовия. Этого оказывается достаточно, чтобы увидеть всю сложность и загадочность генофонда восточного фланга Европы.

Каждая из первых двух серий карт обнаружила свой собственный генетический ландшафт, охватывающий обширные пространства Европы – ландшафт Северо-Восточной Европы (раздел 5.1) и ландшафт «северных славян» (раздел 5.2). Причем в каждой из серий все карты за единственным исключением строго следовали общей модели. Однако третья серия – Уральская – несмотря на много меньший охваченный ареал, обнаруживает совершенно разные паттерны для каждого из народов. Более того – генетические ландшафты всех шести рассмотренных этнических групп резко различны даже по размеру ареала генетически сходных с ними популяций – от сродства с популяциями половины Восточной Европы до очень локального варианта. Поэтому общую характеристику ландшафта этой серии можно обозначить как «многообразии восточного фланга Европы».

Все же приведем этот обобщенный ландшафт данной серии (рис. 5.22). Он ничем не сможет помочь нам отслеживать различия каждого из генофондов от общей картины. Мы видим на обобщенном ландшафте практически лишь две градации: а) генетически далекие популяции (оранжевые тона), занимающие огромный ареал Уральского региона, Восточной и Центральной Европы; б) совершенно генетически чуждые популяции остальной Европы

и Кавказа. Эта карта дает понять, насколько генетически своеобразна каждая из шести рассмотренных этнических групп, хотя они и являются географическими соседями на относительно небольшом пространстве Поволжья и Южного Урала. А рассмотрение карт этой серии подчеркнет, что генофонд Уральского региона требует особо детального изучения и вдумчивого подхода.

На всех генетических ландшафтах серии самое нижнее течение Волги окрашено в тона генетически очень далеких популяций. Но на них пока не отражены генофонды астраханских ногайцев и астраханских татар: в момент написания книги их генотипирование нами еще не закончено. Но оно явно внесет коррективы в генетические ландшафты данной серии.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ БАШКИР (рис. 5.23)

По данным переписи 2010 г. в России 1.6 млн. башкир, из них 1.2 млн. – в Башкортостане. Однако, несмотря на столь большую численность, ареал популяций, сходных с генофондом башкир, очерчен очень четко: он тянется по Уральским горам и их западным предгорьям вдоль Камы, занимая только ее левобережье, и далее продолжает

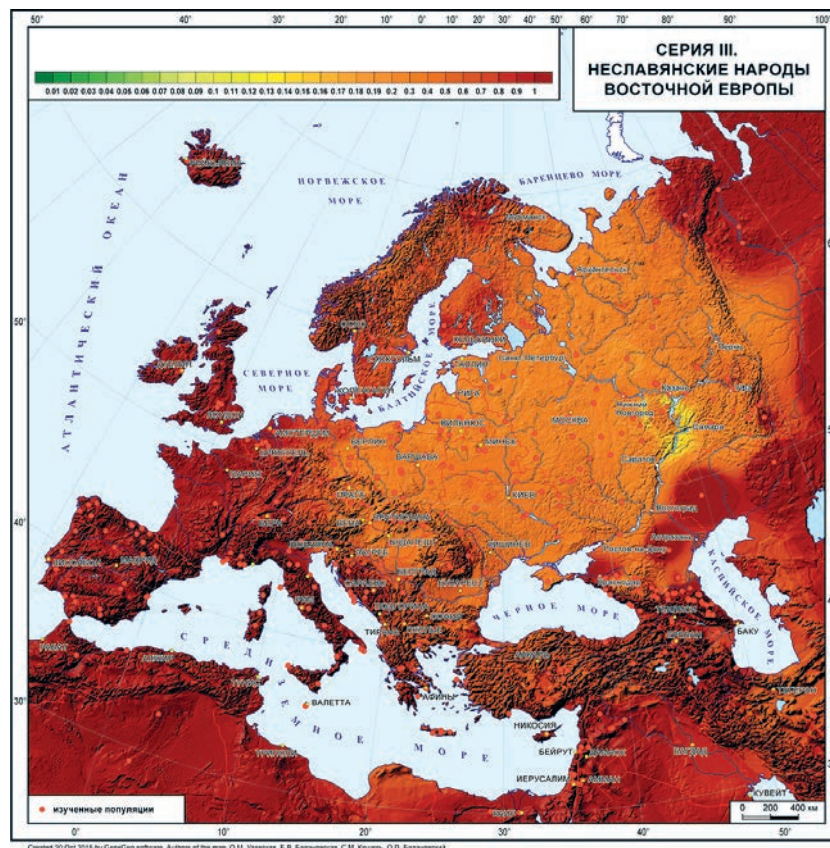


Рис. 5.22. Карта обобщенного генетического ландшафта неславянских народов Восточной Европы по гаплогруппам Y-хромосомы (Построена как средняя по картам генетических расстояний от башкир, казанских татар, мишарей, мокши, чувашей, эрзи).

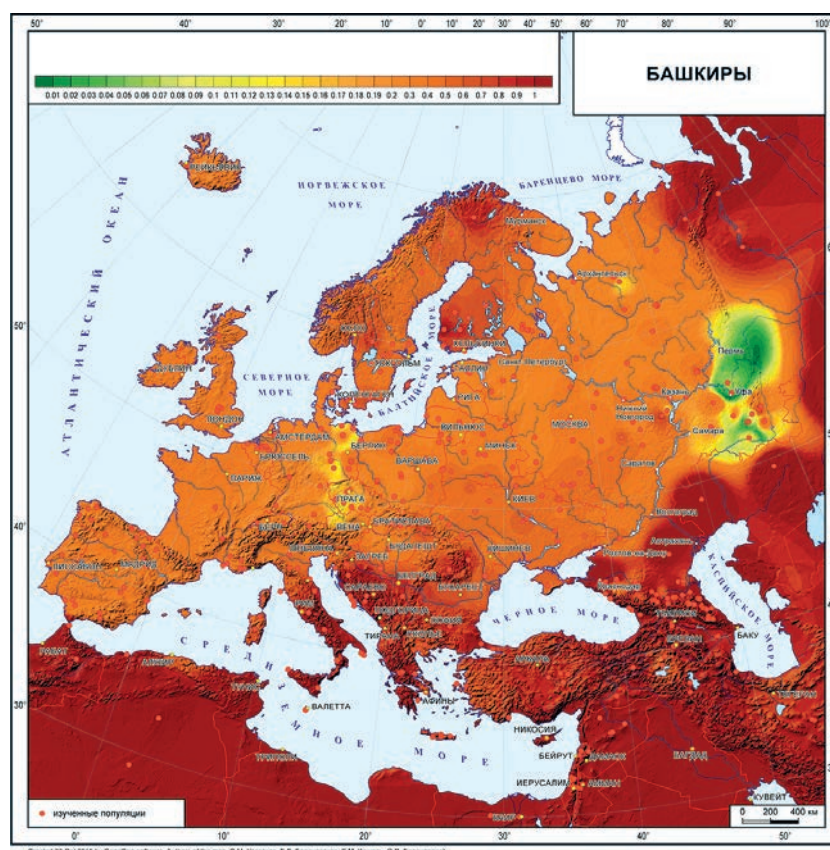


Рис. 5.23. Карта генетических расстояний от башкир (генетический ландшафт по гаплогруппам Y-хромосомы).

путь на юг вдоль этого же меридиана, доходя до Казахстана. При этом южная часть ареала более широка и размыта, протягиваясь на запад почти до Волги, а на востоке вновь до ареала казахов. Максимальные значения генетических расстояний (темно-красные тона) окружают ареал башкир с востока, севера и юга, следуя изгибу Оби и ареалу казахов. На западе же от башкир – от Волги и нижнего Дона до крайнего европейского севера находятся оранжевые тона популяций, генофонды которых, конечно, далеки от башкир, но все же не настолько, как Сибирь и Казахстан.

Детальное изучение генофонда башкир и окружающих народов, проводимое сейчас нашим коллективом в содружестве с уфимскими коллегами (под руководством к.и.н. Ю.М. Юсупова) позволит вскоре значительно уточнить эту картину: уже прослеживаются важные генетические связи с рядом народов евразийской степи.

Из характерных черт генетического ландшафта отметим явные различия между генофондами башкир и татар. Впрочем, планируемое нами детальное изучение генофонда населения Татарстана позволит определить более точно и особенности каждого из этих двух генофондов, и их взаимодействие.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ КАЗАНСКИХ ТАТАР (рис. 5.24)

По данным переписи 2010 г. в Татарстане проживает 2 млн. татар. Но поскольку эти данные включают и мишарей, и кряшен, и тептярей, то численность казанских татар Татарстана, видимо, сопоставима с численностью башкир Башкортостана. Однако казанские татары характеризуются совершенно иным генетическим ландшафтом, чем башкиры – ареал популяций, генетически сходных с казанскими татарами, обширен и весь обращен к Северо-Восточной Европе.

Хотя область популяций, максимально сходных с казанскими татарами (темно-зеленые тона, показывающие минимальные генетические расстояния $0 < d < 0.05$) невелика, ареал популяций, окрашенных в желто-зеленые тона небольших генетических расстояний ($0.05 < d < 0.10$), чрезвычайно обширен (рис. 5.24). Этот ландшафт почти полностью вторит ландшафту Северо-Восточной Европы (рис. 5.10), детально описанному в первой серии карт (раздел 5.1.). Вся северная и западная часть ареала сходных популяций практически та же (за исключением разве что побережья Баренцева моря) – она включает не только западных

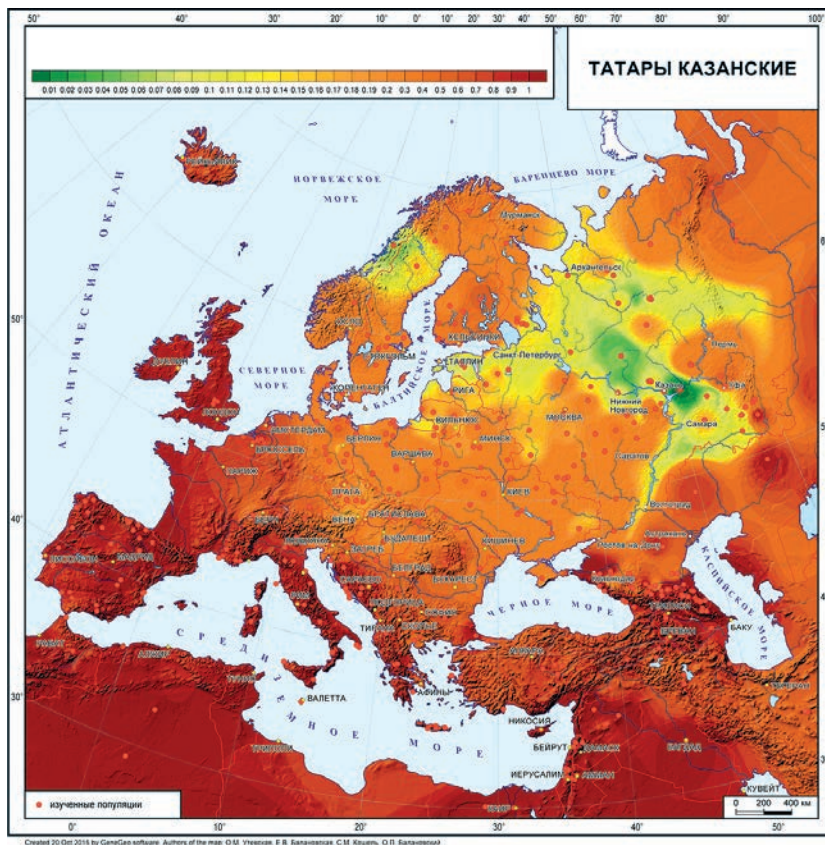


Рис. 5.24. Карта генетических расстояний от казанских татар (генетический ландшафт по гаплогруппам Y-хромосомы).

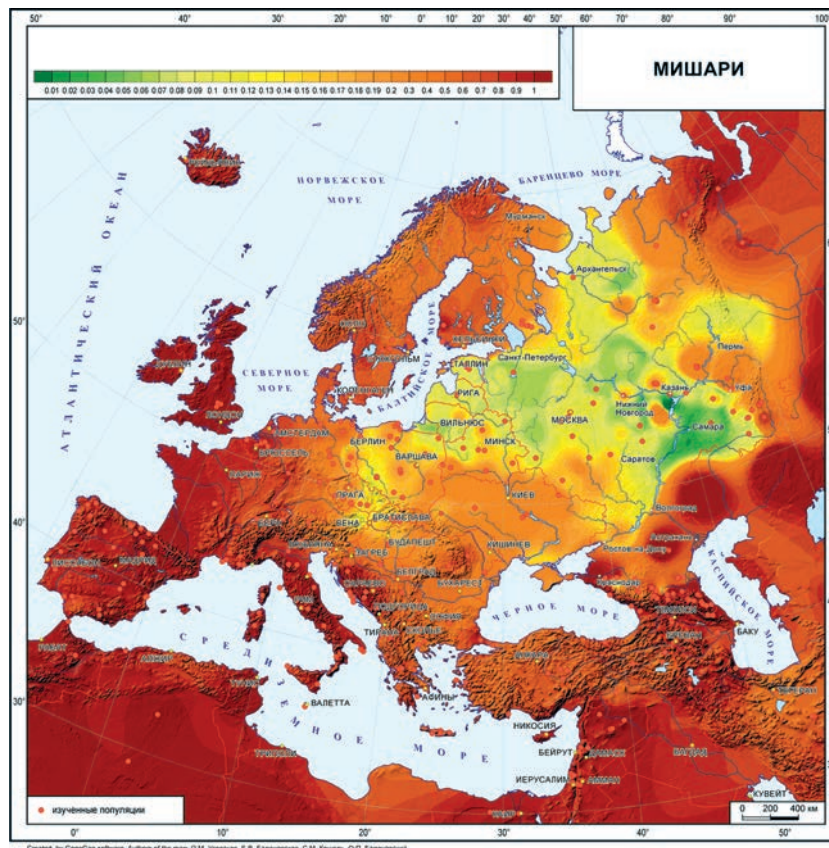


Рис. 5.25. Карта генетических расстояний от мишарей (генетический ландшафт по гаплогруппам Y-хромосомы).

финнов и балтов, но и запад Фенноскандии (рис. 5.10). На юге этого ареала границей вновь служит Волга.

Но есть и разница. В отличие от народов Северо-Восточной Европы, область сходных генофондов охватывает Татарстан и часть популяций Башкортостана, указывая на наличие общего северо-европейского субстрата и у них. Если бы надо было дать выразительное название наиболее характерным чертам этой карты, то ее можно было бы назвать ландшафтом «левого бережья Волги» – так как Волга практически на всем своем течении ограничивает ареал генофондов, сыгравших наиболее важную роль в сложении генофонда казанских татар. И приходится отметить, что Y-хромосомный генетический ландшафт не подтверждает ни болгарскую, ни золотоордынскую версии этногенеза казанских татар, а вместо этого подчеркивает мощный северо-европейский генетический субстрат в их генофонде.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ МИШАРЕЙ (рис. 5.25)

Численность мишарей указать не просто, так как в последних переписях они не выделялись из совокупности татар, но в переписи 1926 года их численность оценивалась около 200 тыс. Из их обширного ареала к настоящему времени по Y-хромосоме изучены только популяции мишарей Татарстана. И хотя в Татарстане мишари явно уступают по численности казанским татарам, однако они обнаруживают новый, ранее еще невиданный, генетический ландшафт. Поражает его обширность – зона генетического сходства (минимальных и небольших расстояний) простирается от южного Урала до Белого и Балтийского морей (рис. 5.25). Но самая удивительная черта – это то, что здесь Волга не служит границей, как мы это видели практически на всех картах. Напротив, Волга является центром – во всем ее течении области генетически сходных популяций расположены на ее обоих берегах. Генетический ландшафт мишарей объединяет карты первой (рис. 5.10) и второй (рис. 5.21) серий. Но при этом генетический ландшафт мишарей ограничен с запада и с юга – в отличие от карт второй серии, зона генетического сходства в западном направлении доходит лишь до Прибалтики, но не покрывает ареал западных славян, не заходит она ни к украинцам, ни на юг России, но тянется широкой полосой светло-зеленых интервалов от Прибалтики до Поволжья и Южного Урала.

Неожиданность генетического ландшафта мишарей позволяет выдвинуть рабочую гипотезу, требующую тщательной проверки. Можно допустить, что «левобережно-волжская» часть ландшафта, практически повторяющая предыдущую

карту (рис. 5.24) – генетического ландшафта казанских татар – связана с потоком генов к мишарям от более многочисленного генофонда казанских татар. В этом случае «правобережно-волжскую» часть можно считать «собственным» генетическим ландшафтом мишарей, который, возможно, несет следы древнего балтского генофонда.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ЧУВАШЕЙ (рис. 5.26)

Следующий генофонд в нашем продвижении с востока на запад – чувашский, который в данных по широкой панели гаплогрупп Y-хромосомы пока представлен единственной популяцией, изученной на территории Татарстана. Но вопреки и ее географическому соседству с мишарями и казанскими татарами, и их общей принадлежности к тюркским языкам, и большой численности (по переписи 2010 года, чувашей в России около 1.5 млн. человек) чувашаи являются собой генетический остров – мы вообще не обнаруживаем других популяций, генетически с ними сходных.

Сразу вспоминается, что чувашский язык – единственный уцелевший из болгарской ветви тюркских языков. Уникален он и тем, что отвечает от общей ветви тюркских языков раньше всех и, похоже, первым проникает в Европу, двигаясь на запад с дальней прародины тюрков. Путь этот был не только длинным, но и долгим – с неизбежными контактами и заимствованиями на пути. Поэтому не удивительно, что в культуре чувашей предполагается значительный пласт, связанный с культурой древних земледельцев Передней Азии. Трудно об этом не вспомнить при взгляде на генетический ландшафт чувашей. Впервые, в отличие всех ранее рассмотренных карт, мы видим, что южные территории – Кавказа, Закавказья и Передней Азии – окрашены вовсе не в максимальные темно-красные тона генетических расстояний как на всех предыдущих генетических ландшафтах, а в оранжевые тона умеренно больших генетических различий. И действительно, своеобразие генофонда чувашей задается в первую очередь резко повышенными (по сравнению со всеми соседями) частотами гаплогрупп **E** и **J**, типичных для Передней Азии (глава 2).

В целом, генетический ландшафт чувашей не противоречит (в отличие от ландшафта казанских татар) «болгарской» версии их этногенеза. Однако для решительных выводов необходимо, конечно, детальное изучение всех трех этнографических групп чувашей, каждая из которых испытала влияние разных соседних этносов. Сложив эту мозаику, мы, возможно, сумеем разглядеть наиболее древний генетический пласт в генофонде чувашей.

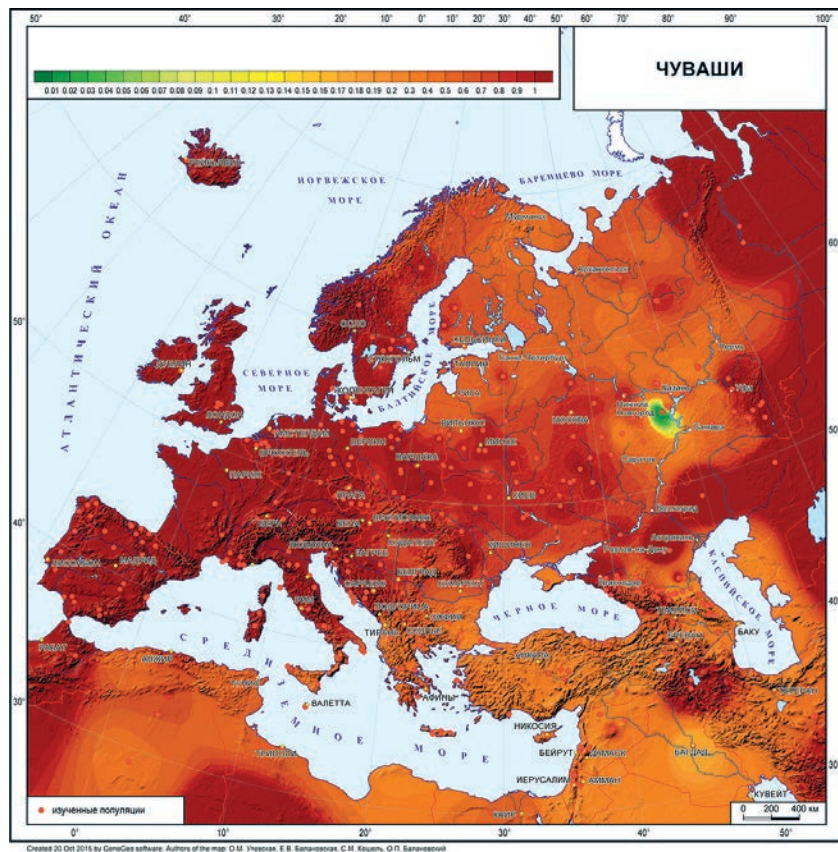


Рис. 5.26. Карта генетических расстояний от чувашей (генетический ландшафт по гаплогруппам Y-хромосомы).

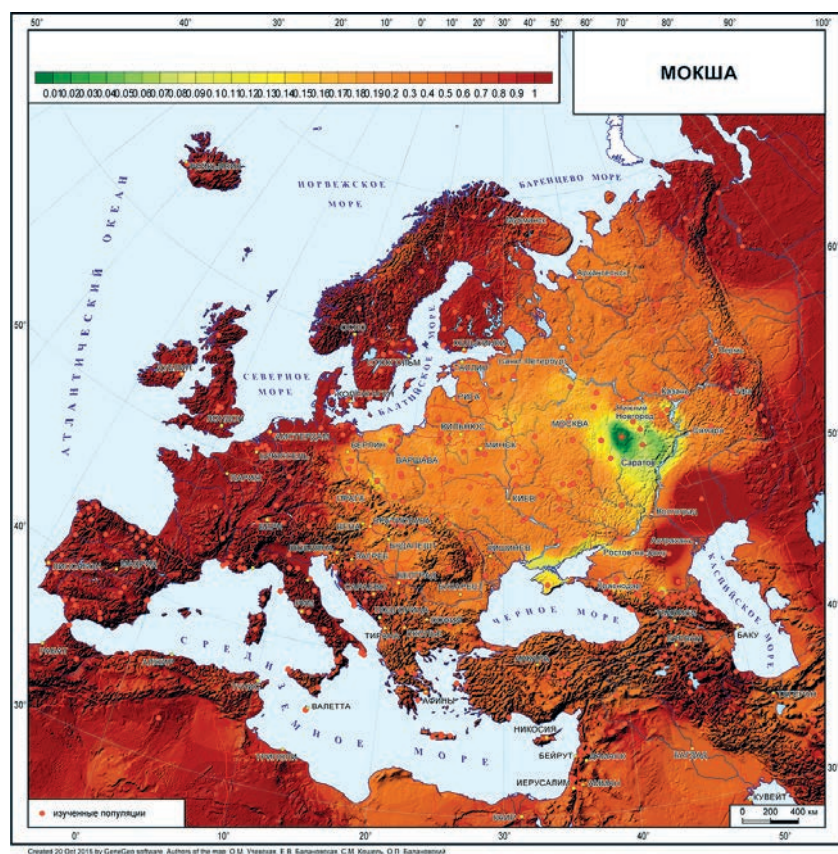


Рис. 5.27. Карта генетических расстояний от мокши (генетический ландшафт по гаплогруппам Y-хромосомы).

**КАРТОГРАФИРОВАНИЕ СХОДСТВА
С ГЕНОФОНДАМИ МОКШИ (рис. 5.27)
И ЭРЗЯ (рис. 5.28)**

Часто возникающий вопрос о том, являются ли мокша и эрзя разными народами или же субэтническими группами одного этноса – мордвы – может быть адресован не к генетикам, а лишь к этнологам. Отметим лишь, что мордва – это экзоэтноним, а мокша и эрзя – самоназвания. Лингвистика выделяет их языки как самостоятельные, а не диалекты – существенная разница в фонетическом строе, лексике и грамматике не позволяет их носителям понимать друг друга. Наиболее близкими к ним являются ныне мертвый мещерский язык, а также марийский и языки прибалтийских финнов. По данным переписи 1989 г. только в республике Мордовия насчитывалось более 180 тыс. мокшан. По данным последней переписи, 50 тыс. человек причислили себя к мокше, 80 тыс. человек – к эрзе, а 700 тыс. человек назвали себя мордвой.

Но генетика изучает популяции всех уровней этнической и субэтнической иерархии, и сходство или же различия их генофондов не могут служить указанием на то, кто из них является этносом, а кто – субэтнической группой. В первой серии карт мы уже видели, что части одного этноса могут быть

генетически не схожи (как северная и южная части русского генофонда), в то время как разные этносы могут быть генетическими близнецами (как, например, центрально-южная половинка русского генофонда и белорусы). Определяющим фактором может быть лишь самосознание самих русских и белорусов, мокши и эрзя, и других подобных общностей. Поэтому, ни в коем случае не обращаясь к решению вопроса, являются ли мокша и эрзя самостоятельными этносами или же частями одного этноса – мордвы, рассмотрим их генофонды отдельно, поскольку для них характерны разные географические ландшафты.

Генетический ландшафт мокши (рис. 5.27) указывает на яркое своеобразие их генофонда – область генетически близких значений охватывает лишь небольшой ареал популяций среднего течения Волги, строго ограничиваясь ее правобережьем.

Генетический ландшафт эрзян (рис. 5.28), напротив, поражает обширнейшим ареалом генетически близких популяций. Этот ареал уже хорошо узнаваем – это все тот же хорошо знакомый нам по второй серии карт (раздел 5.2.) генетический ландшафт «северных славян» (рис. 5.21). При этом ярко-зеленые области генетически близких значений включают в себя не только ареал белорусов,

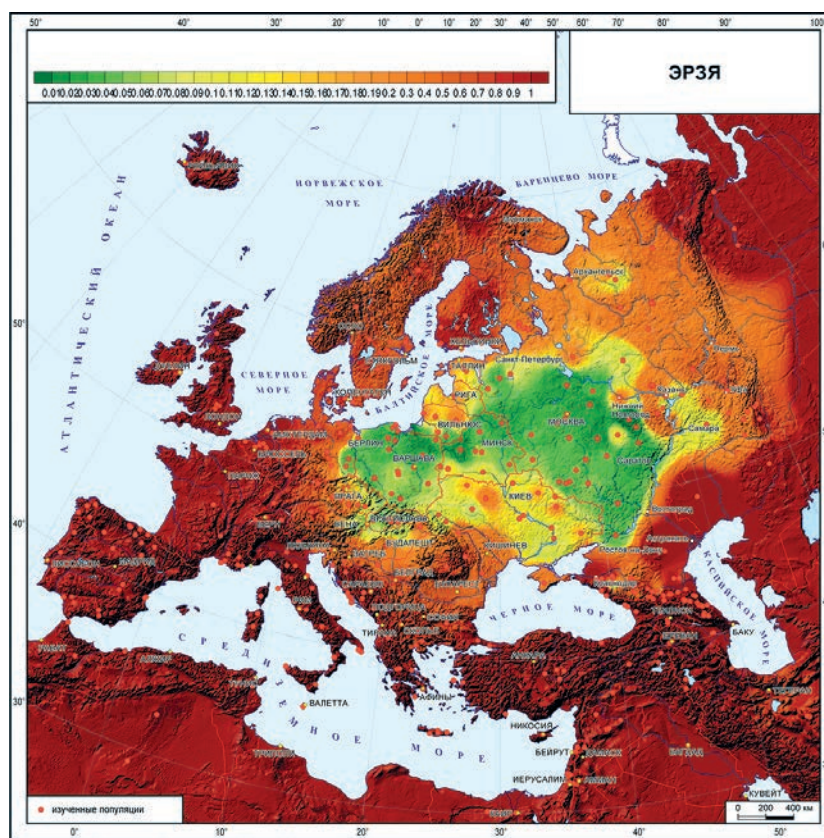


Рис. 5.28. Карта генетических расстояний от эрзян (генетический ландшафт по гаплогруппам Y-хромосомы).

центральных и южных русских популяций, но и Польшу, и запад Германии, и Словакию, оставляя Украине и левобережью Волги область умеренно близких частот, окрашенных желтыми тонами. Обратим особое внимание, что в отличие от мокшан, генофонд эрзян генетически далек от крымских татар.

Такое яркое различие между генетическими ландшафтами мокши и эрзи служит важным аргументом, что при описании характеристик генофонда Мордовии недостаточно просто указание «мордва», и необходимо указывать, популяции мокши или же эрзи были включены в анализ.

5.4. НА СЕВЕРЕ БАЛКАН (СЕРИЯ IV)

Возвращаясь в юго-восточную Европу, рассмотрим пять карт генетических расстояний (рис. 5.30–5.34) от народов, говорящих на языках трех больших лингвистических семей: молдаване и румыны – на языке романской ветви индоевропейской семьи, а словенцы – на ее славянской ветви; гагаузы – на языке тюркской ветви алтайской языковой семьи; венгры – угорской ветви уральской языковой семьи. Эти разноязыкие народы объединяет то, что они заселяют северную окраину Балкан, а генетически их объединяет преобладающий своеобразный ландшафт в виде широкой полосы генетически сходного населения, протянувшейся от Черного моря до Адриатического (рис. 5.29).

Карта генетических расстояний от молдаван (рис. 5.30) – самая яркая иллюстрация этого ландшафта: зеленая «река» генетически близких к молдаванам популяций тянется от Черного моря до Адриатического. По обе ее стороны тянутся желтые «берега» популяций, умеренно близких к генофонду молдаван. Но основная часть Балкан, Центральной и Восточной Европы генетически далеки от них. Отметим, что географически соседний украинский генофонд находится в области умеренных генетических отличий, окрашенных в желтые тона (зеленые оттенки в области, пограничной с Молдовой, связаны с интерполяцией на соседние генетически неизученные области юго-западной Украины). Такой зеленый пояс генетически близких популяций, окаймляющий Балканы с севера, возможно, фиксирует зону миграций, связывающую Адриатику и Причерноморье, но не доходящую до Апеннин.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ МОЛДАВАН (рис. 5.30)

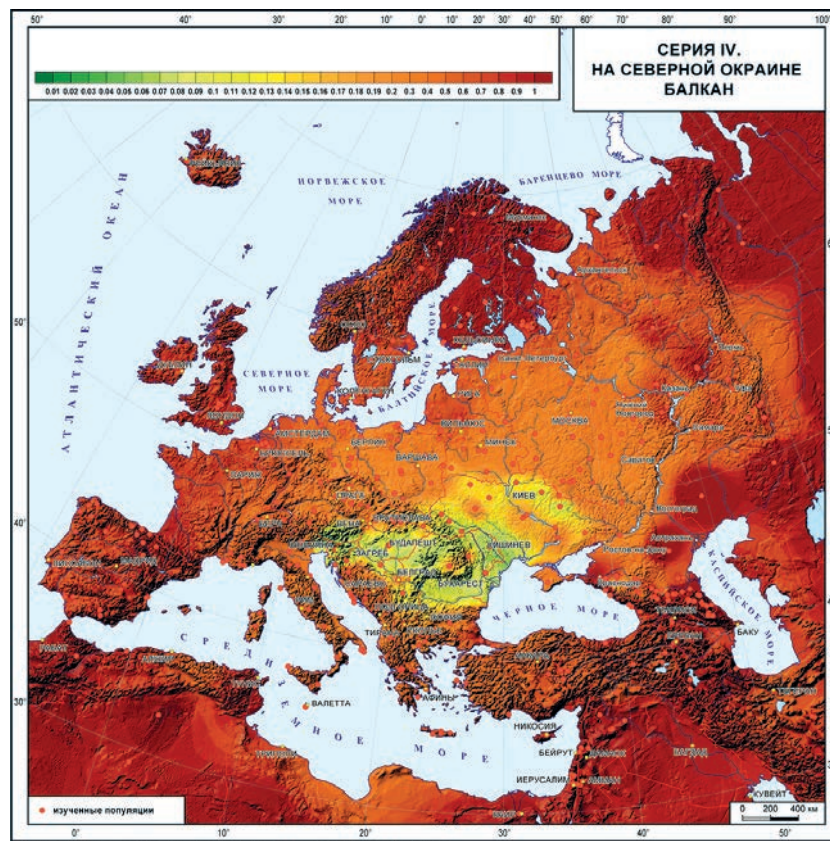


Рис. 5.29. Карта обобщенного генетического ландшафта народов северной окраины Балкан по гаплогруппам Y-хромосомы (Построена как средняя по пяти картам генетических расстояний от венгров, гагаузов, молдаван, румын, словенцев).

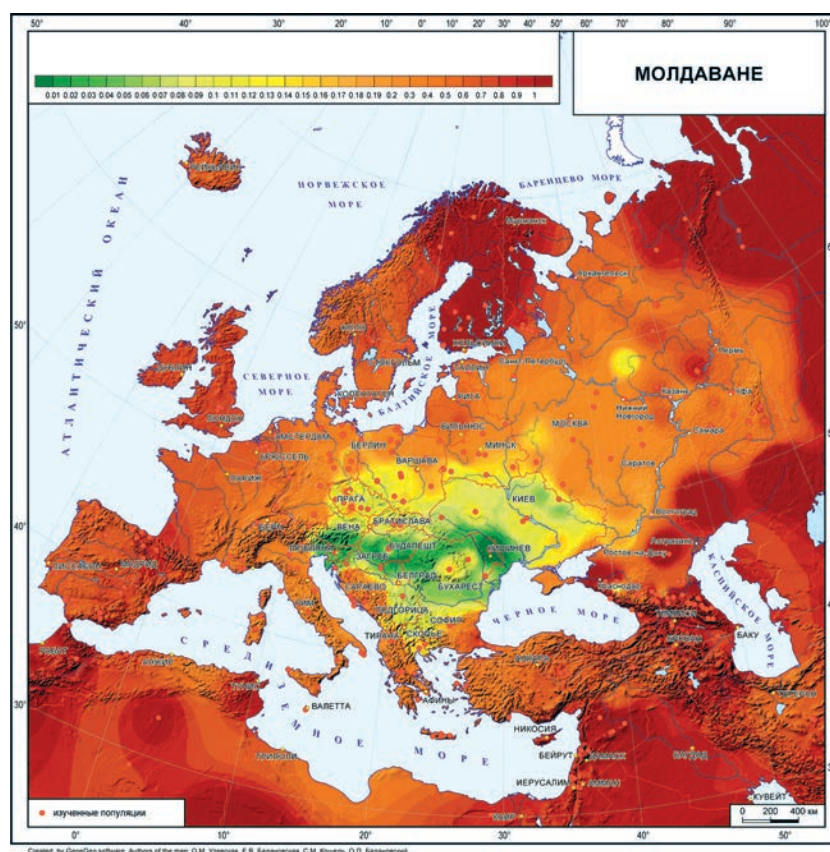


Рис. 5.30. Карта генетических расстояний от молдаван (генетический ландшафт по гаплогруппам Y-хромосомы).

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ РУМЫН (рис. 5.31)

Казалось бы, на карте генетических расстояний от румын (рис. 5.31) мы должны увидеть картину, ничем не отличающуюся от только что виденной у молдаван, ведь не только языки молдаван и румын считаются вариантами одного языка, но и в высказываниях официальных лиц часто утверждается, что молдаване и румыны – один народ. Однако генофонды их различны. Если генофонд молдаван в равной мере принадлежит и южным славянам Балкан, и лежащим к северу и востоку от них популяциям западных и восточных славян, то генофонд румын вторит лишь Балканскому генетическому ландшафту: отвернувшись от Черного моря, он обнимает Адриатику с генетически близкими к румынам и родными по языку популяциями Апеннин.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ГАГАУЗОВ (рис. 5.32)

Еще более впечатляет генетический ландшафт тюркоязычных гагаузов (рис. 5.32).

Вопреки «лингвистическим» ожиданиям, генофонд гагаузов смотрит не на юго-восток (ведь язык гагаузов близок к турецкому, азербайджанскому и туркменскому), а на запад. Вопреки «географическим» ожиданиям он не повторяет генетический

ландшафт соседних молдаван. Не подтверждает он и «исторические» ожидания – одна из заманчивых версий происхождения гагаузов связывает их с протоболгарами, мигрировавшими с Волги на Балканы в середине 1 тыс. н.э., однако генетические связи гагаузов с нынешним Поволжьем на карте отсутствуют. Генофонд гагаузов отличен и от ближайших степных тюркоязычных популяций – крымских татар и ногайцев (достаточно сравнить с генетическим ландшафтом ногайцев на рис. 5.43), и от генофонда турок, близость с которым можно было ожидать, исходя и из истории (вхождения в Османскую империю), и языкового сходства.

Генофонд гагаузов явно обращен к Балканам, к восточной окраине. Видимо, он хранит генетическую память о балканской родине, откуда гагаузы в XVIII–XIX веках переселились в Бессарабию.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ВЕНГРОВ (рис. 5.33)

Генофонд венгров, напротив, обращен к северу. Хотя та же «зеленая река» генетически близких популяций протекает по северной окраине Балкан, но зона генетического сходства с венграми охватывает Австрию и доходит до Польши. Конечно же, генетической связи с лингвистическими родственниками – угроязычными хантами и манси Западной Сибири – мы не видим (рис. 5.33). Мы

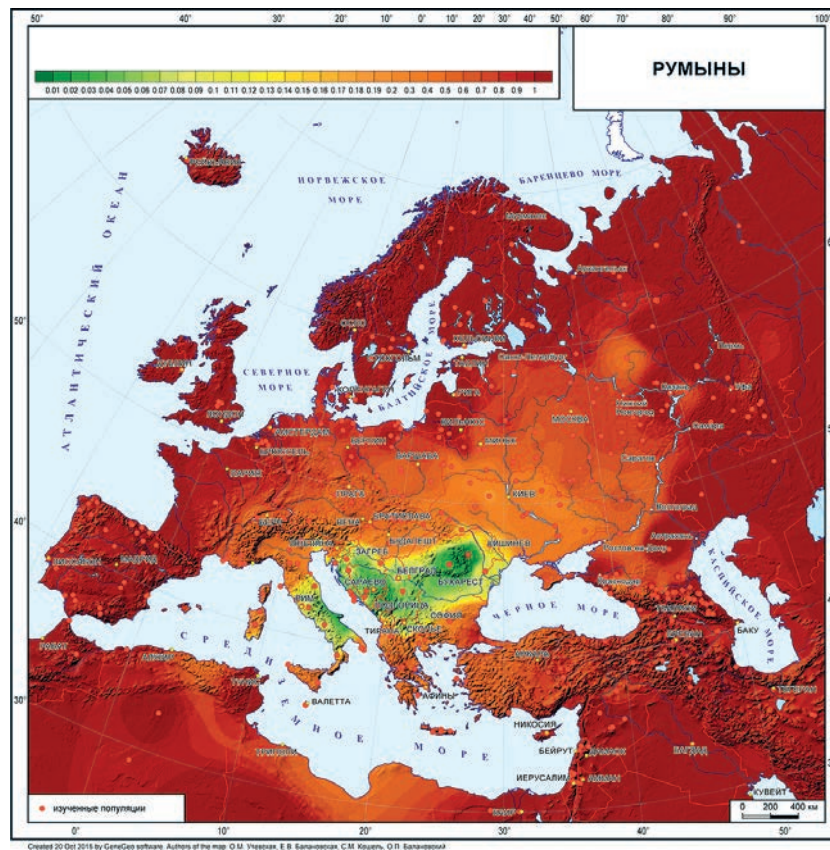


Рис. 5.31. Карта генетических расстояний от румын (генетический ландшафт по гаплогруппам Y-хромосомы).

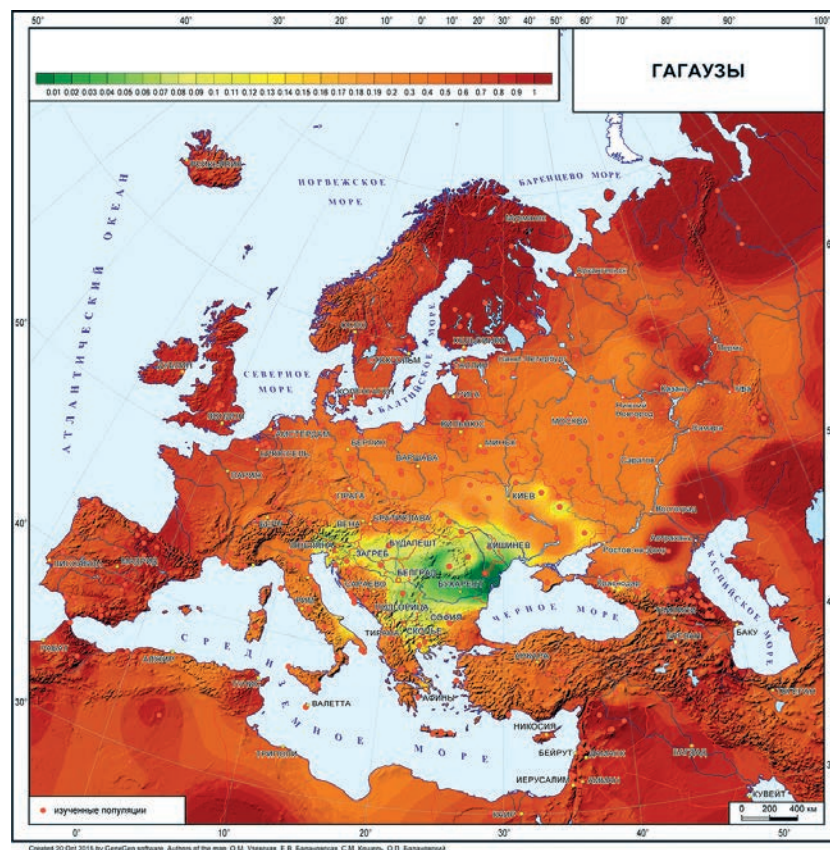


Рис. 5.32. Карта генетических расстояний от гагаузов (генетический ландшафт по гаплогруппам Y-хромосомы).

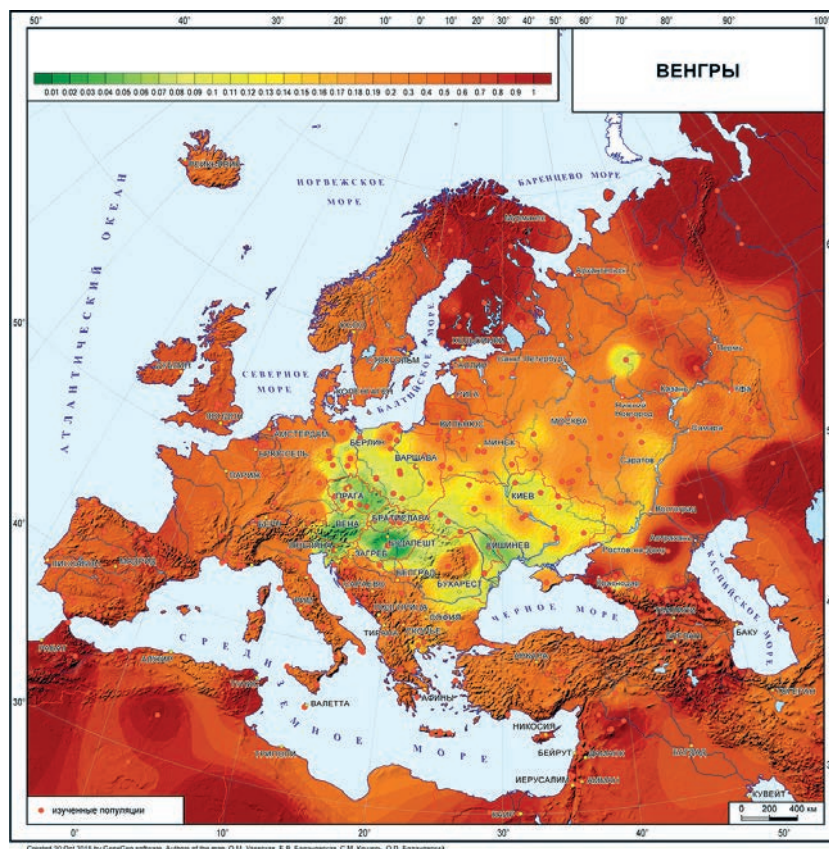


Рис. 5.33. Карта генетических расстояний от венгров (генетический ландшафт по гаплогруппам Y-хромосомы).

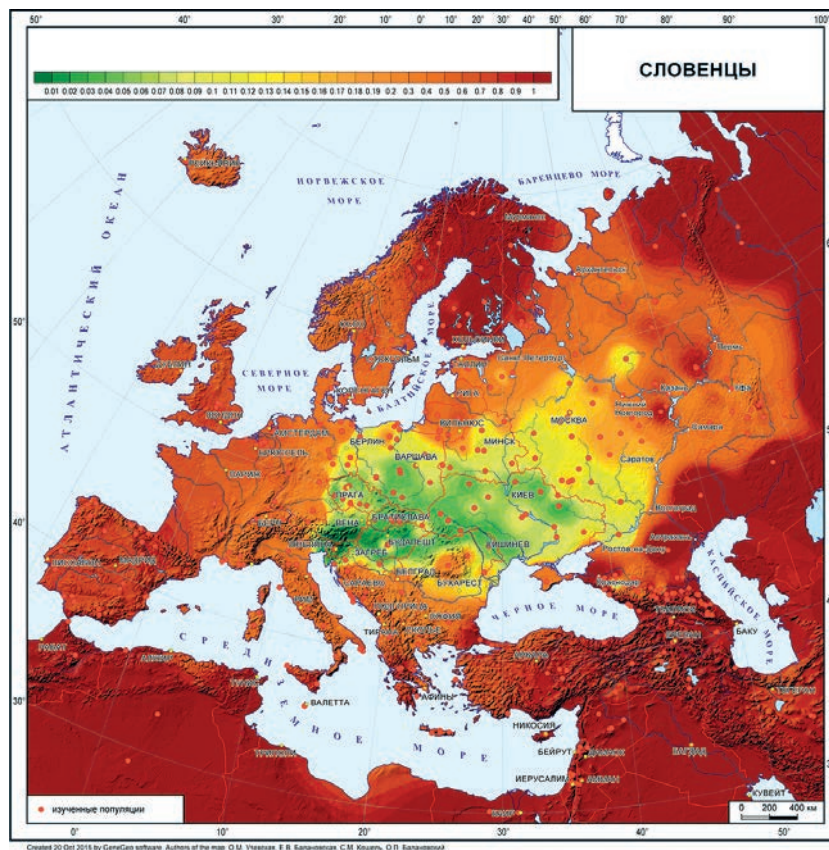


Рис. 5.34. Карта генетических расстояний от словенцев (генетический ландшафт по гаплогруппам Y-хромосомы).

видим окрашенные в желтые тона области умеренного генетического сходства, которые охватывают обширный ареал славянского мира Центральной и Восточной Европы и доходят до Волги. Однако «волжские» и «причерноморские» генетические связи нельзя прямолинейно трактовать как прародину и путь угрозных племен от Волги до нынешней Венгрии. На это указывает генетический ландшафт словенцев (рис. 5.34), полностью включающий в себя и популяции, генетически сходные с венграми (рис. 5.33).

И еще раз отметим, что генетический ландшафт венгров целиком вписывается в этот «славянский» ландшафт, указывая на хорошо известный в антропологии факт ведущей роли местного, автохтонного населения в сложении генофонда венгров.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ СЛОВЕНЦЕВ (рис. 5.34)

Карта генетических расстояний от словенцев объединяет в себе сразу два общих паттерна ге-

нетических ландшафтов – и второй (рис. 5.21), и четвертой (рис. 5.29) серий. Мы видим и зеленую «реку» генетически близких к словенцам популяций, которая тянется от них до Черного моря. Но к северу и востоку находится область генетически близких значений, ареал которых вторит паттерну генетического ландшафта «северных» славян. Чтобы убедиться в этом, достаточно сравнить эту карту от представителя южных славян – словенцев (рис. 5.34), например, с картой генетических расстояний от представителя западных славян – словаков (рис. 5.18). Мы видим тот же ареал, отличия же карт состоят лишь в мере генетических различий – на карте словенцев этот ареал окрашен в желтые тона умеренных генетических отличий.

Такое особое промежуточное положение генофонда словенцев согласуется и с их лингвистическим своеобразием [Kushniarevich et al., 2015], и с их генетическим своеобразием по другим генетическим системам, описанном в главе 6.

5.5. ЮЖНЫЕ СЛАВЯНЕ (СЕРИЯ V)

Македонцы, сербы, хорваты, боснийцы и герцеговинцы (а также рассмотренные в предыдущем разделе словенцы) формируют группу южных славян в пределах индоевропейской лингвистической семьи. Географически эти народы объединяет их принадлежность к Балканам. Их генетическое своеобразие не только по Y-хромосоме, но и по аутосомным широкогеномным маркерам детально рассмотрено в главе 6. Это своеобразие связывается с сохранением субстратного генофонда тех балканских племен и народов, которые стали говорить на славянских языках, но сохранили свои особенности. Их генетический ландшафт можно назвать «южнославянским», помня, что рассмотренные выше словенцы (рис. 5.34) являются исключением (наличие хотя бы одного исключения для каждой серии стало уже правилом).

Основной паттерн их обобщенного генетического ландшафта, приведенного на рис. 5.35, задается тем, что северная граница как раз проходит по границе Балкан, а южная не затрагивает ареалы греков и албанцев.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ МАКЕДОНЦЕВ (рис. 5.36)

Впрочем, карта генетических расстояний от македонцев сразу вносит ожидаемые коррективы в южную границу ландшафта: она отчасти заходит и на северные территории современных Греции и Албании, и на юг Италии. При этом северо-запад южнославянского ареала генетически чужд

македонцам: территории современных не только Словении, но и Хорватии, Боснии и Герцеговины, окрашены на карте в красные тона генетически несходных популяций. Однако популяции сербов генетически близки к македонцам, образуя с ними единый массив сходных генофондов, окрашенных на карте в зеленые тона. Этот массив тянется на восток, охватывая популяции румын, а окрашенные в желтые тона области умеренных генетических различий доходят до значительной части северо-западного побережья Черного моря.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ СЕРБОВ (рис. 5.37)

Карта генетических расстояний от сербов (рис. 5.37) очень сходна с генетическим ландшафтом македонцев (рис. 5.36). Их небольшие отличия соответствуют ожиданиям: южная граница зоны сходства с сербами не заходит в Албанию и Грецию, но зато северо-запад южнославянского ареала оказывается уже генетически более близок к сербам – генофонды Хорватии, Боснии и Герцеговины окрашены на карте в желто-зеленые тона. Этот массив вновь тянется на восток, охватывая популяции румын и северо-западного побережья Черного моря.

В целом сербы и македонцы характеризуются наибольшей площадью ареала генетически сходных с ними популяций среди всех южных славян. Возможно, они в наибольшей степени воспроизводят генетическую общность южных славян и могут служить основной моделью при ее реконструкции.

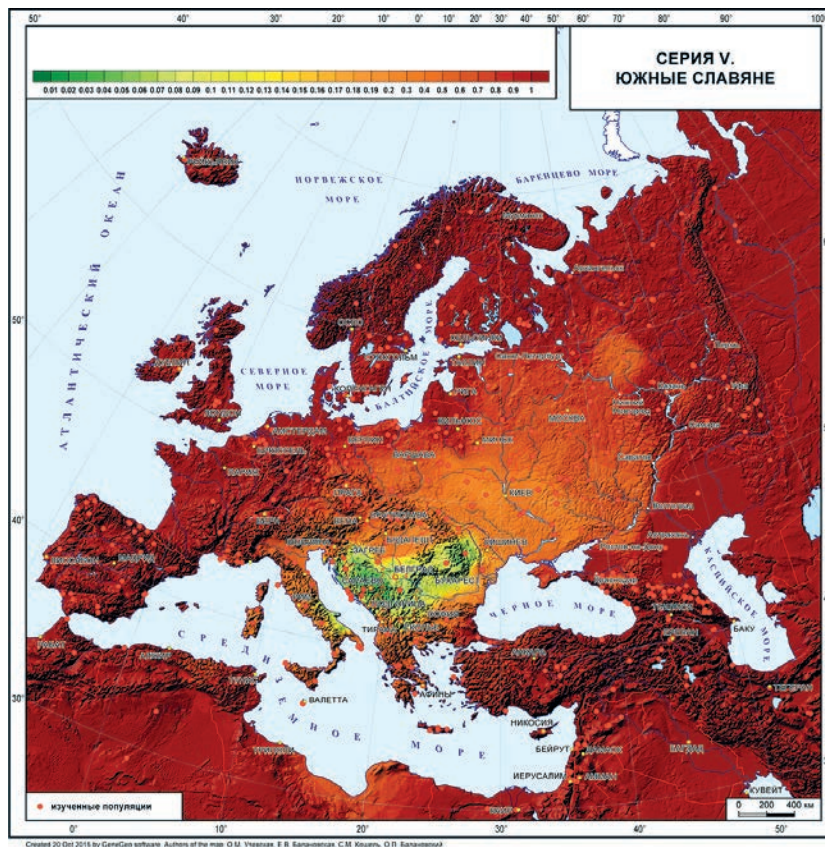


Рис. 5.35. Карта обобщенного генетического ландшафта южных славян по гаплогруппам Y-хромосомы. (Построена как средняя по пяти картам генетических расстояний от боснийцев, популяций Герцеговины, македонцев, сербов, хорватов).

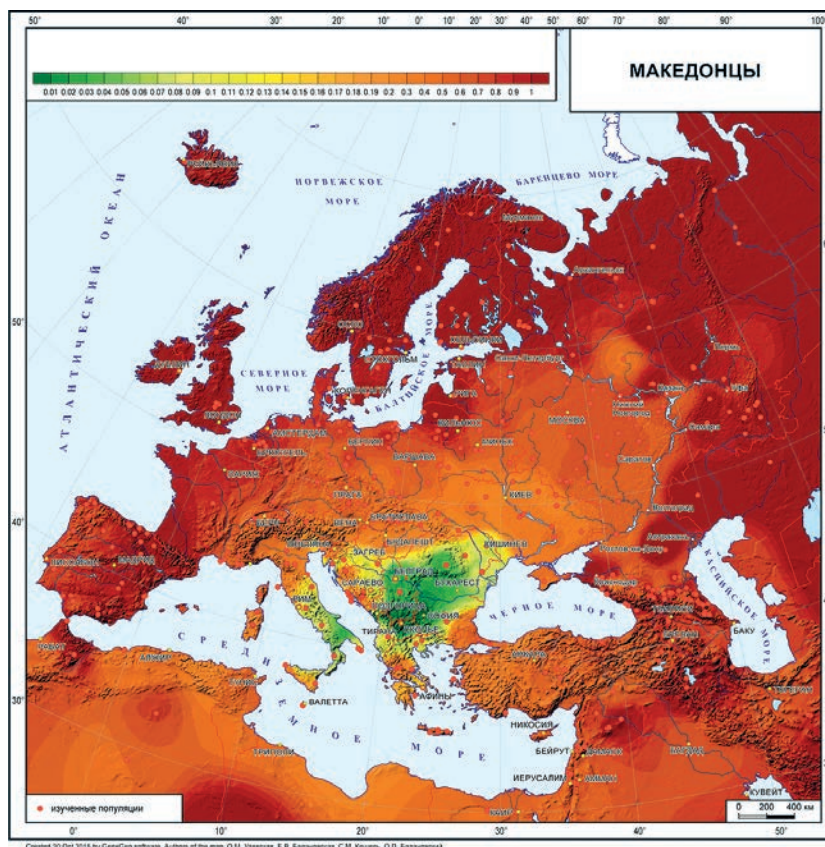


Рис. 5.36. Карта генетических расстояний от македонцев (генетический ландшафт по гаплогруппам Y-хромосомы).

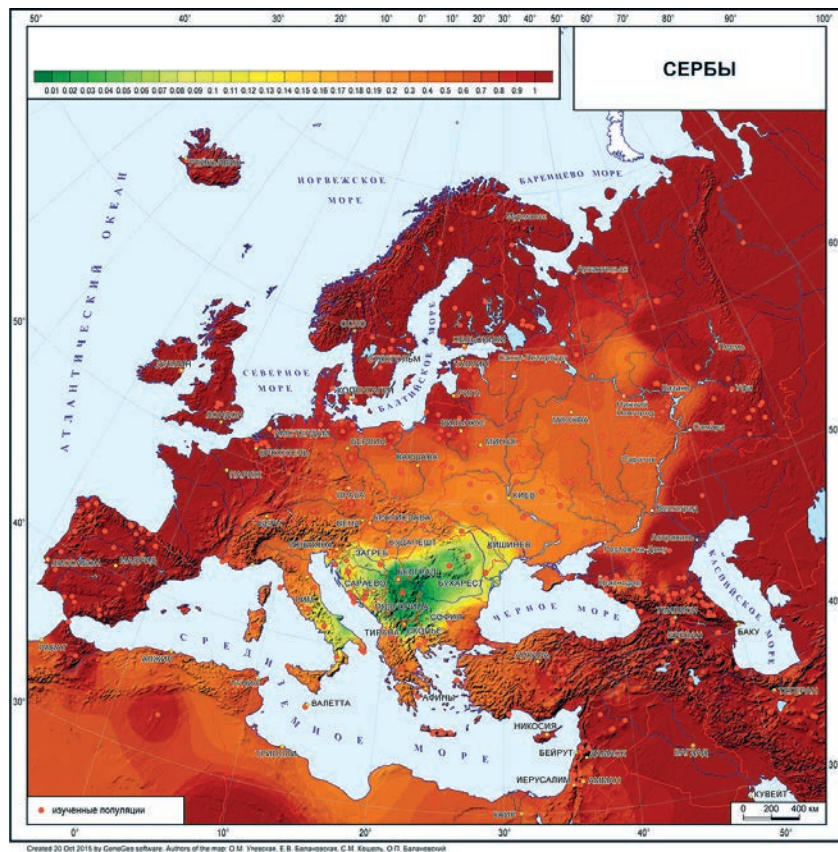


Рис. 5.37. Карта генетических расстояний от сербов (генетический ландшафт по гаплогруппам Y-хромосомы).

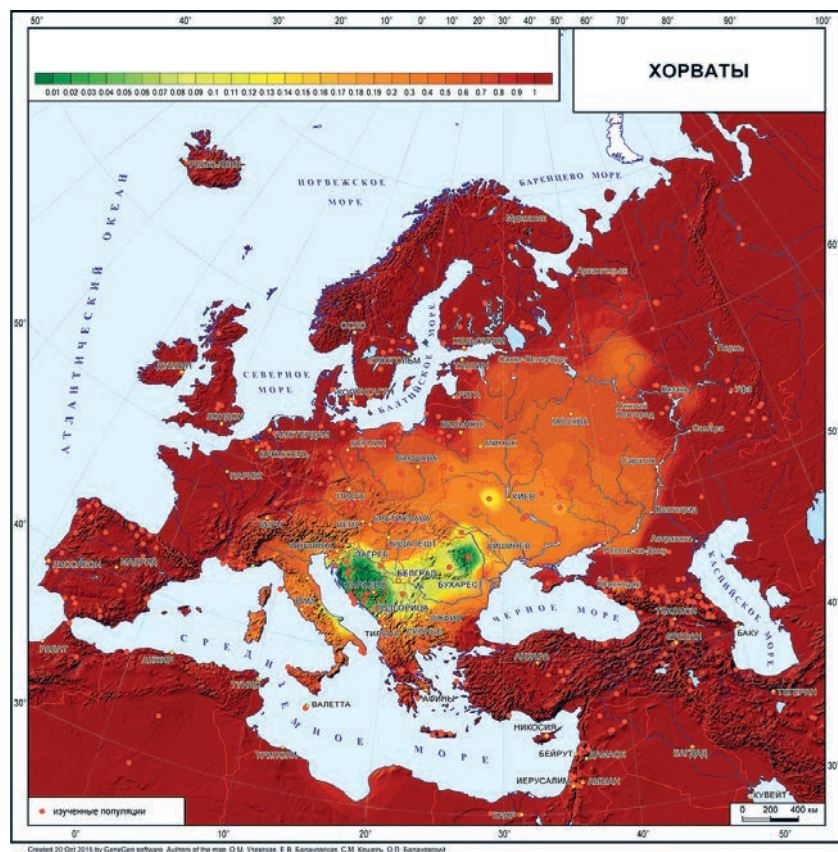


Рис. 5.38. Карта генетических расстояний от хорватов (генетический ландшафт по гаплогруппам Y-хромосомы).

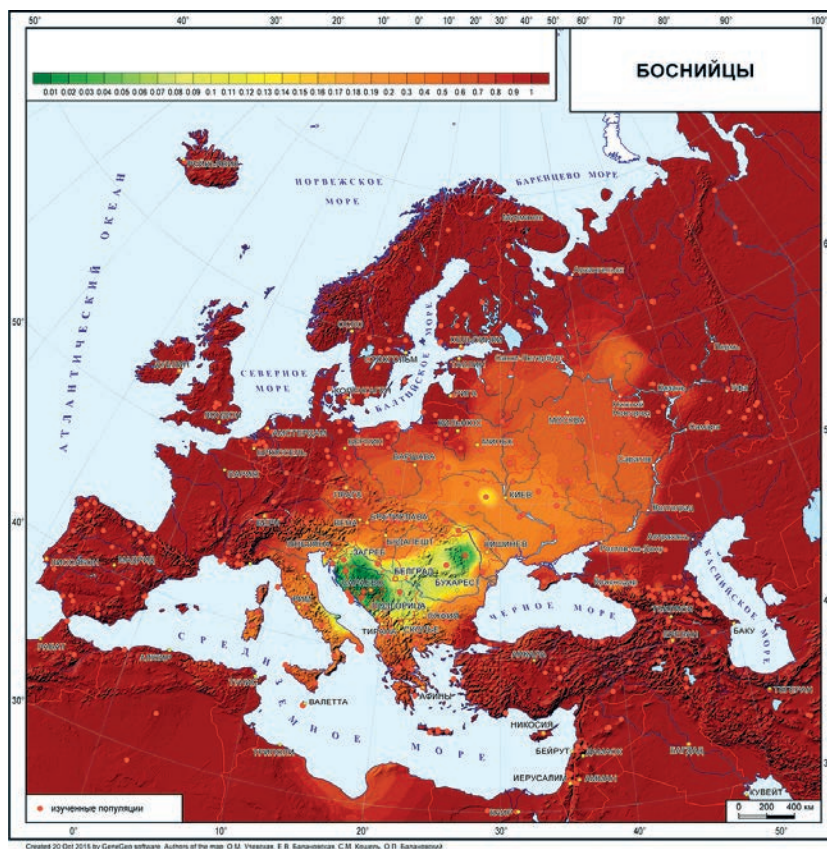


Рис. 5.39. Карта генетических расстояний от боснийцев (генетический ландшафт по гаплогруппам Y-хромосомы).

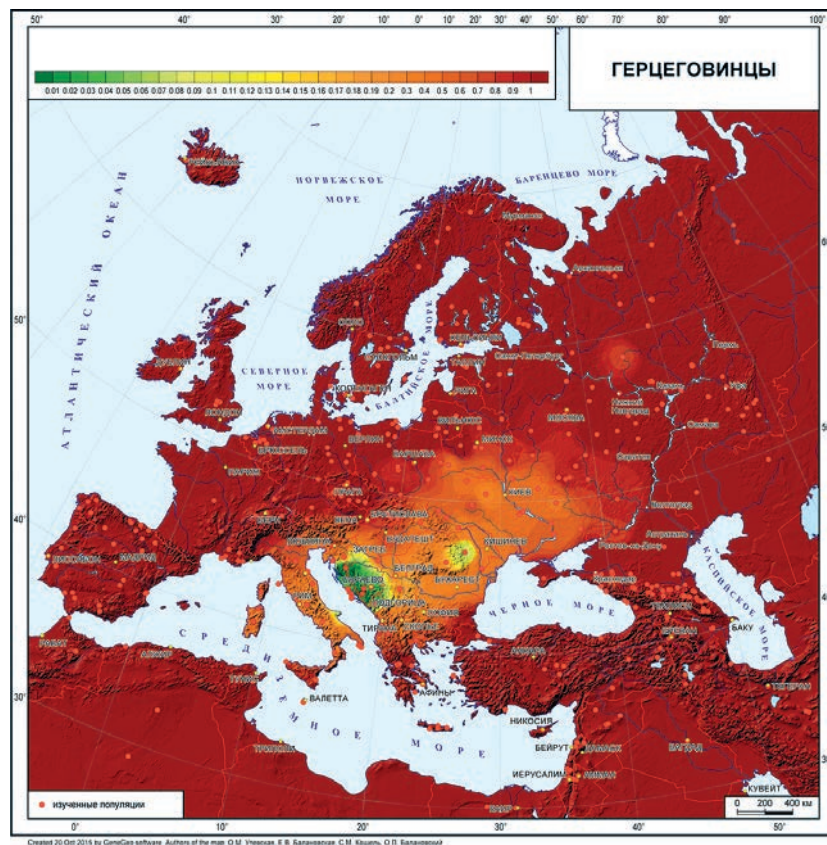


Рис. 5.40. Карта генетических расстояний от народов Герцеговины (генетический ландшафт по гаплогруппам Y-хромосомы).

**КАРТОГРАФИРОВАНИЕ СХОДСТВА
С ГЕНОФОНДОМ ХОРВАТОВ (рис. 5.38),
БОСНИЙЦЕВ (рис. 5.39) И ПОПУЛЯЦИЙ
ГЕРЦЕГОВИНЫ (рис. 5.40)**

Все три карты имеет смысл рассматривать вместе, поскольку их сходство очень велико. Во-первых, все они имеют меньший, чем у сербов и македонцев, ареал генетически сходных популяций (зеленые тона). Во-вторых, они составляют как бы часть генетического ландшафта сербов и македонцев.

Естественно, что зеленая зона максимально близких генетических популяций сосредоточена в их этнических ареалах, однако желто-зеленые тона генетического сходства вновь рукавом простираются на территорию современной Румынии, но не доходят до черноморского побережья. При этом генетические ландшафты хорватов и боснийцев практически неотличимы друг от друга, а карта расстояний от генофонда популяций Герцеговины представляет их локальный вариант, примыкающий к Адриатике.

5.6. ОБРАМЛЕНИЕ ЕВРОПЫ (СЕРИЯ VI)

Рас-стояние: вёрсты, мили...
Нас рас-ставили, рас-садили,
Чтобы тихо себя вели
По двум разным концам земли.
Марина Цветаева

К сожалению, объем книги не позволяет привести и описать генетические ландшафты от каждого из народов Европы – для этого нужна отдельная книга. Но чтобы увидеть, насколько велики различия генофондов даже европейской окраины Евразии, рассмотрим генофонды трех этносов, окаймляющих пространство народов, рассмотрен-

ных в пяти предыдущих сериях. Эти три этноса – албанцы, шведы, ногайцы – не только географически «расставлены» по трем «концам земли», но и генетически они оказываются полярно различными, как бы оформляя масштаб разнообразия генофонда Европы.

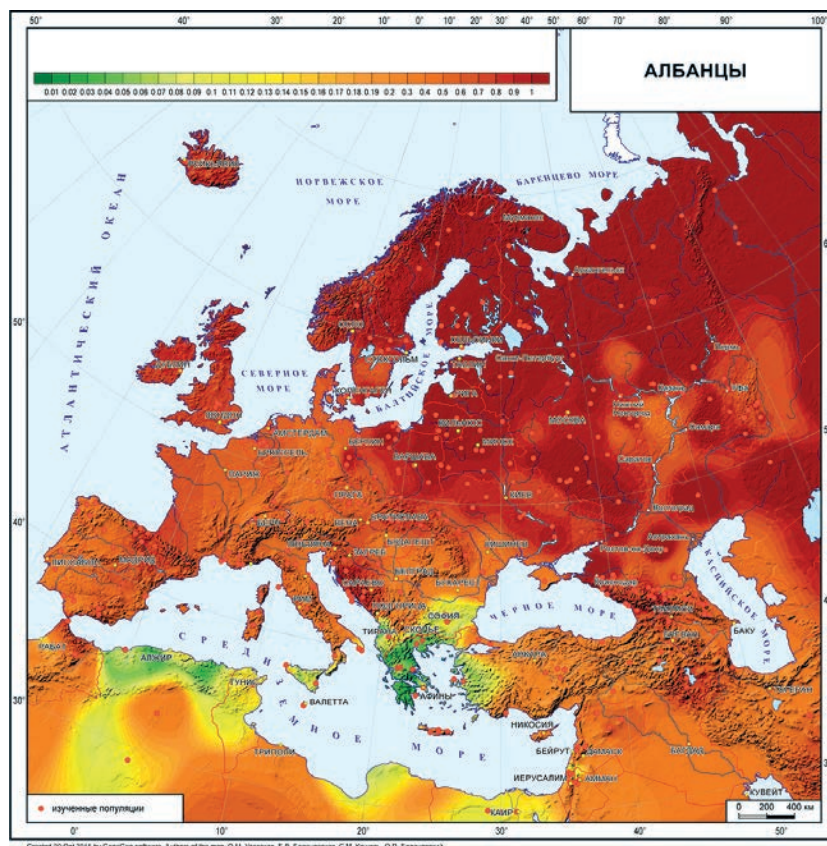


Рис. 5.41. Карта генетических расстояний от албанцев (генетический ландшафт по гаплогруппам Y-хромосомы).

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ АЛБАНЦЕВ (рис. 5.41)

Эта карта является естественным продолжением предыдущей балканской серии (раздел 5.5). Но она являет совсем иной паттерн: он весь обращен не только к югу Балкан, но и к Средиземноморью в целом. Генетически близкими к албанцам оказываются популяции всей Греции, желто-зеленые тона небольших генетических различий захватывают не только Сицилию и запад Малой Азии, но и африканское побережье.

Такой генетический ландшафт можно назвать «средиземноморским», и албанцы могут выступать представителем «средиземноморского генетического полюса» Европы. При этом европейский генофонд в целом оказывается албанцам почти чуждым: европейский континент окрашен в интенсивно красные тона максимальных генетических отличий, в то время как Средиземноморское побережье не только Европы, но и Африки, и части Передней Азии генетически более близки к албанцам, чем Европа в целом.

Нельзя не заметить разительные отличия между генетическими ландшафтами двух соседних генофондов – албанцев (рис. 5.41) и македонцев (рис. 5.36). Это два разных генетических мира, как бы стоящих спиной друг к другу и обращенных один

– к северу, а другой – к югу. Лингвистически они также резко различны: хотя и принадлежат к одной индоевропейской семье, но албанский язык стоит особняком ото всех и представляет наследие палеобалканских языков с многочисленными заимствованиями. Генетический ландшафт албанцев не позволяет утверждать, что их генофонд сохранил генетическое наследие иллирийцев, древние корни которых по данным археологии ведут на север, в центральную Европу, а не на юг. Но о чем приведенные карты ясно свидетельствуют, так это о том, что одновременно быть наследниками генофонда иллирийцев и хорваты (рис. 5.38) и албанцы (рис. 5.41) не могут, настолько различны их генофонды: кто-то из них или же и те, и другие не сохранили это наследство Балкан.

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ ШВЕДОВ (рис. 5.42)

Северо-западный край земли – «скандинавский генетический полюс» – представляет карта расстояний от генофонда шведов. Эта карта (рис. 5.42) почти альтернативна предыдущей (рис. 5.41): темно-красные тона максимальных генетических расстояний переместились из Европы как раз на те территории, к которым были генетически близки албанцы. Теперь уже Балканы, Средиземноморье,

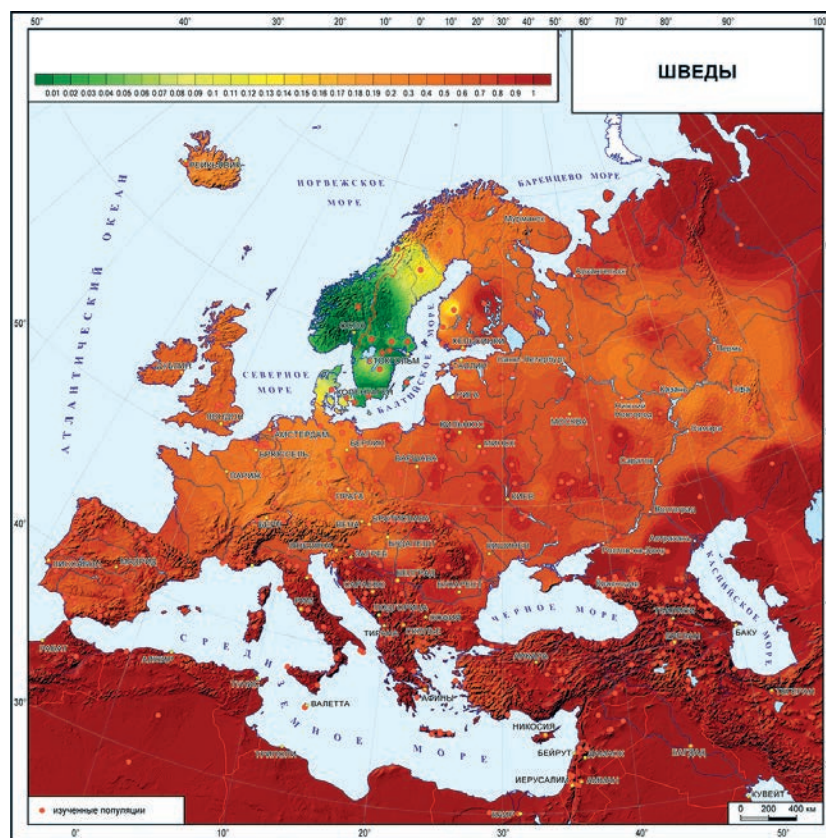


Рис. 5.42. Карта генетических расстояний от шведов (генетический ландшафт по гаплогруппам Y-хромосомы).

Передняя Азия, Кавказ и циркумкаспийский регион окрашены в темно-красные тона максимальных генетических отличий, в то время как вся Европа, за исключением южной, являет оранжевые тона более умеренных генетических отличий.

Темно-зеленая зона генетического сходства четко ограничена Скандинавией: шведы и норвежцы являют единый генетический пласт, близок к ним и генофонд датчан. При этом самый север Скандинавии уже генетически отличен, как и практически все популяции Финляндии. В первой серии карт (раздел 5.1) мы видели, что эти популяции севера Скандинавии генетически очень близки к финнам (рис. 5.9), а также к карелам и вепсам (рис. 5.2), и лишь отчасти – к эстонцам (рис. 5.3).

КАРТОГРАФИРОВАНИЕ СХОДСТВА С ГЕНОФОНДОМ НОГАЙЦЕВ (рис. 5.43)

Юго-восточный край европейской земли – «генетический полюс евразийской степи» – представлен генетическим ландшафтом ногайцев. Зеленые

тона минимальных генетических расстояний обнаруживаются лишь в популяциях самих ногайцев и у крымских татар (рис. 5.43). Однако желтые тона умеренного генетического сходства тянутся по прикавказским степям от Каспия до Причерноморья и отдельным рукавом поднимаются вдоль нижнего течения Волги.

Было бы слишком большим упрощением считать генетический ландшафт ногайцев просто степным. Ведь обращают на себя внимание желтые тона умеренного генетического сходства ногайцев с населением горных областей Передней Азии – от Анатолии до Иранского нагорья. Не менее впечатляет, что практически вся Центральная и Восточная Европа, Урал, Западная Сибирь окрашены в оранжевые тона умеренных генетических отличий. Можно считать, что генофонд ногайцев впитал в себя столь многие евразийские генетические потоки, что может успешно выступать в качестве одного из эталонов «генетического полюса евразийской степи».

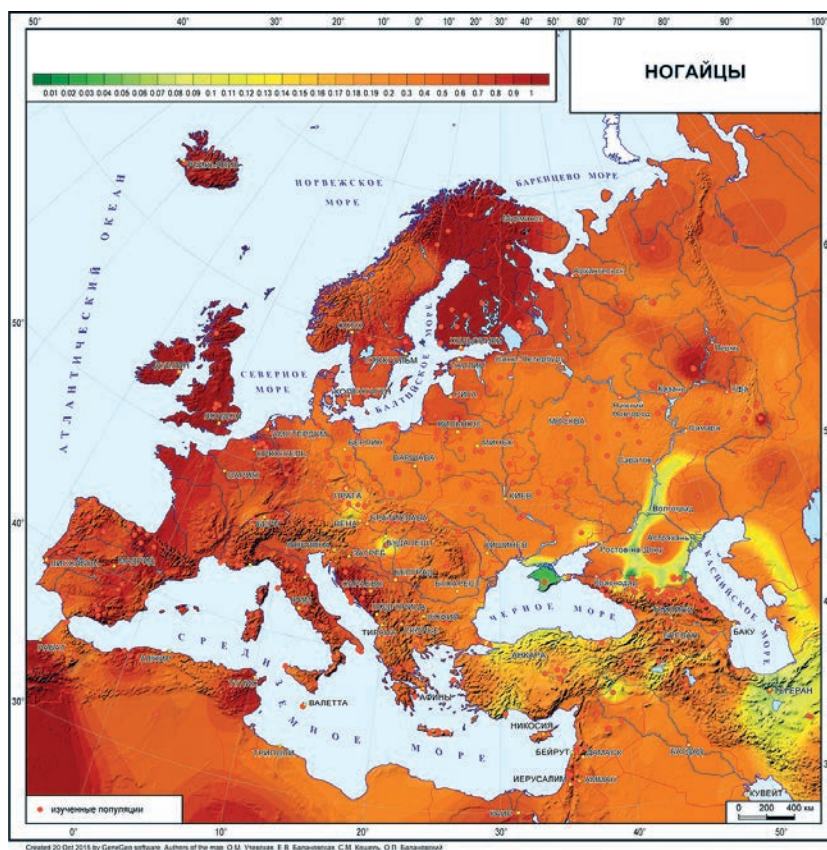


Рис. 5.43. Карта генетических расстояний от ногайцев (генетический ландшафт по гаплогруппам Y-хромосомы).

ГЛАВА 6.

ГЕНОФОНД СЛАВЯН

Изучение этногенеза и этнической истории народов традиционно проводится рядом взаимодействующих дисциплин – археологией, историей, лингвистикой, этнографией, этнической антропологией. Популяционно-генетическое изучение изменчивости ДНК маркеров неотделимо от исследований этнической антропологии, поскольку также опирается на биологическую изменчивость популяций, только имеет дело не с фенотипическими (внешними) признаками, а с ДНК маркерами. Молекулярно-генетическое изучение популяционных генофондов человека с каждым десятилетием занимает все более важное место в системе наук о народонаселении и решает, на наш взгляд, две задачи.

При решении первой (наиболее популярной) задачи информация о генофонде становится еще одним историческим источником. Например, генетическое сходство двух географически соседних популяций может свидетельствовать об их общем происхождении, а генетическое сходство двух удаленных популяций может стать основанием для гипотезы о миграции. Вторая задача является уже собственной задачей популяционной генетики: изучить структуру генофонда и закономерности ее

изменения. Логически эта задача является перво-степенной, поскольку только после получения и осмысления собственных данных популяционная генетика может предоставлять свои результаты и выводы в пользование смежным наукам. В данной главе мы попытались решить обе задачи применительно к славянским народам, изучение которых (особенно восточных славян) традиционно для российской науки. Генетические данные о них на момент начала нашего исследования были представлены в мировой литературе лишь фрагментарно, и в результате огромный славянский мир, занимающий почти половину Европы, исчезал из знаний о генофонде населения.

Полученные данные находят широкий спектр приложений – от эколого-генетического мониторинга и судебно-медицинской экспертизы до преподавания и популяризации науки. В данной же главе анализ полученных данных проводится в аспекте исторической геногеографии. Его цель – приблизиться к пониманию того, какие исторические процессы в народонаселении (в первую очередь, соотношение исторических миграций и частичной изоляции популяций) сформировали наблюдаемую картину генофонда.

6.1. ОСНОВНЫЕ ЧЕРТЫ БАЛТО-СЛАВЯНСКОГО ГЕНОФОНДА

ОБОБЩАЮЩЕЕ ИССЛЕДОВАНИЕ

Исследования генофонда народов балто-славянской лингвистической группы проводятся нашим коллективом уже 15 лет. Пришла пора подводить итоги. Они опубликованы нами в статье [Kushniarevich et al., 2015] на английском языке, а на русском представлены на нашем сайте (генофонд.рф) и в данной главе. Эти исследования почти с самого начала проводились совместно с Эстонским биоцентром. В работе участвовали исследователи и из многих других стран, в которых славянские и балтские народы составляют большинство населения – Украины, Белоруссии, Литвы, Хорватии, Боснии и Герцеговины. Для реконструкции давней истории народов, говорящих на родственных языках, применен междисциплинарный подход. Прослежены пути формирования генофонда

да всех групп славян и балтов одновременно по трем генетическим системам: по Y-хромосоме (отцовские линии наследования), по митохондриальной ДНК (материнские линии наследования) и по широкогеномным данным об аутосомных маркерах (где отцовские и материнские линии представлены равноправно). Это генетическое исследование проведено параллельно с лингвистическим, целью которого было уточнение древа балто-славянских языков и использование новейших лексико-статистических данных для расчета тесноты связи лингвистической и генетической реконструкций балто-славянских народов.

СЛАВЯНИЗАЦИЯ ЕВРОПЫ

На балто-славянских языках говорит примерно треть современного населения Европы, а по пло-

щади балтские и славянские народы занимают около половины Европы. Лингвисты сходятся во мнении, что балтские и славянские языки происходят от общего корня в семье индоевропейских языков (балто-славянский узел присутствовал уже на первой реконструкции индоевропейского древа [Schleicher, 1861]). Прото-балто-славянский язык отделился от других индоевропейских языков в интервале от 7000 до 4500 лет назад [Fortson, 2004; Mallory, Adams, 2006; Rexová et al., 2003; Ringe et al., 2002; Nakhleh et al., 2005; Novotná, Blažek, 2007; Gray, Atkinson, 2003; Bouckaert et al., 2012] и произошло это, вероятнее всего, в Центральной Европе [Bouckaert et al., 2012]. Расхождение балтской и славянской языковых ветвей датируется временем 3500–2500 лет назад [Novotná, Blažek, 2007; Gray, Atkinson, 2003; Bouckaert et al., 2012]. Дальнейшее разделение славянских языков происходило уже относительно недавно – 1700–1300 лет назад

[Novotná, Blažek, 2007; Gray, Atkinson, 2003; Bouckaert et al., 2012; Birnbaum, 1975; Sussex, Cubberley, 2006; Blažek, 2007]. С ранним средневековьем (примерно 1400–1000 лет назад) связывается так называемая «славянизация Европы» – период быстрого распространения славянских языков на огромных территориях. В Восточной Европе славяне распространялись на территории, где проживали балтские, финно-угорские и тюркские популяции, в Западной Европе – на территории носителей германских языков, на Балканах – на территории местных разноязычных популяций. Но как эти изменения в культуре Европы, фиксируемые распространением славянских языков, повлияли на генофонд Европы? Иными словами, какие были основные черты генетической истории балто-славянских популяций и их взаимодействия с генофондами популяций, говоривших на других языках – финно-угорских, германских, тюркских и иных?

ТРИ ЗЕРКАЛА БАЛТО-СЛАВЯНСКОГО ГЕНОФОНДА

КАКИЕ МАРКЕРЫ ИЗУЧАЛИ

Для максимально полного изучения балто-славянских популяций использовали все три «зеркала» – все три генетические системы, которые на данный момент наиболее информативны для исследования генофонда (рис. 6.1).

1) Y-хромосома, которая наследуется по отцовской линии: изучено 6078 образцов из 62 популяций;

2) Митохондриальная ДНК (мтДНК), которая наследуется по материнской линии: изучено 6876 образцов мтДНК из 48 популяций;

3) Широкогеномные (полногеномные) маркеры: 1 297 образцов из 16 популяций. Это полиморфные точки (однонуклеотидного полиморфизма, SNP), которые разбросаны по всему геному и расположены на аутосомах – неполовых хромосомах.

Для анализа методом ADMIXTURE использованы 200 тысяч SNP-маркеров, которые являются общими для трех использованных панелей Illumina (610K, 650K и 660K) и несцепленными друг с другом; для анализа общих фрагментов использованы все 500 тысяч маркеров, которые являются общими для трех использованных панелей Illumina (включая маркеры, сцепленные друг с другом); для анализа главных компонент и расчета генетических расстояний использованы 57 тысяч маркеров, которые являются общими для панелей Illumina и Affimetrix, и при этом не сцепленными друг с другом.

КАКИЕ ПОПУЛЯЦИИ ИЗУЧАЛИ

Значительная часть этих внушительных массивов данных получена впервые – 1254 образцов по Y-хромосоме, 917 образцов по мтДНК, 70 образцов

по широкогеномным маркерам. Остальные данные взяты из ранее опубликованных работ. Для сравнения использованы все накопленные к нынешнему времени данные по другим генофондам Европы.

По всем трем генетическим системам были изучены практически все современные народы, говорящие на языках балто-славянской группы – шестнадцать народов по единой обширной панели маркеров:

балтские народы – латыши и литовцы;

восточные славяне – белорусы, русские, украинцы; **западные славяне** – кашубы, поляки, словаки, сорбы, чехи;

южные славяне – болгары, боснийцы, македонцы, сербы, словенцы, хорваты.

Такие подробные и разносторонние данные по какой-либо группе народов (охват всех этносов, да еще и по всем основным генетическим системам) являются большой редкостью в популяционных исследованиях. Поэтому они позволяют решить не только конкретную, но и более общую методологическую задачу. Конкретная задача – это описание генофонда самих славян и балтов. Общая задача – на их примере изучить, как связаны друг с другом разные признаки, по которым обычно характеризуются популяции: Y-хромосомное разнообразие, митохондриальное разнообразие, полногеномное разнообразие, лингвистическое родство, географическое положение.

ВИЗУАЛИЗАЦИЯ ИТОГОВ СРАВНЕНИЯ

Генетические соотношения друг с другом всех изученных популяций, установленные в результате исследования, показаны на рис. 6.2.

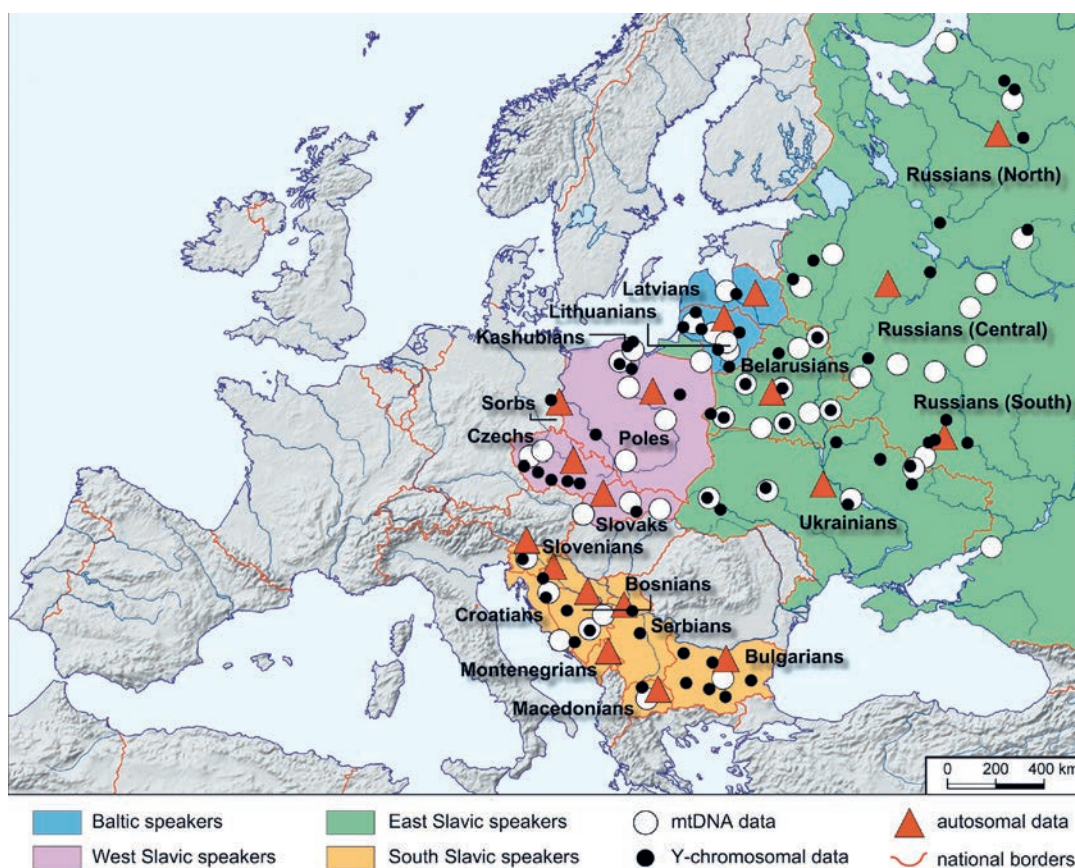


Рис. 6.1. Карта изученных балто-славянских популяций.

Зеленым фоном показаны восточные славяне, розовым – западные славяне, бежевым – южные славяне, голубым – балтские популяции. Значки обозначают местоположение популяций, в которых собраны образцы: проанализированные по мтДНК (белые кружки), проанализированные по Y-хромосоме (черные точки), проанализированные по широкогеномным аутосомным маркерам (красные треугольники).

Чтобы показать относительную близость и удаленность разных популяций на двумерном графике, в популяционной генетике традиционно используются два метода, заимствованные из многомерной статистики: метод главных компонент и метод многомерного шкалирования. По своей сути они близки, но достоинства и недостатки их противоположны. Метод главных компонент показывает положение популяций математически точно, но теряет порой значительную часть генетической информации, содержащейся в исходных данных. А метод многомерного шкалирования, наоборот, использует всю генетическую информацию, но геометрические расстояния между популяциями-точками на графике могут быть несколько искажены относительно рассчитанных генетических расстояний между ними. В данном случае для аутосомных данных был применен метод главных компонент, а для Y-хромосомных и митохондриальных данных – метод генетических расстояний.

Как видно, и по широкогеномным маркерам (рис. 6.2. А), и по Y-хромосоме (рис. 6.2. Б) большинство балто-славянских популяций выстраиваются вдоль оси «север-юг».

Восточные славяне – русские, белорусы и украинцы – отчетливо группируются вместе. Они образуют свой кластер, хотя внутри него русские, белорусы и украинцы не перекрываются полностью друг с другом. Исключение составляют северные русские популяции, которые генетически отдалены от остальных восточных славян и тяготеют к соседним финно-угорским популяциям.

Из западных славян чехи и в меньшей степени словаки отличаются от восточных славян и смещены в сторону немцев и других западноевропейских популяций. А вот поляки наиболее близки к восточным славянам. Фактически на графиках поляки, русские, белорусы и украинцы формируют общий кластер, а словаки и особенно чехи несколько удалены от него.

Южные славяне формируют на графике дисперсную группу, которая внутренне поделена на западный (словенцы, хорваты и боснийцы) и восточный (македонцы и болгары) регионы с сербами посередине. При этом словенцы генетически близки к венграм (географически близкий, но не славянский народ), а восточная ветвь южных славян тоже группируется с неславянскими, но гео-

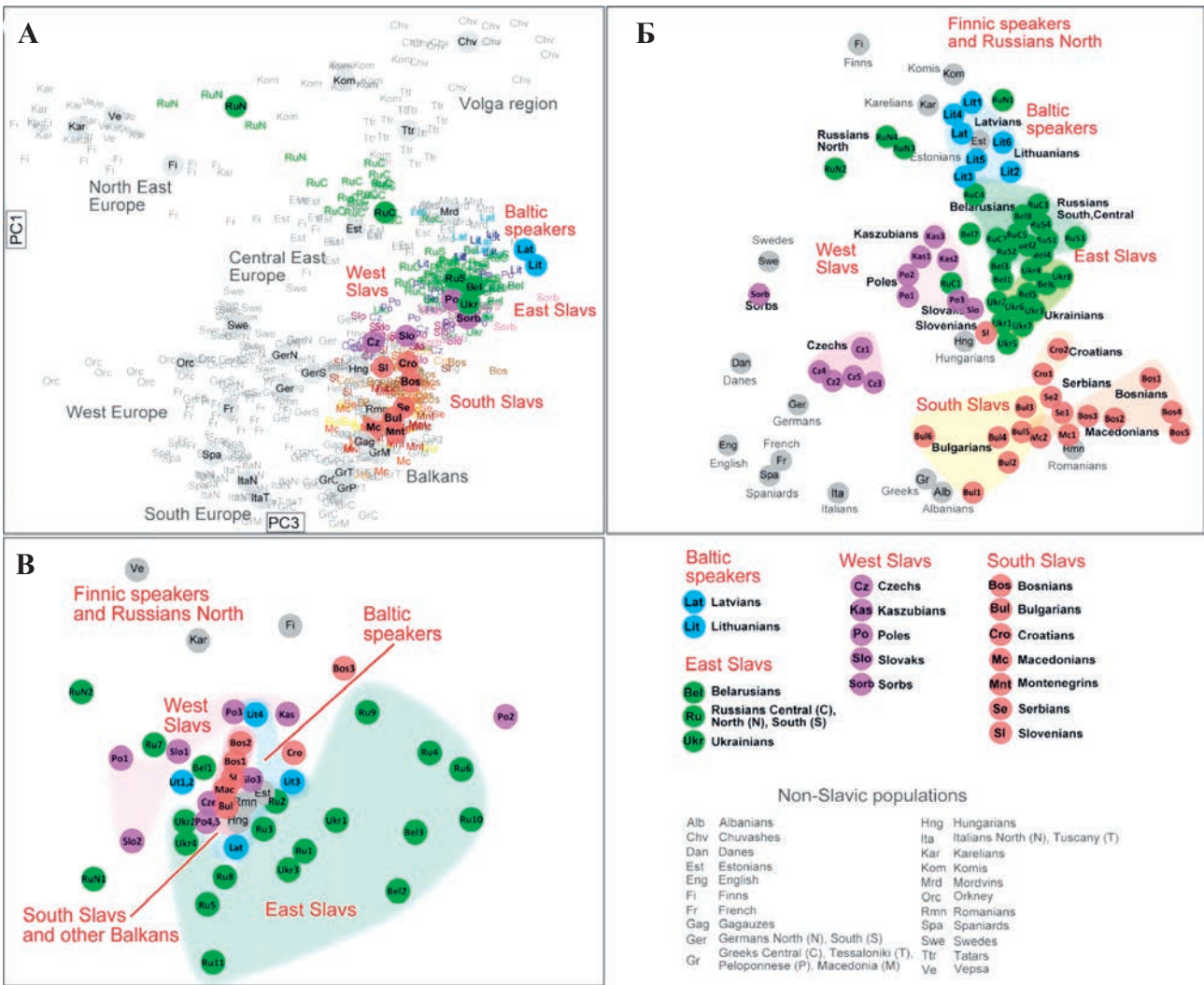


Рис. 6.2. Генетическая структура балто-славянских популяций в сравнении с другими народами Европы по трем генетическим системам:
 А) по аутосомным SNP-маркерам,
 Б) по Y-хромосоме,
 В) по митохондриальной ДНК.

графически близкими румынами и до некоторой степени с греками.

Балтские народы – латыши и литовцы – обнаруживают генетическую близость к эстонцам, говорящим на языке финно-угорской группы, и к популяциям восточных славян (белорусам). Оказалось также, что балтские популяции близки к волжской группе финно-угорских народов (особенно к мордве). Это может отражать исторические события – в древности ареал балтоязычных популяций простирался далеко на восток и почти доходил до нынешнего ареала мокши и эрзи.

Важно, что все перечисленные закономерности выявлены по независимым и совершенно разным генетическим системам – Y-хромосоме и широкогеномным аутосомным маркерам.

По мтДНК (рисунок В), как обычно, степень структурированности генофонда выражена значительно слабее, что связано с более низким филогенетическим разрешением в имеющихся данных по мтДНК. Но, хотя и не так четко выраженные, в результатах по мтДНК проглядывают те же самые закономерности. Например, и на графике по мтДНК большинство восточнославянских популяций перекрываются друг с другом, северные русские отделены от них, а южные славяне генетически сходны со своими не-славяноязычными соседями по Балканам.

Сравнение степени выраженности одних и тех же закономерностей по разным генетическим системам, показывает, что Y-хромосома часто выявляет закономерности более детально, чем не только мтДНК, но и более модные широкогеномные маркеры.

«ПРЕДКОВЫЕ» ГЕНЕТИЧЕСКИЕ КОМПОНЕНТЫ

ПРОГРАММА ADMIXTURE

ШЕСТЬ ПРЕДКОВЫХ КОМПОНЕНТОВ

Чтобы сравнить популяции по составу их предковых компонентов, часто используется программа ADMIXTURE («смешение», или «состав»). Напомним (глава 4), что при использовании этого метода в программу закладывают широкогеномные данные по большому числу популяций и задают число гипотетических предковых популяций, из которых сформировались все эти современные популяции. Программа вычисляет, каков должен быть генетический состав этих предковых популяций (предковых компонентов), и рисует для каждой современной популяции цветной спектр, указывающий на доли этих предков в ее генофонде. Понятно, что такая модель достаточно условна – в реальности вряд ли современные генофонды сформировались в результате смешения фиксированного заданного числа предковых популяций. Но такая упрощающая модель часто оказывается полезна, а выявляемые предковые компоненты обычно имеют реальный смысл, как подробно описано в главе 4.

При исследовании балто-славян мы задавали разное число предковых популяций (графики представлены в [Kushniarevich et al., 2015]), но специальный тест показал, что статистически наиболее обоснованные результаты получены в случае, когда число предковых компонентов задавалось равным шести ($K=6$). В этом случае получена картина, показанная на рис. 6.3.

У балто-славянских популяций почти весь спектр представлен двумя цветами: синим (предковый компонент k3) и голубым (предковый компонент k2), хотя и в разных пропорциях. Если посмотреть на Европу в целом, то видно, что k3 (синий) вносит большой вклад во все европейские популяции и снижается от северо-востока к югу. Этот предковый компонент максимален у балтских популяций, превалирует у восточных славян (80–95%) и снижается у южных славян (55–70%). Напротив, k2 (голубой) более характерен для популяций средиземноморского и кавказского реги-

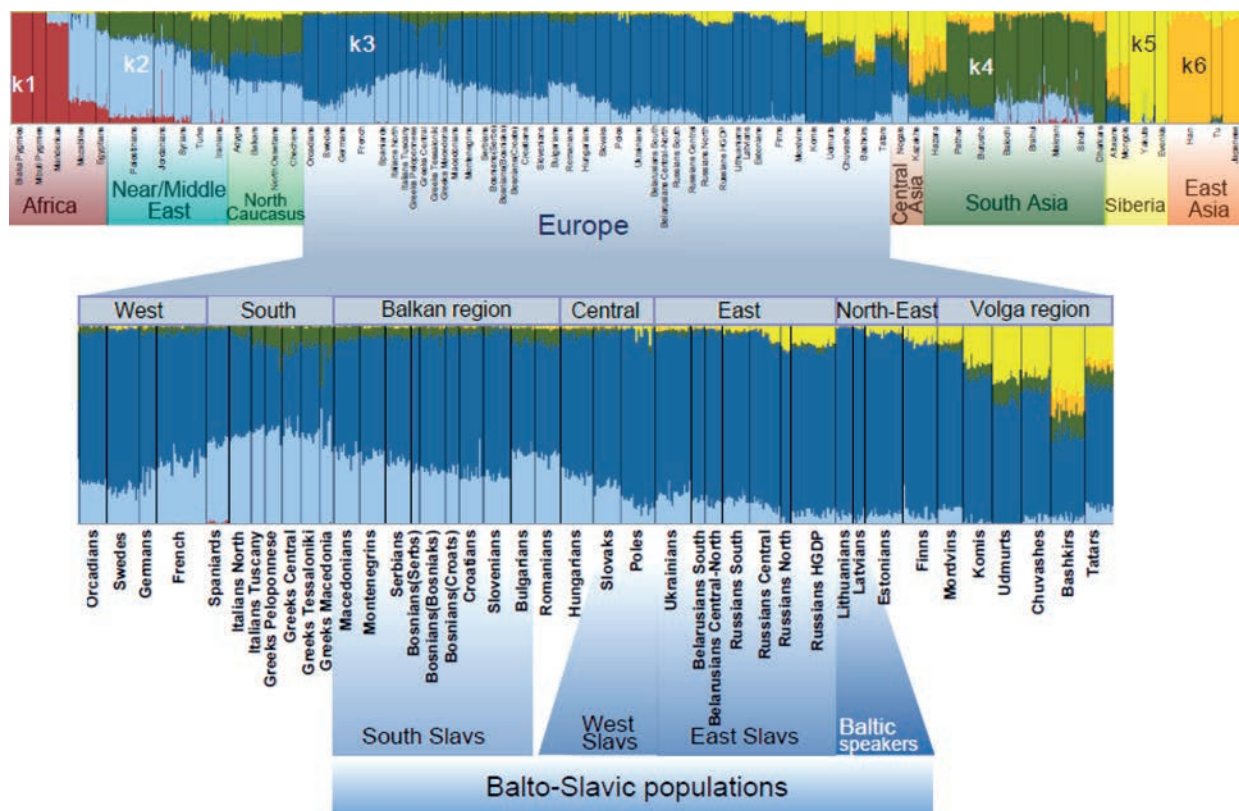


Рис. 6.3. Анализ предковых компонентов у балто-славянских популяций (метод ADMIXTURE).

Число заданных предковых популяций $k=6$. На нижней панели в увеличенном масштабе показаны спектры предковых компонентов для популяций Европы. Каждая вертикальная полоса в пределах каждого этноса – это характеристика одного изученного индивида. Поэтому ширина, занимаемая каждым этносом, соответствует числу изученных индивидов. А различия между вертикальными полосками – различия между индивидами в пределах этноса. (Russians HGDP соответствует выборке русских неизвестного происхождения – возможно, из Вологодской области – опубликованной в рамках проекта Human Genome Diversity Project).

онов и снижается к северу Европы. У южных славян на него приходится примерно 30%, у западных славян он снижается до 20%, а у северных русских и балтских популяций – до 5%.

Видно, что у славян есть еще лимонно-желтый цвет в предковом спектре, это компонент k5, который представлен сколько-нибудь значимо только у восточных славян, а из них больше выражен у северных русских. По происхождению этот компонент сибирский, поскольку, как видно на графике, он составляет основную часть спектра для популяций Сибири. А вот компонент k6 (темно-желтый), который доминирует в Китае, Монголии и на Алтае, у русских и других славян почти на нуле. Это означает, что восточный след в генофонде северных русских связан, скорее, с древними миграциями из лесов и тундр Сибири, чем из степей

Центральной Азии. Темно-зеленый компонент k4 характеризует популяции Южной Азии, он распространен также на Ближнем Востоке и Средиземноморье. Поэтому неудивительно, что он, пусть с небольшой частотой, но встречается у южных славян и других народов Балканского полуострова, но почти сходит на нет у западных и восточных славян.

Из рассмотрения состава предковых компонентов следует вывод о значительном генетическом сходстве большинства западных и восточных славян на большой территории – от Польши на западе до европейской части России на востоке. А южные славяне, географически ограниченные небольшим Балканским полуостровом, существенно отличаются от западных и восточных. Но как возникли эти отличия?

ОБЩИЕ ФРАГМЕНТЫ В ГЕНОМАХ

МЕТОД IBD – «IDENTICAL BY DESCENT»

Для ответа на этот вопрос проведен тонкий анализ генофонда для двух групп славян: в первую вошли западные и восточные славяне (ведь они генетически оказались очень схожи), а во вторую – южные славяне. Сравнение проводили по наличию одинаковых фрагментов хромосом у людей, происходящих из этих групп популяций. Этот метод называется IBD анализ – его название происходит от классического понятия популяционной генетики «*identical by descent*», то есть поиск генетических фрагментов, идентичных по происхождению. Эти фрагменты разными людьми, представителями разных популяций, унаследованы от одного и того же общего для них предка. Понятно, что чуть ли не в любой популяции мира может найтись хоть один потомок представителя западных и восточных славян, и, наоборот, среди восточных славян может найтись хоть один потомок чуть ли не любого народа мира. Но это будут лишь единичные совпадения – поэтому популяционная генетика и изучает популяции, а не отдельных ее представителей. Те популяции, в которых таких совпадений найдено много, находятся действительно в значительном родстве друг с другом, точнее, имеют значительное число общих предков.

Эти общие фрагменты, по сути, являются гаплотипами, похожими на гаплотипы мтДНК и Y-хромосомы тем, что также имеют одного предка, но отличными тем, что с ходом времени разбиваются рекомбинациями – обменом участками между хромосомами, пришедшими от отца и от матери, при делении клетки. А гаплотипы дают возможность датировок – зная скорость рекомбинаций, можно по длине сохранившихся общих гаплотипов оценить, сколько времени прошло от общих предков, то есть, давно ли существовала общность генофондов.

ОБЩИЕ ФРАГМЕНТЫ ГЕНОМОВ «ЗАПАДНО-ВОСТОЧНЫХ» СЛАВЯН И ИХ СОСЕДЕЙ ПО ЕВРОПЕ

Было подсчитано число общих гаплотипов между «западно-восточными» славянами (пришлось пользоваться этим неуклюжим термином за неимением лучшего) и восемью другими группами народов Европы (рис. 6.4):

- 1) южными славянами (болгары, боснийцы, македонцы, словенцы, хорваты);
- 2) популяциями Западной Европы (итальянцы, немцы, французы);
- 3) балтскими популяциями (латыши, литовцы)
- 4) популяциями Северо-Восточной Европы (западно-финские народы – вепсы, карелы, финны, эстонцы);
- 5) популяциями Центральной Европы, ареал которых находится между западно-восточными и южными славянами – они условно названы «между-славянскими популяциями» (это удивительно разноязычные популяции: гагаузы говорят на языке тюркской группы алтайской языковой семьи, венгры – на языке угорской группы уральской языковой семьи, а румыны – на языке романской группы);
- 6) греками;
- 7) популяциями Волжского региона и Приуралья (башкиры, коми, мордва, татары, удмурты, чувашаи);
- 8) северокавказскими популяциями (адыги, балкарцы, ногайцы).

Если принять за некий эталон число общих гаплотипов между западно-восточными и южными славянами, то часть окружающих неславянских популяций будет (по числу общих гаплотипов) выше уровня этого эталона, часть ниже, а часть равняться ему. Ниже уровня эталона (то есть име-

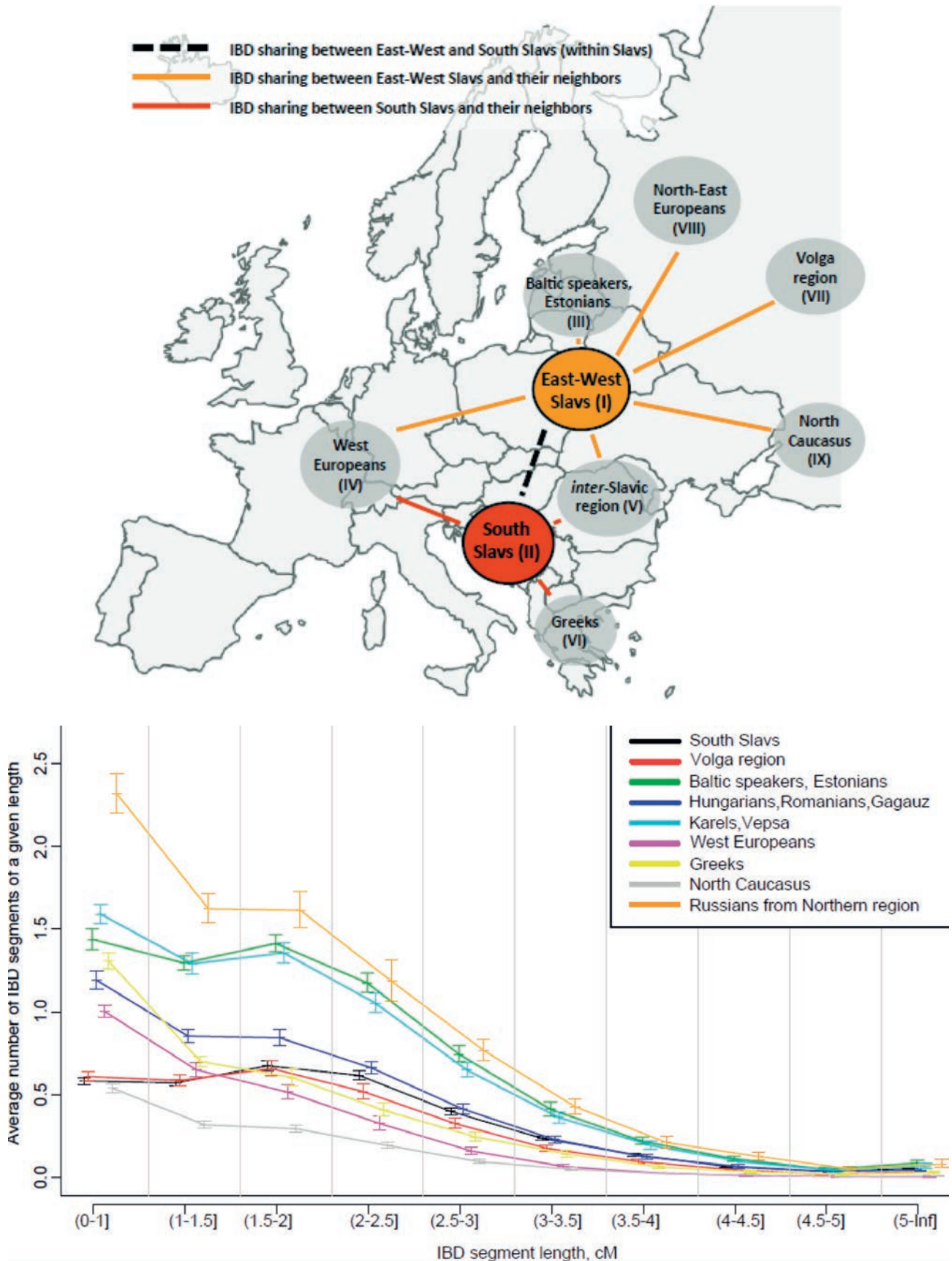


Рис. 6.4. Схема и результаты анализа общих фрагментов генома (IBD).

Вверху: схематическое представление групп популяций, использованных в анализе IBD (подсчет числа общих фрагментов генома). Внизу: графики среднего числа общих фрагментов генома (IBD сегментов) между группами западно-восточных славян и их географическими соседями (северные русские рассматриваются отдельно от группы популяций северо-востока Европы). По оси X показано 10 классов IBD сегментов по их длине (в сантиморганах, cM). По оси Y показано среднее число общих фрагментов для пары индивидов в пределах каждого класса.

ют меньшее родство с западно-восточными славянами, чем южные славяне) оказались народы Поволжья, Западной Европы, Кавказа, а также греки (рис. 6.4.).

Казалось бы, можно говорить о большем родстве славянских генофондов друг с другом, чем с окружающими неславянскими народами. Отчасти это так, но не все так просто – в два раза выше уровня эталона оказалось родство генофондов балтов и популяций северо-восточной Европы (вепсы, карелы, финны, латыши, литовцы, северные русские, эстонцы).

Можно впасть в противоположную крайность и считать, что «западно-восточные» славяне генетически родственны не южным славянам, а только своим географическим соседям, вероятно, за счет ассимиляции родственных им народов. Но картину дополнительно осложняет то, что с народами, живущими сейчас на территориях посередине между «западно-восточными» и южными славянами – то есть с венграми, румынами и гагаузами – у западно-восточных славян число общих фрагментов генома такое же, что и с южными славянами (эти «между-славянские» популяции находятся на уровне эталона).

ОБЩИЕ ФРАГМЕНТЫ ГЕНОМОВ ЮЖНЫХ СЛАВЯН И ИХ СОСЕДЕЙ ПО ЕВРОПЕ

Поэтому был проведен еще один аналогичный анализ, но теперь в центр рассмотрения поставлены южные славяне. Сравнивалось число общих генетических фрагментов у них и окружающих групп популяций. Оказалось, что число общих фрагментов у южных славян с «западно-восточными» славянами примерно такое же, что и число их общих фрагментов с «между-славянскими» популяциями (гагаузы, венгры, румыны). А вот число общих фрагментов с географически соседними греками значительно меньше.

Но учтем и географию. «Западно-восточные» славяне географически расположены дальше от южных славян, чем «между-славянские», поэтому с точки зрения географии число общих фрагментов с «западно-восточными» славянами должно было бы быть меньше. А раз это не так, значит язы-

ковое родство «западно-восточных» и южных славян отчасти проявляется и при этом анализе общих фрагментов генома.

Тем более что, хотя общие фрагменты генома, найденные между двумя группами славян, и различаются по длине, но фрагментов длиной около 2–3 сантиморганид чуть больше, чем других. А именно такой длины фрагменты и должны были сохраниться со времени славянской экспансии второй половины I тысячелетия н.э.

ВЫВОД: ОБЩИЕ ФРАГМЕНТЫ ГЕНОМА У СЛАВЯН ЕСТЬ, НО ИХ ДОЛЯ НЕВЕЛИКА

Эти результаты по славянам, из которых нельзя сделать однозначных выводов, следует сравнить с недавним похожим исследованием тюркоязычных популяций [Yunusbaev et al., 2015]. Казалось бы, в обоих случаях идет быстрое распространение носителей языков (соответственно, тюркских или же славянских) по обширным территориям, которое не может не сопровождаться ассимиляцией местного (дотюркского или же дославянского) населения. Но в случае с тюрками метод анализа общих фрагментов выявил – пусть очень небольшой – компонент генома, который тюрки принесли со своей вероятной алтайской прародины. А в случае славян картина оказалась гораздо более сложной. Возможно, это связано с тем, что тюрки в ходе расселения часто ассимилировали генетически резко отличные от них и друг от друга популяции, а славяне распространялись по территории Европы с ее относительно гомогенным генофондом, и часть ассимилированных ими популяций была генетически родственна, по крайней мере, балтским группам.

В целом, из этого анализа общих фрагментов генома можно сделать два вывода. Прежде всего, явно видны результаты смешения генофонда западно-восточных славян с другими популяциями северной части Восточной Европы. Во-вторых – пусть и далеко не столь выразительно – видна и несколько большая степень родства западно-восточных и южных славян друг с другом, чем можно было бы ожидать, исходя просто из географического расстояния между ними.

ДРЕВО СЛАВЯНСКИХ ЯЗЫКОВ И ЕГО ПРОЕКЦИЯ НА ГЕНОФОНД

РЕКОНСТРУКЦИЯ ДЕРЕВА ЯЗЫКОВ

Лексикостатистика занимается выявлением скорости языковых изменений, определением времени разделения языков и степени родства между ними. Исходным материалом послужили лексические списки (списки Сводеша) 20 современных балто-славянских языков и диалектов. Эти списки

были перепроверены и уточнены член-корр. РАН А.В. Дыбо и д.б.н. А.С. Касьяном, что позволило им получить уточненное древо родства балто-славянских языков (рис. 6.5.).

После разделения балтской и славянской ветвей самая первая развилка на славянской ветви оказалась тройной – разделение славян на западную, восточную и южную ветви – и датированной око-

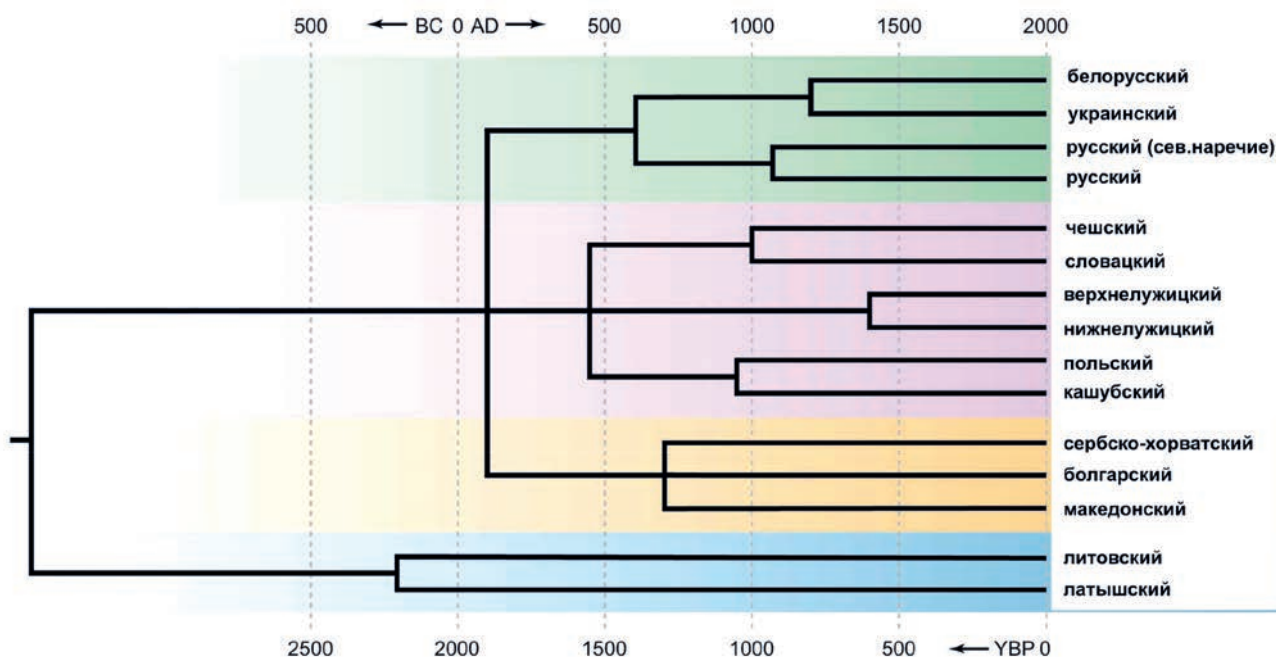


Рис. 6.5. Дерево родства балто-славянских языков.

Дерево построено путем сочетания нескольких методов (StarlingNJ, NJ, BioNJ, UPGMA, BayesianMCMC, UMP). Тройные узлы образованы путем объединения соседних двойных узлов, если временное расстояние между ними было < 300 лет.

ло 1900 лет назад. Дальнейшее разделение славянских языков началось в V–VI веках (около 1300–1500 лет назад): восточная ветвь разделилась на русский и украинский/белорусский, западная ветвь – на чешско/словацкий, протосорбсколужицкий и польский/кашубский, южная ветвь – на сербско-хорватский, болгарский, македонский. Выделение современных языков произошло 1000–500 лет назад. Такая датировка дерева соответствует историческим и археологическим данным, которые говорят о быстром распространении славян по Европе во второй половине 1-го тысячелетия н.э.

ГЕНЕТИЧЕСКОЕ РАЗНООБРАЗИЕ НА РАЗНЫХ УРОВНЯХ ЛИНГВИСТИЧЕСКОГО ДРЕВА

Поскольку лингвистическое дерево славянских языков имеет столь четкую и надежно реконструированную структуру, появилась возможность проанализировать, как распределено по этому дереву генетическое разнообразие славянских популяций, оцененное по частотам гаплогрупп Y-хромосомы. Этот анализ проведен с помощью стандартного метода AMOVA.

Оказалось, что генетические различия между популяциями, говорящими на одном и том же языке, хотя и варьируют от почти нулевых значений (для говорящих на чешском или македонском) до значения 0,05 (для говорящих на северных диалектах русского языка), в среднем составили только 0,01.

Далее, частоты во всех этих популяциях одного народа были усреднены и получены среднеэтнические частоты гаплогрупп Y-хромосомы. А затем были рассчитаны генетические различия между этими среднеэтническими характеристиками народов в пределах каждой ветви славянских языков. Эти различия оказались не совсем одинаковыми для разных ветвей: например, для западных славян различия больше, чем для восточных, но это и можно было ожидать, глядя на графики их генетических взаимоотношений. Однако в среднем различия между этносами оказались в три раза больше – 0,03.

Наконец, были рассчитаны средние частоты гаплогрупп для трех ветвей славянских языков – западных, восточных и южных – и различия между ними возросли еще в два раза – до 0,06.

Согласно принципу эквидистантности, разработанному отечественной школой геногеографии, если система популяций развивается сама по себе, без больших внешних влияний, то постепенное разделение популяций приводит к линейному накоплению и лингвистического, и генетического разнообразия. В результате генетическое разнообразие примерно одинаково на всех уровнях – что между популяциями одного народа, что между этносами одной ветви, что между разными ветвями (их усредненными характеристиками). Действительно, ведь предки разных групп славян некогда были лишь близкими друг к другу популяциями одного народа, и их языки отличались не больше,

чем сейчас отличаются диалекты одного языка. А усредняя частоты по всем современным популяциям ветви, мы находим ее центр тяжести, точку происхождения, реконструируем генофонд этой предковой популяции.

Но все это, как сказано выше, работает лишь тогда, когда популяции предоставлены самим себе и мало взаимодействуют с соседями. Однако для славян величины генетического разнообразия на разных иерархических уровнях не одинаковы: при эквидистантности они должны быть 0,03, 0,03, 0,03, а они резко различаются – 0,06, 0,03, 0,01. Это говорит о том, что популяции славян как раз активно смешивались с окружающими народами. А то, что наибольшая изменчивость приходится на самый древний уровень (различия между тремя ветвями славянских языков) указывает, что эти взаимодействия были особенно сильными на ранних этапах истории славянских популяций.

РОДСТВО ИЛИ СОСЕДСТВО?

Сравнить роли, которые сыграли география и языка в формировании генетического разнообразия балто-славянских популяций, можно с помощью теста Мантеля. География играет двойную роль. Конечно – это фактор географического соседства, который сближает генофонды через смешанные браки между соседями. Но с другой стороны, география может отражать и происхождение, когда родственные народы не уходят далеко друг от друга, а расселяются на соседние территории. Языки – это фактор изначального родства генофондов или частей генофондов, унаследованных от общих предков вместе с общим языком (или унаследованных без языка, если язык сменили, а генофонд остался почти прежним).

Тест был независимо проведен для трех генетических систем: Y-хромосома, мтДНК и аутосомные маркеры. Все три варианта теста показали чрезвычай-

Таблица 6.1. Анализ молекулярной изменчивости (AMOVA) в балто-славянских популяциях

Лингвистическая классификация	Разнообразие			
	Между группами (ветвями)	Между этническими популяциями в пределах групп	Между локальными популяциями, говорящими на одном языке	Внутри локальных популяций
А. Дерево языков 1 – балты: латыши, литовцы; 2 – восточные славяне: белорусы, русские, украинцы; 3 – южные славяне: македонцы, болгары, носители сербско-хорватских языков (хорваты, сербы, боснийцы); 4 – западные славяне: чехи, поляки, кашубы, словаки, сорбы.	5,93	3,53	1,2	89,34
В. Дерево с диалектами 1 – балты: латыши, литовцы; 2 – восточные славяне: белорусы, русские Большое Давыдовское, русские Смехново, русские Архангельск, украинцы Тисив, украинцы литературный; 3 – южные славяне: македонцы, болгары, носители сербско-хорватских языков (хорваты, сербы, боснийцы); 4 – западные славяне: чехи, поляки, кашубы, словаки, сорбы.	6,54	3,64	1,6	88,22

Примечания: Локальные популяции были сгруппированы по их языкам (или диалектам), которые были затем объединены в 4 группы, в соответствии с лингвистической классификацией. Состав этих групп указан в первом столбце таблицы. Для вычисления разнообразия на первом иерархическом уровне (локальные популяции в пределах языка) включали только языки с двумя и более изученными локальными популяциями. Поэтому латыши, сорбы и словаки, представленные только одной популяцией, не включались в этот анализ. Для вычисления на втором уровне рассчитывали разнообразие между среднеэтническими частотами в пределах каждой из четырех ветвей, а на третьем уровне эти среднеэтнические частоты усреднялись в частоты по лингвистическим ветвям и рассчитывалось разнообразие между ветвями.

Анализ был выполнен в двух вариантах: в первом популяции на первом уровне были объединены в пределах языка, во втором – в пределах диалекта.

Таблица 6.2. Тест Мантеля для генетических, лексикостатистических и географических расстояний

Тип корреляции	Параметры корреляции	Аутосомные данные	Y-хромосомные данные	Данные мтДНК
парная корреляция	Генетические и лингвистические	0.78*	0.76*	0.741*
парная корреляция	Генетические и географические	0.946*	0.924*	0.801*
частная корреляция	Генетические и лингвистические (при постоянной географии)	0,171	0,137	0.298*
частная корреляция	Генетические и географические (при постоянном языке)	0.861*	0.812*	0.525*

Примечание: * – статистически достоверный коэффициент корреляции (p-level < 0.05).

чайно высокую корреляцию между генетикой и географическим положением популяций (0,80–0,95). Но очень высокая корреляция обнаружена и между генетикой и лингвистикой (0,74–0,78).

Поскольку лингвистические показатели сами по себе высоко коррелируют с географией, рассмотрели частную корреляцию, чтобы различить прямое и непрямое влияние географии на две другие системы. При исключении географического фактора, частная корреляция с лингвистикой ста-

ла намного ниже (0,3 для мтДНК и 0,2 для остальных двух систем), в то время как для всех трех генетических систем корреляция с географией при исключении фактора лингвистики, осталась большой (0,5 для мтДНК и 0,8 для остальных двух систем). Это указывает на то, что связь с географическим фактором – основная, а высокая связь с лингвистикой часто определяется тем, что народы, говорящие на родственных языках, являются и географическими соседями.

СИНТЕЗ

ПЯТЬ СИСТЕМ ПРИЗНАКОВ

В генетических работах на каждом шагу встречается слово «анализ», и очень редко – «синтез». Что же означает «синтез» для исследования славянского генофонда?

Сравнение результатов генетических и лингвистических реконструкций с географией изучаемых популяций является традицией в популяционной генетике [Cavalli-Sforza et al., 1994; Рычков, Ящук (Балановская), 1980, 1983, 1985]. Уже говорилось о том, что проведенное исследование уникально тем, что одни и те же популяции (все современные народы балто-славянской лингвистической группы) были охарактеризованы по пяти системам: трем генетическим системам («широкогеномные» аутосомные данные, Y-хромосома, мтДНК), по лингвистической (лексикостатистической) системе признаков и по их географическому положению. Тем самым впервые в мире появилась – на примере балто-славянских популяций – возможность оценить скоррелированность всех пяти систем друг с другом. Это позволило на примере славян посмотреть, как связаны между собой три разные генетические системы, лингвистика и география – и синтезировать эти разнородные данные в общие выводы.

СВЯЗЬ КАЖДОЙ СИСТЕМЫ ДРУГ С ДРУГОМ

Материалом сравнения пяти систем послужили пять матриц расстояний, рассчитанных для одного и того же набора балто-славянских популяций по каждой системе. Далее с помощью теста Мантеля были рассчитаны попарные корреляции этих матриц друг с другом и получена матрица 5x5 корреляций (сходства) пяти систем друг с другом. Корреляции всех пяти систем (трех генетических, лингвистической и географической) друг с другом показаны на рис. 6.6.

Бросается в глаза очень высокое сходство всех пяти систем: ни один из коэффициентов корреляции не опускается ниже 0,68 – то есть фактически 0,7, что считается в популяционной генетике очень высокой корреляцией. А самые высокие коэффициенты почти достигают максимально возможного потолка (корреляция 0,95). Особенно удивительное соответствие обнаружено для Y-хромосомных и аутосомных маркеров и географического положения. Можно сказать, что эти три характеристики балто-славянских популяций образуют взаимосвязанную триаду (коэффициенты корреляции выше 0,9, темно-рыжий цвет на рис. 6.6 вверху).

Далее эта матрица сходства была преобразована в матрицу расстояний (бралась обратная величина сходства). Полученная матрица расстояний

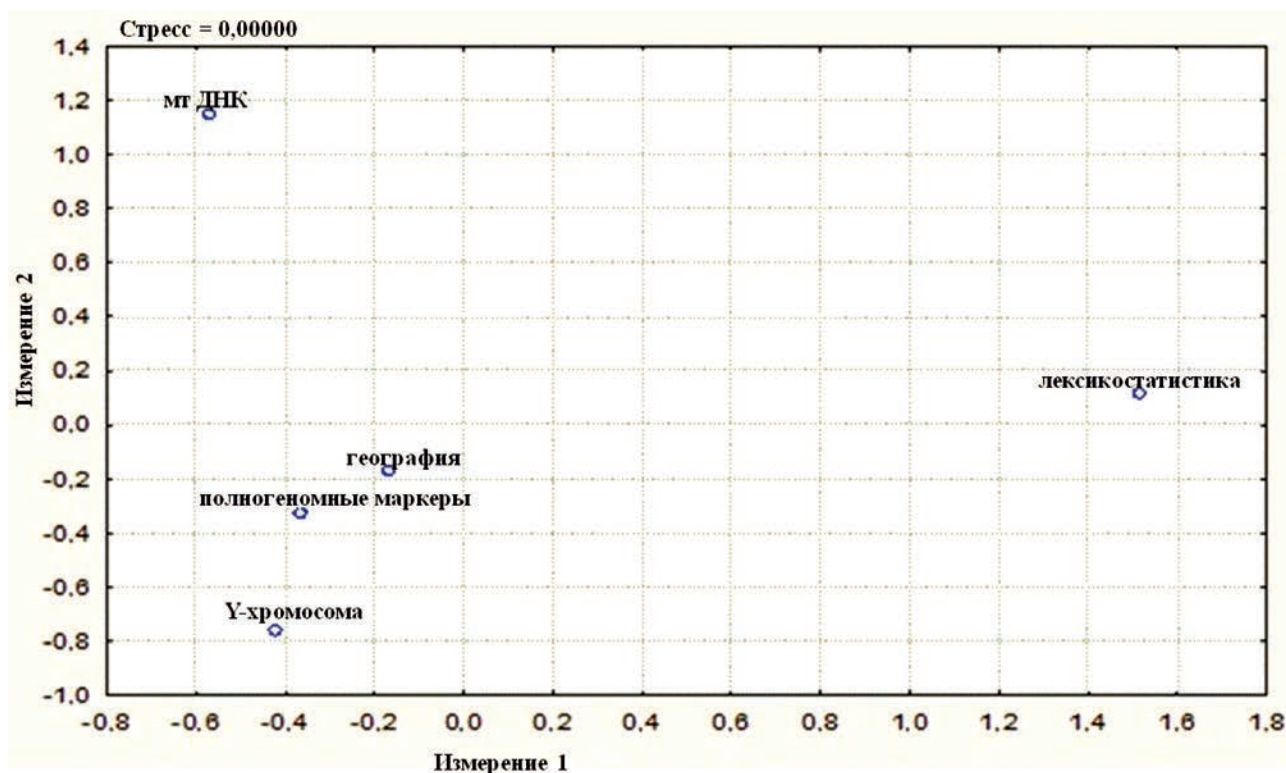
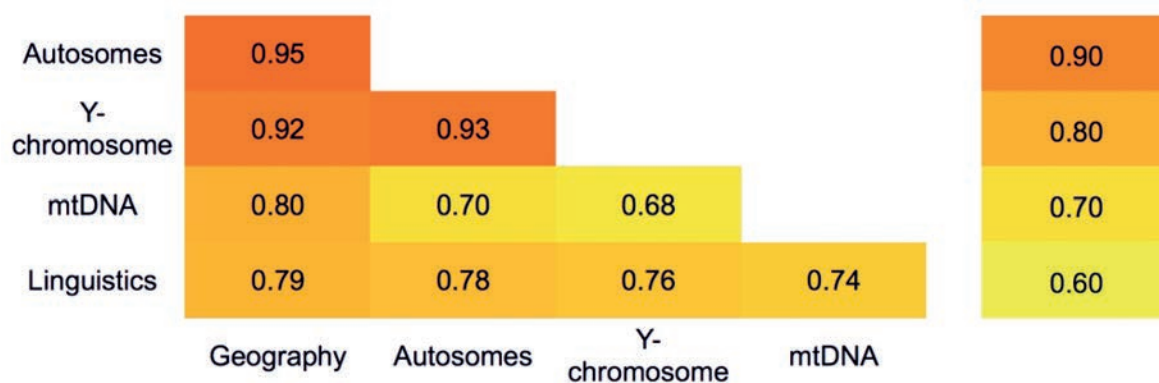


Рис. 6.6. Взаимосвязь изменчивости пяти систем признаков (Y-хромосома, мтДНК, полногеномные маркеры, география, лексикостатистика) в балто-славянских популяциях. Вверху: коэффициенты корреляции (столбец справа – цветовые обозначения интервалов коэффициентов корреляции). Внизу: график многомерного шкалирования на основе этих корреляций.

была визуализирована методом многомерного шкалирования (рис. 6.6 внизу). Этот график наглядно показывает тот же вывод о триаде особенно сходных систем: «аутосомная» реконструкция генофонда, «Y-хромосомная» реконструкция и географическое положение популяций. Остальные две системы – лингвистические характеристики и мтДНК – на графике отдалены, то есть рисуют чуть отличающиеся версии популяционной структуры.

ПОЛИСИСТЕМНЫЙ ПОДХОД

Соответствие результатов по разным системам признаков подтверждает надежность таких результатов. Оно также указывает на перспективность полисистемного подхода. Этот подход [Балановская, Балановский, 2007] состоит, во-первых, в па-

раллельном анализе разных систем признаков; во-вторых, в безусловном доверии лишь тем закономерностям, которые выявляются не по какой-то одной системе, а по большинству систем; в-третьих, во внимательном рассмотрении случаев, когда какая-то система выбивается из общего паттерна. В исследовании балто-славянского генофонда мы широко использовали полисистемный подход: формулируя утверждения о генетическом сходстве или различии тех или иных народов, каждый раз проверяли, подтверждаются ли они по большинству использованных систем. А то, что из общего паттерна выбивается лингвистика, послужило одним из аргументов в пользу гипотезы преобладания субстрата.

Такой паттерн почти полного совпадения трех генетических систем друг с другом, их совпаде-

ния и с географией, но лишь частичного сходства с лингвистикой может служить маяком и для будущих исследований генофондов других регионов мира. В то же время этот паттерн не универсален для всего мира: для популяций с контрастным происхождением мужской и женской частей популяции данные по Y-хромосоме и мтДНК могут сильно различаться (что показано, например, в статье [Quintano-Murci et al., 2008]), а для популяций, в которых процессы постепенного роста и дробления преобладали над метисацией, генетика может быть больше скоррелирована с лингвистикой, чем с географией (что показано, например, в статье [Balanovsky et al., 2011]).

Таким образом, хотя нет указаний на универсальность выявленного паттерна для всех регионов мира, можно предполагать, что выявленная для славянских популяций триада систем признаков окажется тесно связанной друг с другом и для других групп народонаселения Европы. Положение лингвистической реконструкции, как будет рассмотрено в главе 7 (раздел 7.5) зависит от истории народонаселения исследуемого региона (преобладание дифференциации одной общей прапопуляции или же преобладание экспансии с генетической ассимиляцией предшествующего населения, перешедшего на язык пришлого населения) и от того, насколько оба «зеркала» – лингвистическое и генетическое – будут «незамутненными».

Положение реконструкции по митохондриальному геному будет зависеть от характера брачных связей (интенсивности и широты ареала перемещения женщин при брачной миграции) и от четкости генетического зеркала мтДНК. Его четкость будет значительно повышаться при переходе к анализу полных геномов мтДНК, однако из-за того, что размер мтДНК невелик по сравнению с ядерным геномом, эта четкость в пределах все равно значительно ниже, чем Y-хромосомы. При этом показательна максимальная близость реконструкции по аутосомным маркерам и географическому соседству популяций.

Можно предполагать, что в общем случае степень «исторической» информативности генетических маркеров будет зависеть от степени их межпопуляционной изменчивости и следовать такой приоритетности: Y-хромосома, аутосомный геном, митохондриальная ДНК.

ДВА СУБСТРАТА В СЛАВЯНСКИХ ГЕНОФОНДАХ

Распространяясь по Европе, славяне должны были ассимилировать местные популяции, которые жили на данных территориях в дославянские времена. Это тот генетический субстрат, который они впитали в себя, и этот субстрат различается на разных территориях. Судя по описанным выше результатам, можно выделить два основных субстрата.

«Центрально-восточноевропейский субстрат» приняли в себя западные и восточные славяне (этот субстрат можно охарактеризовать различными способами – например, на спектре предковых компонентов он выражается синим цветом, а в данных по Y-хромосоме эти популяции несут высокие частоты гаплогруппы R1a). Другой, «южно-восточноевропейский субстрат», впитали в себя южные славяне (это голубой цвет в спектре предковых компонентов, а особенностью Y-хромосомного генофонда являются высокие частоты гаплогруппы I2a).

В пользу этого вывода о важности субстрата в формировании генофонда славян говорят три аргумента.

Во-первых, тот факт, что объединенная группа западных и восточных славян имеет меньшее число общих фрагментов генома с южными славянами, чем с популяциями северо-восточной Европы, включая балтские и финно-угорские народы. Особая генетическая близость финно-угров с балтами видна и на графиках главных компонент, и на графиках многомерного шкалирования. А как раз народы балтской и финно-угорской языковых групп и были расселены на той части Восточно-Европейской равнины, которая потом вошла в ареал славян.

Во-вторых, тест AMOVA также указывает на важную роль субстрата, поскольку генетическое разнообразие между разными ветвями славян намного превышает разнообразие внутри ветвей; такая картина и должна была сформироваться, если восточные и южные ветви славян ассимилировали генетически различные популяции.

В-третьих, преобладающая роль географии в формировании генофонда славян говорит о том же. Ведь если бы включения субстрата не было, то общее происхождение, фиксированное в языке, не могло не сказаться и на сходстве генофондов, даже когда какие-то группы славян мигрировали на далекое расстояние от своих родственников. Но такой роли лингвистического родства выявлено не было. И напротив: генетическое сходство между дославянскими популяциями, жившими на территории половины Европы, должно было быть примерно пропорционально географическим расстояниям между ними, но никак не связанным с языковым родством между славянскими группами, которые потом пришли на эти земли. Тогда, если в современных славянских генофондах преобладает субстрат, то и сходство этих генофондов должно следовать географическим расстояниям. Что и было выявлено.

ИСТОРИЯ СЛАВЯНСКИХ ГЕНОФОНДОВ: ИТОГИ

Итак, для генетического изучения славян использованы обширные генетические данные и воз-

возможность синтеза разных наук. Какие же выводы позволил сделать такой синтез?

Прежде всего, это преобладание в славянских популяциях дославянского субстрата – двух ассимилированных ими генетических компонентов – центрально-восточноевропейского для западных и восточных славян и южно-восточноевропейского для южных славян.

Но, несмотря на то, что в генофонде западных и восточных славян велик ассимилированный компонент их соседей по Восточно-Европейской равнине, эти славянские популяции формируют генетически довольно целостную группу, отличающуюся как от своих западных соседей (германоязычных популяций), так и от соседей восточных и северных (финно-угорских народов). Конечно, из этого правила можно найти пару исключений, но они сосредоточены на периферии ареала западных и восточных славян. Например, у своеобразного генофонда чехов есть определенное генетическое сходство с их немецкими соседями на западе, однако другие западнославянские популяции (поляки и сорбы) генетически четко отделяются от своих соседей-немцев. Аналогично, на другом конце славянского ареала, северные русские имеют ярко выраженное сходство с финно-угорскими и балтскими популяциями, но такого явного сходства не наблюдается для центральных или южных русских, не говоря уже о других славянских народах.

Поэтому можно предполагать, что после того, как прошел основной этап распространения славянских языков и ассимиляции дославянского субстрата, началось формирование местных особенностей генофонда. Оно протекало по-разному для разных частей обширного ареала западных и восточных славян, но изначальное родство (общий субстрат плюс общий славянский суперстрат) и, вероятно, интенсивный последующий обмен генов

внутри славянского ареала, сцементировали западных и восточных славян в единую генетическую общность.

Можно с осторожностью предположить, что ассимилированный субстрат мог быть представлен по преимуществу балтоязычными популяциями. Действительно, археологические данные указывают на очень широкое распространение балтских групп перед началом расселения славян. Балтский субстрат у славян (правда, наряду с финно-угорским) выявляли и антропологи. Полученные нами генетические данные – и на графиках генетических взаимоотношений, и по доле общих фрагментов генома – указывают, что современные балтские народы являются ближайшими генетическими соседями восточных славян. При этом балты являются и лингвистически ближайшими родственниками славян. И можно полагать, что к моменту ассимиляции их генофонд не так сильно отличался от генофонда начавших свое широкое расселение славян. Поэтому если предположить, что расселяющиеся на восток славяне ассимилировали по преимуществу балтов, это может объяснить и сходство современных славянских и балтских народов друг с другом, и их отличия от окружающих их не балто-славянских групп Европы.

Что же касается южных славян, то история их генофонда могла протекать схожим образом, хотя и независимо от западных и восточных славян. Южные славяне ассимилировали значительную часть дославянского населения Балкан, которая обладала иным генофондом, чем ассимилированное восточными и западными славянами население Восточно-Европейской равнины. Потому южнославянские популяции и обнаруживают большее сходство с неславянскими популяциями Балкан (румынами и венграми), чем с другими славянскими народами.

6.2. ОТДЕЛЬНЫЕ ЧЕРТЫ ГЕНОФОНДА ВОСТОЧНЫХ СЛАВЯН

Изложенные в предыдущем разделе результаты изучения генофонда славян являются обобщающими в обоих смыслах – и итоговыми, и порой слишком общими. Чтобы уравновесить эту некоторую отвлеченность и лаконичность предыдущего раздела, в этом разделе я опишу ряд подробностей – особенности тех или иных славянских генофондов. Эти результаты получены в разные годы и по разным маркерам. Поэтому не стоит на них основывать глобальные выводы о генетической истории славян – для этого есть предыдущий раздел. Но их можно использовать для уточнения тех или иных частных вопросов, а также для придания объемности уже обрисованному генетическому профилю славян.

ВЫБОРКИ, ИЗУЧЕННЫЕ ПО мтДНК

В ходе работы, потребовавшей многих лет и неоднократных экспедиций, нами созданы уникальные по объему и географическому охвату коллекции биологических образцов от коренного населения – русских, украинцев и белорусов, причем собранных и проанализированных строго по единой технологии. Экспериментальный анализ митохондриальной ДНК и Y-хромосомы в созданных ДНК коллекциях выполнен во многом при поддержке Российского фонда фундаментальных исследований, Российского гуманитарного научного фонда, международного проекта «Genographic» и благодаря сотрудничеству с Эстонским биоцент-

ром. Это позволило сделать генофонды восточнославянских народов одними из наиболее подробно изученных этнических генофондов мира.

По мтДНК исследованы тринадцать русских, пять украинских и пять белорусских популяций. В таблице 1.6 приведена характеристика изученных популяций. В ней (и в последующем анализе) две популяции из Западной Украины рассматриваются суммарно: увеличение объема выборки обеспечивает более высокую статистическую достоверность результатов. Суммарно проанализирована изменчивость мтДНК у 22 восточнославянских популяций, причем общий объем собственных данных составил $N=2396$ образцов.

Эта внушительная выборка была дополнительно расширена за счет привлечения опубликованных данных других авторов. Сравнительно с объемами нашего исследования их немного - суммарно 360 образцов. Это данные по двум северным русским популяциям [Belyaeva et al., 2003; Tonks et al., 2006], по населению Ростовской области [Kornienko et al., 2004; Richards et al., 2000] и белорусам Бобруйска [Behar et al., 2006]. В работах Б.А. Мальярчука также проанализировано около 500 образцов из различных областей европейской части России, но, к сожалению, стратегия формирования этих выборок (анкетирование лишь по родителям или даже по месту рождения самого пробанда) делает их несопоставимыми с нашими данными. По сходным причинам мы не рассматриваем и 101 образец, опубликованный в работе В.А. Орехова [Orekhov et al., 1999]. То же относится и к данным из работы [Mogozova et al., 2011]. Две другие имеющиеся в литературе небольшие выборки (русские Магадана – 50 образцов, и русские Башкирии – 83 образца) не включены в наше исследование, поскольку представляют русские популяции, проживающие за пределами исторического русского ареала. С учетом используемых литературных данных общая анализируемая выборка по восточным славянам составляет $N=2756$ образцов. Таким образом, работа основана преимущественно на собственных результатах (90% от суммарной анализируемой выборки).

Принципиальной особенностью анализируемого материала является не столько его объем, сколько качество сформированных выборок и их география. Для выявления географической структурированности восточнославянского генофонда было необходимо охватить всю обширную территорию расселения восточных славян, для чего требуется анализ многочисленных популяций. Исследование проведено в пределах исторического ареала восточнославянских народов. Та же цель – выявить не сиюминутный портрет генофонда, а его исторически устойчивые черты – определила и подход к формированию выборок. В обследование включались только те представители популяций, оба

дедушки и обе бабушки которых родились на территории данного изучаемого региона и относили себя к данному народу. Этот принцип «анкетирования до третьего поколения» в последние годы стал общепринятым в мировой науке, а нашим коллективом применялся, начиная с наших первых же экспедиций на Кавказе в 90^х годах. Также в обследование не включались родственники (как минимум до второй степени родства) – их заведомое генетическое сходство могло бы сместить и исказить получаемые генетические характеристики популяции. Во всех популяциях проводился забор венозной крови, который всегда сопровождался письменным информированным согласием.

Столь строгие требования к сбору первоначального биологического материала потребовали значительных и постоянных усилий для организации и проведения множества экспедиций, которые для восточных славян начались в 2000 году и проводятся нашим коллективом ежегодно по настоящее время.

СПЕКТР ГАПЛОГРУПП мтДНК У ВОСТОЧНЫХ СЛАВЯН

Большинство гаплогрупп мтДНК у восточнославянских народов принято объединять в два класса: восточно-евразийские гаплогруппы (наиболее распространенные у народов Восточной Азии и Сибири) и западно-евразийские (свойственные народам Европы, Передней и Южной Азии).

При этом восточные славяне целиком относятся к западному кругу популяций, поскольку западно-евразийские гаплогруппы составляют подавляющее большинство в их генофонде, а суммарная доля восточно-евразийских гаплогрупп не достигает даже двух процентов. В пределах западно-евразийских популяций по спектру гаплогрупп можно выделить две основных региональных группы – переднеазиатские и европейские народы (см., например, [Naak et al., 2010]). При таком делении восточные славяне полностью примыкают к генофонду европейских популяций (что, например, видно из невысокой частоты типично «переднеазиатской» гаплогруппы **J**). Далее в пределах Европы простое рассмотрение частот гаплогрупп неинформативно, поскольку все европейские народы практически идентичны по спектру и очень похожи по частотам гаплогрупп мтДНК. Такая гомогенность митохондриального генофонда Европы была даже предметом специальных публикаций [Richards et al., 2002; Simoni et al., 2000] и подробно рассмотрена в главе 3.

ГЕНЕТИЧЕСКИЕ ВЗАИМООТНОШЕНИЯ МЕЖДУ ВОСТОЧНЫМИ СЛАВЯНАМИ И ИХ СОСЕДЯМИ ПО мтДНК

Применение базы данных MURKA, содержащей как наши собственные данные, так и практи-



Рис. 6.7. Структура митохондриального генофонда восточных славян и их соседей. График многомерного шкалирования.

чески всю опубликованную информацию по митохондриальным генофондам народов Европы, позволило получить четкую картину их генетических взаимоотношений (рис. 6.7). При построении этого графика избегали его чрезмерной перегрузки точками-популяциями и потому оставили лишь те, которые имеют непосредственное отношение к этногенезу и этнической истории восточной ветви славян. Эти группы перечислены в легенде к рисунку и занимают обширный ареал от Балканского полуострова до Урала.

Общая структура митохондриального генофонда, выявляемая этим анализом, сводится к трем основным кластерам: Приуральскому, Восточно-Европейскому и Северно-Европейскому. Условные границы (пунктирные линии) на рис. 6.7 показывают примерное генетическое пространство, популяции которого относятся к каждому из трех перечисленных кластеров.

В Приуральский кластер вошли восточные финно-угорские (марийцы, коми) и тюркские народы Приуралья (татары, чуваша), в генофонде которых предполагается значительный финно-угорский субстрат. В Восточно-Европейский кла-

стер вошли западные и восточные славяне, а также народы Балканского региона. В этот же кластер вошли и татары-кряшены – они генетически удалены от всех других изученных нами популяций татар, но приближены к русским, что, возможно, является результатом смешанных браков с единоверным русским населением. В Северно-Европейский кластер вошли западные финно-угорские народы и северные русские. Из народов Кавказа два народа, не принадлежащие к северокавказской языковой семье (представители индоевропейской семьи на Кавказе: осетины и армяне) в целом близки к восточно-европейскому кластеру, но народы северокавказской языковой семьи оказались ближе к народам Передней Азии, генетически весьма удалены от генофонда Восточной Европы и потому не могут быть отображены в «генетическом масштабе» данного графика. Итак, выявляется единая, генетически гомогенная балто-славянская общность, к которой тяготеют и популяции Балканского региона. От этой общности отличаются финно-угорские и тюркские народы Восточной Европы, генофонд которых, в свою очередь, подразделяется на западную и восточную части.

Сравнение расположения популяций в генетическом пространстве и на географической карте показывает сходство в основных чертах (чтобы увидеть это сходство, достаточно допустить, что «восток» в генетическом пространстве находится слева, а «север» – внизу). Лишь русские популяции, которые географически должны бы тяготеть к Приуральскому кластеру, показывают сходство с расположенными намного западнее группами балтов и западных славян. Конечно, такая картина могла сформироваться в ходе славянской колонизации Восточно-Европейской равнины:двигающееся с запада славянское население несло с собой «восточно-европейский» митохондриальный генофонд, который замещал ранее существовавший на этих территориях автохтонный «приуральский» генофонд. Однако результаты полногеномных исследований, дающие очень слабый сигнал собственно общеславянского генофонда, позволяют выдвигать гипотезу, что пришлое славянское население и дославянское население в пределах основной части исторического ареала русских (центральные и южные русские популяции) было генетически сходным и, возможно, генетически было наследником волны расселения балтов по Восточно-Европейской равнине.

Эту гипотезу, возможно, подтверждает и своеобразие генофонда северных русских популяций, сформировавшихся на ином генетическом субстрате – Северо-Восточной Европы. Вопрос об ассимиляции славянами автохтонного населения, многократно обсуждавшийся в антропологической литературе, также находит отражение в данных по митохондриальной ДНК. Митохондриальный генофонд северных русских популяций более сходен не с основной массой русских, украинских и белорусских популяций, а с неславянскими народами севера Восточной Европы. Можно предполагать, что славянская колонизация на своих поздних этапах (Русский Север) включала в себя ассимиляцию генетически несходного со славянами местного населения, и в генофонде северных русских до сих прослеживается этой компонент. Это и приводит к тому, что их размещение на графике – промежуточное между восточнославянскими и неславянскими популяциями. Вопрос о том, что представлял этот дославянский субстрат, который впитали северные русские, рассмотрен в конце этой главы (раздел 6.4).

Что касается украинских популяций, то три из четырех изученных популяций располагаются практически в той же области графика, что и западнославянские, и балтоязычные популяции. Это согласуется с установленным по картам генетических расстояний тяготением украинского митохондриального генофонда к популяциям, расположенным севернее – поляков, белорусов, южных русских (см. раздел 6.3).

СООТНОШЕНИЯ МЕЖДУ МИТОХОНДРИАЛЬНЫМИ ГЕНОФОНДАМИ ВОСТОЧНЫХ СЛАВЯН

Определив место восточнославянских групп в генофонде Европы, перейдем к рассмотрению внутренней структуры восточнославянского генофонда. Для этого построен аналогичный график многомерного шкалирования, но на него выведены только восточнославянские популяции (рис. 6.8).

График наглядно показывает сходство митохондриального генофонда всех восточнославянских популяций друг с другом: все они образуют единый малодифференцированный кластер. Такая генетическая гомогенность восточных славян находит параллели в их антропологическом и лингвистическом сходстве друг с другом. Однако при более пристальном рассмотрении начинает проглядывать структура генофонда: все-таки существующие различия между популяциями по мтДНК. Так, большинство белорусских популяций образуют скопление в левой части графика. Далеко в правой части графика, отдельно от всех прочих групп, располагается популяция белорусов Бобруйска (Bbobr). Эта выборка была изучена другими авторами и в частотах гаплогрупп «прочие» гаплогруппы достигают очень высокой величины – 17%, что видимо, связано с методическими особенностями генотипирования. Это означает, что особенность этой популяции – кажущаяся, и вызвана она, скорее всего, техническими причинами генотипирования.

Все три северные русские популяции (Ошевенск, Пинега, поморы) смещены вправо. Шесть русских популяций образуют единое облако со всеми украинскими популяциями, а также белорусами и кубанскими казаками. Эта группа оказывается ядром восточнославянского генофонда (к ней относятся 12 из 26 изученных популяций). Остальные шесть русских популяций занимают различное положение на графике. В большинстве случаев, видимо, это объясняется статистическими случайностями, неизбежными при многомерном шкалировании большого числа популяций (поскольку двумерные геометрические расстояния никогда не могут в точности соответствовать реальным многомерным генетическим расстояниям).

Так, парадоксально расположение тамбовских русских (Rpetr) рядом с северными; кажется странным положение Курской и Орловской групп (Rche, Rbol), а также одной из двух рязанских популяций (Rspas), поскольку вторая рязанская популяция вошла в основную восточнославянскую генетическую общность. Лишь особенность русских Ростовской области (Rros) и терских казаков (Rter) легко может быть объяснима их периферийным положением на самом юге исторического русского ареала. Любопытно, что группа кубанских казаков

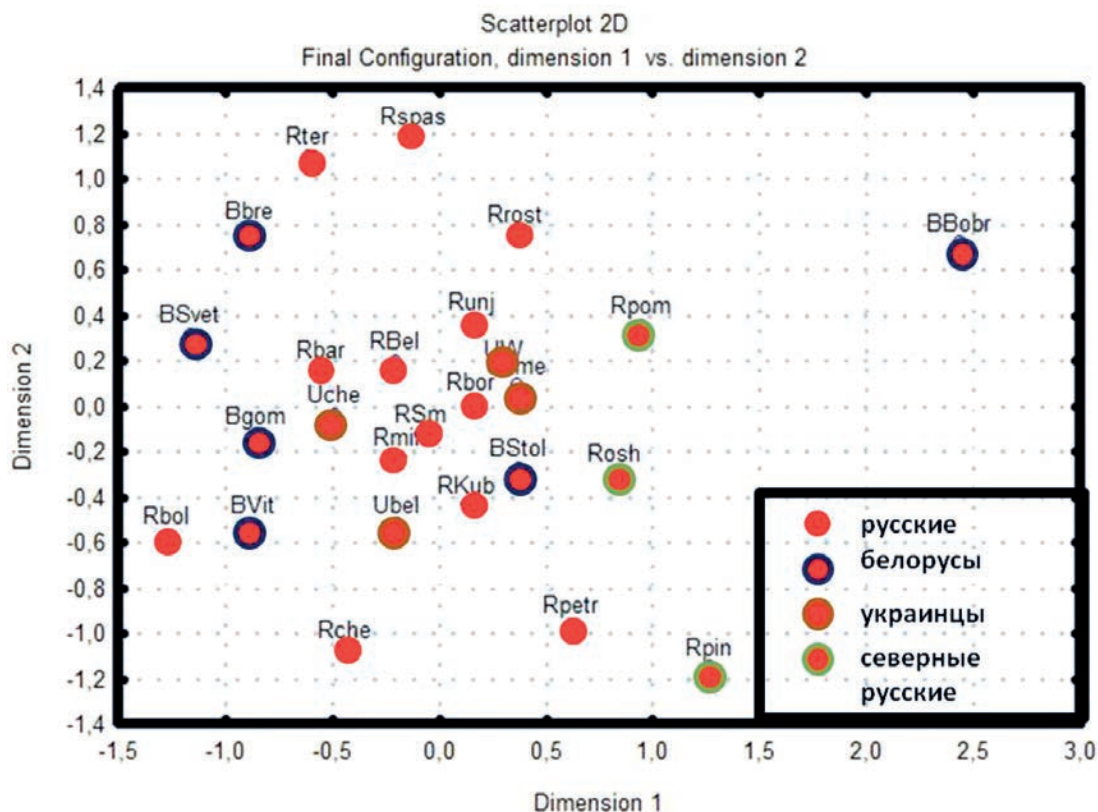


Рис. 6.8. График многомерного шкалирования русских, украинских и белорусских популяций по мтДНК.

Обозначения: Русские: Rspas – Спасск-Рязанский р-н, Rrost – Ростовская обл., Runj – Унжа (Костромская обл.), RBel – Белгородская обл., Rbar – Брятинский р-н, RSm – Смоленская обл., Rbor – Боровский р-н, Rbol – Болховский р-н, Rche – Черемисиновский р-н, Rkub – кубанские казаки, Rter – терские казаки, Rpetr – Петровский р-н, Rpin – Пинежский р-н, Rosh – Ошевенск (Каргопольский р-н), Rrom – поморы. Белорусы: Bgom – Гомельская обл., BVit – Витебская обл., BStol – Столинский р-н, BSvet – Светлогорский р-н, Bbre – Брестская обл., BBobr – Бобруйский р-н. Украинцы: UW – западные, Uche – Черкасская обл., Ubel – Белгородская обл., UHme – Хмельницкая обл.

(Rkub) неотличима от основного восточнославянского ядра. Аналогичное поведение тех же групп казаков выявлено нами по маркерам Y-хромосомы [Balanovsky et al., 2008].

Итак, рассмотрение генетических взаимоотношений двадцати шести восточнославянских популяций по маркерам митохондриальной ДНК показывает:

- выходящую на первый план гомогенность восточнославянского генофонда;
- умеренное своеобразие большинства белорусских популяций (что может объясняться наличием балтского компонента в составе их генофонда);
- выраженное своеобразие северных русских популяций;
- возможное своеобразие терского казачества.

МЕЖПОПУЛЯЦИОННАЯ ИЗМЕНЧИВОСТЬ ТРЕХ ТИПОВ МАРКЕРОВ

Отметим особенности трех генетических систем – Y-хромосомы, мтДНК и аутосомных ДНК маркеров, важные для интерпретации результатов.

Однородительские маркеры гаплоидны, в отличие от аутосомных маркеров, что уменьшает генетически эффективный размер популяции (N_e) для однородительских маркеров в два раза. К тому же, являясь однородительскими, они отражают генофонд только одной – мужской или женской – половины генофонда, что уменьшает N_e еще в 2 раза. В результате N_e для однородительских маркеров в четыре раза меньше, чем для аутосомных, а дрейф генов ($1/N_e$), соответственно, действует на них в четыре раза интенсивнее, чем на аутосомные маркеры.

Вторая, более важная особенность связана с уровнем дифференциации популяций. Для населения с патрилокальной традицией межпопуляционное разнообразие по маркерам Y-хромосомы значительно выше, чем разнообразие по мтДНК и по аутосомным маркерам. В изученных нами восточнославянских популяциях (рис. 6.9. А) межпопуляционное разнообразие Y-хромосомы в 7 раз выше, чем разнообразие по мтДНК и по аутосомным ДНК маркерам [Пшеничнов, 2007]. Те же закономерности выявлены и для коренного населения Кавказа (рис. 6.9. Б). На Кавказе межпопуля-

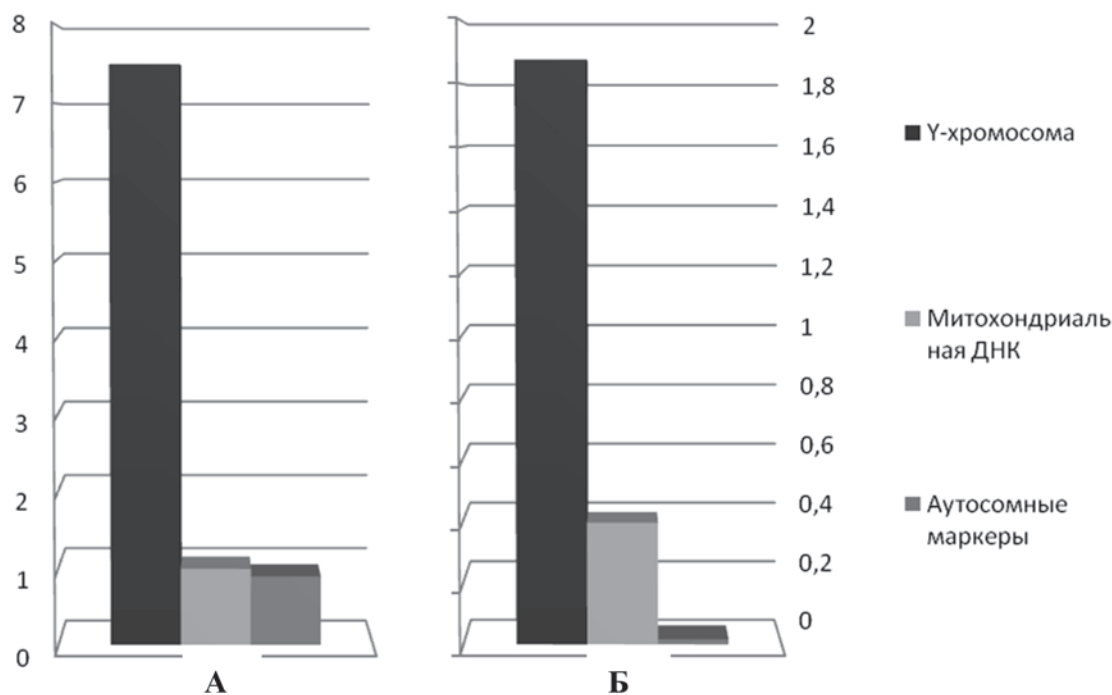


Рис. 6.9. Межпопуляционное разнообразие по Y-хромосоме, мтДНК и ауtosомным (преимущественно ALU) маркерам. А – для восточнославянских популяций (межпопуляционное разнообразие G_{ST}); Б – для коренных народов Кавказа (генетические расстояния d_{Nei}).

ционное разнообразие Y-хромосомы в 4 раза выше, чем по мтДНК (размах изменчивости которой на Кавказе повышен за счет народов Дагестана, сочетающих, благодаря кровнородственным бракам, патри- и матриликальные традиции) и на 2 порядка больше, чем по ауtosомным ДНК маркерам [Дибирова, 2011; Почешхова, 2008].

СРАВНИТЕЛЬНЫЙ АНАЛИЗ С АУТОСОМНЫМИ ДНК МАРКЕРАМИ

Хотя в «полногеномную эру» исследования, основанные на малом числе маркеров, вышли из моды, рискну все же предложить читателю результаты нашего исследования пяти ауtosомных маркеров – Alu-инсерций ACE, APOA1, B65, PV92, TRA25 [Соловьева и др., 2010]. Ведь значение имеет не только число маркеров, но и число популяций, и качество формирования выборок, а по этим параметрам полногеномные исследования не всегда могут похвалиться первенством. В описываемое исследование были включены наши собственные данные по десяти русским популяциям (суммарная выборка 1088 человек, табл. 6.3), охватывающих весь исторический ареал русского народа, сопоставленные с другими народами Восточной Европы, изученными по той же панели маркеров.

По разным локусам русские популяции проявляют сходство то со своими западными (народы Западной Европы), то с восточными соседями (на-

роды Уральского региона). По частотам изученных Alu-инсерций русский генофонд оказывается малодифференцированным: различия между десятью региональными популяциями составляют лишь $d=0.007$, тогда как по классическим маркерам, митохондриальной ДНК и Y-хромосоме гетерогенность русского генофонда заметно выше (0.013, 0.033 и 0.142, соответственно). Таким образом, данный набор из пяти Alu-инсерций характеризуется сниженной изменчивостью на внутривнутриэтническом уровне – в пределах русского этноса.

При кластеризации популяций проявляется не только их географическое соседство, но и этническое (лингвистическое) сходство. Так, на графике многомерного шкалирования (рис. 6.10) все русские популяции образуют единый кластер, хотя географически некоторые из них расположены ближе к народам других регионов. Впрочем, кубанские и терские казаки, которые географически живут чересполосно с народами Северного Кавказа, и в генетическом пространстве расположились вблизи северокавказских популяций.

Однако при межэтнических сравнениях обнаруживается четкая картина: тринадцать рассмотренных народов Восточной Европы образуют три кластера в соответствии с их историко-географическим положением: восточнославянский (ВС), кавказский и южноуральский кластеры (рис. 6.10). Отметим, что генетические различия между восточнославянскими, кавказскими и уральскими популяциями выявлялись и по данным об ауtosом-

Таблица 6.3. Частоты аутосомных Alu-инсерций и величины межпопуляционного разнообразия русских популяций

Популяции	N	ACE	АРОА1	В65	PV92	ТРА25
Лешуконье	91	0,533	0,962	0,489	0,181	0,467
Пинега	142	0,520	0,914	0,492	0,179	0,510
Красноборск	114	0,504	0,944	0,400	0,212	0,550
Вологда	118	0,487	0,912	0,552	0,175	0,504
Кострома	75	0,506	0,905	0,590	0,173	0,514
Кашин	101	0,531	0,913	0,524	0,209	0,530
Псков	111	0,494	0,930	0,514	0,078	0,496
Смоленск	144	0,466	0,922	0,401	0,137	0,445
КазакИ терские	50	0,500	0,939	0,440	0,290	0,610
КазакИ кубанские	142	0,504	0,950	0,602	0,219	0,546
В среднем	1088	0,505	0,929	0,500	0,185	0,517
Межпопуляционные различия	0,007	0,002	0,000	0,020	0,006	0,008

Обозначения: N – объем выборки в среднем по 5 локусам.

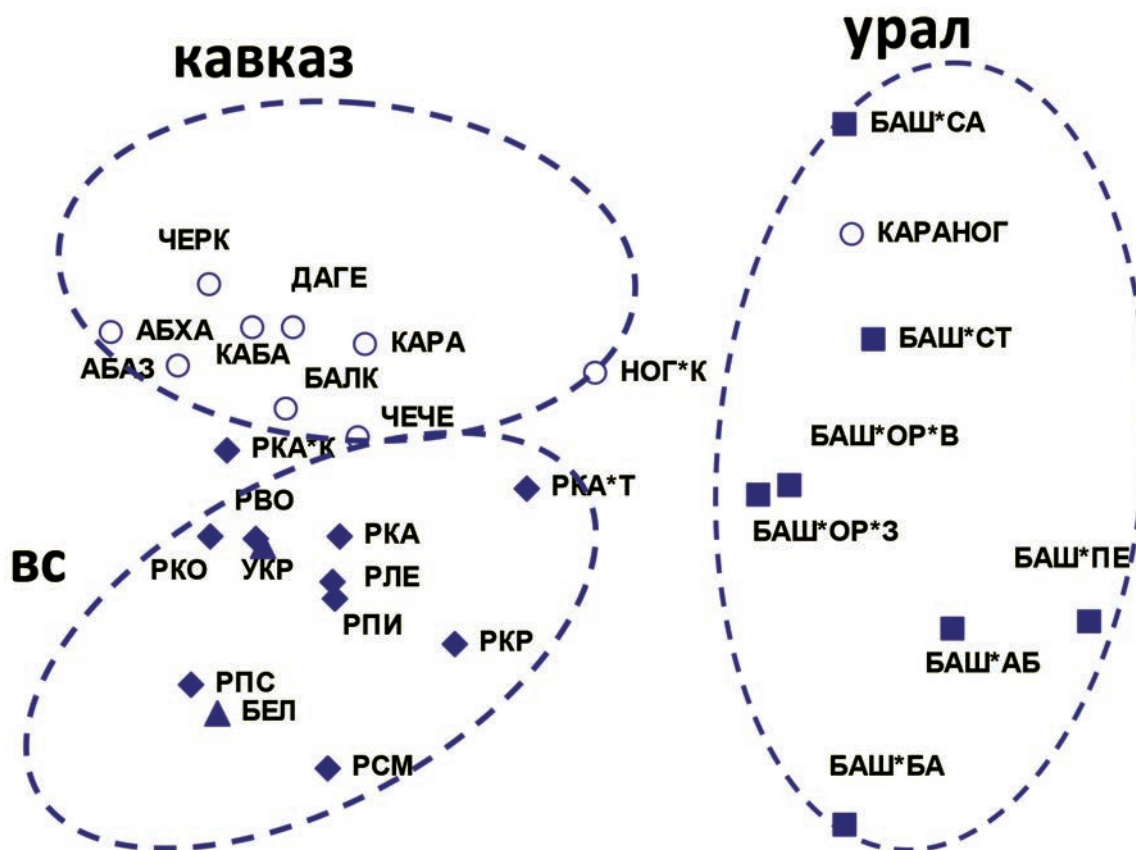


Рис. 6.10. Генетические взаимоотношения русских популяций и соседних народов по аутосомным STR-маркерам (график многомерного шкалирования). Обозначения: русские популяции обозначены ромбами, другие восточнославянские (ВС) народы – треугольниками, народы Урала – квадратиками, народы Кавказа – кружками. РВО – Вологда, РКО – Кострома. РСМ – Смоленск, РКА – Кашин, РПИ – Пинега, РПС – Псков, РЛЕ – Лешуконье, РКР – Красноборск, РКА*Т – КазакИ терские, РКА*К – КазакИ Прикубанья, УКР – Украинцы, БЕЛ – Белорусы, ЧЕРК – Черкесы, КАБА – Кабардинцы, АБХА – Абхазы, АБАЗ – Абазины, КАРА – Карачаевцы, БАЛК – Балкарцы, НОГ*К – Кубанские ногайцы, КАРАНОГ – Караногайцы, ДАГЕ – Дагестан, ЧЕЧЕ – Чеченцы. БАШ*ОР*З Башкиры Оренбургской области западные, БАШ*ОР*В – Башкиры Оренбургской области восточные, БАШ*СА – Башкиры Самары и Саратова, БАШ*БА – Башкиры Баймакского района, БАШ*СТ – Башкиры Стерлибашевского района, БАШ*АБ – Башкиры Абзелиловского района, БАШ*ПЕ – башкиры Пермской области.

ных классических маркерах, и по данным об аутосомных микросателлитных локусах [Лимборская и др., 2002]. Такое сходство изменчивости Alu-инсерций с паттернами, выявленными по другим маркерам, является дополнительным аргументом в пользу объективности выявляемых закономерностей. Итак, три проявившихся кластера популяций соответствуют трем этногеографическим регионам (восточные славяне, Кавказ, Южный Урал).

Таким образом, и по аутосомным маркерам, как и по гаплогруппам мтДНК и Y-хромосомы также вырисовывается единый гомогенный кластер вос-

точных славян, удаленный и от коренного населения Кавказа, и от народов Волго-Уральского региона.

При этом по всем использованным генетическим системам – мтДНК, Y-хромосома, отдельные аутосомные маркеры, широкогеномные панели – рельефно проявляется генетическое своеобразие северных русских популяций. Это отличие формирует основную – широтную – закономерность в изменчивости русского генофонда. Поэтому важно рассмотреть причины генетического своеобразие Русского Севера, чему посвящен раздел 6.4.

6.3. МИТОХОНДРИАЛЬНЫЙ ГЕНОФОНД УКРАИНЦЕВ

Задача этого раздела заключается в выявлении степени связи митохондриального генофонда одного из восточнославянских народов – украинцев – с генофондами других народов Европы. Для решения этой задачи предложен метод, выявляющий группы гаплотипов мтДНК по принципу сходного географического распространения.

КАК УЛОВИТЬ РАЗЛИЧИЯ В ГОМОГЕННОСТИ?

Территория современной Украины в древности входила в зону масштабных демографических процессов. Около 45 тысяч лет назад, когда люди современного антропологического типа заселяли Европу, область к северу от Черного моря находилась на путях интенсивных миграций. Предполагается, что позднее, во время максимума последнего оледенения, эта область входила в зону одного из рефугиумов – тех территорий Европы, где сохранялось население. Другие рефугиумы находились – около Пиренеев, Альп и на Балканах. Возможно, что из этих рефугиумов – в том числе и из причерноморского – в мезолите была заново заселена Европа. В период от неолита до славянской экспансии Причерноморье приняло множество миграционных волн, которые привели к смешению популяций и замещению одних другими. Эти сложные демографические процессы и сформировали генофонд народов Причерноморья, в том числе украинцев.

Митохондриальный генофонд Европы изучен вдоль и поперек. И главный вывод, который был

сделан – о гомогенности этого генофонда, огромном генетическом сходстве популяций, живущих даже на разных концах Европы. Хотя генофонд Европы относительно гомогенен по всем генетическим маркерам, но по митохондриальной ДНК эта гомогенность особенно сильна. Поэтому при изучении генофонда какого-то одного народа заранее известно, что он будет очень похож на генофонды остальных европейских народов и трудно будет выделить, на кого он похож в особенности. Отчасти эта проблема преодолевается полным секвенированием мтДНК – тогда выделяются субветви, имеющие не совсем повсеместное распространение, а приуроченные хотя бы к крупным регионам Европы. Но для украинцев таких данных по полному секвенированию мтДНК пока очень мало. Выход был в том, чтобы получить максимально подробные данные по частичному секвенированию (ГВС1 и ГВС2) и постараться выявить закономерности с помощью эффективных методов анализа данных. Чтобы в общей картине гомогенности европейских генофондов мтДНК все же уловить различия в степени сходства украинцев с разными народами Европы, были применены два наиболее информативных подхода. Во-первых, один из наиболее чувствительных методов геногеографии – картографирование генетических расстояний. Во-вторых, был разработан новый метод – анализ географической скоррелированности гаплотипов.

КАК ФОРМИРОВАЛИСЬ ПОПУЛЯЦИОННЫЕ ВЫБОРКИ УКРАИНЦЕВ

Поскольку изучение украинского генофонда вызывает в последние годы пристальное внимание научной и лженаучной общественности, опишем подробнее методы формирования популяционных выборок украинцев в нашем исследовании.

ГЕОГРАФИЯ ВЫБОРОК

Мы провели анализ мтДНК украинцев по обширной выборке 607 образцов [Pshenichnov et al., 2013], включившей основные этнорегиональные подразделения украинцев [Алексеева, Дяченко,

2000; Пономарев, 2000, Гриценко, 2000]: это группа западных украинцев (Львовская и Ивано-Франковская области, N=142), подольских (Хмельницкая область, N=181), днепровских (Черкасская область, N=174) и восточных (Белгородская область, N=96). Остальные 15 индивидов отнесены к тем выборкам, к которым по данным антропологии, лингвистики и историко-этнографического районирования населения они оказались ближе. Поскольку изучить весь массив популяций украинцев невозможно, при тщательном планировании экспедиционного обследования было решено охватить все разнообразие украинского генофонда, которое следует линии запад-восток (рис. 6.11). Изучение генофонда Карпатского региона представляет собой отдельную научную проблему. Поэтому было решено их представить общей популяцией, частично состоящей из представителей закарпатско-верхнеднепровского антропологического типа (обследован юг Львовской области), частично – карпатского антропологического типа (обследован регион Ивано-Франковска). Такая выборка дает представление об основном спектре гаплогрупп мтДНК населения Предкарпатья и Карпат. При планировании было решено не обследовать население украинского Полесья, поскольку, соглас-

но данным антропологии оно чрезвычайно близко к населению белорусского Полесья, для которого у нас уже имелись данные по мтДНК [Balanovsky et al., 2008] (рис. 6.11). Итоги изучения генофонда восточных славян подтвердили правильность такого планирования для Y-хромосомы – как мы видели в предыдущей главе (раздел 5.2), генофонд белорусского Полесья (рис. 5.14), сохраняет историческое сходство с популяциями Украины: паттерн их генетического ландшафта по Y-хромосоме сходен с паттерном карты генетических расстояний от украинцев (рис. 5.15). Также было решено на данном этапе не обследовать население южной части Украины, поскольку оно сформировалось относительно недавно при смешении многих разнородных групп населения. И вновь отметим, что население Крыма, в соответствии с правилами популяционной генетики, представлено только наиболее коренным его населением – крымскими татарами. Они были изучены сообществом украинских и российских генетиков при заинтересованной поддержке разных организаций крымских татар задолго до того, как Крым стал объектом мирового внимания. Ни русские, ни украинские популяции – в полном соответствии с правилами геногеографии – в Крыму не анализируются.



Рис. 6.11. Карта изученных украинских популяций (западные, подольские, днепровские и восточные), совмещенная с картой говоров украинского языка. Обозначения: Желтые круги – изученные региональные популяции украинцев. Зеленые круги – изученные популяции белорусского Полесья, привлеченные для характеристики Полесского генофонда.

МЕТОДЫ СБОРА ВЫБОРОК

Экспедиционное обследование проведено в 2002 году под руководством профессора Е.В. Балановской содружеством российских и украинских исследователей по той же единой программе, что и обследование всех других этносов Северной Евразии, и под контролем Этической комиссии Медико-генетического научного центра РАМН. На местах сбор образцов осуществлялся силами двух экспедиционных групп: на Украине под руководством профессора Харьковского университета Л.А. Атраментовой, в Белгородской области РФ – под руководством профессора Белгородского университета М.И. Чурносова. Сбор образцов проводился не только с ведома, но и при содействии местных администраций и органов здравоохранения Украины и России.

Для обследования отбирались неродственные в трех поколениях лица, все предки которых на протяжении трех поколений (т.е. включая бабушек и дедушек) родились в обследуемой популяции и считали себя украинцами. Такая выборка исключает влияние недавних миграций. Если более далекие предки индивида пришли в данный регион из других областей и, соответственно, оставили в этом регионе потомство, то такая миграция считается генетически эффективной и отражает уже устоявшиеся миграционные потоки между регионами. Такие индивиды (с предками из других регионов более 3 поколений назад) уже включаются в обследование. Поскольку генетическая близость популяций определяется накопившимся за длительное время эффектом генетических потоков между ними, крайне важно отделить эффективные

миграции от демографических сдвигов, произошедших одно или два поколения назад, и в отношении которых пока нельзя сказать, насколько они окажутся генетически эффективны. В тех редких случаях исключений, когда образцы все же были взяты от индивидов, у которых бабушка по материнской линии (т.е. линии мтДНК) и дед по отцовской линии (т.е. линии Y-хромосомы) оказались из разных популяций, то данный индивид при изучении мтДНК относился к той популяции, откуда родом его бабушка по материнской линии, а при изучении Y-хромосомы – к той популяции, откуда родом его дед по отцовской линии. В изучение аутосомного генома такие образцы не включались.

Информация о происхождении индивидов собиралась путем заполнения специальных анкет, в которых отражалось место рождения и этническая принадлежность самих добровольцев-участников, их родителей, бабушек и дедушек. Образцы собирались у отобранных вышеописанным способом представителей коренного населения (достигших 18-летнего возраста), добровольно желающих принять участие в обследовании, при получении их письменного информированного согласия. Образцы собирались в виде проб венозной крови с помощью системы вакуумных пробирок Vacuette. Для самой процедуры взятия крови привлекался местный медицинский персонал.

Нашу выборку мтДНК из 607 образцов украинцев объединили с другими данными: 240 образцов украинцев, опубликованных в работе [Gusar et al., 2006] и 159 образцов украинцев, опубликованных в статье [Mielnik-Sikorska et al., 2013], так что общая база анализируемых далее данных включала 1006 образцов мтДНК украинцев.

УКРАИНЦЫ НА ФОНЕ ЕВРОПЫ: КАРТЫ ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ

Карты генетических расстояний (как они строятся – подробно описано в главах 1 и 5) между митохондриальными генофондами украинцев и других народов Европы созданы в двух вариантах: и по данным о частотах 30 гаплогрупп мтДНК (рис. 6.12), и по данным о частотах 8160 гаплотипов мтДНК (рис. 6.13). На черно-белых картах (см. шкалу в левом верхнем углу) более темные тона означают большие генетические расстояния. И наоборот – территории, население которых наиболее генетически сходно с митохондриальным генофондом украинцев, окрашены в светлые тона. Обе карты (рис. 6.12 и 6.13) описывают одни и те же генофонды, но изученные на разных уровнях разрешения – более общем (гаплогруппы мтДНК) и более детальном (гаплотипы мтДНК). Сходство обеих карт подтверждает устойчивость выявленных закономерностей.

На уровне гаплогрупп (рис. 6.12) максимальное генетическое сходство по мтДНК (интервалы

генетических расстояний $d < 0.2$) украинцы обнаруживают с генофондом западных славян, а также с популяциями восточных областей Германии и Австрии. Несколько меньшее (но все же значительное) сходство (интервалы генетических расстояний $0.2 < d < 0.3$) проявляется с генофондами восточных и южных славян, балтов, а также рядом народов центральной и южной Европы. Исключение составляют популяции западного Причерноморья, максимально близкие к генофонду украинцев и образующие с ними единый ареал минимальных генетических расстояний. Генетически далекие от украинцев популяции обнаруживаются лишь на периферии Европы – юг Пиренейского полуострова, север Фенноскандии, север Восточной Европы, Урал и Предкавказье.

На уровне гаплотипов (рис. 6.13) область максимального сходства несколько расширяется – добавляются некоторые западноевропейские и российские популяции, некоторые исчезают (напри-

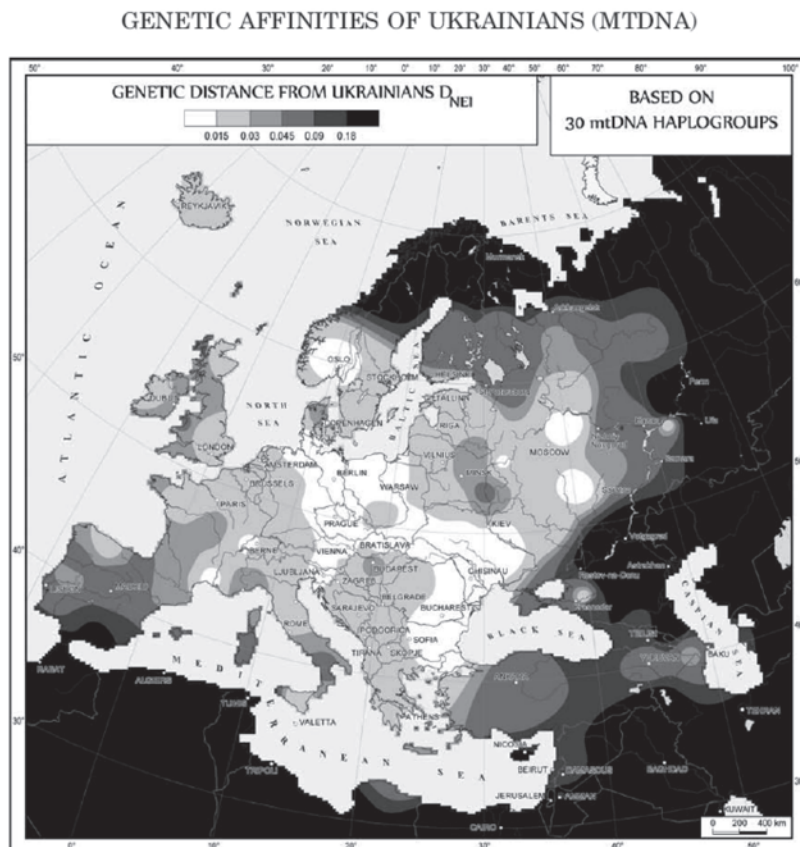


Рис. 6.12. Карта генетических расстояний от украинцев до популяций Европы (генетический ландшафт по 30 гаплогруппам мтДНК).

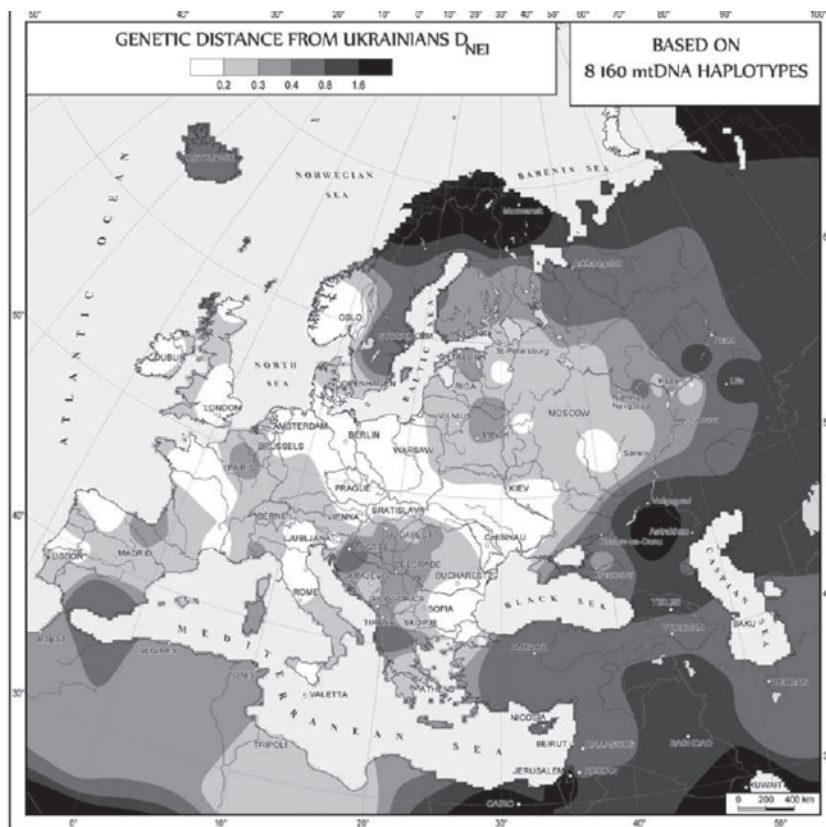


Рис. 6.13. Карта генетических расстояний от украинцев до популяций Европы (генетический ландшафт по 8160 гаплотипам мтДНК).

мер, Австрия), но они разбросаны не закономерно, а случайными пятнами.

Отметим любопытную черту обеих карт: митохондриальный генофонд Полесья обнаруживает примерно тот же уровень отличий от украинского генофонда, что и популяции белорусов. При этом паттерн карты генетических расстояний от белорусов Полесья по Y-хромосоме (рис. 5.14) оказался сходным с паттерном карты генетических расстояний от украинцев (рис. 5.15), а не от белорусов (рис. 5.13). Это вновь подчеркивает, что разные генетические системы могут описывать раз-

личные траектории генофонда, и только их сравнение позволяет выявлять и наиболее устойчивые закономерности структуры генофонда, и обнаруживать особенности его динамики.

В целом, картографирование генетических расстояний подтвердило ожидаемое сходство украинцев с другими европейскими народами, добавив к этой ожидаемой картине новый штрих – повышенное сходство с западными славянами, восточными областями Германии и западным Причерноморьем. Использование нового метода принесло больше новостей.

АНАЛИЗ МЕТОДОМ «ГЕОГРАФИЧЕСКИХ» ГРУПП ГАПЛОТИПОВ

Для того, чтобы в гомогенном митохондриальном генофонде Европы все-таки выявить, к каким популяциям наиболее генетически близки украинцы, был разработан новый метод – анализ географической скоррелированности гаплотипов.

ГЕОГРАФИЧЕСКАЯ СКОРРЕЛИРОВАННОСТЬ ГАПЛОТИПОВ

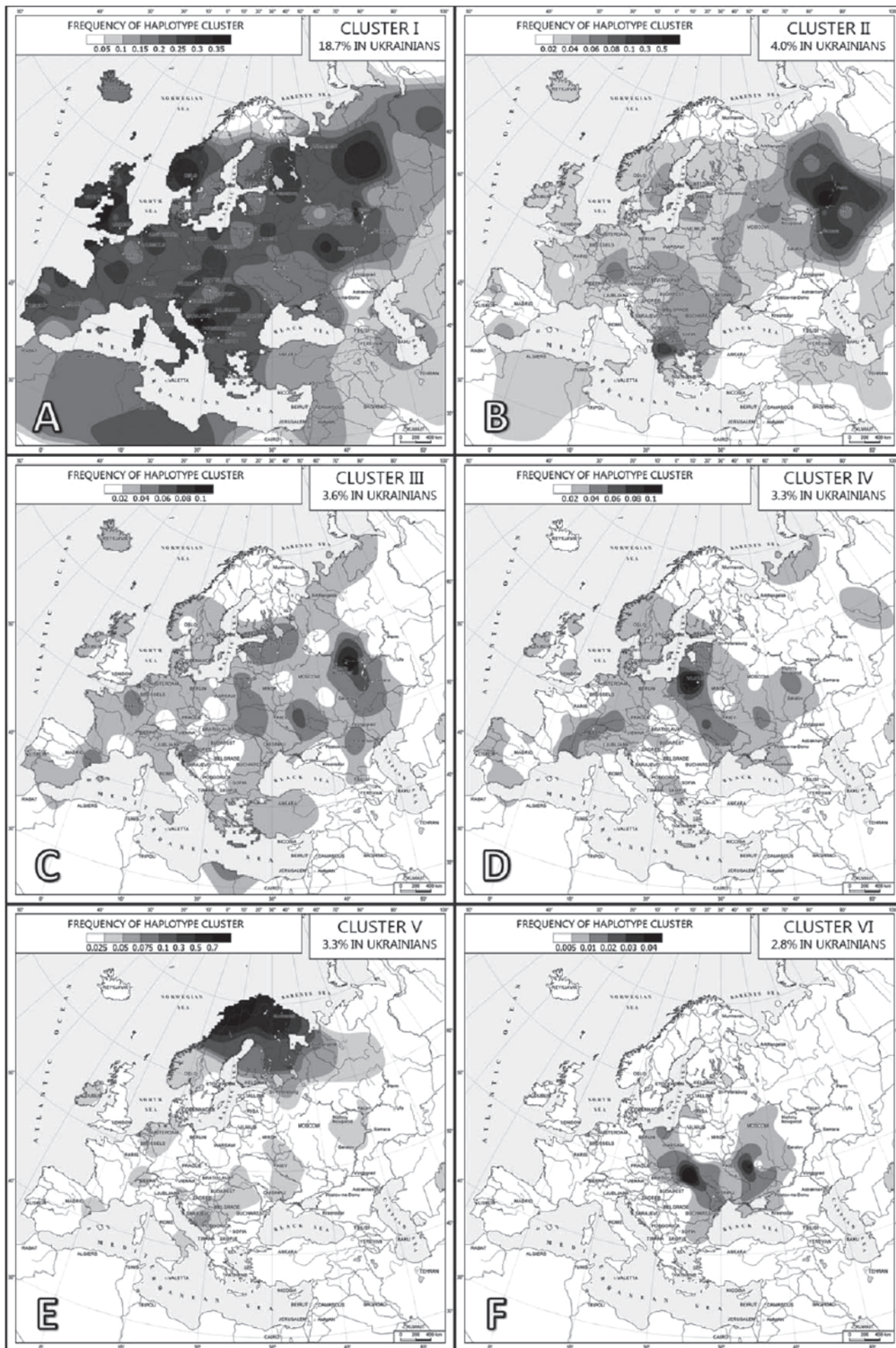
Метод состоит в следующем. Обычно гаплотипы объединяют в группы (в гаплогруппы и их субветви) по признаку общего происхождения, родства друг с другом. В данном же случае гаплотипы объединялись в один тип по сходству географического распределения, а не филогенетического родства. Иными словами, если гаплотипы, пусть даже неродственные друг другу, распределены по карте по принципу «где один, там и другой», – появляются в одних и тех же популяциях и отсутствуют в одних и тех же популяциях, то их относили к одному типу. Если другие гаплотипы были распределены не похоже на первый тип, но зато похоже друг на друга, их относили ко второму типу, потом аналогично выделялся третий тип и так далее. Можно предполагать, что такая географическая скоррелированность гаплотипов («где один, там и другой») возникла исторически и сохраняет следы исторического взаимодействия популяций.

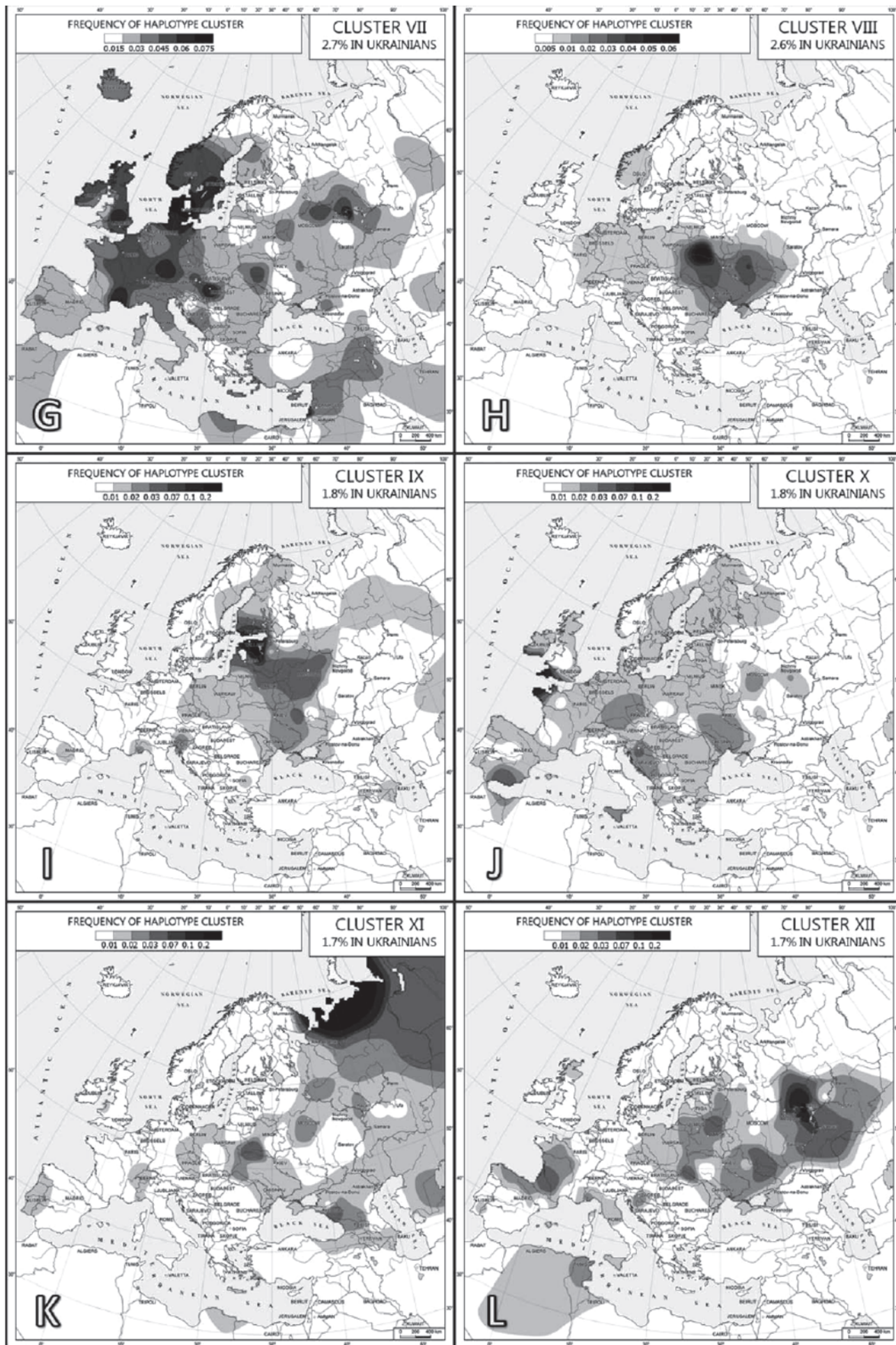
Технически этот подход был реализован с помощью кластерного анализа. Была взята обычная таблица частот встречаемости гаплотипов в разных популяциях. Но обычно по этой таблице рассчитываются генетические расстояния популяций друг от друга и выделяются группы популяций, сходные по частотам гаплотипов. А тут матрица была как бы развернута «наоборот»: по той же исходной таблице были рассчитаны расстояния между гаплотипами (а не популяциями) и выделены группы гаплотипов, сходные по «частотам популяций».

КАРТЫ «ГЕОГРАФИЧЕСКИХ» ГРУПП ГАПЛОТИПОВ УКРАИНЦЕВ

Чтобы можно было сравнить базу данных из 1006 образцов мтДНК украинцев с данными по другим 129 популяциями Европы, длину анализируемой части гипервариабельного сегмента пришлось уменьшить до интервала позиций нуклеотидов 16069–16365. Для каждого образца украинцев определили его принадлежность к одной из 30 гаплогрупп, по которым были изучены 129 популяций Европы, пригодные для сравнения по гаплогруппам мтДНК. В итоге оказалось, что 416 гаплотипов мтДНК, выявленных в генофонде украинцев, принадлежат к 28 из этих 30 гаплогрупп. При этом 130 из 416 гаплотипов не найдены в популяциях других народов и потому названы «условно уникальными украинскими гаплотипами». Условно – потому что по мере дальнейшего увеличения объемов выборок эти гаплотипы, скорее всего, обнаружатся и у других народов Европы. Но даже условно уникальными гаплотипами обладают только 15% из 1006 украинцев. Остальные 286 гаплотипов украинцев, охватывающие 85% украинского генофонда, являются общими с другими народами Европы. То есть митохондриальный генофонд украинцев, как и подавляющего большинства других народов Европы, сложен из тех же кирпичиков, что и генофонд Европы в целом.

Для анализа методом географически скоррелированных гаплотипов использованы 94 популяции, а из 8160 анализируемых гаплотипов были отобраны «неуникальные» – каждый из них встречается, по меньшей мере, в двух популяциях. Таких «неуникальных» гаплотипов оказалось 2098 (26%), и далее определяли, как они кластеризуются друг с другом по принципу сходного географического распространения. Среди выявленных кластеров отфильтровали те, частота которых у украинцев превышала 1,5% и которые состояли хотя бы из трех гаплотипов. В результате выделилось тринад-





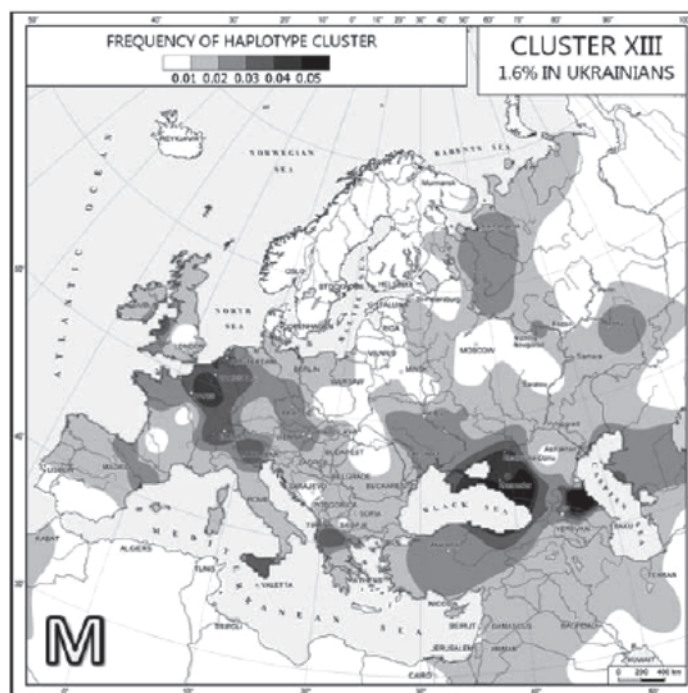


Рис. 6.14. Распределения частот тринадцати наиболее частых «географически скоррелированных» групп гаплотипов мтДНК, встречающихся и в украинском генофонде.

цать таких «географических» групп гаплотипов, и эти группы (кластеры) гаплотипов обозначены под номерами от I до XIII. Географическое распределение этих 13 кластеров картографировано (рис. 6.14). Таким образом, карты указывают на популяции, которые связаны родством по материнской линии и образуют «географически скоррелированные» группы гаплотипов, встречающихся с частотой выше 1.5% в украинском генофонде (примерно 1% критерий полиморфизма).

Оказалось, почти пятая часть украинского митохондриального генофонда (18,7%) принадлежит к кластеру № I, который распространен по всей Европе, причем с низкой частотой в Центральной Европе и с высокой частотой на периферии (рис. 6.14.A). Наиболее частые гаплотипы, относящиеся к гаплогруппам мтДНК H и J, составляют более половины этого кластера в Украине.

Кластеры VI и VIII (рис. 6.14 F, H) чаще всего встречаются в Украине, реже – в центральных областях Европы и практически отсутствуют в других регионах. Они вместе составляют 5,4% украинского митохондриального генофонда. В совокупности с 14,5% носителей уникальных украинских гаплотипов получается, что 19,9% – то есть пятая часть – украинцев относятся к линиям мтДНК, высокоспецифичным для населения Украины. Среди 148 этих украинско-специфичных линий мтДНК 44 принадлежат к гаплогруппе H, 26 к гаплогруппам J или T, 19 – к U4 или U5, 23 принадлежат к другим ветвям гаплогруппы U, 10 принадлежат к различным ветвям HV, 6 – к гаплогруппам восточ-

ного происхождения (A, B, C, D и G), остальные – к гаплогруппам I, N1b, W, X или V.

Примерно десятая часть (9,2%) гаплотипов мтДНК украинцев принадлежат к линиям, типичным для Волжского региона. Это кластеры II, III и XII (рис. 6.14. B, C, L), имеющие высокую частоту у финно-угорских и тюркоязычных популяций Поволжья и Приуралья (удмурты, мари, коми, чуваша, татары и башкиры).

Около 13% образцов ДНК украинцев демонстрируют генетическое сродство то с одной, то с другой популяцией Европы. Так, кластер IV (3,3% у украинцев) с наибольшей частотой встречается в Литве, кластер V (тоже 3,3% у украинцев) – типичен для саамов, а редкие у украинцев кластеры (по 1,5–2% каждого) IX, X, XI, и XIII включают линии, наиболее частые в Эстонии, Корнуолле, у ненцев и на Северном Кавказе. Кластер VII (2,7% у украинцев) (рис. 6.14 G) встречается в самых различных частях Европы.

Подводя итог, мы видим, что метод «географически скоррелированных гаплотипов» и карты генетических расстояний демонстрируют в целом сходную картину генетических связей украинской популяции с другими европейскими популяциями. Но предложенный метод «географически скоррелированных гаплотипов» позволяет точнее выявить детали, точнее дифференцировать генетически сходные регионы. Например, этим методом выявлены связи (около 10% линий мтДНК украинцев) с Волго-Уральским регионом. Показано, что пятая часть передающегося по материнской линии укра-

инского генофонда представлена линиями, высокоспецифичными для украинцев (почти не встречающимися в других популяциях). Остальные гаплотипы (их большинство) либо тяготеют к отдельным популяциям Европы (редкие кластеры), либо распределены почти по всей Европе, либо вообще не кластеризуются на основе «географически скоррелированных» гаплотипов.

При интерпретации результатов, полученных этим методом, важно помнить, что он не оценивает прямой поток генов между украинской и другими европейскими популяциями. Кластеры, полученные методом «географически скоррелированных» гаплотипов, указывают лишь на наличие общих предков украинцев и других популяций, живших на той или иной территории в прошлом.

6.4. ПРОБЛЕМА СВОЕОБРАЗИЯ ГЕНОФОНДА РУССКОГО СЕВЕРА

ОТЛИЧИЯ СЕВЕРА ОТ ЮГА

Освоение Русского Севера признается одной из важнейших вех в этногенезе русского народа. При этом многие исследователи указывают, что генофонд и фенотип северных русских сформирован не только (или даже не столько) древнерусской колонизацией этого края, сколько мощным субстратом дославянского (в основном финно-угорского) населения. Наши предыдущие исследования русского генофонда выявили по всем изученным системам – соматологии, дерматоглифики, классических генетических маркеров, фамилий – широтную изменчивость русского генофонда, задаваемую различиями севера и юга [Балановская, Балановский, 2007]. Дальнейшие исследования изменчивости Y-хромосомы [Balanovsky et al., 2008] и митохондриальной ДНК [Балановский и др., 2010] только подтвердили эту закономерность. Близкие ре-

зультаты получены и в исследованиях других авторов [Malyarchuk et al., 2002, 2004; Morozova et al., 2011]. При этом сравнительный анализ с генофондами народов Европы по маркерам Y-хромосомы [Balanovsky et al., 2008] выявил, что популяции Архангельской и Вологодской областей приближаются к кластерам не только финских, но также шведских и немецких популяций.

ЧЕТЫРЕ ЗЕРКАЛА ГЕНОФОНДА

Для создания возможно более полного генетического портрета населения Русского Севера в этом заделе использована вся доступная нам информация о полиморфизме четырех систем признаков: Y-хромосомы, мтДНК, аутосомных ДНК маркеров и квазигенетических маркеров (фамилий). На рис. 6.15 показано географическое положение изученных северных русских популяций.



Рис. 6.15. Карта расположения пяти изученных районов Архангельской области.

Отметим, что в контексте данного раздела термины «генофонд населения Архангельской области»

и «генофонд Русского Севера» использованы как синонимы.

КВАЗИГЕНЕТИЧЕСКИЕ МАРКЕРЫ

ВЫБОРКА – ДВА МИЛЛИОНА ЧЕЛОВЕК

Данные о квазигенетических маркерах (фамилиях) не могут быть использованы для межэтнических сравнений, зато неоднократно показана их эффективность для исследования структуры генофонда в пределах одного этноса, в том числе и для русских популяций [Чурносов и др., 2004; Сорокина и др., 2007; Балановский, Балановская, 2007]. В данной работе использована собранная нами информация об 1.9 млн. представителей 68 районов 14 областей России.

После отсева редких фамилий (являющихся обычно следствием миграции из других районов и областей), в анализ было включено 1,6 млн. человек, в том числе 57 тыс. человек из пяти районов Архангельской области. Частоты фамилий анализировались методами многомерной статистики как аналоги множественных аллелей одного локуса.

ВЗАИМООТНОШЕНИЯ ПОПУЛЯЦИЙ

Соотношения изученных русских популяций хорошо иллюстрируются графиком многомерного шкалирования (рис. 6.16): разнообразие фонда фамилий Русского Севера задается в основном двумя пограничными районами – Пинежским и Лешуконским. Важно, что своеобразие фамилий северно-русских популяций четко выражено, но оно вовсе не максимально для русских фамилий. Средние отличия фамилий Северного региона от остальных регионов примерно такие же ($d=1.2$), как и отличия фамилий Восточного региона ($d=1.1$), и заметно меньше, чем отличия фамилий кубанских казаков (Кавказский регион, $d=1.6$).

На графике многомерного шкалирования выявляется довольно четкая структура русского генофонда: практически все районные русские популяции кластеризуются в региональные группы.

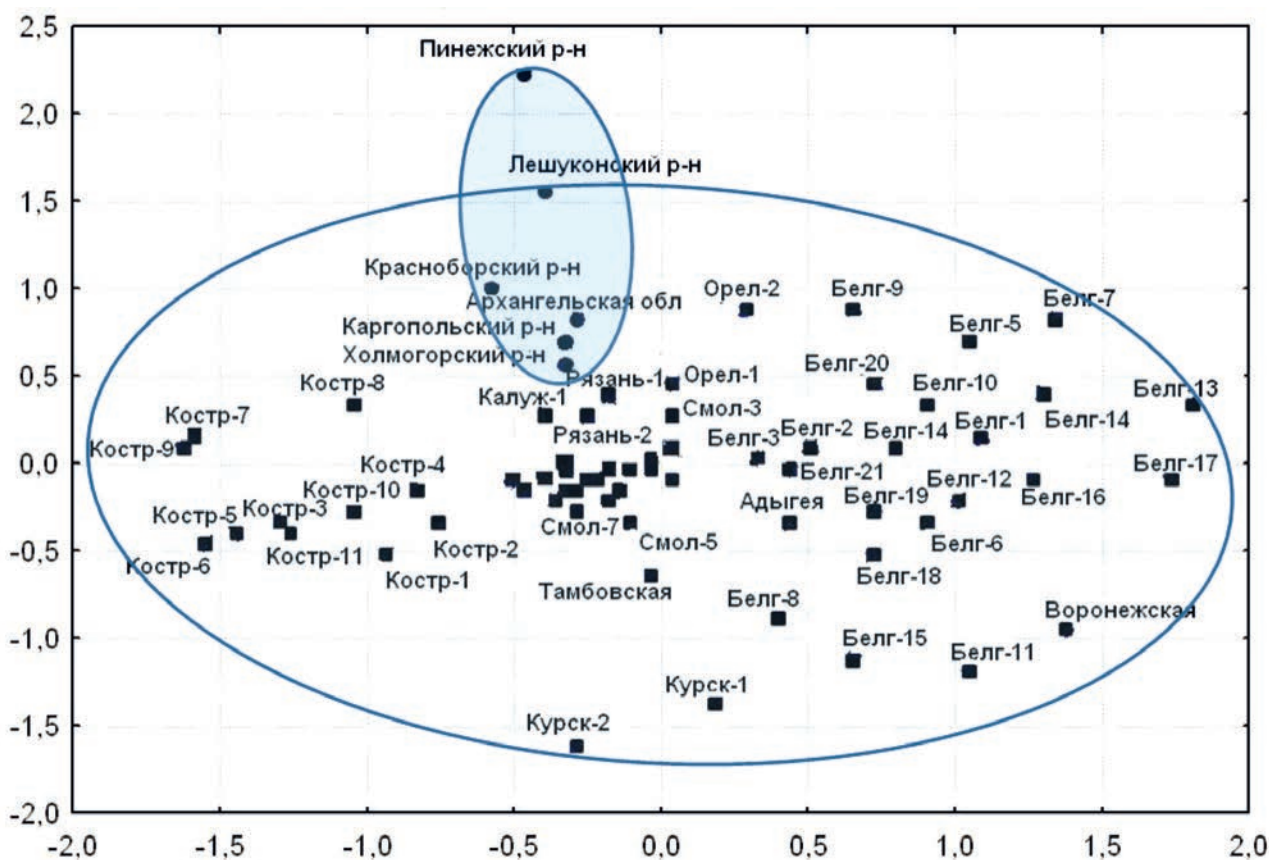


Рис. 6.16. Взаимное расположение на графике многомерного шкалирования генофондов 68 районов 14 областей по данным о распределении частот фамилий. (Стресс = 0,22, алиенация = 0,23).

Обозначения: в названии районов указано сокращенное название области и условный номер района (напр. Костр-11 обозначает район № 11 Костромской области). Для ряда районов в центре графика названия привести не удалось из-за большого сходства фамильного состава районов. Названия районов Архангельской области приведены полностью и выделены на графике овалом.

При этом график демонстрирует, что по распределению фамилий население Русского Севера целиком принадлежит к общему массиву популяций, представляющих исконный ареал русского народа.

Таким образом, детальный анализ русского генофонда по данным о фамилиях позволяет утвер-

ждать, что в целом генофонд Русского Севера является неотъемлемой составной частью русского генофонда: он не выходит за пределы изменчивости отдельных регионов и районов исторического русского ареала, а его своеобразие сформировано результатами дрейфа генов в периферийных популяциях этого ареала.

АУТОСОМНЫЕ МАРКЕРЫ

ДАННЫЕ ПО ALU-ИНСЕРЦИЯМ

Для анализа аутосомных ДНК маркеров использованы собственные данные по 10 русским популяциям (три из них представляют Архангельскую область), изученным по единой панели Alu-инсерций (ACE, APOA1, B65, PV92, TRA25), а также данные по СПИД-протекторной делеции в локусе CCR5 (CCR5del32), изученной в 14 русских популяциях.

Большинство русских популяций исследовано, к сожалению, по разным панелям аутосомных ДНК маркеров [Лимборская и др., 2002; Балановская, Балановский, 2007], что затрудняет анализ структуры русского генофонда. Наиболее широкое панорамное исследование, охватывающее весь исторический ареал русского народа и включающее три популяции Архангельской области, проведенное по панели Alu-инсерций (ACE, APOA1, B65, PV92, TRA25), описано выше (раздел 6.2). При этом обнаружено, что по панели ДНК маркеров русский генофонд слабо дифференцирован, причем популяции Архангельской области расположены компактно в пределах кластера русских популяций.

Таким образом, как по фонду фамилий, так и по панели аутосомных ДНК маркеров генофонд Русского Севера является составной частью русского генофонда: он не выходит за рамки отличий, характерных для общего массива русских популяций, а его гетерогенность на порядок меньше, чем гетерогенность, например, башкирских популяций, занимающих значительно меньший географический ареал.

ДАННЫЕ ПО СПИД-ПРОТЕКТОРНОЙ МУТАЦИИ CCR5del32

Однако по одному из наиболее изученных аутосомных ДНК маркеров – СПИД-протекторной делеции CCR5del32 – ярко выявляется своеобразие северных русских популяций. Изменчивость этого маркера активно изучалась многими российскими коллективами [Libert et al., 1998; Сломинский и др., 1997; Limborska et al., 2002; Кожекбаева и др., 2004]. Поэтому рассмотрим его особо. На карте (рис. 6.17) видна четкая закономерность: частота максимальна на севере Европы (на побережьях Балтийского и Белого морей), и во всех на-

правлениях от этой зоны частота плавно снижается. Такой элегантно четкий и простой пространственный тренд крайне редок для аутосомных маркеров. Тем удивительнее, что распределение мутации в русском ареале носит довольно сложный характер. Наибольшие частоты отмечены на Русском Севере – во всех трех изученных популяциях Архангельской области. Именно они формируют беломорскую часть беломоро-балтийского мирового максимума частоты этого аллеля. К югу и западу русского ареала частота снижается. Изменчивость частоты CCR5del32 в русских популяциях столь велика, что заметна даже в глобальном масштабе [Balanovsky et al., 2005].

Однако, хотя генетически эффективные размеры популяций Архангельской области невелики, необычно высокие частоты **CCR5del32** нельзя объяснить лишь генетическим дрейфом: ведь все три Архангельские популяции, географически и генетически изолированные друг от друга, отклонились от среднерусской частоты ($p=0,11$ с учетом Архангельских популяций, без их учета $p=0,10$) в сторону увеличения частоты делеции, причем все три – на значительную величину (Пинежский район $p=0,16$, Лешуконский район $p=0,18$, Красноборский район $p=0,19$), тогда как действие дрейфа предполагает случайное по направлению и силе отклонение.

Поэтому можно утверждать, что наблюдаемое распределение (рис. 6.17) указывает на существование мощного общего генетического пласта в населении Русского Севера и народов Северной Европы, где тоже наблюдаются высокие частоты этого аллеля.

Причем эту генетическую общность никак нельзя назвать финно-угорской, так как она включает европейские народы самых разных ветвей трех лингвистических семей – индоевропейской, уральской и алтайской. С частотой более 10% делеция CCR5del32 встречается у следующих народов (перечисленных в пределах языковой ветви по убыванию частоты делеции): германской ветви (исландцы, шведы, фламандцы, датчане, англичане, немцы, норвежцы); славянской (русские, поляки, белорусы, словаки, украинцы, чехи); финской (эстонцы, мордва, финны, удмурты), угорской (венгры); балтской (литовцы); романской (французы) и даже

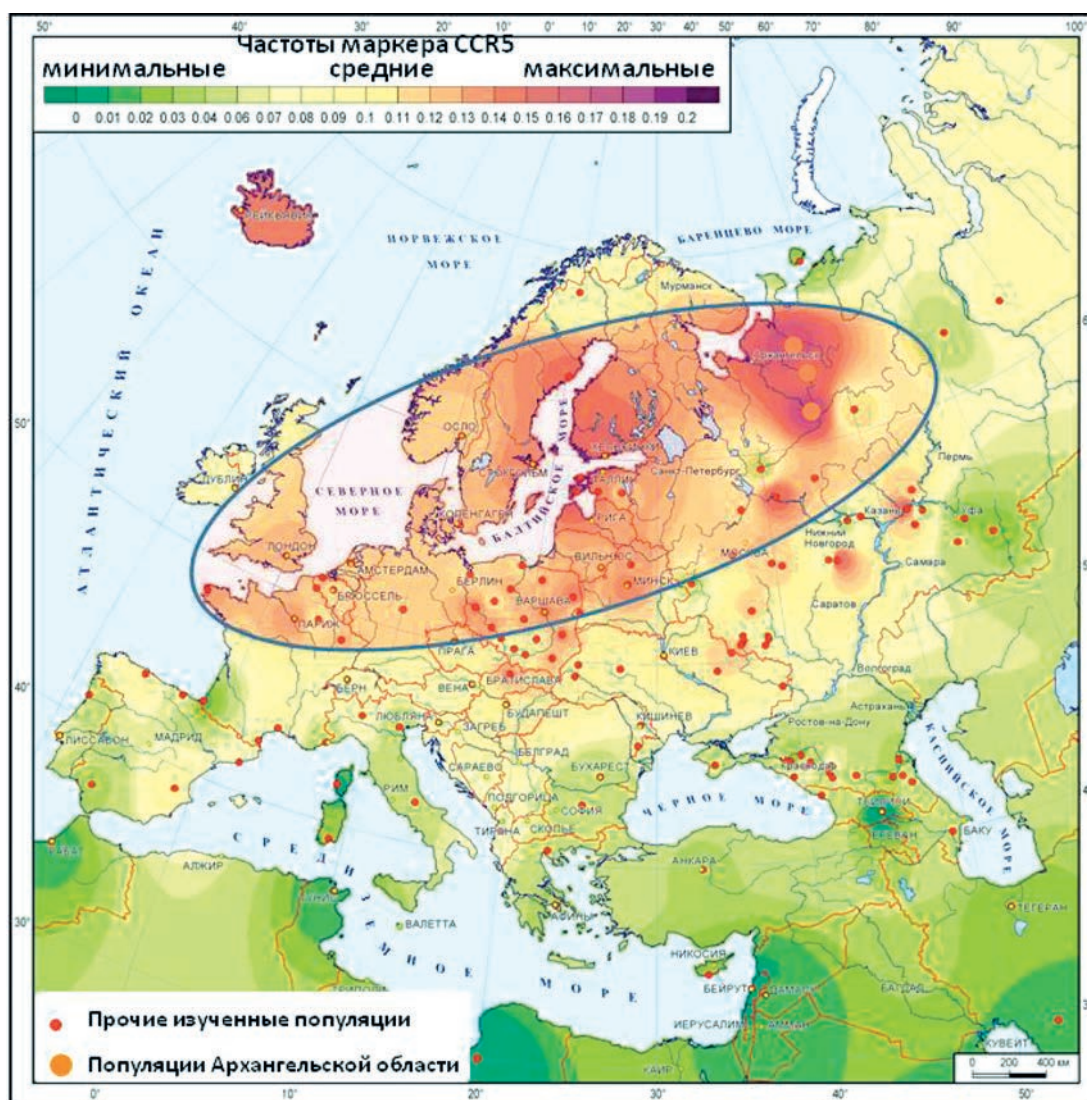


Рис. 6.17. Карта распространения частоты СПИД-протекторной мутации *del32* в аутосомном гене *CCR5*. Область наиболее высоких частот выделена овалом.

тюркской (татары). Как видим, по числу народов с высокой частотой делеции финноязычные народы уступают германоязычным и славянским.

В целом, область регулярно встречаемых высоких частот *CCR5del32*, выделенная на карте ова-

лом (рис. 6.17), указывает на общность генофонда населения северной части Европы (в широком географическом смысле этого термина), к которому полностью принадлежат и популяции Архангельской области.

МИТОХОНДРИАЛЬНАЯ ДНК

ГЕНЕТИЧЕСКИЕ ВЗАИМООТНОШЕНИЯ ПОПУЛЯЦИЙ

Для анализа митохондриальной ДНК использованы данные по частотам 16 гаплогрупп (A, C, D, H, HV, I, J, K, T, U2, U3, U4, U5a, U5b, V, W). Для максимально корректного сравнения, при отборе популяций для многомерного анализа мы стремились к тому, чтобы возможно большее число популяций было изучено по гаплогруппам обеих однородительских систем – и Y-хромосомы, и митохондриальной ДНК.

Население Русского Севера представлено по мтДНК тремя популяциями – Пинежского района (восточная окраина области на границе с республикой Коми), Каргопольского района (запад области у истоков Онеги [Belyaeva et al., 2003]) и суммарной группы поморов (северное побережье). Размах различий по мтДНК между этими тремя северно-русскими популяциями велик: генетическое расстояние между западом и востоком Архангельской области $d=0,025$; между западом и поморами $d=0,028$, между востоком и поморами в два раза выше $d=0,049$.

Взаимное расположение популяций показывает дендрограмма (рис. 6.18). Она указывает на возможность выделения двух близких кластеров европейских генофондов и резко отличающихся генофондов финноязычных популяций.

В общий кластер с популяциями Русского Севера, помимо норвежцев и немцев, вошли другие германоязычные народы (австрийцы, швейцарцы, шотландцы) и ирландцы, славяне (поляки, словенцы и боснийцы), а также балты (литовцы). Второй европейский кластер включил в основном славянские популяции – все остальные русские популяции, белорусов, украинцев, чехов, словаков, а также венгров, эстонцев, латышей и шведов. Таким образом, митохондриальный генофонд Русского Севера тяготеет в большей степени к Западной и Центральной Европе, а остальные русские популяции – к генофондам Восточной Европы.

Но более важен иной вывод: все финноязычные популяции (за исключением эстонцев) находятся в иных кластерах, чем популяции Русского Севера, причем и западные (финны, карелы) и восточные (коми, марийцы, мордва) финноязычные народы генетически удалены от Русского Севера.

Таким образом, изучение митохондриального генофонда свидетельствует против распространенного мнения, что генофонд Русского Севера представляет собой финно-угорский субстрат с тонкой

амальгамой славянских влияний. Напротив, по материнским линиям наследования мы вновь видим, как и по аутосомным маркерам, сходство с генофондом северной половины Европы, в том числе с географически далекими популяциями Западной и Центральной Европы.

КАРТА ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ ПО мтДНК

Эти выводы полностью подтверждаются картой генетических расстояний от генофонда Русского Севера (реперные популяции Архангельской области обозначены крупными оранжевыми кружками), позволяющей включить в анализ много больший массив популяций (рис. 6.19).

Мы видим, что область регулярно встречаемых минимальных генетических расстояний от Русского Севера охватывает страны Центральной и Западной Европы – побережье Северного моря, южную часть Балтики, Центральную Европу и даже узким языком доходит до Балкан. Популяции этого ареала оказываются генетически более близкими к Русскому Северу, чем популяции Восточной Европы, окрашенные в тона средних генетических расстояний. Приуралье и север Фенноскандии окрашены в красные тона максимальных генетических отличий.

В целом, вся совокупность приведенных результатов по мтДНК указывает на существова-

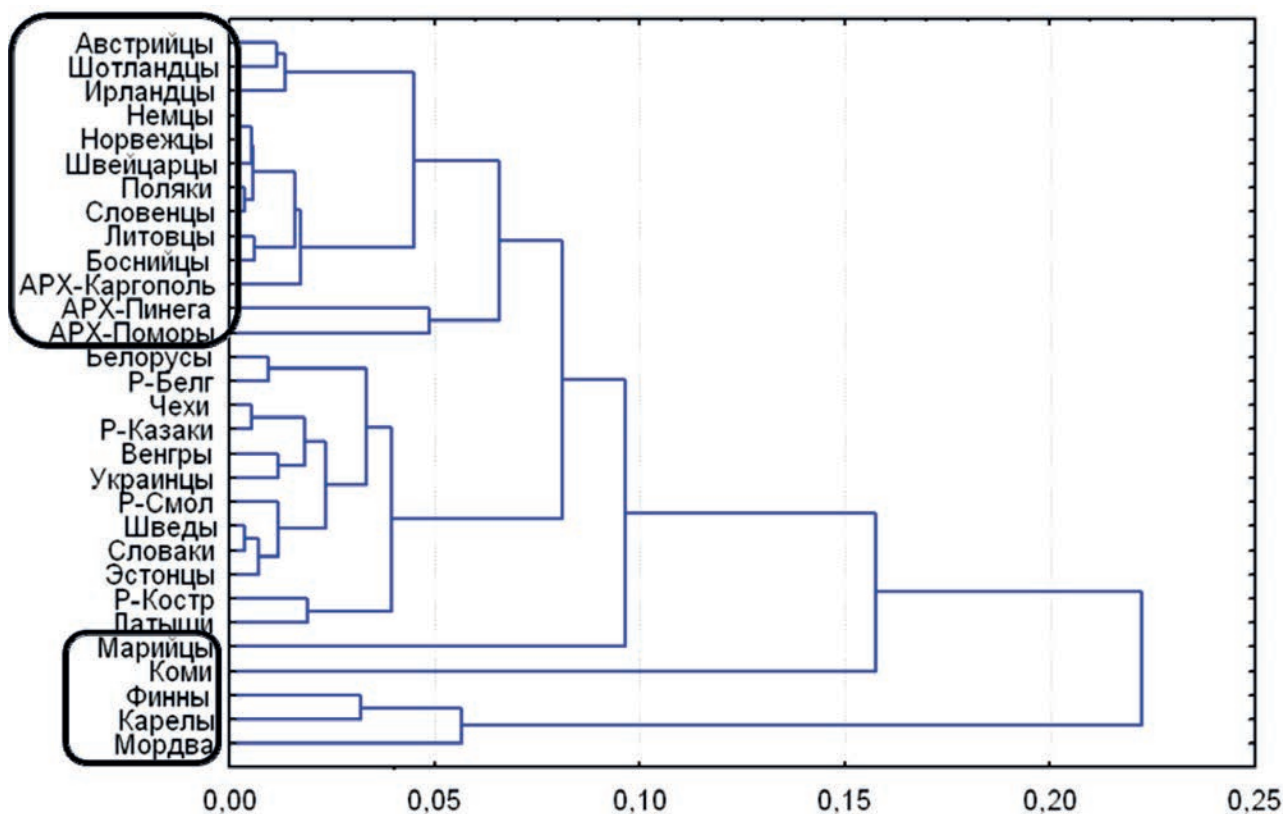


Рис. 6.18. Взаимное расположение на дендрограмме генофондов народов Европы по данным о частотах гаплогрупп мтДНК. Обозначения: популяции Русского Севера начинаются с обозначения «АРХ» (Архангельская область), остальные русские популяции – с обозначения «Р».

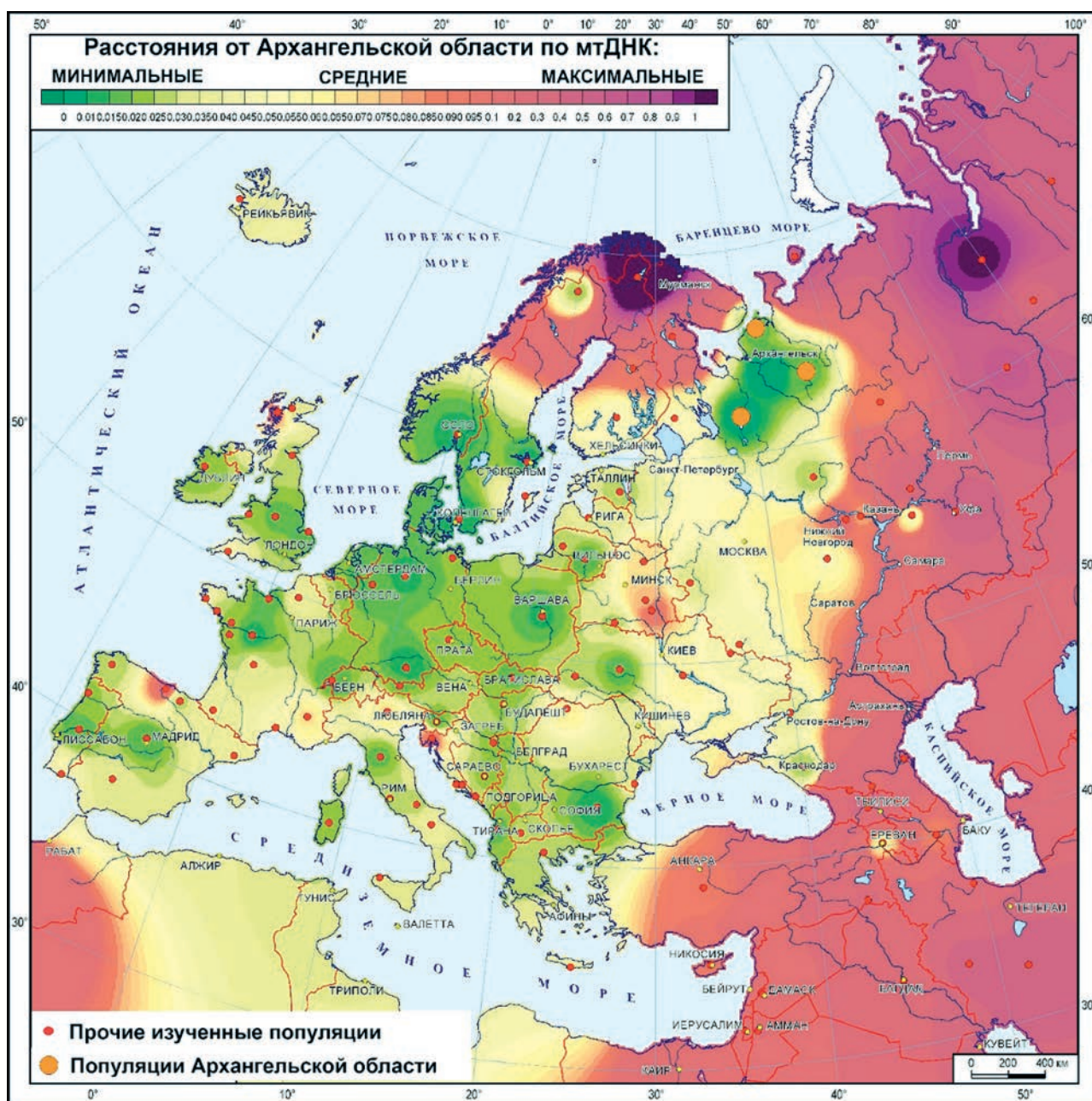


Рис. 6.19. Карта генетических расстояний от Русского Севера по мтДНК.

Территории, население которых наиболее генетически близко к генофонду Русского Севера, окрашено в зеленые тона; средние генетические расстояния от Русского Севера – в желтые тона; максимальные генетические отличия – в красные тона. Генетически изученные популяции Европы обозначены красными точками, популяции Архангельской области – оранжевыми кружками.

ние общего мощного пласта в генофондах популяций Русского Севера и народов Центральной и Западной Европы. При этом как западные, так и восточные финны резко отличаются

от севернорусских популяций, что позволяет отрицать их преобладающее участие в формировании митохондриального генофонда Русского Севера.

У-ХРОМОСОМА

КАРТА ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ ПО У-ХРОМОСОМЕ

Для анализа истоков своеобразия генофонда Русского Севера отобраны именно те шесть гап-

логрупп У-хромосомы (I1-M253, I2b1-M223, N1b-P43, N1c1-M178, R1a1-M198, R1b1a2-M269), которые наиболее распространены на севере Европейской России: в изученных нами популяциях Архангельской области они составляют суммарно

почти 95% генофонда. Карта генетических расстояний (рис. 6.20) демонстрирует, насколько каждая ее точка генетически близка к средним параметрам популяций Русского Севера и читается аналогично карте расстояний по мтДНК (рис. 6.19). Напомним, что на карте распространения СПИД-протекторной делеции **CCR5del32**, представляющей аутосомные ДНК маркеры (рис. 6.17), шкала читается иначе: она отражает диапазон частот от минимальных (зеленые тона) до максимальных (красные тона) частот.

Карта генетических расстояний от популяций Русского Севера по Y-хромосоме выявляет иной

ареал, чем мы видели по мтДНК и по аутосомным ДНК маркерам. Этот ареал теперь наряду с Прибалтикой охватывает также север Фенноскандии и доходит вплоть до Печоры, а также узкой полосой тянется вдоль Приуралья – эта полоса образована популяциями коми, удмуртов и одной из нескольких изученных популяций тюркоязычных башкир (рис. 6.20). Таким образом, генофонд отцовских линий – в отличие от материнских – связан в основном с финно- и балтоязычными народами.

Однако даже этот результат не может свидетельствовать в пользу популярного мнения «по-

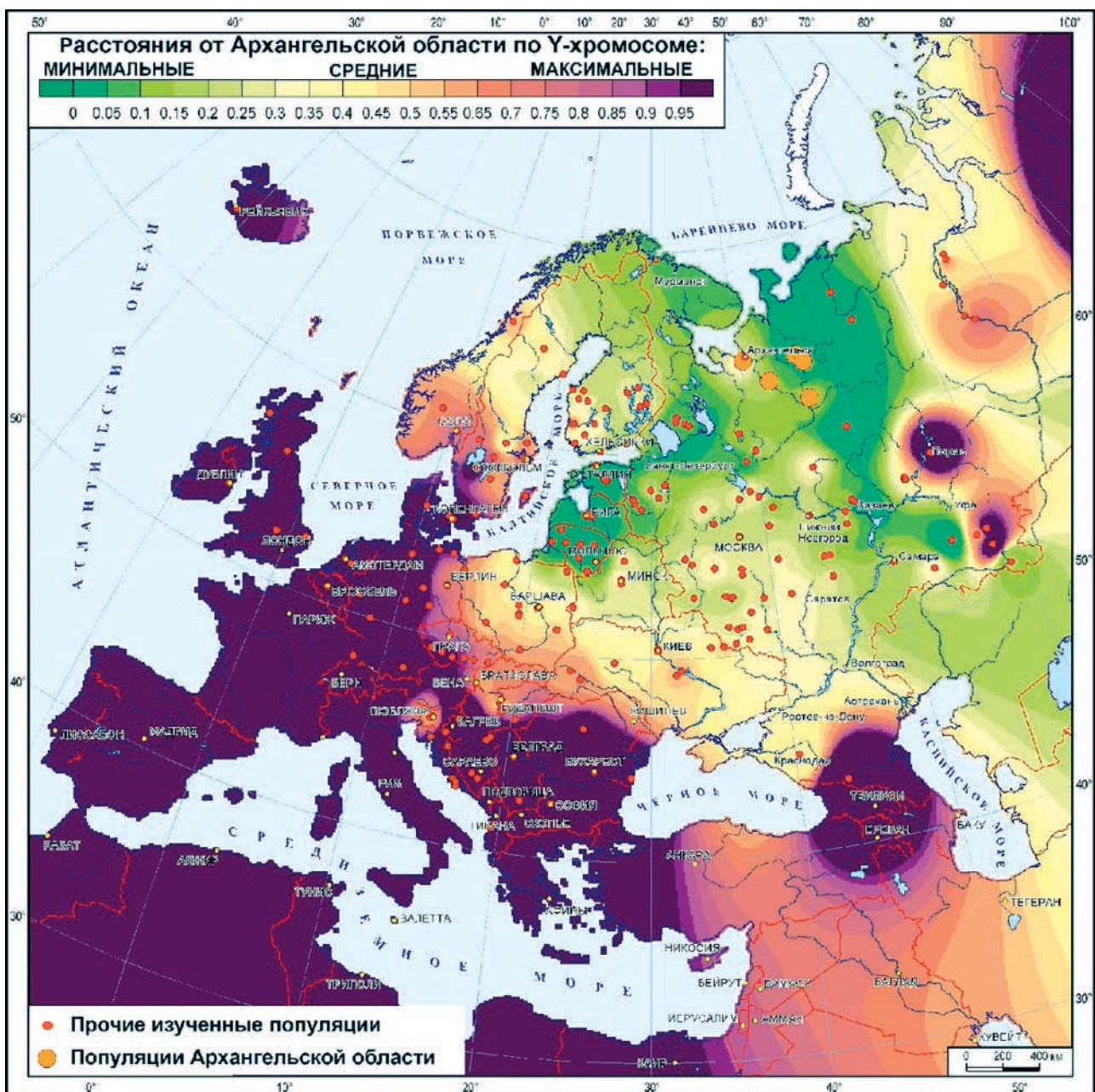


Рис. 6.20. Карта генетических расстояний от Русского Севера по Y-хромосоме.

Территории, население которых наиболее генетически близко к генофонду Русского Севера, окрашено в зеленые тона; средние генетические расстояния от Русского Севера – в желтые тона; максимальные генетические отличия – в красные тона. Генетически изученные популяции Европы обозначены красными точками, популяции Архангельской области – оранжевыми кружками.

скреби северного русского и увидишь финна». Во-первых, это мнение должно быть тогда распространено и на балтов (литовцев и латышей), и на тюркоязычных башкир, и даже на германоязычных шведов. Во-вторых, этот результат отражает миграции лишь мужского населения. В-третьих, степень генетического сходства с финноязычными популяциями различна для разных северно-русских популяций.

ГЕНЕТИЧЕСКИЕ РАССТОЯНИЯ ОТ ОТДЕЛЬНЫХ СЕВЕРНОРУССКИХ ПОПУЛЯЦИЙ

Например, самая восточная популяция Архангельской области (Пинега) оказывается генетически ближе к географически далеким балтам (до латышей $d=0,10$, до литовцев $d=0,13$), чем к своим географическим соседям – коми ($d=0,14$). При этом генетические расстояния до финно-угров в целом намного выше, чем до балтов: до эстонцев $d=0,14$, до карелов, вепсов, ижорцев $d=0,30$, до венгров $d=0,46$, до финнов Карелии $d=0,63$, до финнов Финляндии $d=0,65$. Среднее генетическое расстояние до финно-угров составляет $d=0,39$ (включая соседей коми), в то время как среднее расстояние до других русских популяций почти в два раза меньше ($d=0,22$). При этом наибольшее отличие дает географически близкая к Пинеге костромская популяция русских ($d=0,30$), демонстрируя эффекты дрейфа генов.

Северо-восточная популяция Архангельской области (Мезень) еще более генетически близка к географически далеким балтам (расстояние до литовцев $d=0,03$, до латышей $d=0,03$) и к соседним коми ($d=0,03$). Генетические расстояния до других финно-угров вновь выше, чем до балтов: до эстонцев $d=0,06$, до карелов, вепсов, ижорцев $d=0,10$, до финнов Карелии $d=0,28$, до финнов Финляндии $d=0,37$, до венгров $d=0,74$.

Среднее генетическое расстояние до финно-угров составляет $d=0,26$, в то время как отличия от балтов на порядок ниже ($d=0,03$).

При этом среднее генетическое расстояние от Мезени до других русских популяций такое же, как и до финно-угров $d=0,29$.

Популяция юго-востока Архангельской области (Красноборский и Ленский районы) вновь генетически ближе всего к балтам ($d=0,06$). Расстояния до финно-угров в три раза больше ($d=0,20$), а средние расстояния до других русских популяций по сравнению с Пинегой и Мезенью возросли в два

раза ($d=0,48$), несмотря на то, что географически Красноборск ближе к другим русским популяциям. Поэтому своеобразие красноборской популяции отражает не ее «оторванность» от русского генофонда по происхождению, а эффект дрейфа генов – частота мажорной для русского генофонда гаплогруппы **R1a1-M198** в красноборской популяции составляет только 19,8%.

В целом анализ гаплогрупп Y-хромосомы обнаруживает более устойчивое сходство Русского Севера с балтами, а не с финно-уграми, то есть выявляет западные, а не восточные влияния.

ПАЛЕОЕВРОПЕЙСКИЙ СУБСТРАТ

Однако мы далеки от того, чтобы просто поменять финно-угорский субстрат на балтский. Выявленное генетическое сходство и с балтскими, и с финноязычными, и с германоязычными популяциями севера Европы указывает, на наш взгляд, на сохранение более древних пластов генофонда, существовавших задолго до разделения балто-славянских и финноязычных общностей и роднящих разноязыкие ныне генофонды севера Европы.

Отметим, что по данным антропологии (свидетельства которой подробно рассмотрены в работе [Балановская и др., 2011]) ильменско-белозерский антропологический тип не был единственным в ряду реликтовых расовых вариантов Европы, сохранившихся до настоящего времени. Об этом свидетельствуют и материалы по псковским поозерам, где представлен очень архаичный западно-балтийский антропологический тип, и материалы М.В. Витова [Беневоленская, Давыдова, 1986; Витов, 1997]. В современном населении Архангельской области также фиксируются популяции, для которых характерно заметно более высокое лицо, достаточно широкий нос и средняя длина тела. К ним относятся и русские Лешуконского и Мезенского районов. Характерно, что морфологическое своеобразие некоторых современных популяций Архангельской области находит себе аналогии в палеоантропологическом материале с территории северо-запада. Особенно хотелось бы обратить внимание на то, что это те же популяции, что выделяются и по данным ДНК полиморфизма. Как по данным молекулярной генетики, так и по данным палеоантропологии они тяготеют к западно-европейским популяциям – можно полагать, что в силу общего генезиса на палеоевропейской антропологической основе.

ГЛАВА 7.

ГЕНОФОНДЫ И ЯЗЫКИ

Анализ связи между генетической и лингвистической структурой народонаселения проведен нашим коллективом для нескольких регионов мира, но наиболее подробно – на многообещающем примере Кавказа. Такие особенности этого региона, как наличие географических барьеров (горных хребтов), детальная изученность родства языков северокавказской лингвистической семьи и подробные собственные данные о генофондах этих народов позволяют корректно исследовать вопрос о взаимосвязи генетической, географической и лингвистической изменчивости популяций чело-

века. Попытки решить эту проблему традиционны для популяционной генетики человека [Рычков, Ящук, 1980, 1983, 1985; Cavalli-Sforza et al., 1994; Rosser et al., 2000; Nasidze et al., 2003, 2004; Спицын и др., 2009; Кутуев, 2010]. Ее решение легче осуществить на Кавказе, поскольку горный рельеф способствовал изолированности даже соседних популяций. Это создавало условия для формирования как большого лингвистического разнообразия путем дифференциации из единого праязыка северокавказской семьи, так и для формирования четкой популяционной структуры генофонда.

7.1. СТРУКТУРА И ИСТОРИЯ НАРОДОНАСЕЛЕНИЯ КАВКАЗА

НАРОДОНАСЕЛЕНИЕ КАВКАЗА ПО ДАННЫМ АРХЕОЛОГИИ И АНТРОПОЛОГИИ

ПАЛЕОЛИТ-МЕЗОЛИТ

Территория Кавказа была прочно освоена людьми современного антропологического типа еще в верхнем палеолите: археологические памятники этой эпохи обнаружены во всех ландшафтных зонах Кавказа [Бжания, 1996; Бадер, Церетели, 1989]. Следующая эпоха – мезолит – на Кавказе развивается преемственно и настолько связана с верхним палеолитом, что «критерии разделения нечетки, нижняя граница мезолита условна» [Бадер, Церетели, 1989, с. 93] и ориентировочно может быть проведена на уровне 15 000 лет назад. В голоцене открываются от ледников перевалы, связывающие население Северного Кавказа с Передней Азией не только приморскими, но и горными путями.

МЕСТНЫЕ КОРНИ КАВКАЗСКОГО НЕОЛИТА

Уже 8000 лет назад высокоразвитый мезолит сменяется неолитом. Причем для всех регионов Северного Кавказа – от Западного до Восточного – характерна преемственность культуры (и, следовательно, населения) на этапах мезолит-неолит. Важно, что на Северном Кавказе уже в начале VI тыс. до н.э. (8000 лет назад) появляются местные очаги скотоводства и земледелия, включая одомашнивание скота и окультуривание местных сортов ячменя и пшеницы [Массон и др., 1982; Бжания, 1996].

В это же время в центральное Закавказье проникли из Передней Азии племена энеолитической культуры, которые принесли навыки земледелия, скотоводства, керамики и плавки меди. Под их влиянием местный неолит Закавказья очень быстро перешел в стадию энеолита (раннего металла) V–IV тыс. до н.э. (7000–6000 лет назад).

Таким образом, в отличие от неолита Закавказья, для неолита Северного Кавказа археологи подчеркивают его преемственность и то, что к северу от Большого Кавказского хребта «развитие земледельческо-скотоводческой культуры ... шло своим, особым путем» [Массон и др., 1982]. В неолите отмечаются налаженные контакты племен Кавказа с Закавказьем и Крымом, но не обнаруживаются связи с Передней Азией. Даже неолитическая культура Прикаспия (8000 тыс. лет назад) тяготеет к прикавказским степям, но не к Передней Азии [Бжания, 1996]. Для неолита Северного Кавказа «характерно сочетание ... орудий ..., весьма близких к мезолитическим... и примитивной керамикой. Некоторые из этих поселений свидетельствуют о знакомстве их обитателей со скотоводством и земледелием уже в начале VI тыс. до н.э., до появления в Закавказье энеолитических культур переднеазиатского типа с глинобитной архитектурой и развитым производящим хозяйством» [Бжания, 1996].

Таким образом, тезис о том, что Кавказ являлся одним из первичных очагов формирования зем-

ледельческо-скотоводческого хозяйства, сейчас (особенно после открытия высокогорной дагестанской стоянки Чох, фиксирующей момент перехода от собирательства и охоты к производящей экономике) уже не вызывает дискуссий. Если неолит Закавказья попал под влияние Передней Азии и быстро сменился энеолитом, то неолит Северного Кавказа продолжал идти своим путем, вырастая из традиций мезолита и создавая свое земледелие, керамику и одомашнивая животных.

ВЛИЯНИЯ ПЕРЕДНЕЙ АЗИИ В РАННЕБРОНЗОВОМ ВЕКЕ

Однако на следующем этапе развития культуры населения Кавказа – в бронзовом веке – отмечается смена векторов миграций. В это время в значительной части Северного Кавказа распространяется Майкопская культура раннебронзового века (ее расцвет приходится на период 5200–4300 лет назад), «представляющая собой из ярчайших феноменов бронзового века Европы и всего евразийского пограничья» [Мунчаев, 1994, с. 225]. Именно в это время предполагается проникновение на Северный Кавказ инородных этнических и культурных элементов.

«Ни в предшествующую эпоху, ни в последующий период развития бронзового века связи Северного Кавказа и Передней Азии не были столь интенсивными, как в конце IV–III тысячелетия до н.э. Не исключено, что на Северный Кавказ проникли и осели здесь отдельные группы населения из Передней Азии, которые и занесли сюда культурные достижения Передней Азии. Одним из основных путей, по которому шло это проникновение на Северный Кавказ, являлся, по-видимому, морской – по Черному морю. Нужда в металле, и прежде всего в золоте, была, как нам представляется, одной из главных причин, привлечших внимание населения Передней Азии к Северному Кавказу» [Мунчаев, 1994, с.225]. Исходная территория переднеазиатских миграций включала ареал от Тигра на востоке до северной Сирии и смежной части Восточной Анатолии на западе [Мунчаев, 1994].

Позднее, около 4000 лет назад в значительной части Северного Кавказа развивается местная кобанская археологическая культура, расцвет которой приходится на период 3200–2400 лет назад [Крупнов, 1960, 1969], а постепенное развитие прослеживается с эпохи поздней бронзы до сарматского времени, а в горных районах – до раннего средневековья. Эта культура местная и восходит к древнему автохтонному населению Центрального Кавказа. Причем «большинство кавказоведов придерживаются мнения, что эта древняя кавказская этническая общность являлась мощным субстратом в последующем формировании почти всех со-

временных народов Северного Кавказа» [Мелюкова, 1989, с. 295]. Кобанская культура занимала горные области центрального Кавказа по обе стороны Большого Кавказского хребта и простиралась от Кубани на западе до Дагестана на востоке [Жоженкова, 1989].

ПОЗДНИЕ МИГРАЦИИ С СЕВЕРА

Около 2500 лет назад впервые важную роль начинают играть миграции на Северный Кавказ с севера, из степей Евразии – сначала ираноязычных народов (скифов, сарматов и их потомков – алан), затем тюркоязычных (готов, гуннов и т.д.). Под их давлением не только местное население оттесняется в горы, но и кочевое население проникает вплоть до высокогорий Кавказа, постепенно смешиваясь с автохтонным населением и переходя на оседлый образ жизни [Абрамова, 1989; Мелюкова, 1989; Агеева, 2000].

Разгром Алании татаро-монголами (XIII век, 1222–1229 гг.) и Тамерланом (конец XIV века, 1395–1396 гг.) вновь изменил этническую карту Кавказа: в освободившийся ареал, ранее занятый ираноязычными аланами, стали проникать с запада племена адыгских народов – будущих кабардинцев и черкесов. Завершили формирование этнической картины Северного Кавказа тюркоязычные ногайцы, осевшие в XVI–XVII веках в его степных предгорьях – от Дагестана до Кубани.

В начале XVI века начинается последняя мощная волна миграций – славянского населения: на Кавказе появляются первые славянские поселения – терские казаки, а, начиная с XIX века, в результате Кавказской войны славянское население Кавказа (преимущественно русские и украинцы) становится все более и более многочисленным.

ИТОГИ

Итак, в формировании населения Кавказа участвовали четыре основных компонента:

- (1) автохтонное население, восходящее к палеолиту;
- (2) миграции из Передней Азии (5–4 тыс. лет назад);
- (3) миграции из степей Евразии ираноязычных и тюркоязычных народов (2,5–0,5 тыс. лет назад);
- (4) миграции славянского населения (0,5 тыс. лет назад – современность).

Роль каждого из этих сценариев в формировании генофонда Северного Кавказа неизвестна, и даже данные антропологии дают весьма противоречивую картину формирования населения Кавказа [Абдушелишвили, 1964; Алексеев, 1974; Герасимова, Рудь, Яблонский, 1987]. Так, академик В.П. Алексеев считал кавкасионный массивный тип древнейшим, а грацильный понтийский – произ-

водным. Его позицию в отношении кавкасионного типа можно в нескольких словах охарактеризовать как «древность, преемственность, масштабность, консервация» [Кашибадзе, 2006]. Другой, не менее знаменитый, антрополог М.Г. Абдушелишвили настаивал, что, напротив, в ареале кавкасионного типа шло локальное увеличение массивности, только лишь имитирующее палеоевропеидность. Его позицию можно в нескольких словах охарактеризовать как «древность, преемственность, автохтонность, динамизм» [Кашибадзе, 2006].

При этом по данным одонтологии формирование народонаселения Кавказа представляло собой

развитие единой популяционной системы, связанной общностью происхождения и сохраняющей в совокупности исходный предковый фенотип западных групп Евразии [Кашибадзе, 2006].

Таким образом, велика потребность в новом историческом источнике, чтобы решить, какие из миграций оказали основное влияние на его генотип, а какие проявились лишь в культуре, не затронув генетические основы. Таким новым источником, дополняющим антропологические и археологические данные, могут стать современные генетические и лингвистические исследования.

НАРОДОНАСЕЛЕНИЕ КАВКАЗА ПО ДАННЫМ ЛИНГВИСТИКИ

СЕВЕРОКАВКАЗСКИЕ ЯЗЫКИ ПО ДАННЫМ ГЛОТТОХРОНОЛОГИИ

В результате многочисленных исследований языков северокавказской языковой семьи [Гигенишвили, 1977; Шагиров, 1977; Талибов, 1980; Бокарев, 1981; Chirikba, 1996] был накоплен большой массив лингвистических данных и выявлены закономерные фонетические соответствия между языками. В результате была создана классификация [Николаев, Старостин, 1994], которая стала наиболее общепринятой с некоторыми частными изменениями [Kuipers, 1963; Comrie, 1987; Ruhlen, 1987; www.ethnologue.com].

Эта классификация основана на распространенном методе анализа инноваций и на глоттохронологическом методе [Старостин, 1989], который в настоящее время широко используется в проекте «Эволюция языков человека» и координируется институтом Санта-Фе, США (<http://ehl.santafe.edu/intro1.htm>, <http://starling.rinet.ru/main.html>). Например, применение этого метода к современным языкам романской группы вывело дату их разделения от общего праязыка примерно 1600 лет назад. Это соответствует времени распада Римской империи, когда вульгарная латынь (*Latina Vulgata*, общий язык в провинциях империи) стала подразделяться на региональные диалекты [Blazhek, Novotna, 2008]. Этот пример подтверждает правомерность и точность такой методологии. Достоверность глоттохронологического метода датирования языков продемонстрирована на временных отрезках до 10 тысяч лет [Gell-Mann и др., 2008; Starostin, 2003].

Отметим, что имеются и иные лексико-статистические методы, в которых используется байесовский подход к увеличению статистической надежности языковой классификации [Gray, Atkinson, 2003; Kitchen et al., 2009; Greenhill et al., 2010]. Использование этих методов для сравнительного анализа лингвистического и генетического разнообразия Кавказа, видимо, будет предпринято в будущих исследованиях лингвистов.

РАСПРЕДЕЛЕНИЕ ОСНОВНЫХ ЯЗЫКОВЫХ ГРУПП НА КАВКАЗЕ

С точки зрения лингвистики Кавказ представляет собой мозаику, состоящую из более чем 50 языков, подавляющее большинство которых принадлежит к северокавказской языковой семье. Наиболее многочисленными представителями двух других языковых семей являются: а) ираноязычные осетины и пришлые славянские народы (индоевропейская семья); б) тюркоязычные народы (алтайская семья) [Ruhlen, 1987].

Обширные пространства от побережья Черного моря до Центрального Кавказа занимают народы абхазо-адыгской группы северокавказской лингвистической семьи. Еще несколько столетий назад почти все они пользовались собирательным самоназванием «адыге», а европейцами столь же собирательно именовались «черкесами». Огромное разнообразие племенных и родовых групп, говоривших на абхазо-адыгских языках, в XIX веке резко сократилось в результате Кавказской войны, и к настоящему времени выделяются только пять народов: абазыны, абхазы, адыгейцы, кабардинцы, черкесы.

К востоку от абхазо-адыгской группы расселены осетины Центрального Кавказа – единственный народ на Кавказе, который говорит на иранской ветви индоевропейской языковой семьи и который оседлал Большой Кавказский хребет, образовав по его северному и южному склонам Северную и Южную Осетию.

На востоке они граничат с нахскими народами северокавказской языковой семьи – ингушами, чеченцами и бацбийцами.

Еще восточнее, у Каспийского моря проживают народы Дагестана, столь многочисленные и лингвистически разнообразные, что идут споры о количестве языков и народов Дагестана – от 30 до 50. Однако все они, за исключением тюркоязычных карангайцев и кумыков, говорят на языках дагестанской ветви северокавказской языковой семьи.

Тюркоязычные народы алтайской языковой семьи на Кавказе не имеют приуроченности к определенному региону и представлены отдельными вкраплениями: на Западном и Центральном Кавказе балкарцы, карачаевцы и кубанские ногайцы, на Восточном Кавказе – кумыки и караногайцы.

Однако для оценки роли лингвистического родства в формировании генетического разнообразия такая классификация может рассматриваться лишь как вспомогательная. Она является качественной, а не количественной, и очерчивает лишь основные

контуры лингвистического древа. Тем более, что далеко не все лингвисты согласны с таким делением и для северокавказской языковой семьи предлагают новые ветвления и подразделения. Впрочем, ни одна из этих классификаций не позволяет провести прямое количественное сравнение генетики и лингвистики, что является одной из задач нашего исследования. Поэтому для выявления роли лингвистического родства мы использовали достижения современных направлений лингвистики – лексикостатистики и тесно с ней связанной глоттохронологии.

МЕТОДОЛОГИЯ ЛИНГВИСТИЧЕСКИХ ДАТИРОВОК

МЕТОДЫ ЛЕКСИКОСТАТИСТИКИ

Для реконструкции родства языков необходимо рассмотреть, какими методами языки могут быть классифицированы как точки на древе, отображающем их происхождение от праязыка. Языковые “знаки”, как правило, двухсоставные: они включают и семантику (значение), и текст, а связь между ними задается традицией конкретного языка. Случайное совпадение текста и значения в двух разных языках при рассмотрении большого набора слов является статистически маловероятным. Поэтому основной постулат сравнительно-исторического языкознания состоит в том, что такие совпадения показывают, что эта пара знаков в двух языках является двумя различными отражениями одного прото-знака. Две формы слов на двух языках могут быть выведены из третьей (гипотетической) прото-формы этого слова, следуя регулярным фонетическим правилам. Эти прото-знаки могут быть или унаследованы от общего праязыка, или же могут быть результатом контакта языков (заимствований). Чтобы различать эти два случая, используется определенный набор правил, касающихся набора слов (знаков) с их прото-формами.

Основное предположение специальной ветви сравнительно-исторического языкознания – глоттохронологии – состоит в следующем: процесс замены слов в языке может быть описан как случайная вероятность, и может быть вычислен средний коэффициент замещения.

Для этого необходимо включить в анализ слова со значениями, принадлежащими к так называемой предварительной (базовой, докультурной) лексике. Такие слова, как правило, не заимствуются при контакте народов, поскольку они являются базовыми для любой человеческой культуры, присутствуют в любом языке, и потому нет необходимости в заимствовании их из другого языка. Таким образом, списки слов, используемые в лексико-статистических расчетах, составляют из тех тематических групп, в которых заимствование слов про-

исходит редко [Swadesh, 1955; Старостин, 1989]. Редкие заимствованные слова считаются “информационным шумом” и исключаются из расчетов [Старостин, 1989].

Этот подход в его первоначальном варианте был предложен Сводешем [Swadesh, 1955]. В современных исследованиях вместо формулы линейной скорости замещения слов используется модель “релаксирующие часы” (“relaxed clock”) [Kitchen et al., 2009] или экспоненциальная формула. Списки слов Сводеша недавно были успешно использованы для реконструкции языков индоевропейской [Gray, Atkinson, 2003], семитской [Kitchen et al., 2009], австронезийской [Greenhill et al., 2010] и других языковых семей. В этих исследованиях большинство усилий было предпринято в направлении статистического анализа уже существующих наборов данных, в том числе математического моделирования и использования передовых подходов расчета (например, байесовские цепи Маркова, метод Монте-Карло), аналогичные тем, которые используются в эволюционной биологии.

Другая исследовательская группа – московской школы глоттохронологии – применила списки слов Сводеша для классификации тюркской [Дыбо, 2006], палеосибирской [Мудрак, 2008], северокавказской [Николаев, Старостин, 1995], алтайской [Старостин, 1991] и ряда других языковых семей. В этих исследованиях значительные усилия были вложены в создание новых высококачественных наборов данных, в том числе проведены этимологические исследования, установлены фонетические правила, а также исключены заимствования. Аналитические методы Сводеша были скорректированы С.А. Старостиным [Старостин, 1989], и с тех пор существенно не изменялись.

МЕТОД ЛИНГВИСТИЧЕСКОГО ДАТИРОВАНИЯ

Лингвистические датировки рассчитываются по числу замененных слов, которые накопились после распада праязыка [Старостин, 1989;

Embleton, 2000]. Этот глоттохронологический подход был впервые использован С.А. Старостиним для северокавказских языков [Николаев, Старостин, 1995].

Распад стабильных лексических элементов из первоначального списка может быть математически описан формулой, предложенной С.А. Старостиним [Старостин, 1989]:

$$C(t) = C_0 \cdot e^{-\lambda C(t)t^2}$$

где $C(t)$ – частота слов из основного списка, сохранившихся в данном языке в момент времени t ,

C_0 – первоначальный основной список,

λ – коэффициент скорости замены слов из основного списка.

Для лексической глоттохронологии используют 100-словники (список Сводеша), для которого $\lambda = 0.05$ [Старостин, 1989].

При расчете датировки распада двух языков, эта зависимость приобретает вид:

$$C_2(t) = C_0 \cdot e^{-2\lambda t^2 \sqrt{C_2(t)}}$$

где $C_2(t)$ частоты слов, выживших в обоих языках-потомках.

На основе этой формулы можно вычислить время t , используя коэффициент λ и частоту слов, сохранившихся в языковой паре:

$$t^2 = \frac{\ln \frac{C_2(t)}{C_0}}{-2\lambda \sqrt{C_2(t)}}$$

В нашем исследовании [Balanovsky et al., 2011] ведущими российскими лингвистами – член-корреспондентом РАН Анной Владимировной Дыбо и профессором Олегом Алексеевичем Мудраком – приложены особые усилия для создания высококачественной, расширенной и переработанной языковой базы северокавказских языков. К этим данным А.В. Дыбо и О.А. Мудраком был применен глоттохронологический подход. Благодаря этому в нашем исследовании появилась уникальная возможность провести прямое – полностью количественное – сопоставление языковых и генетических данных для одних и тех же этнических групп.

СХОДСТВО МЕЖДУ ЛИНГВИСТИЧЕСКИМИ И ГЕНЕТИЧЕСКИМИ МЕТОДАМИ ДАТИРОВАНИЯ

Расчет лингвистических дат основывается на подсчете различий, которые накопились после многочисленных разделений языков [Старостин, 1989]. Генетические даты рассчитываются по числу мутаций, которые накопились после роста дре-

ва гаплотипов в пределах гаплогруппы, как подробно описано в главе 1. Оба подхода используют коэффициент (скорость возникновения изменений), который преобразует разнообразие (генетическое или же языковое) во время. И лингвистика, и генетика получают этот коэффициент по калибровке на конкретных примерах с известным разнообразием и известным временем, за которое накопилось это разнообразие. Столь выраженное методологическое сходство лингвистических и генетических датировок позволяет напрямую сравнить генетические и языковые даты. Эти даты далее можно сопоставлять с событиями в истории популяции, которые привели к разделению языков и к появлению генетического разнообразия. Кроме сравнения генетических и лингвистических датировок, возможно сравнение генетических и лингвистических расстояний между одними и теми же популяциями. Для этого потребовалось только преобразовать лингвистическое сходство в расстояние, взяв обратные величины.

Отметим еще одну особенность, общую генетическим и лингвистическим датировкам.

С генетической точки зрения, когда прапопуляция распадается, и в одной из дочерних популяций возникает кластер гаплотипов Y-хромосомы, он будет ограничен этой популяцией (предполагается, что миграция мужчин между популяциями незначительна). Возраст кластера является не датировкой распада прапопуляции, а датировкой распространения кластера, которое могло произойти в любое время после распада прапопуляции.

С лингвистической точки зрения, нововведения, представленные в одном языке/диалекте, но отсутствующие в другом, возникают после разделения прото-популяции на частично изолированные субпопуляции. Языковая датировка разделения, таким образом, покажет тот момент, когда языковые различия уже стали фиксированными, а не момент самого разделения. Таким образом, начало популяционной дивергенции могло произойти за несколько поколений до даты, определяемой по лингвистическим данным. Возможно и дополнительное отставание языковой датировки на величину времени, необходимого для включения вновь вводимых слов из устной речи в письменные источники, которые используются для лексико-статистического анализа.

С учетом этих особенностей и языковой, и генетический методы датирования рассматривались в нашем исследовании как наиболее поздняя (нижняя, более близкая к современности) оценка времени разделения популяции.

7.2. СРАВНЕНИЕ ЛИНГВИСТИЧЕСКОЙ И ГЕНЕТИЧЕСКОЙ СТРУКТУРЫ КАВКАЗА

Изучение взаимосвязи лингвистической и генетической изменчивости народонаселения Кавказа проведено на обоих уровнях анализа Y-хромосомы – гаплогрупп и гаплотипов. На уровне гаплогрупп (данный раздел) разными методами сравнивается распределение частот гаплогрупп и распределения языков в одних и тех же популяциях. На уровне гаплотипов (раздел 7.3) выделяются и датируются кластеры гаплотипов, определяется их популяционная «принадлежность», и затем генетические датировки гаплотипов данной группы популяций сравниваются с лингвистическими датировками языков, на которых говорят эти популяции. Эти два вида анализа (на уровне гаплогрупп и на уровне гаплотипов) независимы друг от друга, что дает дополнительный контроль правильности выводов.

Отметим, что если анализ на уровне гаплогрупп на Кавказе проводился во многих исследованиях [Nasidze et al., 2003, 2004a,b; Bulaeva et al., 2006; Caciagli et al., 2009; Tofanelli et al., 2009; Кутуев, 2010; Дибирова и др., 2010; Yunusbaev et al., 2012; ряд других работ тех же коллективов], то анализ многосотенной выборки на уровне гаплотипов впервые был проведен в нашем исследовании и опубликован в [Balanovsky et al., 2011].

Выборки, изученные по Y-хромосоме, были охарактеризованы в главе 2 (табл. 2.3, рис. 2.1). Планирование этих выборок было проведено таким образом, чтобы охватить и основные регионы Северного Кавказа, географически подразделяемого на Западный, Центральный и Восточный, и ос-

новные лингвистические группы этих регионов. Исследование, охватившее все народы тюркской ветви алтайской семьи Северного Кавказ и Закавказья, является самостоятельным направлением, разрабатываемым нашим коллективом [Схалыхо и др., 2013; Схалыхо, 2013] в рамках взаимодействия народов Кавказа и евразийской степи, и потому в этой книге не рассматривается. Остальные изученные нами тринадцать народов относятся к трем ветвям (абхазо-адыгской, нахской и дагестанской) северокавказской языковой семьи, к трем ветвям (иранской, армянской и славянской) индоевропейской языковой семьи и представляют население четырех географических регионов: Западного, Центрального, Восточного Кавказа и Закавказья, как следует из приводимого ниже перечня.

Северокавказская языковая семья:

– *адыго-абхазская группа* (шапсуги, абхазы, черкесы);

– *нахская группа* (ингуши, чеченцы);

– *дагестанская группа* (аварцы, даргинцы, кубачинцы, кайтагцы и лезгины).

Индоевропейская языковая семья:

– *иранская группа* (осетины);

– *армянская группа* (армяне);

– *славянская группа* (терские и кубанские казаки).

Сравнение генетической и лингвистической структурированности популяций на уровне частот гаплогрупп проведено четырьмя независимыми методами. Результаты всех четырех видов анализа поочередно описываются в данном разделе.

СРАВНЕНИЕ МЕТОДОМ ДЕНДРОГРАММ

Самым простым методом является кластерный анализ – построение дендрограмм отдельно по генетическим и по лингвистическим данным, и затем их сравнение. Такой подход традиционен для популяционной генетики человека [Cavalli-Sforza et al., 1994].

Эти две дендрограммы для генетически изученных нами народов Кавказа представлены на рис. 7.1. На генетической дендрограмме кубачинцы и кайтагцы оказались наиболее схожими популяциями, образуя тесный кластер, к которому последовательно присоединяются аварцы и даргинцы, а затем лезгины (рис. 7.1, слева). На лингвистическом древе (рис. 7.1, справа) мы видим ту же топологию, за исключением того, что даргинцы присоединились к кубачино-кайтагскому кластеру до аварцев.

Следующий кластер идентичен на генетическом и лингвистическом деревьях (рис. 7.1) и содержит только народы нахской ветви – чеченцев и ингушей.

Третий кластер объединяет сперва шапсугов и черкесов, а затем к нему присоединяются абхазы. И вновь эта топология идентична на обеих дендрограммах.

Наконец, осетины, говорящие на индоевропейском языке, далеко отстоят на лингвистической дендрограмме от других народов. Аналогично и на генетической дендрограмме они занимают особое место, обнаруживая лишь небольшое сходство с абхазо-адыгским кластером.

Таким образом, дендрограмма, основанная на генетических расстояниях, точно отражает лингвистическую дендрограмму практически со всеми ее деталями.

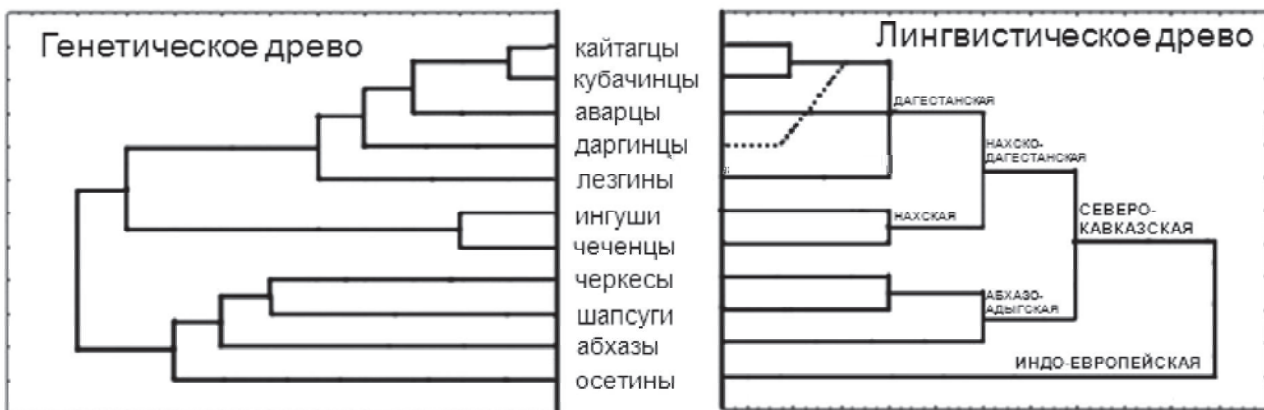


Рис. 7.1. Сравнение генетической и лингвистической дендрограмм родства народов Кавказа.

СРАВНЕНИЕ МЕТОДОМ ГЕНЕТИЧЕСКИХ ГРАНИЦ

Выявление генетических границ, то есть зон резких изменений между соседними генофондами, является традиционной задачей для популяционной генетики [Womble, 1951; Cavalli-Sforza et al., 1994]. Анализ, проведенный с помощью программы *Bagrier 2.2*, позволил объективным методом выделить положение генетических границ на карте Кавказа (рис. 7.2). Эти границы подразделяют генофонд Кавказа на отдельные сегменты.

Первая – самая резкая – граница «А» (рис. 7.2) разделяет генофонд Кавказа на народы Восточного Кавказа (нахско-дагестанская лингвистическая группа) и все прочие популяции.

Следующая по значимости граница «В» проходит внутри Восточного Кавказа и отделяет четыре популяции чеченцев и ингушей от народов Дагестана.

Следующая граница «С» разделяет ираноязычных осетин Центрального Кавказа и абхазо-адыгские народы Западного Кавказа.

Таким образом, первые три основные генетические границы разделили Кавказ на четыре региона. Важно, что каждый из этих регионов совпадает с зоной преобладания одной из четырех основных гаплогрупп Y-хромосомы. Рассмотрим это совпадение поочередно для каждого региона.

1. **На Западном Кавказе** в генофонде народов абхазо-адыгской группы мажорной является гаплогруппа **G-P303**. Ее средняя частота составляет 54%, варьируя с частотой от 21% у черкесов до 86% у шапсугов. Во всех других популяциях Кавказа частота этой гаплогруппы не поднимается выше 10%, в среднем составляя всего лишь 2%.

2. **На Центральном Кавказе** в генофонде осетин мажорной является гаплогруппа **G-P18**.

Ее средняя частота составляет 63%, варьируя от 60% у дигорцев до 73% у иронцев. У других народов Кавказа частота этой гаплогруппы не поднимается выше 12%, в среднем составляя всего лишь 3%.

3. **Для генофонда нахских народов Восточного Кавказа** – ингушей и чеченцев – характерна гаплогруппа **J-M67(xM92)**. Ее средняя частота составляет 70%, варьируя от 51% у чеченцев из Ингушетии до 84% у ингушей. У остальных этносов Кавказа частота этой гаплогруппы не превышает 9%, в среднем составляя всего лишь 3%.

4. **У народов Дагестана** мажорной является гаплогруппа **J-M267(xP58)**. Ее средняя частота составляет 70%, варьируя у аварцев, даргинцев, кайтагцев, кубачинцев и лезгин от 44% до 99%. С частотой менее 25% она встречается у чеченцев и ингушей (нахская языковая группа), а у остальных народов Кавказа ее частота менее 5%.

Таким образом, приуроченность каждой из четырех гаплогрупп к одному из четырех доменов на Кавказе выражена очень ярко. Эту закономерность можно интерпретировать двумя способами.

С одной стороны, она показывает четкую связь с географией, так как каждая из четырех гаплогрупп Y-хромосомы преобладает в конкретном географическом регионе Кавказа.

С другой стороны, обнаруженная закономерность не менее четко вписывается в языковую классификацию, поскольку каждая из четырех гаплогрупп приурочена к популяциям, говорящим на языках одной из четырех основных лингвистических групп Кавказа.

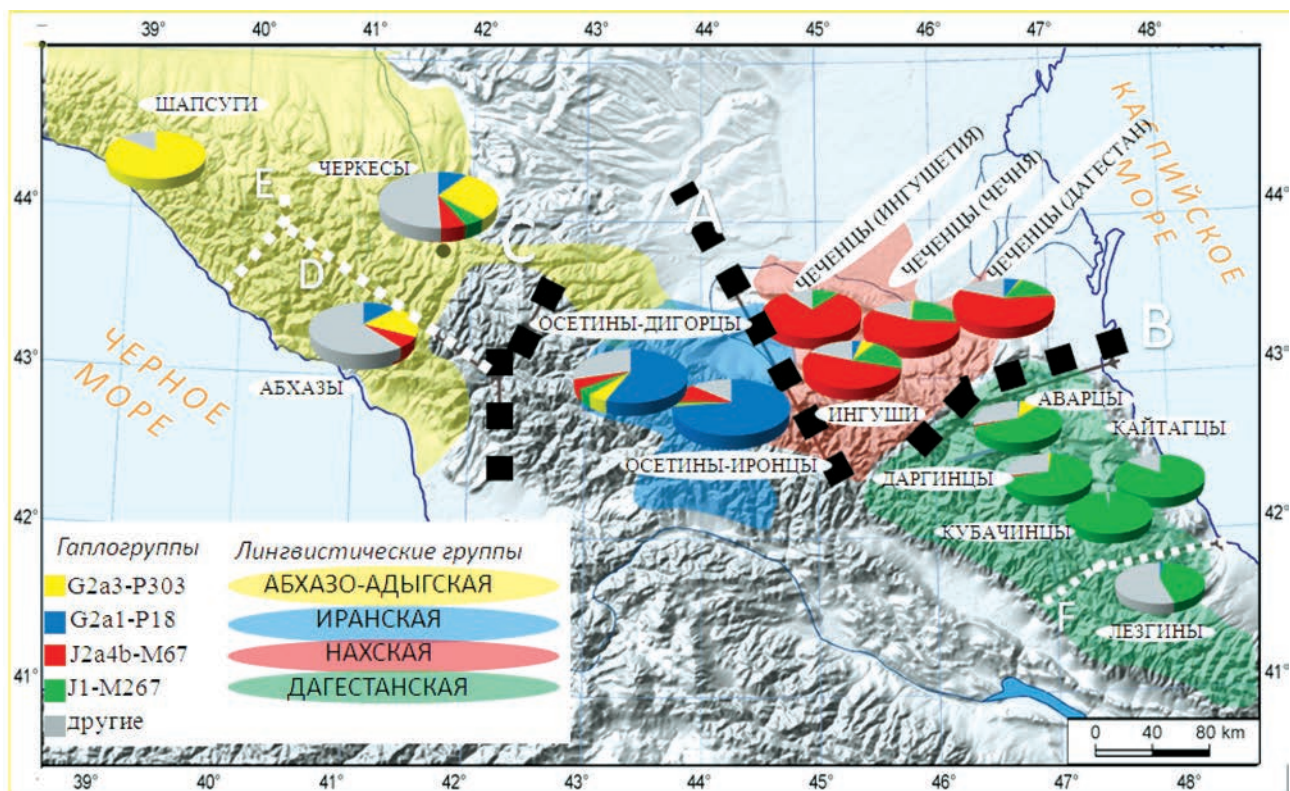


Рис. 7.2. Географическая приуроченность четырех гаплогрупп и анализ генетических границ.

Обозначения. Каждая изученная популяция показана круговой диаграммой, размер секторов разного цвета отражает частоту соответствующей гаплогруппы Y-хромосомы согласно легенде. Пунктирные линии обозначают объективные генетические границы, полученные с помощью программы *Barrier 2.2* и обозначенные буквами латинского алфавита по степени убывания значимости границ (черным пунктиром обозначены первые три наиболее значимые границы – А, В, С). Цветным фоном показаны ареалы лингвистических групп.

СРАВНЕНИЕ МАТРИЦ РАССТОЯНИЙ

В предыдущем подразделе получено идеальное соответствие распределения гаплогрупп Y-хромосомы и с географической, и с лингвистической структурой Кавказа. Необходимо решить, какой из этих факторов является определяющим. Предыдущие исследователи Кавказа пришли к выводу о том, что основным фактором в формировании структуры генофонда выступал географический фактор, а не лингвистический [Nasidze et al., 2003; 2004; Кутуев, 2010]. Однако в нашем исследовании мы не только располагали намного более обширными выборками для всех основных этносов Северного Кавказа, генотипированными по намного более широкой панели SNP-маркеров, но и, благодаря тесному сотрудничеству с ведущей школой лексикостатистики, впервые имели возможность провести не качественный, а точный количественный анализ лингвистического сходства.

Для сравнения роли географического и лингвистического факторов мы вычислили три матрицы попарных расстояний между всеми изученными народами: матрицу генетических расстояний (табл. 7.1), матрицу лингвистических расстояний (табл. 7.2) и матрицу географических расстояний

(табл. 7.3). Из всех изученных коренных народов Северного Кавказа в анализ не были включены лишь осетины, говорящие на языке индоевропейской семьи, поскольку расчет лингвистических расстояний между народами, относящимися к разным языковым семьям, проблематичен. Данные о популяциях, говорящих на одном языке, были для этого вида анализа усреднены.

Сравнение скоррелированности этих трех матриц проводилось с помощью теста Мантеля. В результате (табл. 7.4) обнаружены достоверные ($p=0.002$) корреляции – как между генетикой и лингвистикой ($r = 0.64$) так и между генетикой и географией ($r = 0.60$). Отметим, что корреляция с лингвистикой оказалась несколько выше.

При проведении корреляционного анализа нужно учитывать зависимость от взаимодействия факторов. Корреляция двух показателей может быть следствием того, что они оба скоррелированы с третьим признаком. В этом случае необходим анализ корреляций между двумя признаками при фиксированных значениях остальных признаков (анализ частных корреляций). Поэтому частные корреляции рассчитаны меж-

Таблица 7.1. Генетические расстояния (по частотам гаплогрупп) между народами Кавказа

	Абхазы	Аварцы	Ингуши	Даргинцы	Лезгины	Черкесы	Шапсуги	Чеченцы
Абхазы	0	1.40	1.42	3.39	1.02	0.07	0.04	1.15
Аварцы	1.40	0	1.50	0.05	0.06	1.14	1.66	0.81
Ингуши	1.42	1.50	0	1.88	1.87	0.74	2.62	0.03
Даргинцы	3.39	0.05	1.88	0	0.19	1.80	4.32	1.02
Лезгины	1.02	0.06	1.87	0.19	0	1.01	1.38	1.03
Черкесы	0.08	1.14	0.74	1.80	1.01	0	0.17	0.61
Шапсуги	0.04	1.66	2.62	4.32	1.38	0.17	0	1.83
Чеченцы	1.14	0.81	0.034	1.02	1.03	0.61	1.83	0

Таблица 7.2. Лингвистические расстояния (доля общих слов*) между народами Кавказа

	Абхазы	Аварцы	Ингуши	Даргинцы	Лезгины	Черкесы	Шапсуги	Чеченцы
Абхазы	0	4.55	5.26	5.26	5.26	2.04	2.08	5.26
Аварцы	4.55	0	3.23	3.33	3.23	4.76	4.76	3.03
Ингуши	5.26	3.23	0	4	3.85	4.35	4.35	1.08
Даргинцы	5.26	3.33	4.00	0	3.23	4.35	4.17	4.35
Лезгины	5.26	3.23	3.85	3.23	0	4.76	4.76	3.85
Черкесы	2.04	4.76	4.35	4.35	4.76	0	1.08	4.55
Шапсуги	2.08	4.76	4.35	4.17	4.76	1.08	0	4.55
Чеченцы	5.26	3.03	1.08	4.35	3.85	4.55	4.55	0

*Лингвистические расстояния рассчитаны по методу С.А. Старостина [Старостин, 1989] в сотрудничестве с представителями московской школы глоттохронологии (член-корр. РАН А.В. Дыбо и проф. О.А. Мудраком). Расстояния между парой языков представляют собой величину, обратную доле совпадающих слов из стословного стандартного списка М. Сводеша, после исключения лингвистических заимствований.

Таблица 7.3. Географические расстояния (в километрах) между народами Кавказа

	Абхазы	Аварцы	Ингуши	Даргинцы	Лезгины	Черкесы	Шапсуги	Чеченцы
Абхазы	0	470	330	520	540	190	280	370
Аварцы	470	0	170	60	120	440	650	140
Ингуши	330	170	0	230	280	280	490	40
Даргинцы	520	60	230	0	70	500	710	200
Лезгины	540	120	280	70	0	550	760	260
Черкесы	190	440	280	500	550	0	210	310
Шапсуги	280	650	490	710	760	210	0	520
Чеченцы	370	140	40	200	260	310	520	0

Таблица 7.4. Корреляция между генетическими, лингвистическими и географическими расстояниями

Тип корреляции	Корреляция между параметрами	Коэффициент корреляции	Уровень значимости
Парная корреляция	Генетика и лингвистика	0.64	0.002
Парная корреляция	Генетика и география	0.60	0.001
Частная корреляция	Генетика и лингвистика (без влияния географии)	0.34	0.120
Частная корреляция	Генетика и география (без влияния лингвистики)	0.21	0.180

ду генетикой и каждым из других признаков – лингвистики и географии – по отдельности, при фиксированных значениях второго фактора. К сожалению, анализ осложняется тем, что лингвистическая изменчивость на Кавказе сама тесно связана с географией ($r = 0.78$). Поэтому частные корреляции с обоими факторами не дости-

гают уровня достоверности (табл. 7.4). Но важна тенденция: связь генетики с лингвистикой ($r = 0.34$) почти в полтора раза сильнее, чем с географией ($r = 0.21$). Таким образом, коэффициенты и общей, и частной корреляции указывают на более выраженную связь генетики с лингвистикой, чем с географией.

СРАВНЕНИЕ МЕЖПОПУЛЯЦИОННОГО РАЗНООБРАЗИЯ

Для проверки этой тенденции был использован независимый метод AMOVA (табл. 7.5). Метод AMOVA позволяет сгруппировать популяции согласно различным классификациям и сравнить получившиеся величины генетического межпопуляционного разнообразия. В качестве одной классификации послужили географические регионы (популяции группировались согласно географическому соседству), в качестве альтернативной классификации взята лингвистическая (популяции группировались по их лингвистическому родству). Немаловажно, что в этом виде анализа мы можем учесть генофонды не только народов северокавказской языковой семьи (как при анализе матриц расстояний), но и осетинские популяции, представляющие индоевропейскую языковую семью в коренном населении Северного Кавказа.

В первом случае доля генетической изменчивости между географическими группами составила $F_{ST} = 0.15$. Во втором варианте доля генетической изменчивости между лингвистическими группами оказалась почти в два раза выше: $F_{ST} = 0.27$. Таким образом, лингвистическая классификация популяций выявляет намного большую часть межпопуляционной изменчивости генофондов, чем классификация по географическим регионам. То есть именно лингвистическое сходство, а не географическое соседство объясняет основную часть изменчивости Y-хромосомы на Кавказе.

Этот результат был получен впервые нашим коллективом [Дибирова и др., 2010; Balanovsky et al., 2011]. В предшествовавших работах достоверные

корреляции первоначально вообще не удавалось выявить [Nasidze et al., 2004]. Это могло быть связано с малыми выборками и недостаточным уровнем филогенетического разрешения (анализировались всего десять гаплогрупп, тогда как в нашем исследовании – в четыре раза больше). В работе [Кутуев, 2010] была снова предпринята попытка оценить роль географического и лингвистического фактора в структурировании генофонда Кавказа по Y-хромосоме, и обнаружена несколько более высокая связь с географией. Вероятно, это объясняется отчасти вновь недостаточной дробностью гаплогрупп (в особенности отсутствием подразделения гаплогруппы **G-P15** на субтипы, которые как раз и дифференцируют осетин от популяций Западного Кавказа), а отчасти использованием не количественных, а менее точных качественных лингвистических данных. В более поздней статье того же коллектива [Yunusbaev et al., 2012] использованы большие и подробно охарактеризованные выборки, но вопрос внутренней структуры генофонда Кавказа по Y-хромосоме специально не ставился, поскольку основная задача работы состояла в сравнении Кавказа с другими регионами мира, причем главным образом по аутосомным маркерам.

Итак, все четыре вида проведенного анализа (кластерный анализ, анализ генетических границ, тест Мантеля и AMOVA) убедительно демонстрируют, что для народонаселения Кавказа лингвистическое родство является более важным, чем географическое соседство, фактором формирования разнообразия Y-хромосомы.

Таблица 7.5. Сравнение генетической информативности географической и лингвистической классификаций методом AMOVA

Уровни популяционной системы	Лингвистические группы	Географические группы
Генетические различия между группами	0.268	0.146
Генетические различия между популяциями внутри групп	0.099	0.235
Суммарные межпопуляционные генетические различия	0.367	0.381
Генетические различия внутри популяций	0.633	0.619

7.3. ФИЛОГЕНЕТИЧЕСКИЕ ДАТИРОВКИ НА КАВКАЗЕ

Если предыдущий раздел был посвящен анализу на уровне гаплогрупп Y-хромосомы, то теперь мы переходим к анализу на уровне гаплотипов. Поскольку все образцы изучались не только по панели SNP-маркеров, но и по панели 17 STR-маркеров, то был впервые получен обширный массив данных о STR-гаплотипах кавказских популяций. На основании этого массива данных были постро-

ены филогенетические сети всех основных кавказских гаплогрупп, выделены кластеры гаплотипов и определена их популяционная принадлежность. Наиболее существенно, что генетические датировки этих популяционно-специфичных кластеров гаплотипов были сопоставлены с лингвистическими датировками языков, на которых говорят эти популяции.

ДАТИРОВКИ ПОПУЛЯЦИОННО-СПЕЦИФИЧНЫХ КЛАСТЕРОВ ГАПЛОТИПОВ

Были построены филогенетические сети для одиннадцати основных гаплогрупп Кавказа: **G2a*-P15(xP16,M406,P303)**, **G2a1a-P18**, **G2a3b1-P303**, **J1*-M267(xP58)**, **J2*-M172(xM47,M67,M12)**, **J2a3b*-M67(xM92)**, **L-M20**, **R1a1*-M198(xM458)**, **R1a1a1g-M458**, **R1b1a2-M269** и **Q-M242**. Почти на всех сетях обнаружено явление популяционной специфичности кластеров гаплотипов.

Наиболее яркой в этом отношении является филогенетическая сеть гаплогруппы **G2a1a-P18** (рис. 7.3). Эта гаплогруппа встречается почти исключительно на Кавказе. Сеть состоит из ветвей STR-гаплотипов, соединенных в центральной ретикулярной зоне. На рисунке видно разделение на два крупных и один меньший кластер, обозначенные, соответственно, α , β и γ .

Кластер α включает в основном STR-гаплотипы, встреченные у осетин-иронцев. Структура кластера звездчатая, с частым центральным гаплотипом основателя и несколькими основателями второго порядка. В кластер β попали преимущественно гаплотипы осетин-дигорцев. Для него тоже видна звездчатая структура. Третий, менее многочисленный кластер γ включает обе осетинские популяции и имеет более сложную структуру.

Если бы эта гаплогруппа распространялась на Кавказе долгое время, она имела бы совершенно другой вид. Звездчатая же структура обоих кластеров указывает на их сравнительно молодой возраст. Для кластеров α и β среднее число мутационных шагов сходно (1,46 и 1,41), что указывает на одно время возникновения.

Таким образом, филогенетическая сеть гаплогруппы, наиболее частой у осетин, демонстрирует подразделение осетин на две субпопуляции. Мы видим запечатленную в филогенетической сети микроэволюцию осетинской популяции: ее разделение на две дочерних, возникновение в каждой из них своего гаплотипа-основателя, и постепенное нарастание генетического разнообразия обеих субпопуляций в ходе мутационного процесса. Совпадение датировок показывает, что, скорее всего, оба кластера возникли почти сразу же после разделения на субпопуляции.

Подобное явление обнаруживается и на филогенетических сетях для остальных гаплогрупп Кавказа (рис. 7.4). Для каждой гаплогруппы обнаружено наличие специфичных кластеров гаплотипов, характерных для отдельной популяции и отсутствующих у других народов.

Таких популяционно-специфичных кластеров обнаружены десятки, что позволяет провести статистически обоснованный анализ. Обилие популяционно-специфичных кластеров, вероятно, вызвано высокой эндогамией и изоляцией кавказских популяций. Для последующего анализа отбирались лишь наиболее надежные кластеры, которые соответствовали следующим критериям:

а) кластер не должен быть слишком мелким – входящие в него гаплотипы должны быть встречены суммарно у не менее чем 10 человек;

б) кластер должен четко идентифицироваться на медианной сети: все гаплотипы кандидатного кластера должны соединяться с центральным гаплотипом-основателем, и все дистальные по отношению к нему гаплотипы обязательно должны быть включены в кластер;

в) кластер должен обладать высокой популяционной специфичностью: не меньше 85% образцов в кластере гаплотипов должны относиться к данной популяции. У многих кластеров этот показатель составлял 100%.

Несмотря на строгость этих требований, выявлено восемнадцать соответствующих им кластеров. Их характеристика представлена в табл. 7.6. Кроме обозначения кластера и указания на его специфичность для той или иной популяции, в таблице также указаны генетические датировки кластера и лингвистические датировки выделения того языка, на котором говорит данная популяция (или группы языков, на которых говорит группа популяций, обладающих данным кластером). Последний столбец таблицы указывает на то событие в истории формирования популяционной структуры Кавказа, которое, на наш взгляд, маркируется данным кластером.

Как указано в табл. 7.6, датировки кластеров были рассчитаны с применением двух скоростей

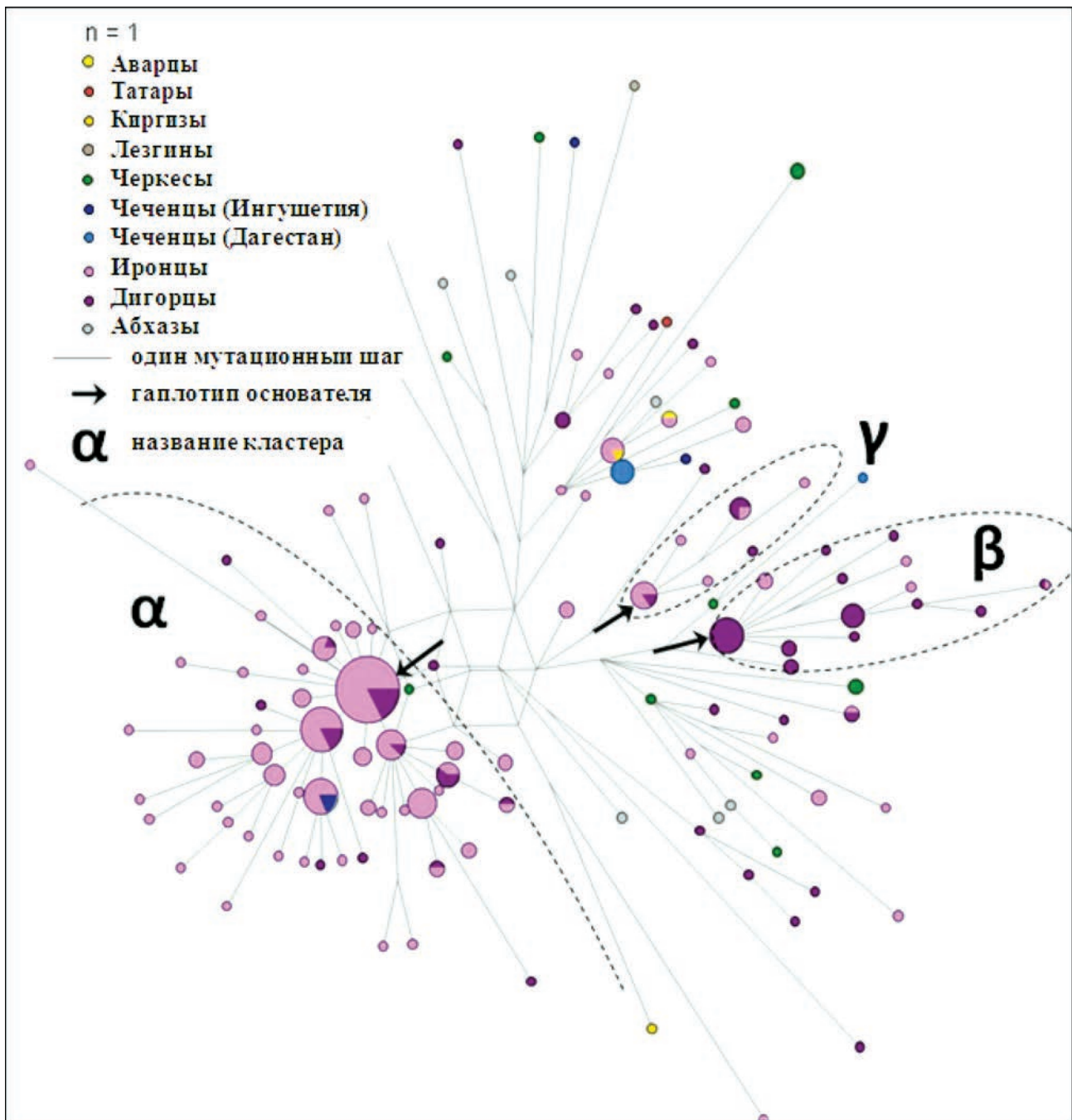


Рис. 7.3. Филогенетическая сеть гаплогруппы G2a1a-P18.

Обозначения: площадь кружка пропорциональна числу носителей данного STR-гаплотипа. Цветовые обозначения указаны в легенде. Описание кластеров приведено в тексте.

мутирования – эволюционной скорости [Zhivotovskiy et al., 2004] и генеалогической скорости, полученной при прямом сравнении на парах «отец-сын» [Ge et al., 2009]. Близкие значения генеалогической скорости получены и в других работах. Вопросы, связанные с датировками и оценкой скоростей мутаций подробно рассмотрены в специально посвященном этим вопросам разделе 1.5 первой главы.

Напомним, что наша интерпретация популяционно-специфичных кластеров основывается на предположении их появления *после* отделения дан-

ной популяции от родственных групп. Ведь в противном случае (если кластер возник до разделения популяций) кластер не был бы специфичным, и встречался бы у всех потомков той прото-популяции, в которой он возник. Поэтому, как указывалось в начале этого раздела, совпадением генетической и лингвистической (или исторической датировки) можно считать случаи, когда генетическая дата меньше или равна лингвистической, а несовпадением – случаи, когда генетическая дата более древняя. (В разделе 7.1 обсуждалось, что и лингвистическая дата может быть более поздней

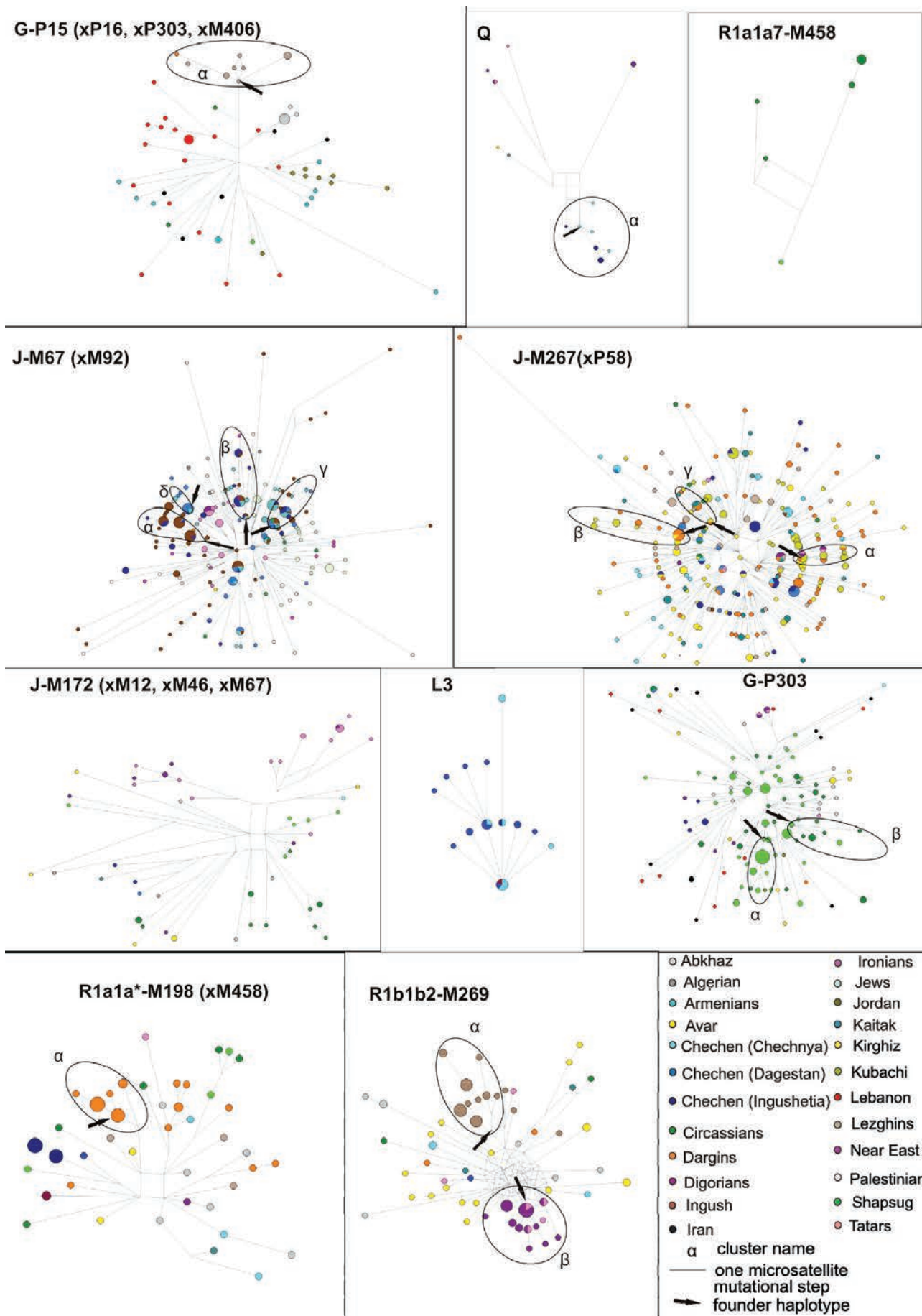


Рис. 7.4. Филогенетические сети для десяти основных кавказских гаплогрупп.

Таблица 7.6. Генетические датировки кластеров гаплотипов и лингвистические датировки языков (в числе лет от современности)

КЛАСТЕРЫ	ПОПУЛЯЦИИ	ДАТИРОВКА по «генеалогической» скорости мутаций	ДАТИРОВКА по «эволюционной» скорости мутаций	Лингвистическая датировка	Примечание
P303-α	Шапсуги	600±200	1900±700	800	Отделение адыгов (шапсугов) от черкесов
P303-β	Шапсуги и черкесы	1400±500	4300±1400	3600	Отделение черкесов от абхазов
P18-α	Иронцы	1400±500	4200±1400	1300	Деление осетин на иронцев и дигорцев
P18-β	Дигорцы	1300±500	4100±1400	1300	
R1b1b2-β	Дигорцы	800±300	2300±800	1300	
P18-γ	Иронцы и дигорцы	1200±500	3500±1400	-	-
M67-β	Чеченцы и ингуши	1300±700	4100±2100	5600	Отделение популяции нахской языковой группы от собственно дагестанской
M67-γ	Чеченцы и ингуши	1600±600	4800±1800	5600	
M67-α	Ингуши	1100±500	3300±1600	1400	Деление нахской языковой ветви на чеченскую и ингушскую
M67-δ	Чеченцы	100±100	400±300	1400	
L3	Чеченцы	900±400	2900±1100	1400	
Q-α	Чеченцы	1100±500	3300±1600	1400	
M267(xP58)α	Даргинцы Кубачинцы	1000±500	3100±1400	3400	Отделение даргинцев, кубачинцев и кайтагцев от других дагестанских популяций
M267(xP58)β	Даргинцы Кубачинцы	1300±600	3900±1900	3400	
M267(xP58)γ	Даргинцы Кубачинцы Кайтагцы	1000±600	3200±2000	3400	
R1a1a*-α	Даргинцы	900±400	2700±1300	1900	Отделение кубачинцев от даргинцев
R1b1b2-α	Лезгины	2000±700	6100±2300	4300	Отделение лезгин от даргинской популяции
P15*-α	Лезгины	2400±700	7400±2200	4300	

по сравнению со временем разделения популяций, но это запоздание касается лишь несколько поколений и носит систематический характер, тогда как генетическое запоздание может быть на тысячелетия и является случайным; поэтому целесообразно брать за ориентир лингвистическую дату и смотреть относительно нее датировки генетических кластеров).

По сравнению с соответствующими лингвистическими датировками «эволюционные» датировки в большинстве случаев оказались слишком завышены, тогда как датировки, основанные на «генеалогической» скорости мутаций, в основном хорошо соответствовали лингвистическим датировкам. Справедливость именно этих дат как минимум в двух случаях подтверждается и историческими сведениями, как описано в следующем подразделе. Отметим, что к похожим выводам пришли и авторы работы [King, Jobling, 2009], где исследовалась связь Y-хромосомы с фамилиями: полученная ими скорость мутирования близка к «генеалогической», а не к «эволюционной». Оценки скоростей мутирования, основанные на полногеномном секвенировании Y-хромосомы, также дают результаты, близкие к «генеалогической» скорости по

STR-маркерам; этот вопрос наиболее основательно исследован в [Karmin et al., 2015] и авторы приходят к выводу, что «эволюционная» скорость хорошо работает лишь для гаплогрупп старше 30 000 лет, а на меньшей глубине времени (особенно для гаплогрупп возрастом около 5 тысяч лет) применяемая скорость мутирования должна быть значительно более быстрой. (Фактически это означает, что для подавляющего большинства гаплогрупп быстрая «генеалогическая» скорость дает более точные оценки, чем медленная «эволюционная»).

Одно из возможных объяснений этого наблюдения – зависимость оценки возраста от размера кластера и способа выбора гаплотипа-основателя. При калибровке эволюционной скорости [Zhitovitsky et al., 2004] использовались кластеры с 2–4 гаплотипами, и выбор гаплотипа поэтому не был формализован. В нашем исследовании, благодаря изобилию кластеров, отбирались кластеры с не менее чем 10 гаплотипами, исходящими из одного основателя. Эти методологические различия могли сказаться на среднем уровне разнообразия выявляемых кластеров, а, следовательно, и на получаемых оценках времени накопления этого разнообразия.

СРАВНЕНИЕ ИСТОРИЧЕСКИХ, ЛИНГВИСТИЧЕСКИХ И ГЕНЕТИЧЕСКИХ ДАННЫХ

Из серии популяционных событий, перечисленных в табл. 7.6, рассмотрим те два события, по которым имеются исторические данные, что поможет независимо проконтролировать правильность совпадающих друг с другом генетических и лингвистических дат.

ОСЕТИНЫ

Первое событие – разделение осетин на два основных субэтноса – иронцев и дигорцев. Существует три гипотезы этого распада. Первая гипотеза состоит в происхождении иронцев и дигорцев от различных, хотя и родственных друг другу сарматских племен, обитавших в степях за пределами Кавказа. Вторая гипотеза связывает появление иронцев и дигорцев на Центральном Кавказе в III–II веках до н.э. с ираноязычными сираками и аорсами до нашествия гуннов (III–IV века н.э.). Третья гипотеза, имеющая наибольшее число сторонников, постулирует происхождение иронцев и дигорцев от алан в VI веке н.э. в горах и предгорьях Центрального Кавказа [Ковалевская, 1981]. Эта гипотеза подтверждается и первым письменным упоминанием о дигорцах, которое находится в летописях “Армянской географии” в VII веке н.э. [Агеева, 2000]. Напомним (табл. 7.6), что лингвистической датировкой является VII век и та же дата получена по обоим осетинским кластерам гаплотипов в нашем исследовании. В итоге все три источника – исторический, лингвистический и генетический – согласуются с распадом осетинского этноса на иронцев и дигорцев в период с VI в. по VIII в. нашей эры.

ЧЕРКЕСЫ

Второе событие – это разделение черкесов и шапсугов. Черкесы и шапсуги выделились из надэтнической общности адыгов, состоявшей из многих племен [Федоров, 1983; Бетров, 1991]. При этом предки генетически изученных нами популяций шапсугов и черкесов проживали на смежных территориях в низовьях реки Кубани [Агеева, 2000]. В начале XI в. в Предкавказье стали проникать кочевники-тюрки, под давлением которых часть адыгов стала продвигаться на юго-восток. Разгром алан монголо-татарами способствовал дальнейшему продвижению в XII–XIV веках адыгских племен на обширные пространства Центрального Кавказа [Федоров, 1983]. Позднее, вместе с некоторыми другими родственными племенами в результате обособления и ассимиляции части коренного населения Центрального Кавказа образовался черкесский этнос [Федоров, 1983; Агеева, 2000].

Таким образом, разделение шапсугской и черкесской популяции, согласно историческим датам, происходит между X и XIV вв., а лингвистические датировки устанавливают XIV век, то есть 700 лет назад.

Для шапсугов популяционно-специфическим кластером является **P303-а**. Его генетическое разнообразие составляет $r=0.64$, что (при использовании «генеалогической» скорости) дает датировку 500 ± 200 лет назад. Таким образом, и в этом примере можно видеть соответствие генетической датировки историческим и лингвистическим данным.

ПАРАЛЛЕЛИЗМ ЛИНГВИСТИЧЕСКОЙ И ГЕНЕТИЧЕСКОЙ МИКРОЭВОЛЮЦИИ

Впервые обнаруженное нами явление популяционной специфичности кластеров гаплотипов народов Северного Кавказа позволило совместить генетическую и лингвистическую реконструкции их микроэволюции. На рис. 7.5 приведено реконструированное лингвистами древо северокавказских языков, иллюстрирующее разделение праязыка на его ветви, языки и диалекты. На это лингвистическое древо были наложены все 18 обнаруженных популяционно-специфичных генетических кластеров: каждый кластер был наложен на ту ветвь, специфичным для которой он является. Можно видеть многочисленные примеры соответствия лингвистической и генетической эволюции, что позволяет предположить, что генетическая структура Кавказа развивалась параллельно с формированием многообразия языков Северного Кавказа.

Предполагаемой нами причиной такой параллельной эволюции генетических и лингвистических изменений являются демографические процессы в популяциях Кавказа. Имеются в виду обычные процессы увеличения численности и расширения ареала, которые приводят к географическому разделению прото-популяции и последующей изоляции ее дочерних групп, чему в немалой степени должен был способствовать горный рельеф региона. Вследствие такой изоляции, новые гаплотипы Y-хромосомы (неизбежно возникающие в силу мутационного процесса) в каждой из дочерних групп оказываются различными, специфичными для нее, и не распространяются к соседям. Однако, когда при увеличении численности этой дочерней популяции она, в свою очередь, делилась на несколько «внучатых», то и внутри этих попу-

ляций со временем возникали кластеры гаплотипов. И эти новые возникающие кластеры окажутся специфичны уже для этих изолированных популяций третьего порядка.

Перечисленные события постепенного разделения прото-популяций и изоляции дочерних групп хорошо известны и лингвистам - они являются необходимым условием для разделения языков. Именно поэтому те, демографические по своей сути, процессы, которые приводят к образованию в каждой популяции своего генетического кластера, приводят и к образованию в ней своего языка. При таком подходе к микроэволюции популяций следует удивляться скорее не тому поразительно-

му совпадению общей картины и многих деталей, которая выявлена между генетическими и лингвистическими данными на Кавказе, а тому, что такое совпадение не было до сих показано ни для одного другого региона мира. На наш взгляд, это объясняется отчасти продуманной методологией синтеза наук – использованием количественных лингвистических данных, единообразной организацией лингвистической и генетической информации, формализованным анализом филогенетических сетей. Но во многом этому способствовала и специфика самого региона, характеризующегося, благодаря горному рельефу, особой консервативностью и устойчивостью популяционной структуры.

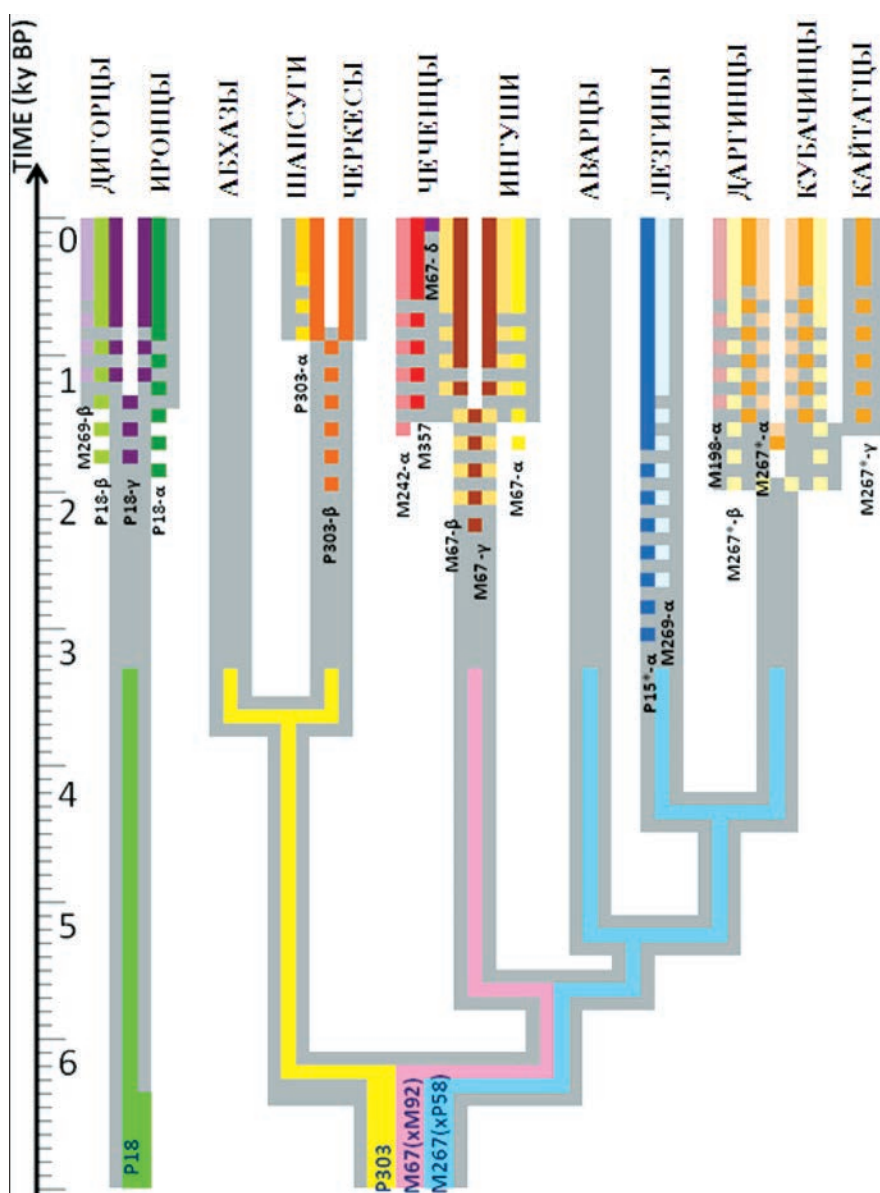


Рис. 7.5. Комбинированная схема микроэволюции генофонда Кавказа.

Серым фоном показано древо ветвления языков. На это древо наложены генетические кластеры (выделены цветом), характерные для отдельных популяций. В целом схема показывает постепенное разделение единой прапопуляции населения Кавказа на отдельные группы. В каждой из этих групп формировался особый диалект или язык, и параллельно возникали генетические кластеры. В большинстве случаев прослеживается совпадение генетических датировок (кластеров гаплотипов Y-хромосомы), полученных для данной популяции (доверительный интервал показан цветным пунктиром) и времени обособления языка, на котором говорит данная популяция.

7.4. ВОЗМОЖНОСТИ ПОЛНОГО СЕКВЕНИРОВАНИЯ Y-ХРОМОСОМЫ

Из предыдущих глав читатель уже мог прийти к выводу об особой роли Y-хромосомы в современном арсенале популяционной генетики – не заменяя собой ни полногеномные, ни митохондриальные данные, Y-хромосома все же характеризуется максимальной межпопуляционной изменчивостью и информативностью для реконструкции истории популяций. Y-хромосома – это подзорная труба туманным днем: то, что исследователь разглядит вдалеке с ее помощью, может оказаться неверным, но без подзорной трубы он вообще не сможет смотреть вдаль и будет ограничен тем, что видит вблизи невооруженным глазом. Поэтому тому, что видно в подзорную трубу, не стоит безоговорочно доверять, а нужно потом подойти ближе и проверить – но еще менее стоит отвергать эти данные с порога. Точно так же и

результаты по Y-хромосоме нуждаются в последующей проверке другими системами, когда накопленные по ним данные и методы анализа позволят «подойти поближе и проверить», но при этом показания Y-хромосомы заслуживают самого пристального внимания исследователя.

Именно поэтому эта глава по связи генофондов и языков основана в основном на данных Y-хромосомы. И важно, что увеличительная сила и четкость изображения этой подзорной трубы резко возросла в последние два-три года благодаря появившейся возможности полного секвенирования Y-хромосомы. В этом разделе я кратко опишу эти возможности и приведу один из примеров – возможную связь «полногеномных» результатов по Y-хромосомной гаплогруппе **G1-M285** и распространения языков иранской группы.

ОБЗОР МЕТОДОВ И ПЕРВЫХ РЕЗУЛЬТАТОВ

«ПОЛНОЕ» СЕКВЕНИРОВАНИЕ Y-ХРОМОСОМЫ

Среди всех генетических систем Y-хромосома характеризуется наибольшей межпопуляционной изменчивостью и, следовательно, наиболее высокой разрешающей способностью при анализе генофондов. Использование технологий полногеномного секвенирования для углубленного анализа Y-хромосомы стало новейшей тенденцией в этой области: в 2013–2014 годах по этой тематике вышли 2 статьи в *Science*, статьи в *Nature Communications* и *Genome Research*, не считая менее знаковых журналов.

Эти статьи основывались на небольшом числе образцов, поскольку данные по Y-хромосоме поступали только в рамках дорогостоящего полного секвенирования геномов. Но начиная с 2014 года целый ряд коллективов (в том числе и наш) начал широкомасштабные исследования. Это стало возможным благодаря тому, что в конце 2013 г. стало коммерчески доступным избирательное секвенирование только Y-хромосомы – соответствующие технологии ее обогащения были независимо разработаны двумя компаниями: FtDNA (США) и BGI (Китай), а в 2014 году появилась и работа английских специалистов, разработавших третью технологию. Отметим, что приоритетность полного секвенирования Y-хромосомы в изучении истории человечества была очевидна для мирового сообщества – потому-то ведущие коллективы стремились к решению этой задачи. Однако полное секвенирование отдельной хромосомы является технически более сложной задачей, чем секвенирование всего генома, поэтому решение этой задачи и

было осуществлено позднее (2013–2014 гг.), причем на основе различающихся технологий.

Поскольку действительно полное секвенирование всей длины Y-хромосомы существующими методами практически невозможно (по причине избытка повторяющихся элементов вдоль большей части хромосомы), существует так называемый «золотой стандарт» избранных для ресеквенирования регионов Y-хромосомы. К этому стандарту различные технологии добавляют дополнительные регионы, поэтому суммарная длина ресеквенирования варьирует от 9 млн. до 17 млн. п.н. из примерно 60 млн. пар нуклеотидов общей длины Y-хромосомы. Именно эти последовательности (формально это не полная Y-хромосома, а лишь примерно ее шестая или четвертая часть) принято обозначать как «полные сиквенсы Y-хромосомы».

Впрочем, само по себе ресеквенирование является не наиболее интересной, а напротив, лишь технической стадией работы. Значительно больший интерес представляют последующие, собственно научные этапы: биоинформационная обработка результатов, построение филогенетических деревьев, популяционный скрининг новых открытых маркеров (то есть мутаций, маркирующих новые открытые ветви на древе Y-хромосомы) на сотнях и тысячах образцов разных популяций, картографический анализ, получение и интерпретация данных по тонкой структуре генофонда.

АЛГОРИТМ АНАЛИЗА ДАННЫХ

О логике анализа данных полного секвенирования Y-хромосомы мы уже говорили в главе 1 (раз-

дела 1.3 и 1.5). Чтобы реализовать эту логику, необходимо создать алгоритм анализа данных. К настоящему времени такие алгоритмы разработаны уже целым рядом коллективов, активно работающих с данными по полным Y-хромосомам, и все эти алгоритмы довольно похожи. Кратко опишу алгоритм, который был разработан, апробирован и отлажен нашим коллективом.

Первичные данные о секвенированных коротких фрагментах (риды) были картированы на геноме человека (сборка 37й версии) и определены полиморфные сайты (отличия от референсной последовательности). Для целей контроля качества отслеживалось общее число ридов, среднее покрытие и среднее качество секвенирования на нуклеотид. В результирующих файлах, кроме обнаруженных в секвенированных образцах отличий от референсной последовательности, также был указан статус 37 000 филогенетически информативных SNP-маркеров, присутствовавших на момент анализа в базах данных International Society of Genetic Genealogy (www.isogg.org) и Family Tree DNA (www.familytreedna.com). Также применялись дополнительные фильтры, среди которых важнейшим был фильтр по отсутствию гетерозиготности. Учитывалась и согласованность результатов прямого и обратного прочтений: если все генотипы данного сайта были получены только при чтении в одном направлении, они рассматривались как менее достоверные. Генотипы, не поддающиеся однозначной автоматической обработке, направлялись на ручную корректировку оператору, которому предоставлялись все исходные данные с указанием числа и качества ридов, перекрывающих данный нуклеотид, и результатов каждого рида по данной позиции.

Кроме генотипирования каждого образца, для решения общей задачи необходимо, чтобы каждый обнаруженный SNP маркер был генотипирован во всех образцах. Это представляет собой техническую проблему, поскольку большинство методов определения полиморфных сайтов оставляют неоднозначность в интерпретации позиций, для которых не обнаружено наличие «варианта» (т.е. отличия от референсной последовательности): это может быть как надежно показанное соответствие с референсом, так и отсутствие надежных данных в силу, например, недостаточного покрытия данного участка в данном образце. Чтобы обеспечить различение этих двух ситуаций, для каждого образца был определен список надежно генотипированных регионов на основании данных о качестве секвенирования (вероятности правильности указанного аллельного состояния на основании phred-показателя) каждого нуклеотида в данном образце. Суммарная длина этих надежно секвенированных участков составляет, в среднем, около 10 млн. п.н., т.е. примерно на 10% меньше, чем полный

секвенируемый регион (11 млн. п.н.). Благодаря наличию этих данных, при последующей обработке можно считать, что если для данного образца в данной позиции не указано отличия от референсной последовательности и данная позиция попадает в интервал надежно секвенированных регионов для данного образца, то в данной позиции надежно показано соответствие с референсной последовательностью.

По итогам анализа были созданы отдельно по каждой анализируемой гаплогруппе итоговые таблицы генотипов по данным секвенирования Y-хромосомы. Эти таблицы были обработаны филогенетическими программами построения деревьев (TNT, MURKA, BEAST) и получены итоговые филогенетические деревья.

УЖЕ ПРОВЕДЕННЫЕ ИССЛЕДОВАНИЯ

Несмотря на молодость этого научного направления (менее трех лет) уже вышли или готовятся к печати целая плеяда работ ведущих популяционно-генетических коллективов мира, включая статьи в наиболее престижных научных журналах. Эти исследования можно условно разделить на три направления: работы по оценке «полногеномной» скорости мутирования Y-хромосомы, работы по отдельным гаплогруппам и работы по оценке общих параметров генофонда мира. Правда, многие статьи развивают сразу два или даже все три направления.

Попытки определить «полногеномную» скорость мутирования уже описывались в разделе 1.5 и суммированы на рисунке 1.22. Первым методом был классический метод калибровок, соответствующий понятию «эволюционной» скорости в терминах, разработанных при дискуссии о скоростях по STR-гаплотипам Y-хромосомы. Это были калибровки по заселению Сардинии [Francalacci et al., 2013] и Америки [Poznik et al., 2013]. Вторым методом был анализ числа мутаций в известных родословных, соответствующий понятию «генеалогической» скорости. Это работы по многочисленным исландским родословным [Helgason et al., 2015] и с более ранней оценкой по одной английской родословной [Xue et al., 2008]. Третьим, промежуточным способом является «клановая» оценка – хотя все члены клана восходят к общему предку (как при «генеалогической» оценке), клан представляет собой популяцию, эволюционирующую в течение многих поколений (как при «эволюционной» оценке). Этот подход реализован в работе нашего коллектива [Balanovsky et al., 2015]. Четвертый подход основан на анализе древней ДНК образцов, надежно датированных радиоуглеродным методом, и оценке генетического разнообразия современных потомков линий, близкородственных этим древним образцам. Этот подход реали-

зован в работе [Karmin et al., 2015]. Принципиально важно, что эти скорости, оцененные шестью различными коллективами на основе четырех различных подходов, оказались близки и варьируют около величины $0,8 \cdot 10^{-9}$ мутаций на нуклеотид в год (рис. 1.22).

Исследования, сосредоточенные на отдельных гаплогруппах, почти столь же многочисленны, и, несомненно, их число будет стремительно увеличиваться в самые ближайшие годы. Среди них надо назвать статью [Rootsi et al., 2013], в которой подробно анализируется гаплогруппа **R1a** у евреев-ашкенази и выявляется ее ближневосточное (а не восточно-европейское, как предполагалось ранее) происхождение в популяциях евреев Европы. О втором исследовании – нашей собственной работе по гаплогруппе **G1** [Balanovsky et al., 2015] – речь пойдет чуть ниже. Третьим исследованием, уже представленном в научном мире (хотя пока только в форме доклада на конференции, а не полноценной статьи), является работа В. Trombetta (из коллектива под руководством Scuciani) по филогеографии гаплогруппы **E-M35** в Африке. В этой работе четко видны все три составляющие полногеномных исследований отдельных гаплогрупп:

1) полное секвенирование Y-хромосомы исследуемой гаплогруппы с построением детального дерева;

2) выбор для популяционного скрининга некоторых ветвей дерева (выделившихся субгаплогрупп), выбор для них SNP-маркеров, скрининг на наличие этих маркеров большого числа популяций и определение частот тестируемых субгаплогрупп во всех этих популяциях;

3) совместный анализ данных по географическому распространению (картографирование частот) и по возрасту (анализ дерева) выбранных субгаплогрупп и выводы по истории миграций популяции (и более широко – их демографической истории).

Кроме этих исследований, Эстонским биоцентром с участием нашего коллектива завершается работа по филогеографическому анализу гаплогруппы **N1**, нашим коллективом завершаются исследования гаплогрупп **C3-M217** и **R1b**, и, несомненно, другими коллективами сейчас тоже ведутся интенсивные исследования других гаплогрупп, которые будут опубликованы в ближайшие пару лет.

Третье направление – исследования, нацеленные на общие выводы о демографической истории человечества по полным сиквенсам Y-хромосомы, – пока представлены только одной работой [Karmin et al., 2015], выполненной под руководством Тоомаса Кивисилда. Но она одна стоит десятка. В ней не только обнаружено «бутылочное горлышко», через которое в конце неолита прошла мужская часть популяции (именно мужская, потому что общий эффективный размер в это время только возрастал), но и дано подробнейшее филогенетическое дерево как в целом, так и по отдельным гаплогруппам, оценены параметры разнообразия для отдельных регионов мира, примирены «эволюционная» и «генеалогическая» скорости мутирования по Y-STR маркерам и многое другое. Можно надеяться, что это обобщающее направление вскоре привлечет внимание и других исследователей.

ИССЛЕДОВАНИЕ ГАПЛОГРУППЫ **G1**: СКОРОСТЬ МУТАЦИЙ И МИГРАЦИИ ИРАНСКОЙ ГРУППЫ

Y-хромосомная гаплогруппа **G1-M285** известна уже много лет, но не привлекала большого внимания исследователей из-за ее в среднем низкой частоты. В то же время было известно, что в некоторых популяциях Юго-Западной и Центральной Азии она достигает частоты 80%. Когда в 2011 году мы создавали Атлас изменчивости Y-хромосомы в мире, то картографировали все изученные к тому времени гаплогруппы Y-хромосомы. И обнаружили, что паттерн географического распределения гаплогруппы **G1** очень своеобразен: в отличие от подавляющего большинства евразийских гаплогрупп ее нельзя назвать однозначно восточноевразийской или западноевразийской. К тому же ее ареал интригующим образом объединял горы Юго-Западной Азии и Евразийскую степь – то есть регионы, которые в историческом плане объединяет только эпоха широкого расселения народов иранской группы. Поэтому, когда появилась возможность

полного секвенирования Y-хромосомы, мы решили применить ее к этой любопытной гаплогруппе и заодно отладить методологию анализа на этом не слишком сложном примере гаплогруппы **G1**.

КАРТА ЧАСТОТЫ

Прежде всего, мы провели генотипирование в дополнительных популяциях, собрали новые опубликованные данные и построили уточненную карту распространения гаплогруппы **G1** (рис. 7.6). Карта основывается на данных о ненулевых частотах гаплогруппы **G1-M285** в 27 популяциях, изученных нашим коллективом, на данных из литературы о 33 популяциях, а также на данных по нескольким сотням популяций, где эта гаплогруппа не была обнаружена (информация о нулевых частотах важна для точного картографирования). Карта показывает, что гаплогруппа **G1** имеет доста-

точно широкий ареал: от Италии до Монголии, но с наивысшими частотами обнаружена в центральноазиатских степях и на Переднеазиатском нагорье. Так, один пик ее частот приходится на Армению, причем среди армян она наиболее распространена лишь в одной группе – у амшенских армян (42%). Второй пик приходится на Казахстан, где она преобладает у родоплеменной группы аргын. Среди 288 казахских индивидов, принадлежащих к **G1**, 90% оказалось аргынами, и, наоборот, среди аргынов 80% несли эту гаплогруппу. Как аргыны, так и амшенские армяне являются сейчас многочисленными группами, поэтому подобное увеличение частоты **G1** у них нельзя объяснить недавним действием дрейфа генов.

Таким образом, уточненная карта подтвердила, что гаплогруппа **G1** связывает народонаселение степей и гор, причем как раз в древнем ареале ираноязычных популяций. Миграция ираноязычных популяций между центральноазиатскими степями и возвышенностями Юго-Западной Азии – это один из важных вопросов в поиске прародины индоевропейцев. В соответствии с курганной ги-

потезой, носители иранских языков распространились с территории евразийских степей на территорию современного Ирана. Альтернативная анатолийская гипотеза помещает прародину индоевропейцев в восточную Анатолию, предполагая, что ираноязычные популяции именно оттуда мигрировали в степи. Таким образом, две гипотезы предполагают противоположные направления миграции

КАРТА РАЗНООБРАЗИЯ

Вторым шагом исследования было построение карты гаплотипического разнообразия (показатель, пропорциональный числу выявленных гаплотипов в пределах гаплогруппы). Эта карта (рис. 7.7) показывает исключительно четкий тренд убывания разнообразия от максимальных значений (выше 0,8) на западе Ирана до нулевого разнообразия в Монголии. Можно предполагать, что этот тренд соответствует направлению распространения гаплогруппы, поскольку в общем случае разнообразие максимально в предковой популяции, а мигри-

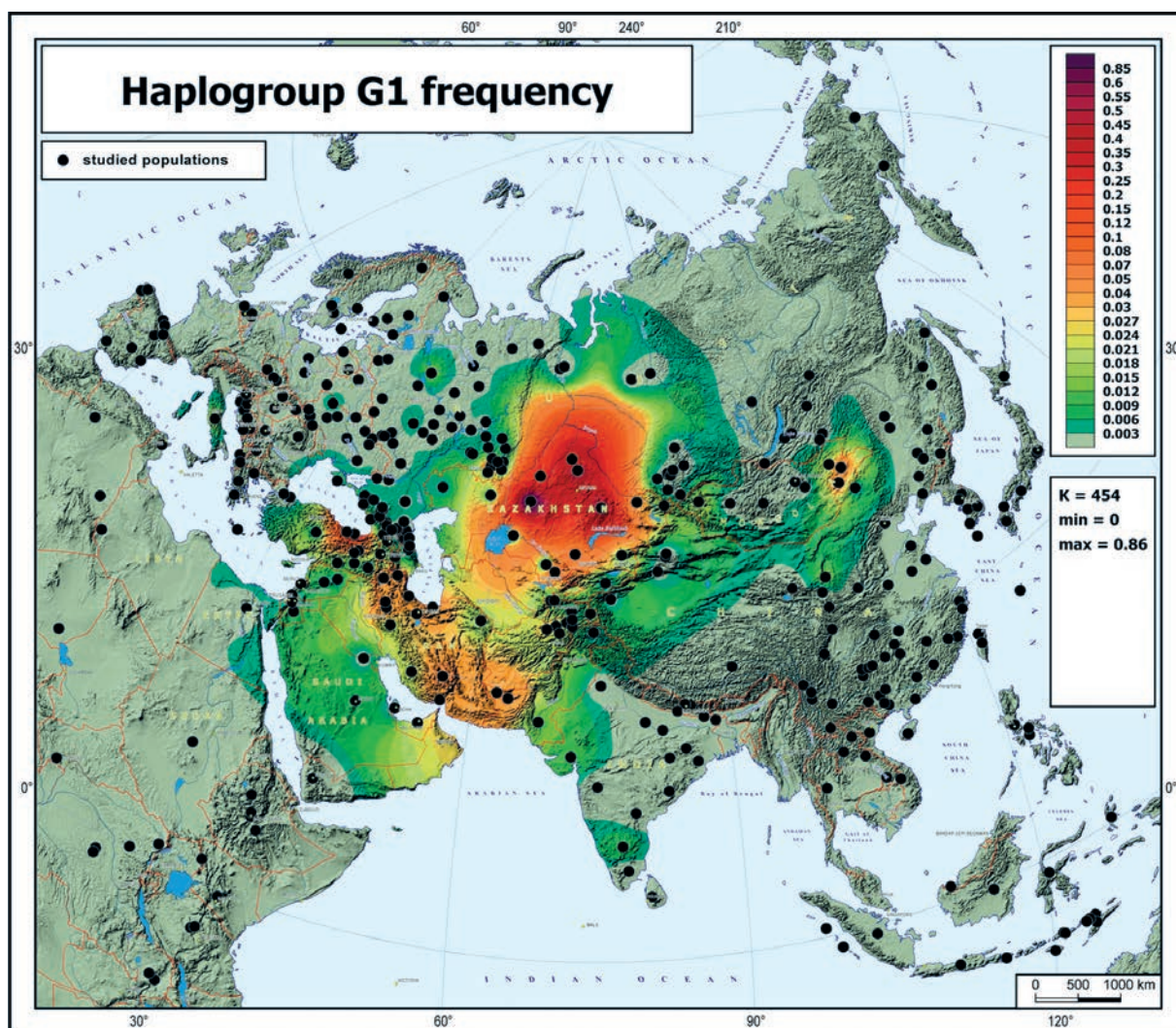


Рис. 7.6. Карта распространения гаплогруппы G1-M285.

ровавшие дочерние популяции захватывают лишь часть разнообразия из родительской популяции. И потому их гаплотипическое разнообразие – в общем случае – меньше, чем в родительской популяции. Конечно, в частном случае возможна утрата гаплотипического разнообразия и в родительской популяции, например за счет эффекта бутылочно-горлышка. Но такие случаи могут сказываться при локальных исследованиях и маловероятны при изучении обширных географических регионов.

ФИЛОГЕНЕТИЧЕСКАЯ СЕТЬ ПО STR-ГАПЛОТИПАМ

Следующим логическим шагом должен был стать полногеномный анализ. Но поскольку он является слишком дорогостоящим, чтобы проводить его для всех образцов, возникает вопрос, как отобрать образцы на полное секвенирование так, чтобы при минимальном числе образцов охватить максимальное филогенетическое разнообразие. Ведь если какая-либо филогенетическая ветвь не будет представлена хотя бы одним образцом (а на самом деле необходимы хотя бы два образца), мы никак

не узнаем о том, что эта ветвь существует. Поэтому к отбору образцов мы подошли очень тщательно и использовали 4 критерия:

- географический и этнический (образцы из разных частей ареала гаплогруппы и из разных этносов);
- филогенетический (образцы из всех кластеров, выявленных на сети по STR-маркерам, поскольку эти кластеры могут отражать реальные филогенетические ветви);
- «клановый» (критерий применим только к популяциям с сохранившейся родовой структурой и состоит во включении образцов не из одного-двух, а из возможно большего числа разных родов).

По сути, все эти критерии основаны на той информации об образцах, которая известна до их секвенирования и которая может оказаться скоррелированной с распространением разных ветвей гаплогрупп.

Для применения филогенетического критерия было нужно сперва построить эту сеть по STR-гаплотипам. Обратим внимание на эту ситуацию – хотя полногеномное дерево является более точной и надежной филогенетической реконструкцией и по-

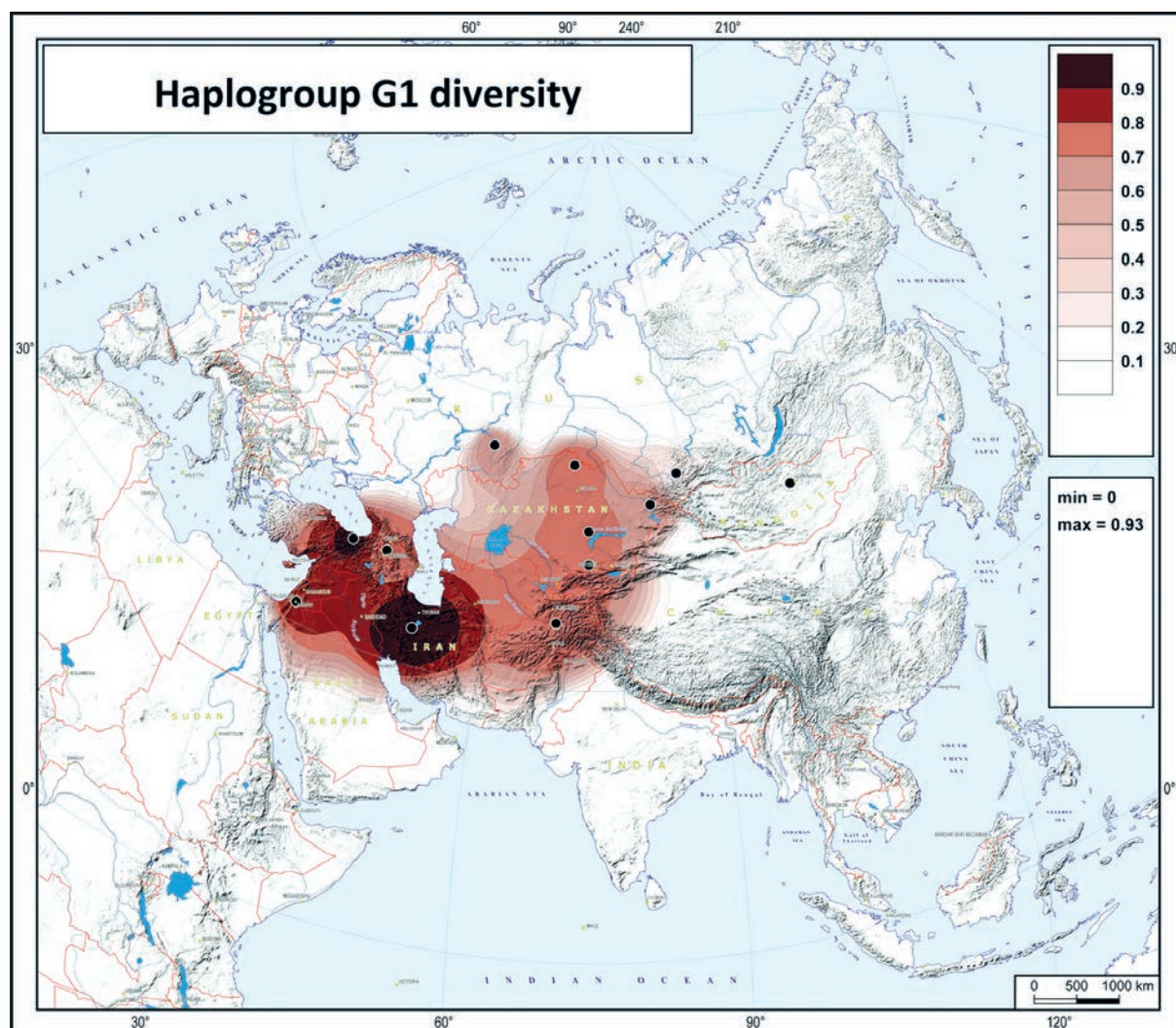


Рис. 7.7. Карта гаплотипического разнообразия в пределах гаплогруппы G1-M285.

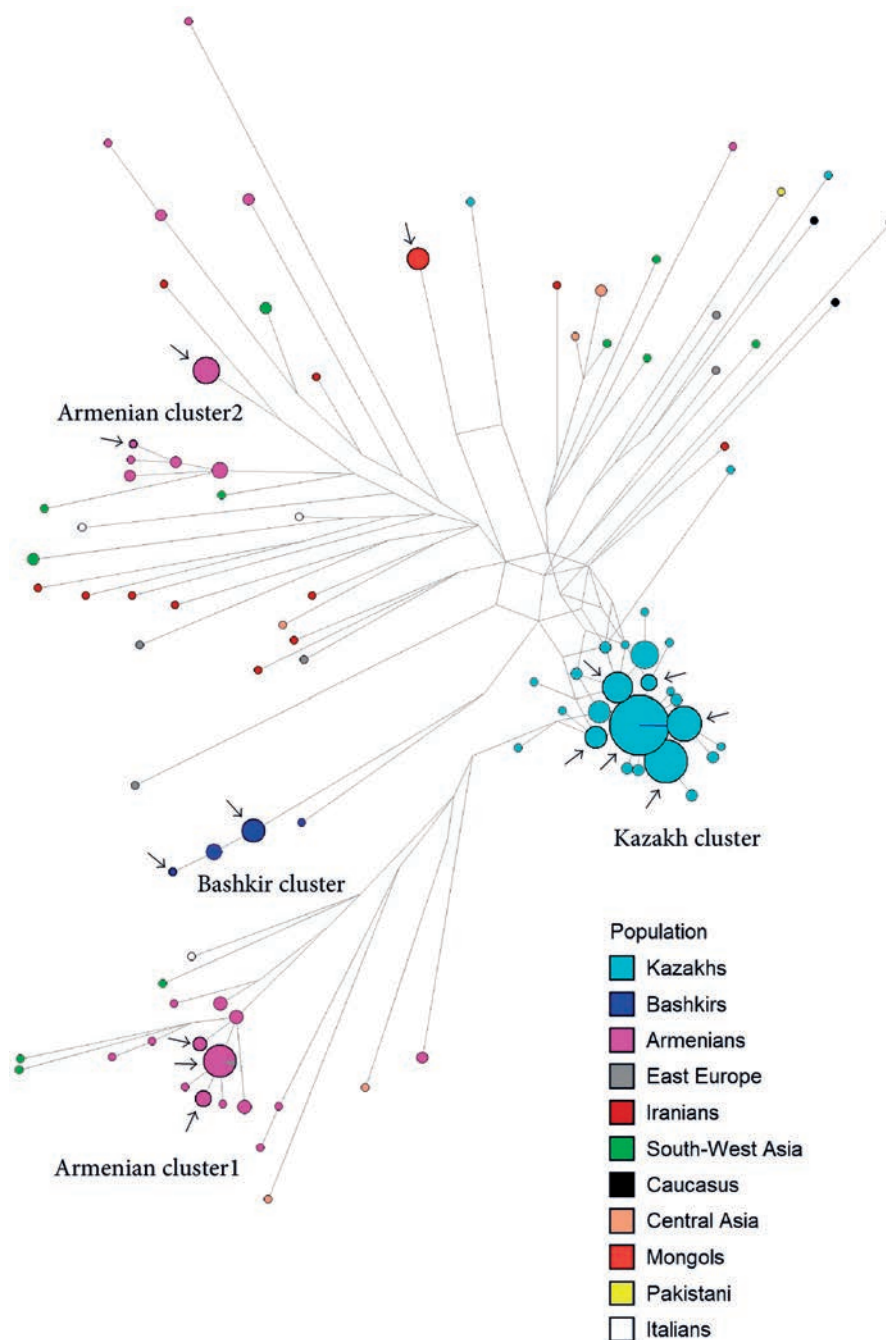


Рис. 7.8. Филогенетическая сеть STR-гаплотипов в пределах гаплогруппы G1. Стрелки указывают на образцы, отобранные для полногеномного секвенирования. Гаплотипы представителей разных народов обозначены разными цветами. Выделяются: казахский кластер, два армянских кластера и башкирский кластер гаплотипов.

этому заменяет собой сеть по STR-маркерам, для наилучшего построения полногеномного дерева весьма желательно построить сперва эту сеть по STR-гаплотипам (рис. 7.8).

На сети четко выделились четыре кластера: два из них были представлены образцами из Армении, один – самый большой – образцами из Казахстана, и один – из Башкирии. Гаплотипы из других популяций представлены в сети разрозненно и не сгруппировались в кластеры. Из каждого кластера были отобраны хотя бы по два образца для полногеномного секвенирования Y-хромосомы: анализ одиночных образцов не позволил бы отличить филогене-

тически информативные SNP-маркеры от «частных» (private) маркеров, свойственных только одной семье.

Применение остальных критериев побудило нас включить в анализ также монгольский образец, а для казахов включить различные родовые подгруппы в пределах аргынов. Кстати, клановая (более широко – субэтническая) приуроченность гаплогруппы G1 проявилась у всех народов, для которых обнаружилось ее накопление: у казахов она типична для рода аргынов, у армян она часта в группе амшен, у башкир все носители гаплогруппы G1 принадлежат к роду канглы.

ФИЛОГЕНЕТИЧЕСКОЕ ДЕРЕВО ПО ДАННЫМ ПОЛНОГО СЕКВЕНИРОВАНИЯ

В итоге для полного секвенирования были отобраны 19 образцов гаплогруппы **G1** и в качестве внешнего таксона (outgroup) был взят один образец «братской» гаплогруппы **G2**. Для секвенирования отобранных образцов использовали метод BigY, разработанный компанией FamilyTreeDNA. Как указывалось выше, этот метод позволяет секвенировать участки Y-хромосомы, составляющие в сумме 11 миллионов нуклеотидов. Эти участки секвенировали с высоким покрытием (в среднем

каждый нуклеотид был прочитан 67 раз). Данные по еще двум образцам **G1** (из популяций Западной Индии) были взяты из проекта «1000 геномов», но эти образцы были секвенированы с низким покрытием.

После того, как к полученным данным был применен наш алгоритм, описанный в предыдущем разделе, и была получена таблица, содержащая для каждого из 20 секвенированных нами образцов значения 636 обнаруженных нами маркеров, которые в этом массиве данных, во-первых, надежно генотипированы для каждого образца и, во-вторых, полиморфны среди данного набора образцов (и,

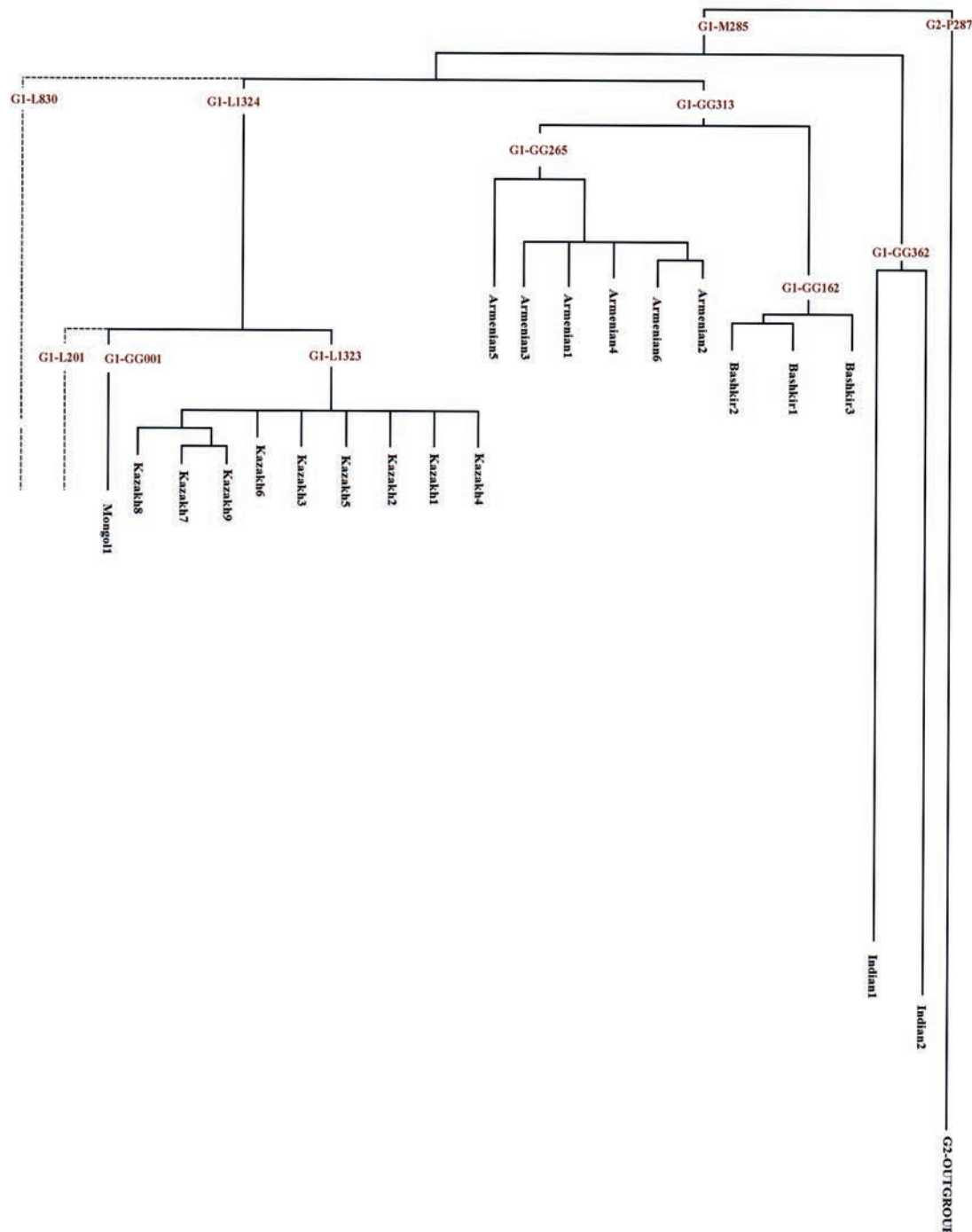


Рис. 7.9. Филогенетическое древо гаплогруппы G1-M285, построенное по данным полного секвенирования Y-хромосомы.

следовательно информативны для выявления внутренней филогенетической структуры). Для построения дерева на основе этой таблицы были применены как методы, основанные на парсимонии, так и байесовский подход, и полученные деревья практически совпали. Это показывает, что данные, полученные при полном секвенировании, настолько обширны и надежны, что реконструируемое на их основе дерево перестает зависеть от применяемых методов и их настроек, и однозначно определяется самими исходными данными.

Полученное филогенетическое древо представлено на рис. 7.9. На нем очень четко видно наличие четырех ветвей, условно названных как «армянский», «казахский», «башкирский» и «индийский» кластеры. Каждый из них характеризуется 100% специфичностью для соответствующих этнотерриториальных групп популяций.

Армянский и башкирский кластеры гаплогруппы **G1** имеют общего предка, тогда как казахи и индийцы располагаются на отдельных ветвях. Ветвь гаплогруппы **G1**, образованная единственным монгольским образцом, объединяется с казахским кластером, что хорошо согласуется и с географическим положением этих народов, и с их историей. Казахская ветвь соответствует ранее описанной субгаплогруппе **G-L1323**, в то время как армянская (**G1-GG265**), башкирская (**G1-GG162**), монгольская (**G1-GG1**) и индийская (**G1-GG362**) ветви выявлены в нашем исследовании впервые. Аббревиатура **GG** в названиях обнаруженных нами маркеров означает геногеографию (Gene Geography) – принято, что имена маркеров начинаются с букв, показывающих, каким научным коллективом они были открыты (так, маркеры, начинающиеся на букву **M**, открыты коллективом Peter Underhill, P – Michael Hammer, L – FamilyTreeDNA и т.д.).

Если сравнить обе филогенетические реконструкции – сеть по STR-гаплотипам (рис. 7.8) и древо по полногеномным данным (рис. 7.9) видно согласие во всех основных чертах. Разумеется, есть и отличия: несколько армянских кластеров, обнаруженных по STR маркерам, оказываются вариантами единого кластера по полногеномным данным; иное положение занимает и монгольский образец.

Отметим, что все ветви на дереве (рис. 7.9) сходны по длине, как и должно быть при постоянных скоростях мутаций. Единственное исключение – очень длинные ветви из образцов проекта «1000 геномов» – вероятно, это связано с тем, что используемые фильтры не совсем подходят к образцам, секвенированным с малым покрытием.

ГЕНЕТИЧЕСКОЕ И ГЕНЕАЛОГИЧЕСКОЕ ДЕРЕВЬЯ

В этом исследовании у нас появилась редкая возможность сравнить данные по Y-хромосомному секвенированию с достоверными исторически-

ми источниками по казахским кланам и откалибровать таким образом скорость мутирования Y-хромосомы.

Родоплеменная группа аргынов, в которой доминирует гаплогруппа **G1**, по преданиям восходит к общему предку (Аргыну) и впоследствии разделилась на 12 кланов. Исторических свидетельств существования Аргына нет, но Караходжа (согласно легенде – правнук Аргына) – это историческая личность. Например, он упоминается как посланник хана Тохтамышша – предводителя Золотой орды – к Тамерлану в 1405 году. Большинство родов аргынов считают себя потомками Караходжи, а другие кланы считают, что происходят от его брата Сомдыка.

Пока вся эта картина генеалогии аргынов – лишь исторические сведения, достоверность которых неизвестна. Но в нашем исследовании, как упоминалось, одним из критериев отбора образцов был охват различных кланов, поэтому в полученных данных из двенадцати родов аргынов было представлено девять. Это позволило получить объективную генетическую реконструкцию реального биологического родства по мужской линии девяти изученных представителей казахов-аргынов. И достоверность этой реконструкции уже очень велика – по сути, можно подвергать сомнению только мелкие детали дерева (случайно произошла дополнительная мутация, что немного удлинило ветку в числе мутаций по сравнению с ее реальной длиной в числе поколений), но основная структура реальной генеалогии этих 9 человек, несомненно, такова, какой она предстает на генетическом дереве.

Сравнение «легендарного» генеалогического родства изученных 9 казахов и генеалогии, реконструированной генетическими методами, представлено на рис. 7.10. Их сходство – практически совпадение – очевидно. Действительно, мы видим, что все люди, восходящие к одному предку по традиционной генеалогии, и по генетическим результатам имеют общего недавнего предка (а сравнение с деревом на рис. 7.9 показывает, что, например, башкиры и монголы, не восходящие к общему предку с аргынами по традиционной генеалогии, и генетически относятся совсем к другим ветвям). Более того, сама топология генетического дерева соответствует топологии генеалогического дерева: образцы 7, 8 и 9, генеалогически представляющие одну линию (восходящую к Сомдыку), и на генетическом дереве группируются вместе, восходя к общему предку. Более того, даже внутри этой ветви образцы 7 и 9 относятся к одной субветви и по генеалогическому, и по генетическому дереву (рис. 7.10).

Конечно, для идеального соответствия двух деревьев было бы желательно, чтобы образцы 1–6 (потомки Караходжи) несли бы и общую мутацию, формируя отдельную ветвь и на генетическом де-

реве. Но для этого необходимо, чтобы хотя бы одна мутация отличала Караходжу от его отца, а судя по опубликованным скоростям мутирования, при использованной длине секвенирования мутации происходят примерно 1 раз в 4 поколения. Это значит, что вероятность появления у Караходжи мутации в секвенируемом диапазоне составляет лишь около 25%, то есть куда более вероятно (около 75%), что такой мутации не будет. Вот она и не произошла. Удивительно скорее то, что за 4 поко-

ления (при передаче от Акжолы к Сомдыку и далее к Ермену) произошла не одна, а целых две мутации (замена цитозина на тимин в позиции 23081087 Y-хромосомы и аденина на гуанин в положении 23526483). Тем более удивительно, что за одно поколение (Шагир – Аманжол) произошли целых две мутации. Вероятность этого невелика (как нетрудно подсчитать, около 6%), поэтому можно даже предложить специалистам по шежире (традиционной генеалогии казахов) вопрос: насколько точно

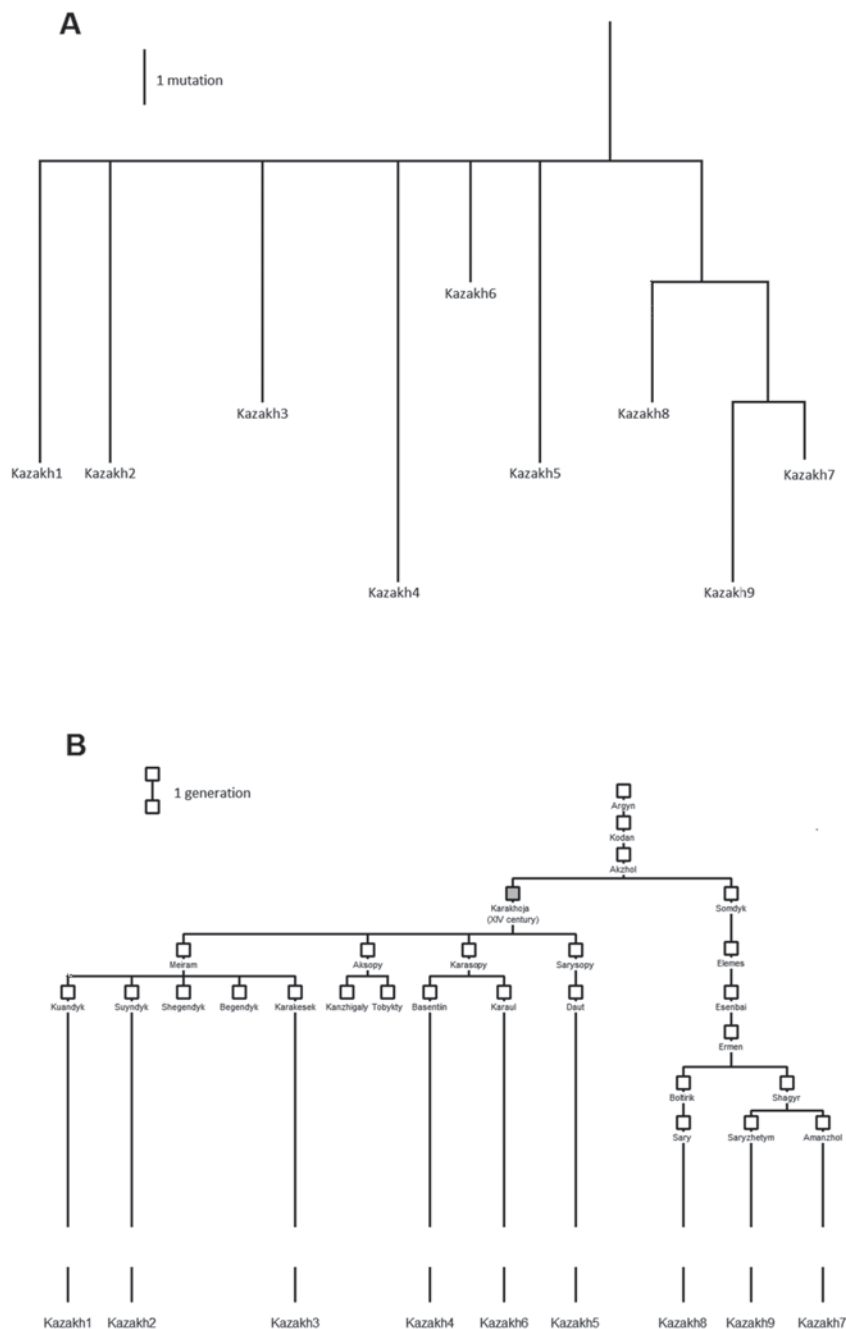


Рис. 7.10. Генетическая и традиционная генеалогическая реконструкции биологического родства связей между представителями родов аргынов, для которых проведено полное секвенирование Y-хромосомы:
 А) Генетическое дерево, построенное по данным полного секвенирования Y-хромосомы;
 В) Генеалогическое дерево, отражающее традиционные генеалогические представления (шежире). Каждый казах, Y-хромосому которого секвенировали, принадлежит к одному из кланов. Серым квадратом отмечен генеалогический предок казахов с номерами 1-6, известный по историческим источникам.

реконструирована генеалогия на этом отдельном отрезке? Но общее сходство генеалогического и генетического деревьев не оставляет сомнений в справедливости традиционного генеалогического дерева в целом.

Получается следующая ситуация. Казахское шежире утверждает: «Изученные представители казахов происходят от Караходжи и его брата Сомдыка». И генетические данные тоже утверждают: «Изученные представители происходят от двух близких родственников». Тогда мы вынуждены признать и то, что они происходят именно от Караходжи и его брата, а это позволяет нам воспользоваться историческими свидетельствами о времени жизни Караходжи для расчета скорости мутирования.

ОЦЕНКА СКОРОСТИ МУТИРОВАНИЯ

Чтобы определить скорость, с какой происходят мутации на Y-хромосоме, мы подсчитали число мутаций, произошедших за известный промежуток времени – от Караходжи до современности. Мы исходили из того, что Караходжа (бывший во время своего посольства в 1405 году уже весьма зрелым мужчиной) вероятно, стал отцом около 1385 года, а время рождения его самого и его брата примерно на 30 лет раньше (30 лет – средняя длина мужских поколений). Средняя дата рождения обследованных ныне живущих казахов пришлась на 1982 год. Таким образом, наблюдаемые на дереве мутации накапливались с 1355 по 1982 годы, т.е. 627 лет. Разумеется, эти расчеты являются лишь наилучшей возможной аппроксимацией, но ошибка вряд ли превышает одно поколение. Общая длина Y-хромосомных сегментов, секвенированных в каждом из 9 образцов казахов, составила 9 972 660 нуклеотидов, а среднее число мутаций в каждом ныне живущем представителе рода по сравнению с родоначальником составило 4,89 мутаций. По этим исходным данным легко вычислить скорость мутаций в изученном участке Y-хромосомы. Скорость составила $0,78 \times 10^{-9}$ на нуклеотид в год (то есть каждый год с вероятностью 0,78 смутит каждый миллиардный нуклеотид).

Тот же подход применили и к Y-хромосомным STR. Для 221 образца аргынов, генотипированных по 15 STR-маркерам, была построена филогенетическая сеть, и среднее число мутаций от гаплотипа-основателя составило 0,68. Время от общего предка до современных изученных образцов для предыдущего случая составило 627 лет, что соответствует 21 поколению (хотя средний год рождения генотипированных по STR-маркерам немного отличался от среднего года рождения секвенированных, отличия в несколько лет тут не имеют значения). Отсюда легко рассчитать скорость мутирования: 0,0022 мутации на locus на поколение.

Это значение очень близко к «генеалогической» скорости мутаций, показанной во многих исследованиях при сравнении пар «отец-сын».

КОМПЛЕКСНЫЙ ФИЛОГЕОГРАФИЧЕСКИЙ АНАЛИЗ И РЕКОНСТРУКЦИЯ МИГРАЦИЙ

Совместный анализ карты частоты гаплогруппы **G1**, карты ее гаплотипического разнообразия, полногеномных данных по ее четкому разделению на ветви и филогенетических оценок их возраста позволил реконструировать историю распространения этой гаплогруппы.

Глядя на карту распространения гаплогруппы **G1**, возникает вопрос, где ее прародина – в горах или степях? Намного большее разнообразие по STR-гаплотипам в западной части Иранского и Армянского нагорий делает горную прародину более вероятным кандидатом. Поэтому можно полагать, что гаплогруппа **G1** выделилась из предкового варианта **G*** в горных областях Передней Азии. Оттуда она затем распространилась в степную зону Евразии и далее была разнесена по обширной территории миграциями степных кочевников. Такая миграция хорошо согласуется с аناлийской гипотезой происхождения индоевропейской языковой семьи. Миграции из Ирана в Центральную Азию надежно подтверждаются также палеоантропологическими данными по древним популяциям Таджикистана и Туркмении.

Таким образом, гаплогруппа **G1** могла быть одним из генетических маркеров миграции некоторых индоевропейских групп в евразийские степи. Но сама по себе гаплогруппа **G1**, конечно, не может служить универсальным маркером распространения даже отдельных ветвей индоевропейцев, поскольку при любой миграции перемещается часть популяции, состоящая из носителей разных гаплогрупп. Ее можно рассматривать только как один из генетических компонентов, принесенный волной ираноязычной миграции,двигающейся на север к степям Евразии. Генетические данные свидетельствуют, что все основные ветви уже существовали к тому времени, когда миграция началась. Действительно, последняя развилка на башкирский и армянский кластеры произошла около 8000 лет назад, а ираноязычные популяции появились значительно позже (армянская лингвистическая ветвь отделилась около 4600 лет назад, а индоиранские языки отделились около 4200 лет назад).

Гаплогруппа **G1**, вероятно, оставалась редким генетическим компонентом среди многих евразийских и особенно ираноязычных популяций. Несколько выше ее частота на предполагаемой прародине – в Иране и Армении. Когда иранские языки в степях были замещены тюркскими (вероятно, в середине 1-го тысячелетия н.э.), носители гаплогруппы **G1** должны были оказаться ассимили-

рованы в тюркские и потом – в монгольские популяции, и в их составе распространились как западнее (Башкирия, Крым), так и восточнее (Монголия). Еще позднее в различных географических регионах независимо и в разное время произошла экспансия этой гаплогруппы – некоторым из ее носителей посчастливилось быть «отобранными» дрейфом генов в быстро растущих по численности популяциях.

Экспансия этой гаплогруппы (быстрый рост численности ее носителей) у казахов генетически датируется временем 470–750 лет назад, в зависимости от того, какую скорость мутации использовать. Генеалогический предок аргынов жил 600 лет назад, как раз посередине этого интервала. Распространение гаплогруппы от одного человека к полумиллиону людей за 600 лет (20 поколений) хотя и удивительно, но в принципе возможно. При предположении, что в каждом поколении два выживших сына оставляют в потомстве снова двух выживших сыновей – и так по всей цепи поколений: в этом случае в 19-м поколении уже имеется полмиллиона потомков. Поскольку у казахов в семьях было в среднем 3,5 ребенка, то такая экспансия

гаплогруппы **G1** у казахов рода аргын оказывается вполне вероятной.

Экспансия этой гаплогруппы у армян-амшенов генетически датируется временем 1150 лет назад. Это хорошо соотносится с историческими источниками, говорящими, что амшены произошли от родственников и слуг князя Шапух Амагуни, который приехал в 791 году из Аббасидского халифата.

Экспансия этой гаплогруппы среди башкир (у рода канглы) генетически датируется XV веком н.э. Считается, что этот род сформировался среди печенегов около VIII века н.э., затем принял участие в этногенезе башкир. Возрастание его численности может быть связано с демографическими изменениями в XIV веке, когда Башкирия стала частью Золотой орды, и в XVI веке, когда она стала частью России.

Стоит отметить, что, несмотря на географическое соседство Башкирии и Казахстана, предки носителей **G1** в этих двух регионах не имеют недавнего общего происхождения. Эти ветви (как и третья ветвь в Монголии) присутствовали в евразийских степях, по всей видимости, начиная с эпохи скифов.

7.5. ТРИ МОДЕЛИ ВЗАИМОСВЯЗИ ГЕНОФОНДОВ И ЯЗЫКОВ: КАВКАЗ, СЛАВЯНЕ, ТЮРКИ

Среди многих наук, исследующих те или иные аспекты истории народонаселения (археология, этнография, физическая антропология, лингвистика, генетика), синтез данных генетики и лингвистики привлекает в последние годы все большее внимание. Одна из причин этого в том, что обе науки являются в большой мере формализованными и хотя бы отчасти количественными – это позволяет проводить статистически обоснованное сопоставление их результатов. Другая причина – бурный рост объема генетических данных, характеризующих генофонды большинства народов мира по все более информативным генетическим системам – митохондриальной ДНК, Y-хромосоме, полногеномным панелям маркеров, полному секвенированию геномов.

ПОДХОД

В популяционной генетике классическим методом определения степени согласованности генетических данных с лингвистическими классификациями является тест Мантеля: расчет парных и частных корреляций между генетическими, лингвистическими и географическими расстояниями для одного и того же набора популяций. География здесь выступает в качестве контроля объективности выявляемых связей между генетикой и

лингвистикой: если лингвистически родственные популяции проявляют генетическое сходство даже в случае их географической удаленности друг от друга, это является сильным аргументом в пользу неслучайности связи генетики и лингвистики; если же генетически сходными оказываются географически соседние популяции безотносительно степени их лингвистического родства, это означает, что поток генов между соседними популяциями являлся более важным фактором формирования генофондов, чем генетический компонент, унаследованный от носителей праязыка. Аппарат частных корреляций, на которых основан тест Мантеля, является статистически строгим инструментом для различения этих двух ситуаций – реальное проявление лингвистического родства в генетической структуре или же связь генетики преимущественно с географическим фактором.

ПРОТИВОРЕЧИВОСТЬ И ЕЕ ПРИЧИНЫ

За последние два десятилетия генетиками проведено множество сравнений получаемых генетических данных с лингвистическими классификациями. Хотя часть из этих оценок была только качественной (самый известный пример – визуальное сравнение генетических и лингвистических деревьев в монографии [Cavalli-Sforza et al., 1994]),

многие другие работы опирались на строго количественные оценки. Тем не менее, результаты разных исследований противоречивы: структура генофонда воспроизводит то географическое положение популяций, то их лингвистическое родство, то рисует совершенно особую, третью картину.

Такая противоречивость отчасти может быть вызвана методологическим разнообразием в генетических исследованиях: одни работы опирались на данные по мтДНК (материнская линия наследования), другие – на Y-хромосому (отцовская линия наследования), третьи – на аутосомные (в последние годы – полногеномные) маркеры. Кроме того, разные исследования часто посвящены населению разных регионов мира, поэтому выводы разных работ о сходстве структуры генофонда с лингвистикой или же с географией оказываются трудно сопоставимы. Наконец, даже исследования одного и того же региона по одной и той же генетической системе могут приводить к противоположным результатам, если одно из них (или оба) опираются на недостаточно обширные или недостаточно подробные генетические данные. Например, в исследовании генофонда Кавказа [Nasidze et al., 2003] был сделан вывод об отсутствии связи генетической структуры как с географическим, так и с лингвистическим факторами. А спустя несколько лет, при накоплении намного более подробных данных, была выявлена достоверная связь с обоими этими факторами, причем по Y-хромосоме она оказалась более тесной с лингвистическим родством [Balanovsky et al., 2011], а по полногеномным (аутосомным) данным – с географическим соседством популяций [Yunusbaev et al., 2012].

СРАВНИМ РЕЗУЛЬТАТЫ ДЛЯ ТРЕХ ЯЗЫКОВЫХ ГРУПП

Поэтому попробуем сопоставить итоги исследований связи генетических и лингвистических данных, проведенных для трех групп народонаселения: популяций северокавказской лингвистической семьи, популяций балто-славянской группы индоевропейской семьи, популяций тюркской группы алтайской семьи.

Изучение генофонда Кавказа проведено под руководством автора по маркерам Y-хромосомы и опубликовано в [Balanovsky et al., 2011]. Изучение тюркоязычных популяций проведено при участии автора по полногеномным (аутосомным) маркерам [Yunusbaev et al., 2015]. Изучение балто-славянских популяций проведено под руководством автора по всем трем генетическим системам (Y-хромосома, мтДНК, полногеномные панели маркеров) и опубликовано в [Kushniarevich et al., 2015]. Таким образом, мы получаем возможность сравнить как результаты, полученные по разным генетическим системам для одних и тех же популяций (балто-

славянское исследование), так и результаты для разных лингвистических семей Евразии (сравнение трех исследований). Такая широкая панорама может помочь лучше понять степень и причины согласованности – или несогласованности – генетических и лингвистических данных.

НАРОДЫ КАВКАЗА

Кратко обрисует итоги исследования, описанного в разделах 7.2 и 7.3 этой главы.

Генетические и лингвистические данные были организованы однотипным образом и потому сопоставимы друг с другом. Было возможно оценить как степень сходства генетической и лингвистической реконструкций родства популяций Кавказа, так и датировки разделений прапопуляций (сопровождавшихся разделением генофондов и языков). Это сравнение структур генофонда и «глотофонда» было проведено четырьмя методами.

Во-первых, визуальное сравнение выявило практически полное сходство полученных генетического и лингвистического деревьев родства.

Во-вторых, тест Мантеля (и по парным, и, что наиболее важно, по частным корреляциям) показал более значимую связь структуры генофонда с лингвистическим, а не с географическим фактором.

В-третьих, анализ методом AMOVA выявил большую генетическую дифференциацию при группировке кавказских популяций по лингвистическому, чем по географическому принципу. Это означает, что сведения о лингвистической принадлежности популяции являются в данном случае более информативными для прогноза ее генофонда, чем ее географическое положение.

В-четвертых, датировки выявленных генетических кластеров гаплотипов (многие из которых оказались специфичны для отдельных народов или лингвистических групп Кавказа) в первом приближении совпали с датировками разделения соответствующих языков, полученными лексикостатистическими методами.

Полное согласие результатов, полученных четырьмя различными методами, делает весьма надежным вывод о том, что генетическая и лингвистическая структуры народонаселения Северного Кавказа жестко связаны друг с другом.

БАЛТО-СЛАВЯНСКИЕ НАРОДЫ

Как описано в главе 6 (раздел 6.1) изучение генофонда балто-славянских народов проведено с высокой степенью подробности: все 16 современных балто-славянских народов изучены по всем трем современным генетическим системам (мтДНК, Y-хромосома, аутосомные полногеномные маркеры), при этом большинство народов изучены в нескольких региональных популяциях. Ис-

пользованные для сравнения лингвистические данные, как и в случае исследования народов Кавказа, представляли собой показатели попарного лексикостатистического сходства балто-славянских языков.

Проведенный тест Мантеля состоял в сравнении скоррелированности друг с другом трех матриц: матрицы генетических расстояний, матрицы лингвистических расстояний и матрицы географических расстояний между изученными балто-славянскими народами. Парные корреляции выявили практически полную согласованность генетики, лингвистики и географии: корреляции генетических расстояний с лингвистическими варьировали от 0,69 до 0,81, а корреляции генетических расстояний с географическими – от 0,80 до 0,96. Однако частные корреляции выявили, что связь генетической и лингвистической структуры является не прямой, а опосредованной географическим фактором – при исключении влияния лингвистики корреляции генофонда с географическим фактором остаются высокими, но при исключении влияния географии корреляции генетики и лингвистики падают почти до незначимых величин.

Такое преобладание географического фактора над лингвистическим в формировании генофонда балто-славян может объясняться быстрой экспансией славян по обширной территории Восточно-Европейской равнины и Балкан, сопровождавшейся интенсивной ассимиляцией дославянского населения. Если этот дославянский компонент численно преобладал, то генетические взаимоотношения современных балто-славянских популяций и должны отражать в большей мере генетические паттерны, предшествовавшие расселению славян; тогда неудивительно, что эти паттерны отражают не (лингвистические) взаимоотношения между расселившимися на эти территории славянами, а скорее (отраженные в географическом положении) складывавшиеся тысячелетиями направления потоков генов между этими территориями.

Отметим все же, что, несмотря на эту общую количественно выявляемую тенденцию, на качественном уровне в ряде случаев виден явный параллелизм генетической структуры все же с лингвистическими данными (например, в генетической гомогенности восточнославянских народов, несмотря на обширный ареал их расселения, или в генетическом сходстве славян с балтами при более выраженных отличиях от генофонда финноязычных народов).

ТЮРКОЯЗЫЧНЫЕ НАРОДЫ

Популяции 22 народов, говорящих на тюркских языках и представляющих различные части обширного ареала тюркской ветви алтайской семьи (общая выборка N=390 человек), были про-

анализированы по полногеномным панелям маркеров [Yunusbaev et al., 2015]. Большинство из тюркских народов оказались генетически сходными не друг с другом, а каждый со своими географическими соседями, относящимися к другим лингвистическим семьям. Таким образом, доминирование географического фактора над лингвистическим в структуре генофонда проявляется в данном случае очень ярко и объясняется преобладанием в генофонде большинства тюркоязычных народов ассимилированного дотюркского компонента (модель «доминирования элиты»). Но в то же время, удается выявить и генетический компонент, свойственный большинству тюркоязычных народов (хотя и составляющий лишь небольшую часть их генофонда). Можно предполагать, что именно этот выявленный компонент и маркирует общность происхождения всех тюркоязычных народов, столь явную лингвистически. Этот компонент является наиболее частым в популяциях Южной Сибири и Монголии, что может рассматриваться как дополнительный – генетический – аргумент в пользу алтайской прародинности тюркоязычных народов. Генетические датировки событий ассимиляции, полученные для тюркоязычных народов, варьируют с V в. до XV в. н.э., что также хорошо согласуется с историческими данными о времени экспансии тюрков.

СТЕПЕНЬ СОГЛАСОВАННОСТИ ГЕНЕТИЧЕСКОЙ И ЛИНГВИСТИЧЕСКОЙ СТРУКТУРЫ ОПРЕДЕЛЯЕТСЯ МОДЕЛЬЮ ИСТОРИИ ПОПУЛЯЦИИ: САМОРАЗВИТИЕ, СМЕШЕНИЕ ИЛИ ДОМИНИРОВАНИЕ ЭЛИТЫ

Характер и степень связи генетических и лингвистических данных оказываются различными для народонаселения разных регионов и зависят от того, по какой модели шло формирование народонаселения.

Так, тесная связь генетической и лингвистической структуры народов Кавказа лучше всего объясняется следующей моделью: народонаселение Северного Кавказа произошло преимущественно из одной общей прапопуляции, саморазвитие которой заключалось в постепенном дроблении на субпопуляции (современные лингвистические группы), далее на субсубпопуляции (современные народы) и так далее. При таком поэтапном разделении популяций происходило одновременное разделение как языков, так и генофондов. Именно этот – демографический по своей сути процесс «дробления» прапопуляции – и отразился, как в двух зеркалах, в генетической и в лингвистической панорамах Северного Кавказа.

Поэтому можно предполагать, что совпадение генетики и лингвистики является естественным и ожидаемым как на Кавказе, так и в других регионах мира при соблюдении двух условий.

1) Оба «зеркала» являются незамутненными: в генетических данных надежно реконструируется реальная структура генофонда (что определяется объемами и качеством сбора выборок, надежностью их генотипирования, адекватностью методов анализа и т.д.), а лингвистические данные, аналогично, надежно реконструируют реальное родство языков.

2) История народонаселения исследуемого региона сводится преимущественно к постепенной дифференциации одной общей прапопуляции.

Если одно из этих условий нарушено, совпадения генетических и лингвистических данных ожидать трудно.

Именно это мы и обнаружили в случае балто-славянских народов: отсутствие связи генетики с лингвистикой при соблюдении первого, но нарушении второго условия. В этом регионе народонаселение сформировалось не из единой прапопуляции, а в результате смешения одной группы населения с рядом разнородных групп (экспансии праславян на территории, населенные генетически различными популяциями). И в этом случае фактор лингвистического родства оказывается менее значимым, чем фактор географического соседства. При этом все три генетические системы дают сходные описания структуры генофонда и проявляют одни и те же тенденции в своей связи с географическим и лингвистическим факторами.

Исследование тюркоязычных народов, проведенное по полногеномным данным, добавляет еще один важный штрих в вырисовывающуюся закономер-

ность: даже если формирование народонаселения шло по модели доминирования элиты, т.е. ассимиляции в своем крайнем выражении, общий генетический компонент, унаследованный от предковой популяции вместе с ее языком, хотя и составляет лишь небольшую часть генофонда, но реально существует и может быть выявлен генетическими методами. Правда, для такого выявления, похоже, требуется, чтобы экспансия происходила на генетически контрастные популяции: в случае тюркской экспансии по генетически гетерогенной Евразии такие следы обнаружены, а в случае славянской экспансии по гомогенной Европе их выявить не удастся, хотя миграция славян, скорее всего, была более массовой.

Можно надеяться, что детальные филогеографические реконструкции, основанные на возможностях полного секвенирования Y-хромосомы, позволят найти генетические маркеры, частично ассоциированные с миграциями отдельных лингвистических групп и поэтому позволяющие выявлять их следы в генофонде.

В качестве еще одного – методологического – наблюдения отметим, что тривиальная связь генетической структуры с географическим положением всегда существует, лежит на поверхности и поэтому легко выявляется даже при неполной генетической изученности. Напротив, связь структуры генофонда с лингвистическими данными для своего выявления требует подробного генетического изучения и применения развитого аналитического аппарата, но зато позволяет глубже понять историю популяций.

ГЛАВА 8.

ДРЕВНЯЯ ДНК ЕВРОПЫ

При исследованиях генофонда в явном или скрытом виде всегда присутствует время. В явном – когда речь идет об истории формирования генофонда, а в скрытом – при каждом виде анализа, ведь воздействие на генофонд любого фактора происходит во времени и обычно измеряется поколениями. Для отечественной геногеографии традиционной стала идея Элизе Реклю, что «География является Историей во времени», и анализ пространственной структуры генофонда рассматривается как ключ к реконструкции динамики генофонда во времени [Рычков, Ящук, 1986; Балановская, Рычков, 1990; Рычков и др., 1997]. Этот подход, хотя специально и не формулируемый, свойственен и многим зарубежным исследованиям [Cavalli-Sforza et al., 1994; Semino et al., 2000 и множество других работ]. Такой путь к реконструкции прошлого генофонда – через анализ его современной пространственной структуры – является самым распространенным, но не единственным.

Другой путь, используемый еще в «классическую» эпоху популяционной генетики, но ставший особенно популярным при исследованиях мтДНК и Y-хромосомы в последние два десятилетия, заключается в получении генетических датировок популяционных событий. Хотя и в этом случае начальной точкой являются данные о современном генофонде, но накопленное разнообразие пересчи-

тывается в даты абсолютного времени [Рычков, 1986; Bandelt et al., 1995; Richards et al., 1996; 2000; Zhivotovsky et al., 2004 и другие работы]; в том числе в последние годы по данным полногеномного секвенирования Y-хромосомы [Poznik et al., 2013; Karmin et al., 2015; Balanovsky et al., 2015 и другие работы]. Полученные датировки дают более прямые оценки изменения генофонда во времени, чем при интерпретации географических (пространственных) закономерностей.

Третий путь к изучению изменчивости генофонда во времени является самым прямым. Он состоит в непосредственном генетическом анализе образцов ДНК от представителей давно ушедших поколений, благо такие образцы в виде костных останков в большом количестве предоставляется палеоантропология.

В данной книге, ориентированной на изучение изменчивости генофонда Европы не только в пространстве, но и во времени, использовались все три пути. Догадками о прошлом генофонда, сделанными на основании его современной пространственной структуры, пестрят главы 2, 3, 4, 5 и 6. Использование генетических датировок является основным содержанием главы 7, а ее теории уделено большое внимание в главе 1. А настоящая глава 8 посвящена третьему пути – анализу древней ДНК.

8.1. ОСОБЕННОСТИ АНАЛИЗА ДРЕВНЕЙ ДНК

ИДЕЯ

Идея анализа древней ДНК (палеоДНК) проста: вместо того, чтобы на основании данных о современных популяциях догадываться, какими были генофонды их предков, изучать генофонды древности напрямую – проанализировав ДНК из костных останков (палеоантропологического материала). Таким образом, анализ палеоДНК, продолжая быть генетикой, является одновременно отраслью палеоантропологии, добавляя к используемым в палеоантропологии системам признаков (краниометрия, краниоскопия, одонтология, остеология) еще и «ДНКметрию». При этом подходы к статистическому анализу и осмыслению результатов при анализе древней ДНК те же, что и в классической

палеоантропологии: изучаются те же самые находки, также они сравниваются друг с другом и с современным населением, также делаются выводы о миграциях или же преемственности населения.

ПЕРСПЕКТИВЫ

Анализ древней ДНК и сравнение полученных результатов с исследованиями современного генофонда является одной из наиболее перспективных областей среди современных наук о жизни. Это подтверждается большим числом публикаций в наиболее престижных научных журналах – Science, Nature и других [Рдбо et al., 1989; Handt et al., 1996; Cooper et al., 2000, 2001; Hofreiter et al., 2001; Haak et al., 2005, 2010, 2015; Noonan et al., 2006;

Brotherton et al., 2007; Sampietro et al., 2007; Burbano et al., 2010; Lazaridis et al., 2014; Raghavan et al., 2014; Allentoft et al., 2015; Gynthera et al., 2015]. В настоящее время исследования древней ДНК ведутся во многих лабораториях разных стран мира. Высокая научная значимость и перспективность этих исследований общепризнана в научном сообществе, поскольку позволяет получать прямые данные о генофондах прошлых эпох и их динамике, которые иначе можно реконструировать только менее надежным косвенным путем из данных о современном народонаселении.

ПРОБЛЕМЫ

Это научное направление является технически сложным, и его развитие наталкивается на ряд объективных проблем.

Прежде всего, к ним относятся проблемы контаминации (загрязнения) древних образцов со стороны современной ДНК, малого количества и химической модификации ДНК, выделяемой из древних образцов.

Другой нередко возникающей проблемой является неполнота информации о современном генофонде, сравнение с которым является одним из важнейших инструментов при интерпретации данных по древней ДНК.

Методы сравнения с современным генофондом и выявления структуры генофондов во многих случаях также оказываются недостаточно эффективны.

Наконец, существуют проблемы крайне малых объемов выборок по древней ДНК (как правило, только 5–10 образцов из популяции), и неравномерной представленности различных эпох и территорий в накапливаемых данных по изменчивости древней ДНК в популяциях человека.

Есть еще и методологическая проблема. Ведь когда мы изучаем современное население, то мы знаем, что у него имела цепочка предков на всем протяжении жизни человечества – и пытаемся ее проследить на тех или иных отрезках времени. Но когда мы изучаем древнюю ДНК, у нас нет уверенности, что это не «засохшая» ветвь истории, нет уверенности, что они оставили своих генетических потомков и что их гены дошли до современности или хотя бы до следующей за ними эпохи.

КОНТАМИНАЦИЯ

Основной проблемой при изучении ДНК и основным источником ошибок является контаминация – случайное попадание современной ДНК в исследуемый древний образец. В результате разрушения ДНК в древних костях, количество ДНК столь мало, что для анализа используются методы максимальной чувствительности (вплоть до ана-

лиза единственной молекулы). Поэтому попадание в образец даже ничтожных количеств современной ДНК (неизбежно присутствующей во всех помещениях, где находятся люди) приведет к тому, что фактически будет изучена современная ДНК (например, ДНК с рук самого исследователя), но исследователь будет ошибочно считать, что полученная нуклеотидная последовательность характеризует древний образец.

Эта наиболее технически сложная проблема контаминации в настоящее время хотя и не может быть решена полностью, но минимизация рисков контаминирования и многократное сокращение ожидаемой частоты ошибок генотипирования реально достижимы в условиях специализированных центров анализа древней ДНК. Такие центры создаются с учетом всех потенциальных факторов контаминации и реализуют широкий спектр технических, методологических, методических, кадровых и организационных аспектов всех известных способов минимизации рисков контаминации. Такие центры к настоящему времени созданы в ряде стран мира.

ЛАБОРАТОРИИ ЗА РУБЕЖОМ

В области анализа древней ДНК успешно работает большое число зарубежных исследовательских коллективов. Среди них наиболее известны являются несколько коллективов:

1) коллектив под руководством проф. Сванте Паабо (Professor Svante Paabo, Institute for Evolutionary Anthropology, Лейпциг, Германия); именно этот коллектив выполнил нашумевшие работы по анализу мтДНК неандертальца, затем полного генома неандертальца и недавно генома человека из Денисовой пещеры на Алтае;

2) коллектив под руководством проф. Алана Купера (Professor Alan Cooper, Australian center for ancient DNA, Австралия), работавшего сначала в Оксфорде, а затем создавшего лабораторию в Австралии и известными как методическими работами, так и исследованиями древней ДНК популяций человека и различных видов животных;

3) новый организованный институт под руководством проф. Иоганнеса Краузе (Professor Johannes Krause, Institute for the Science of Human History, Йена, Германия), объединивший целый ряд ведущих специалистов в области археологии, лингвистики и древней ДНК, в том числе известных генетиков Wolfgang Haak из лаборатории Алана Купера и Guigo Brandt из лаборатории Курта Альта из Майнца (Германия);

4) коллектив под руководством проф. Эске Виллерслева (Professor Eske Willerslev, Centre of Excellence in GeoGenetics, Копенгаген, Дания), начавший с секвенирования полного генома эскимоса по единственному волосу (Rasmussen et al.,

2010), а сейчас изучивший уже более сотни полных древних геномов.

Большое число других научных коллективов, работающих в этой области, можно разделить на две категории: коллективы, обладающие высокой научной репутацией (например, David Reich, Harvard Medical School, США; Professor Kurt Alt, University of Mainz, Германия; Dr Carles Lalueza-Fox, Universitat Pompeu Fabra, Испания) и коллективы со спорной или неизвестной репутацией.

Вопрос о научной репутации для данной области исследований является принципиально важным потому, что технические особенности анализа древней ДНК легко могут привести к ложным результатам, и высокую профессиональную репутацию заслуживают только те исследователи и те лаборатории, которые тщательно минимизируют риски ошибок. Особенно важна репутация лабораторий древней ДНК и потому, что полученные результаты иногда нельзя даже попробовать воспроизвести, если уже израсходован весь пригодный для анализа ДНК материал из палеоантропологической находки. В этом случае остается лишь уповать на репутацию лаборатории и призывать коллег из смежных наук с большой осторожностью относиться к сенсационным выводам лабораторий «второй категории».

ЛАБОРАТОРИИ В РОССИИ

В России отсутствуют специализированные центры анализа древней ДНК, признанные на уровне мировой науки, хотя попытки такого анализа периодически предпринимаются отдельными лабораториями. Не рассматривая ряд провинциальных коллективов, осуществлявших такие попытки, но не достигших серьезных научных результатов (в частности, не имеющих значимых научных статей по этой теме), нужно отметить коллектив новосибирских исследователей, ведущих работы по древним популяциям Западной Сибири и Алтая [Молодин и др., 2000; Воевода и др., 2000; Pilipenko et al., 2015]. Хотя оснащение этой лаборатории, насколько можно судить, не вполне соответствует требованиям к полноценному центру анализа древней ДНК, а число научных статей невелико, этот коллектив является единственным (кроме нашего), активно ведущим в России популяционно-генетические исследования древней ДНК, и единственным, проводящим этап собственно анализа древней ДНК в России.

Также в ИОГен РАН коллективом под руководством проф. Е.И. Рогаева ведутся исследования деградированной и древней ДНК человека, результаты которых опубликованы в высокорейтинговых журналах [Rogaev et al., 2009 a,b; Григоренко и др., 2009]. В рамках этих исследований получены результаты, применимые для повышения эффектив-

ности анализа деградированных ДНК в криминалистике и судебно-медицинской экспертизе.

СХЕМА РАБОТЫ

Из изложенного становится понятным, почему в настоящее время эффективное изучение древней ДНК из антропологических серий с территории России возможно только в рамках сотрудничества с зарубежными центрами. Среди них Австралийский центр древней ДНК, возглавляемый проф. Аланом Купером, обладает высокой репутацией и является одним из лидеров данного направления исследований в мире. Поэтому результаты, приведенные в данной главе, были получены во многом в результате нашего многолетнего сотрудничества с Австралийским центром древней ДНК. За это время были проведены взаимные командирования сотрудников (д-р Вольфганг Хаак в Россию в 2008 г., О.П. Балановский в Австралию в 2010 г.), ежегодные встречи и обсуждения на научных конференциях в разных странах (Китай в 2007 г., Эстония в 2008 г., США в 2009 и 2011 гг., Австралия в 2010 г.). Это сотрудничество проходило в рамках международного проекта «Генографик» (2005–2012 гг.), в котором участвовали оба наших коллектива. По окончании проекта центр тяжести исследования древних популяций человека из Австралийского центра древней ДНК переместился в новый мощный центр в Йене (Institute for the Science of Human History), объединивший целый ряд ведущих специалистов, в том числе и д-ра Вольфганга Хаака.

За это время была выработана наиболее эффективная схема сотрудничества, согласно которой зарубежная сторона проводит экспериментальный анализ (2 этап исследования) и статистическое моделирование (3 этап), российская сторона – сбор образцов (1 этап) и картографо-статистический анализ полученных экспериментальных данных (4 этап). Однако мне довелось поучаствовать и в экспериментальной работе (при исследовании генофонда скифов), что было необходимо для понимания особенностей технологии анализа древней ДНК и путей решения проблем, при таком анализе возникающих.

Кроме результатов нашей совместной работы с Австралийским центром древней ДНК, в этой главе описываются опубликованные при участии автора работы лабораторий David Reich и Eske Willerslev, а для полноты картины приводятся и литературные данные, взятые из публикаций других научных коллективов.

ОСОБЕННОСТИ ЭКСПЕРИМЕНТАЛЬНОГО АНАЛИЗА ДРЕВНЕЙ ДНК

Этап экспериментального анализа палеоДНК отличается как от палеоантропологии, так и от эк-

спериментального анализа ДНК современных образцов. От палеоантропологии он отличается тем, что проводятся не измерения или описания формы или химического состава костей, а экстрагирование из костного материала ДНК и затем прямой анализ ДНК теми же методами и на таком же оборудовании, что и при анализе современной ДНК. А от анализа современного генофонда анализ древней ДНК отличается чрезвычайно жесткими требованиями к чистоте и стерильности эксперимента. Дело в том, что чувствительность методов ДНК анализа очень высока, и единичных молекул может быть достаточно для получения результата. В древнем образце, как правило, сохраняются лишь единичные молекулы. Но такие же молекулы ДНК примерно в тех же концентрациях витают в воздухе любого помещения и находятся на любых предметах (в том числе ДНК костных останков из рядом лежащих коллекций, если та ДНК сохранилась лучше), которых касается сам исследователь или которых когда-либо касались другие люди, включая археологов, антропологов и работников музеев. И чрезвычайно велика опасность, что анализироваться будет не ДНК древнего образца, а посторонняя ДНК совсем другого человека, случайно попавшая в пробирку из какого-либо современного источника.

ПРОБЛЕМА КОНТАМИНАЦИИ

Именно поэтому в тех лабораториях, где занимаются анализом современной ДНК и где поэтому концентрация «витающей в воздухе» ДНК (в том числе амплифицированной) особенно велика, противопоказаны попытки анализа древней ДНК. Для анализа древней ДНК необходимо создавать специальные лаборатории, вся организация которых специально продумывается так, чтобы свести к минимуму риск контаминации древних образцов со стороны современной ДНК. Среди специалистов существует даже негласное ранжирование лабораторий древней ДНК по степени, в которой в них соблюдаются требования к чистоте эксперимента и, соответственно, по степени достоверности получаемых результатов.

В настоящее время разработано большое число правил для минимизации риска такой контаминации. К ним относятся:

1) технические требования (разделение лаборатории по зонам в соответствии с этапами анализа, боксирование этих зон, повышенное давление воздуха внутри лаборатории, выделение отдельных зданий, в которых никогда не проводились работы с современной ДНК, территориальная отдаленность этих зданий от лабораторий, изучающих современную ДНК и т.д.);

2) методические требования, состоящие главным образом в неукоснительном соблюдении по-

вышенной чистоты и точности при постановке экспериментов (на практике это выглядит как «работа в скафандрах»);

3) технологические требования, связанные с особенностями ПЦР-анализа фрагментированной ДНК.

Соответственно, лаборатории, обладающие полным комплектом технического оснащения для минимизации контаминации и строго следующие правилам тщательной постановки экспериментов, со временем заслуживают высокую репутацию, что подтверждается как получаемыми ими результатами, так и проведением независимого дублирующего анализа в других центрах, получивших мировое признание.

Например, в Австралийском центре древней ДНК применяются различные рекомендуемые при работе с палеоДНК подходы для минимизации риска контаминации: боксирование лабораторных помещений; защита образца от контаминации со стороны ДНК самого исследователя; удаление поверхностного слоя образца; работа в многослойной спецодежде - маски, перчатки, очки, комбинезоны (рис. 8.1); непрерывное обеспечение стерильности помещений и рабочих мест; положительное давление воздуха для исключения воздушного потока извне; шлюзовые отсеки при входе в лабораторию и между основными рабочими зонами внутри лаборатории; тщательный выбор поставщиков реактивов и пластика (гарантия от случайной контаминации при фасовке и т.д.); специальное обучение персонала; исключение работы с современной ДНК в той же лаборатории; пространственная отдаленность здания лаборатории древней ДНК от любых зданий, в которых проводятся или проводились в прошлом работы с современной ДНК; многочисленные другие требования. Хотя каждый из этих методов по отдельности описан и рекомендован в специализированных научных работах, но комбинация всех этих методов, осуществляемая в Австралийском центре древней ДНК, имеет очень немного аналогов, обеспечивая максимально возможный уровень снижения рисков контаминации.

Дополнительным методическим приемом, применяемым нашим коллективом, является контрольное сравнение всех получаемых нуклеотидных последовательностей древней ДНК со специально полученными последовательностями ДНК лиц, которые могли контаминировать древние образцы (археологов, палеоантропологов, сотрудников музеев, самих исследователей-генетиков).

При исследовании митохондриальной ДНК важным аспектом является разработанный австралийскими коллегами метод одновременного мультиплексного анализа панели SNP-маркеров, определяющих принадлежность образца к одной из основных гаплогрупп митохондриальной ДНК. Одновременное получение для древнего образца

данных по всем основным SNP маркерам повышает достоверность результата за счет многократного взаимного контроля и позволяет провести анализ при использовании меньшего количества ДНК.

ПРОБЛЕМА НЕМНОГОЧИСЛЕННОСТИ

Другая проблема состоит в немногочисленности палеоантропологического материала, пригодного для анализа древней ДНК. Поэтому большинство исследований палеодНК основаны на выборках всего лишь в несколько образцов, тогда как для надежных сравнений генофондов нужны выборки в десятки и сотни проанализированных индивидуумов. Эта проблема решается двумя путями.

Во-первых, по мере того как направление анализа древней ДНК набирает силу, совершенствуют свои методы и создает новые лаборатории, постепенно увеличиваются как средние объемы выборок, так и число уже проведенных исследований. И, объединяя результаты сразу нескольких исследований древней ДНК, можно уже получить выборку разумного объема (хотя она и будет представлять не локальную популяцию, а в целом определенную эпоху на определенном континенте).

Во-вторых, широко используется метод сравнения (немногочисленных) данных по древней ДНК с (многочисленными) данными по современному населению. Также, поскольку данные по древней ДНК, очевидно, представляют большую научную ценность, то на их статистический анализ и осмысление затрачиваются и куда большие усилия, чем было бы целесообразно затратить на анализ выборки столь же небольшого объема из какой-либо современной популяции.

Таким образом, преимущества и ограничения анализа древней ДНК и анализа современных генофондов противоположны. При анализе современной ДНК доступны обильные данные о генофондах, но возникает проблема вычленения, какая из черт генофонда сформировалась в какую эпоху. А при анализе древней ДНК доступны прямые данные о генофонде прошедших эпох, но на первый план выходит проблема ограниченности исходных данных. Совместный же анализ как современной, так и древней ДНК позволяет отчасти взаимно компенсировать эти недостатки и приблизиться к более надежной реконструкции истории формирования генофондов и тех событий в истории популяций, которые явились причиной этих генетических процессов.



Рис. 8.1. «Скафандровый» стиль работы в Австралийском центре древней ДНК. На фотографии автор книги высверливает скифский зуб.

8.2. СРЕДНИЙ ПАЛЕОЛИТ (НЕАНДЕРТАЛЬЦЫ)

МИТОХОНДРИАЛЬНЫЙ И ЯДЕРНЫЙ ГЕНОМЫ НЕАНДЕРТАЛЬЦА

Исследования древнейших геномов человека связаны, прежде всего, с именем немецкого исследователя Сванте Паабо. Эти исследования начались с секвенирования митохондриальной ДНК неандертальца в 1997 г. Она оказалась значительно отличающейся от всех вариантов мтДНК современного человека, то есть по мтДНК не было обнаружено никакого смещения между кроманьонцами и неандертальцами. Этот результат был подтвержден и дальнейшими исследованиями других лабораторий мира. Был сделан вывод, получивший самое широкое освещение в СМИ, что неандертальцы полностью тупиковая ветвь эволюции человека, поскольку в формировании генома современного человека нет вклада неандертальцев.

Но совершенствование генетических технологий позволило перейти уже от изучения митохондриального генома – ко всему геному неандертальца. Эта работа, начавшаяся в 2006 году, сначала также не позволяла выявить «неандертальские гены» в геноме современного человека. Но технологии анализа стремительно развивались, и в 2010 году был получен «черновой» геном неандертальца с покрытием около 1, то есть каждый участок генома был секвенирован в среднем только один раз. Это для технологий секвенирования следующего поколения (NGS) крайне недостаточно – при секвенировании современных геномов в последние годы считается «нормой» покрытие в 60–70, и уж в любом случае не менее 30. Однако плохая сохранность древней ДНК приводит к тому, что, как мы увидим из дальнейшего изложения, покрытие при секвенировании древних геномов зачастую меньше даже единицы. Поэтому надо всегда оценивать вероятность биоинформационной ошибки, зависящей от степени «покрытия» – надежности прочтения генома.

В результате анализа этого «чернового» генома неандертальца выяснилось, что около 1–2% генома современного населения Евразии имеют неандертальское происхождение. Дальнейшие исследования показали, что эта доля варьирует от 1% до 3–4% по Евразии (в Азии чуть больше, чем в Европе), но очень невелика (по более ранним данным – даже полностью отсутствует) у современного населения Африки южнее Сахары. Это позволило сделать вывод, что небольшая метисация между неандертальцами и кроманьонцами (неоантропом) в Евразии все же была. Напомню, что ареал неандертальцев простирался от Западной Европы через Крым и Кавказ до Средней Азии и Алтая, включая в себя Переднюю Азию.

В июне 2015 года мне довелось слушать об этом в лекции Сванте Паабо, и в ответах на вопросы он рассказал следующий случай. В одном из наиболее известных музеев мира на экспозиции, посвященной неандертальцам, на основании первой статьи (по мтДНК) утверждалось, что они совсем не смешивались с человеком современного вида. Но после выхода следующей статьи сотрудники музея оперативно сменили табличку. Это хороший урок и самим генетикам, и тем представителям смежных наук, которые часто доверяют всему, что «показано по ДНК». Дело в том, что оба результата с точки зрения генетики верны, точны и подтверждены многочисленными исследованиями. Просто молекула митохондриальной ДНК короткая и не сохранила следов общих генов с неандертальцами, обнаруженных несколькими годами позже при исследовании всего остального генома – ядерной ДНК.

В СОВРЕМЕННЫХ ГЕНОМАХ ЕВРАЗИИ ЦИРКУЛИРУЮТ 20% НЕАНДЕРТАЛЬСКОГО ГЕНОМА

Итак, представление о том, что неандерталец нам не родственник, было опровергнуто после того, как в 2010 году Сванте Паабо секвенировал его ядерный геном. В отличие от мтДНК неандертальца, которая не обнаруживает сходства ни с одним из вариантов мтДНК современного человека, в ядерной ДНК обнаружили признаки метисации. Это привело исследователей к выводу, что *Homo sapiens* и *Homo neandertalensis* смешивались, и около 1–4% генома в современных популяциях Евразии имеют неандертальское происхождение.

Дальнейшие исследования и более глубокий анализ позволили вытащить на свет многие последовательности неандертальской ДНК, спрятанные в современных геномах. У разных людей неандертальское генетическое наследие неодинаково как по размеру, так и по набору фрагментов ДНК. Популяционные генетики из Вашингтонского университета [Vernot, Akey, 2014] подсчитали *суммарное* количество неандертальских последовательностей, сохранившихся в современных геномах. Для этого они проанализировали 665 геномов современных людей (379 жителей Европы и 286 – Восточной Азии), которые были секвенированы в рамках проекта «1000 геномов», выделили все участки интрогрессии (включения неандертальской ДНК) и сравнили их с эталонным геномом неандертальца, секвенированным коллективом Сванте Паабо. Проверенные таким образом «неандертальские» последовательности составили около 600 Mb (мегабаз, миллионов пар оснований),

что составляет около 20% неандертальского генома. Иными словами, большая доля – пятая часть генома неандертальцев – хранится и циркулирует в геномах человечества. Эта величина никак не противоречит тому, что в геноме каждого человека (неафриканского происхождения) содержится не более 1–4% неандертальской ДНК. Ведь у разных людей эти фрагменты разные, и только в сумме они приводят к такой большой доле.

Неандертальские последовательности в геноме современных людей довольно короткие – не больше 57 Kb (килобаз, тысяч нуклеотидов). Это объясняется тем, что со времени гибридизации с неандертальцами прошло много времени, и длинные последовательности оказались разбиты рекомбинациями. В разных популяциях и у разных людей в ДНК прячутся разные неандертальские последовательности. В среднем на человека приходится по 23 Mb неандертальской ДНК.

В этом исследовании было также показано, что современные популяции Азии унаследовали в среднем на 21% неандертальских последовательностей больше, чем европейцы.

САПИЕНСЫ И НЕАНДЕРТАЛЬЦЫ – НЕДАВНИЙ ОБМЕН ГЕНАМИ

Одно из самых интересных открытий последнего года [Fu et al., 2015] показывает, что метисация сапиенсов и неандертальцев происходила не только после выхода сапиенсов из Африки, то есть 50–60 тысяч лет назад, но и гораздо позже. И не только на Ближнем Востоке, как до сих пор считали, а и в Европе. Этот результат получен в работе коллектива Сванте Паабо, в которой исследовали ДНК древнейшего современного человека на территории Европы, возрастом 37–42 тыс. лет (по радиоуглеродной датировке). Нижнюю челюсть Oase 1 в 2002 году обнаружили в Румынии, в подводной пещере Peștera cu Oase (название означает «пещера с костями»). Строение челюсти соответствовало анатомически современному человеку, хотя некоторые признаки указывали на его родство с неандертальцами. А вскоре после этого из пещеры был извлечен череп другого индивида (Oase 2), и этот череп подтвердил догадки о неандертальских предках. Оставалось проверить это генетически.

Из челюсти Oase 1 извлекли митохондриальную и ядерную ДНК. Анализ показал, что это мужчина, его Y-хромосомный гаплотип относился к макрогаплогруппе F, которая объединяет большинство гаплогрупп, распространенных сегодня на территории Евразии. Ядерный геном Oase 1 по маркерам однонуклеотидного полиморфизма (SNP) проверили на сходство с большим количеством геномов современных людей. Оказалось, что Oase 1 имеет больше общих аллелей с восточноазиатскими популяциями и с американскими индейца-

ми, чем с европейцами. Этим он отличается от другого почти столь же древнего европейца – со стоянки Костенки-14, возрастом 36–39 тыс. лет, который проявляет большее сходство с современными европейскими популяциями, чем с восточноазиатскими. Эта особенность привела авторов к заключению, что Oase 1 принадлежал к ранней европейской популяции анатомически современных людей, которая практически не внесла вклад в генофонд более поздних европейцев.

Чтобы оценить неандертальский вклад в геном Oase 1, было подсчитано число неандертальских аллелей во многих современных и нескольких древних геномах (взяв за исходный геном алтайских неандертальцев). Доля неандертальских последовательностей у Oase 1 оказалось существенно выше, чем у современных людей и чем у древних людей анатомически современного вида. Подсчет показал, что Oase 1 несет 3746 аллелей неандертальского происхождения, образцы из Усть-Ишима (42–43 тыс. лет назад) и Костенок-14 (36–39 тыс. лет назад) – в два-три раза меньше (1586 и 1121 аллелей соответственно), образцы из Китая и Франции — 1322 и 1033. То есть, неандертальский вклад в Oase 1 оказался в несколько раз выше, чем в остальных геномах. Долю неандертальской ДНК в геноме Oase 1 определили тремя методами, которые дали несколько различающиеся результаты – от 6,0% до 9,4%, в среднем ее оценили в 7,3%. При этом неандертальские последовательности были распределены по хромосомам неравномерно: в семи участках их было особенно много, а когда исключили эти участки, доля неандертальской ДНК упала до 4,8%.

Важнее всего то, что фрагменты неандертальской ДНК в геноме Oase 1 достигали значительной длины, в некоторых участках генома они превышали 50 сМ (сантиморганид). Это означает, что данные фрагменты еще не успели разбиться рекомбинациями. Проанализировав длину фрагментов, определили, что метисация предков Oase 1 с неандертальцами произошла совсем недавно — это случилось всего за 4–6 поколений до его рождения.

Итак, смешение предков этого неантропа произошло примерно за пять поколений до времени жизни этого человека. То есть около 40 тыс. лет назад происходила метисация неантропов с неандертальцами в Европе. И поскольку сто лет – это немного, то и само событие метисации скорее всего случилось в том же регионе, где был обнаружен сам Oase 1, то есть на Балканах. Так что геном Oase 1 показал, что метисация неантропов с неандертальцами происходила не только непосредственно после выхода из Африки и не только на Ближнем Востоке, но случалась и позже, когда сапиенсы одновременно с неандертальцами жили в Европе.

Однако, поскольку Oase 1 генетически не похож на поздних европейцев, нельзя исключать, что он мог быть членом первоначальной ранней популяции сапиенсов, которая смешалась с неандертальцами, но не оставила заметного следа в последующем генофонде. Для более точных предположений нужно проанализировать другие останки из пещеры Oase и исследовать прочие популяции древнейших сапиенсов Европы.

ДЕНИСОВСКИЙ ЧЕЛОВЕК

Следующий виток исследований был связан с обнаружением в Денисовой пещере на Алтае в слоях большой древности фрагмента кости пальца – настолько маленького, что антропологическими методами было невозможно определить, к какому виду гоминид он относится. А когда был секвенирован геном, оказалось, что это третий вид (или подвид – классификации разные) человека разумного. Этот вид назвали «денисовцем» (от названия места находки – Денисовой пещеры, а название, в свою очередь, происходит от имени спасавшегося в ней монаха Дионисия). Происхождение названия «денисовец», таким образом, полностью аналогично происхождению названия «неандерталец», названному по месту первой находки – небольшому гроту около Неандерской пещеры. Денисовцы – единственный вид гоминид, выявленный не морфологическими, а генетическими методами.

На дереве родства популяций гоминид денисовцы объединяются сначала с неандертальцами, а уже потом с анатомически современным человеком, но эти развилки столь близки, что три вида (или подвида *Homo sapiens*) – кроманьонцы, неандертальцы, денисовцы – можно рассматривать и как три равноправных таксона.

В последующих исследованиях было обнаружено, что вклад денисовцев особенно велик в население Меланезии (до 5%). Это удивительно, потому что – где Меланезия, и где Алтай с Денисовой пещерой! Поэтому большинство исследователей, включая и самого Сванте Паабо, предполагают, что ареал денисовцев включал не только Сибирь, но чуть ли не всю Азию – не только к северу, но и к югу от Гималаев. Поэтому возможно, что предки меланезийцев встретили денисовцев и смешались с ними в Юго-Восточной Азии – по пути предков меланезийцев из Африки в Меланезию.

НЕАНДЕРТАЛЬЦЫ ИЗ ДЕНИСОВОЙ

В геномах денисовцев и неандертальцев выявляются и смешения друг с другом, что неудивительно, поскольку в той же Денисовой пещере обнаружены останки и неандертальцев. Обнаружен в их геноме вклад и еще одной – неизвестной – популяции, которую можно гипотетически считать

еще неизвестным, четвертым подвидом *Homo sapiens*. Таким образом, изучение древних геномов начинает прорисовывать сложную картину: в среднем палеолите Евразии были не разделенные друг от друга неандертальцы и кроманьонцы, а существовала метапопуляция гоминид, включающая несколько видов или подвидов.

Если первоначально данные по геному неандертальца ограничивались анализом единственного образца, то впоследствии коллективом Сванте Паабо было проанализировано большое их число, в том числе были проанализированы останки нескольких неандертальцев, найденных в Денисовой пещере. И, по крайней мере, у одного из них по длине не разбитых рекомбинацией кусочков генома стало понятно, что он потомок очень инбридного брака – людей, состоящих примерно во второй степени родства [Prüfer et al., 2014]. Нужно отметить, что сам Сванте Паабо не спешит приписывать практику близкородственных браков всем неандертальцам: он считает, что нужно секвенировать больше неандертальцев, чтобы понять, типичен ли для них высокий инбридинг или это исключение.

САМЫЙ ДРЕВНИЙ ГЕНОМ НЕАНДЕРТАЛЬЦА

Одно из последних громких открытий, связанных с неандертальцами, произошло при совместном исследовании антропологами и генетиками «альтамурского человека», найденного в 1993 году в карстовой пещере Ламалунга возле города Альтамура в Италии [Lari et al, 2015].

Этот скелет был почти полностью погружен в отложения известняка, на поверхности оставалась лицевая часть черепа и некоторые кости. Из-за этого даже спустя много лет после находки исследование ограничивалось лишь описанием внешнего вида, доступного для обозрения. Судя по антропологическим признакам, скелет принадлежал человеку эпохи среднего-начала позднего плейстоцена. И хотя по некоторым чертам черепа он походил на неандертальца, по другим чертам отличался от большинства известных неандертальских скелетов. В 2009 году, фрагмент скелета (часть правой лопатки) удалось осторожно извлечь для антропологического и генетического исследования. Эта работа проведена итальянскими специалистами под руководством Джорджо Манци (Giorgio Manzi) из Римского университета при участии специалистов из Австралии и Испании.

Возраст скелета был определен методом урано-ториевой датировки известковых отложений на поверхности кости. На срезе были последовательно датированы слои этих отложений, от внешних к внутренним, что позволило датировать находку интервалом от 130 до 170 тыс. лет назад, то есть можно условно принять среднюю датировку око-

ло 150 тыс. лет. Антропологи провели детальное исследование фрагмента лопатки и по совокупности признаков поместили ее в неандертальский кластер. «Альтамурский человек» по имеющимся антропологическим признакам занял место между ранними и поздними неандертальцами.

Генетикам тоже удалось извлечь из останков ДНК, но она была очень фрагментирована, что неудивительно для останков такого возраста. Все же один фрагмент митохондриальной ДНК удалось секвенировать. Несколько позиций мтДНК несли типично неандертальские варианты (A16230G, G16244A, C16256A, A16258G) и отличались от гаплотипов как *Homo sapiens*, так и денисовского человека. Прежде всего это означает, что образец при

манипуляциях с ним не был загрязнен ДНК современного человека.

По полученному фрагменту мтДНК было проведено сравнение всех известных митохондриальных геномов неандертальца и денисовского человека и построено филогенетическое дерево (рис. 8.2). Как можно видеть на построенном дереве, «Альтамурский человек» вошел в один кластер с другими неандертальскими образцами из Западной Европы. Таким образом, это самый древний неандерталец, из останков которого была извлечена ДНК, и полученные результаты указывают на то, что популяция неандертальцев около 150 тыс. лет назад была неоднородна.



Рис. 8.2. Филогенетическое дерево изученных образцов неандертальцев, включая «Альтамурского человека» [Lari et al., 2015]. Дерево построено по данным о мтДНК. Геном «денисовца» использован в качестве внешней группы.

ЭФФЕКТЫ ОТБОРА: ОТ НЕАНДЕРТАЛЬЦА К НЕОАНТРОПУ

Данные по геному неандертальца, включенному в генофонд современного человечества, позволяют провести и функциональный анализ этих фрагментов генома. По этому поводу возникает еще больше эмоций и вопросов, чем по поводу родства неандертальцев и людей современного вида.

Даже если только один процент нашего генома унаследован от неандертальцев – этого достаточно, чтобы проводить статистический анализ. Рассматривая разные участки генома вдоль хромосомы, естественно предполагать, что в любом сегменте чисто случайно будет примерно одинаковое количество неандертальских последовательностей. Но геном неантропа (то есть наш геном) оказался очень неоднородным по содержанию неандертальской ДНК: например, 8-я и 17-я хромосомы совсем не содержат неандертальской ДНК, а есть участки и отдельные гены, в которых ее намного больше среднего. Поэтому, если мы где-то видим достоверный «всплеск» – повышение доли неандертальских последовательностей – то можно предполагать действие естественного отбора, который подхватил неандертальские по происхождению аллели, если они были адаптивны и полезны для неантропа. Оказалось [Vernot, Akey, 2014], что примерно 26% от всех кодирующих белки генов имеют кроме чисто сапиентных аллелей еще и такие аллели, в которых один или больше экзонов (участков, несущих информацию о белках) несут неандертальские последовательности. К генам с повышенным неандертальским вкладом относятся гены,

которые экспрессируются в кератиноцитах – клетках кожи – и регулируют ее пигментацию. Возможно, здесь приобретение неандертальских аллелей повысило способность сапиенсов адаптироваться к разным климатическим условиям при расселении по планете. А аллель адаптации к жизни на большой высоте у тибетцев считается заимствованным от денисовцев.

Кроме всплесков, есть и «впадины» – участки хромосомы, в которых последовательности или не передавались от неандертальцев или передавались, но потом отсекались отбором. Например, такие участки типичны вблизи генов, участвующих в сперматогенезе. К участкам генома, обедненным неандертальскими последовательностями, относят и участок на седьмой хромосоме, содержащий ген *FOXP2*, фактор транскрипции, игравший, как считает ряд исследователей, важную роль в появлении человеческой речи.

Некоторые из участков, обогащенных неандертальской ДНК, являются общими для современных популяций Азии и Европы, другие встречаются только в одной из частей Евразии. Поэтому можно рассмотреть две модели гибридизации (рис. 8.3). В соответствии с первой, поток неандертальских генов был принят популяцией – общим предком населения и Западной и Восточной Евразии. Вторая модель предполагает, что помимо этого, основного потока, случилось еще одно, меньшее вливание неандертальских генов в популяции Азии уже после их отделения от общего предка с европейцами. В пользу второй модели говорит большая доля неандертальского генома, циркулирующая в азиатских популяциях.

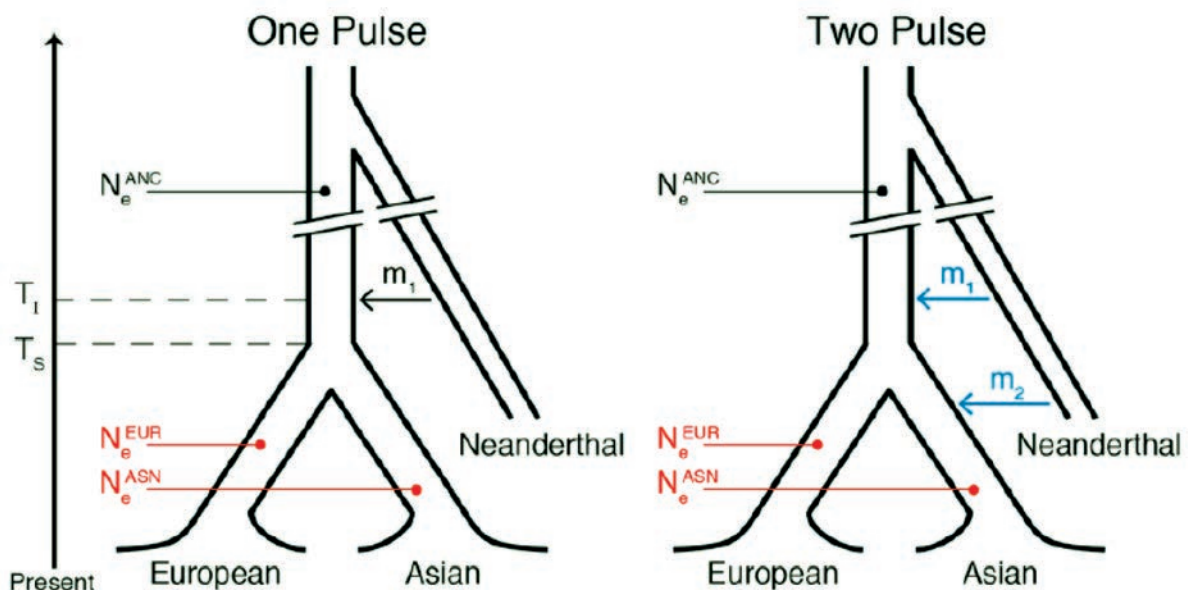


Рис. 8.3. Две модели интрогрессии неандертальской ДНК в геном европейских и азиатских популяций: один поток генов к предковой популяции или дополнительный поток генов в Азии [Vernot, Akey, 2014].

ЭФФЕКТЫ ОТБОРА: ОТ ОБЩЕГО ПРЕДКА К НЕОАНТРОПУ

Еще один вариант анализа – это поиск генетических вариантов, специфичных только для человека современного вида. То есть тех вариантов, которые есть у всех современных людей, но их нет не только у шимпанзе, но и у неандертальцев. Получается, что эти генетические варианты

возникли на самых последних этапах эволюции, ведущей к человеку современного вида. Найдено всего лишь 30 тысяч таких SNP, из них меньше сотни приводят к аминокислотным заменам. К настоящему времени подробно проанализированы и осмыслены не все из них, поскольку в первую очередь исследователи концентрировались на анализе тех из этой сотни генов, которые связаны с работой мозга.

8.3. ВЕРХНИЙ ПАЛЕОЛИТ И МЕЗОЛИТ

ГЕНОМ ИЗ ВОСТОЧНОЙ ЕВРОПЫ (КОСТЕНКИ)

РОДСТВЕННИК ОЧЕНЬ МНОГИМ

Яркий штрих к картине формирования генофонда Европы добавил секвенированный геном верхнепалеолитического человека со стоянки Костенки-14 под Воронежем, на берегу Дона. Это всемирно известная археологическая стоянка, одна из наиболее изученных в Восточной Европе (рис. 8.4). Скелет найденного здесь мужчины датирует-

ся временем 36–39 тыс. лет назад – он один из самых древних неантропов Европе. А его геном – один из самых древних секвенированных геномов *Homo sapiens*: древнее его только геном образца из Усть-Ишима (45 тыс. лет) и практически синхронен рассмотренный выше геном Oase 1 из Румынии (37–42 тыс. лет).

Митохондриальный геном образца Костенки-14 (K 14) был секвенирован еще в 2010 году – он

A Locations of analyzed samples

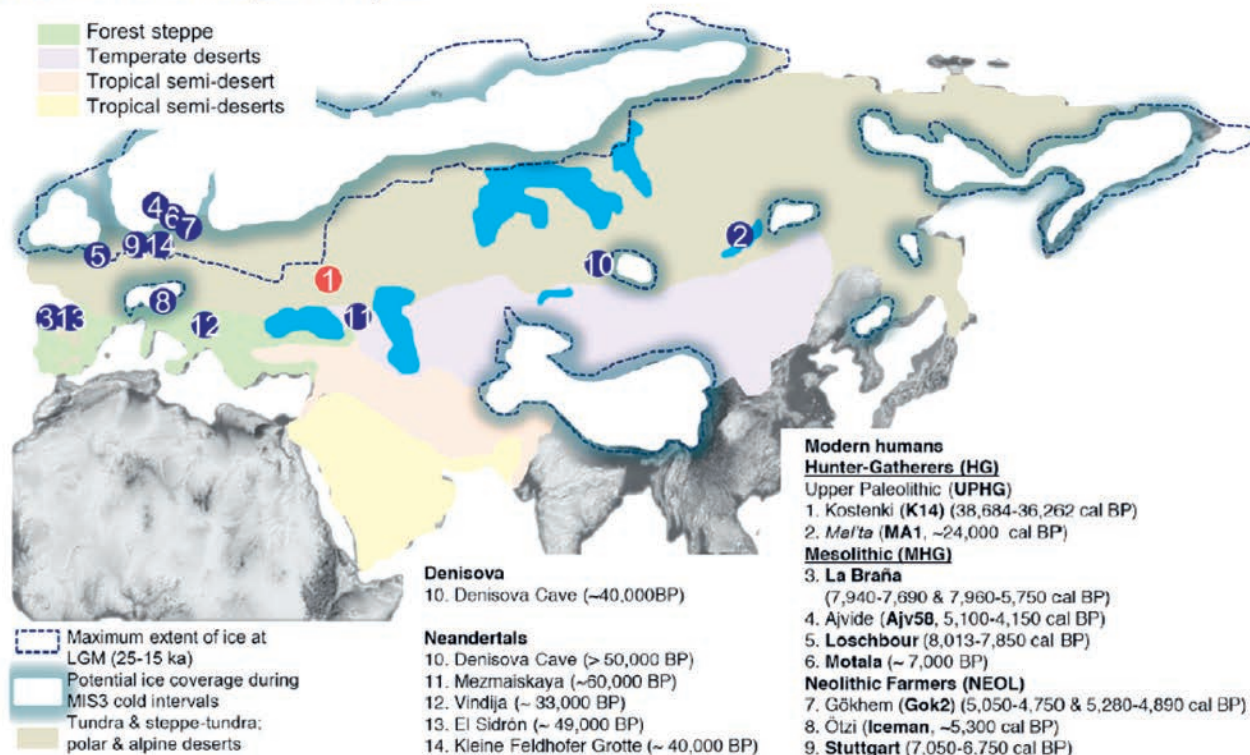


Рис. 8.4. Местоположение стоянки Костенки-14 и мест нахождения образцов других изученных древних геномов [Seguin-Orlando et al., 2014].

ВЕРХНИЙ ПАЛЕОЛИТ: 1. Костенки (K14, 38.7–36.3 тыс. лет), 2. Мальта (MA1, 24 тыс. лет). МЕЗОЛИТ: 3. Ла Бранья (7.9–7.7 тыс. лет и 8–5.8 тыс. лет), 4. Альвиде (5.1–4.1 тыс. лет), 5. Лошбор (8–7.9 тыс. лет), 6. Мотала (7 тыс. лет). НЕОЛИТ: 7. Гёкхейм (Gok2), (5–4.8 тыс.лет и 5.3–4.9 тыс. лет), 8. Отзи (тирольский ледяной человек, 5.3 тыс. лет), 9.Штутгарт (7–6.8 тыс. лет). ДЕНИСОВЕЦ: 10. Денисова пещера (ок. 40 тыс. лет). НЕАНДЕРТАЛЬЦЫ 10. Денисова пещера (>50 тыс. лет), 11. Мезмайская (ок. 60 тыс. лет), 12. Пещера Виндия (ок. 33 тыс. лет), 13. Эль Синдрон (ок. 49 тыс. лет), 14. Малый Фельдхофер грот (ок. 40 тыс. лет).

относится к гаплогруппе U2. А в 2014 году коллективом под руководством Eske Willerslev (Дания) удалось секвенировать и его ядерный геном [Seguin-Orlando et al., 2014]. Y-хромосома образца из Костенок принадлежит к гаплогруппе C-M130, так же как и у мезолитического охотника-собираателя Ла Бранья из Испании. Аутосомный геном Костенок-14 сравнили (рис. 8.5) с другими древними секвенированными геномами (их локализация и датировки даны на рис. 8.4) и геномами современного населения Евразии (2091 человек из 167 популяций).

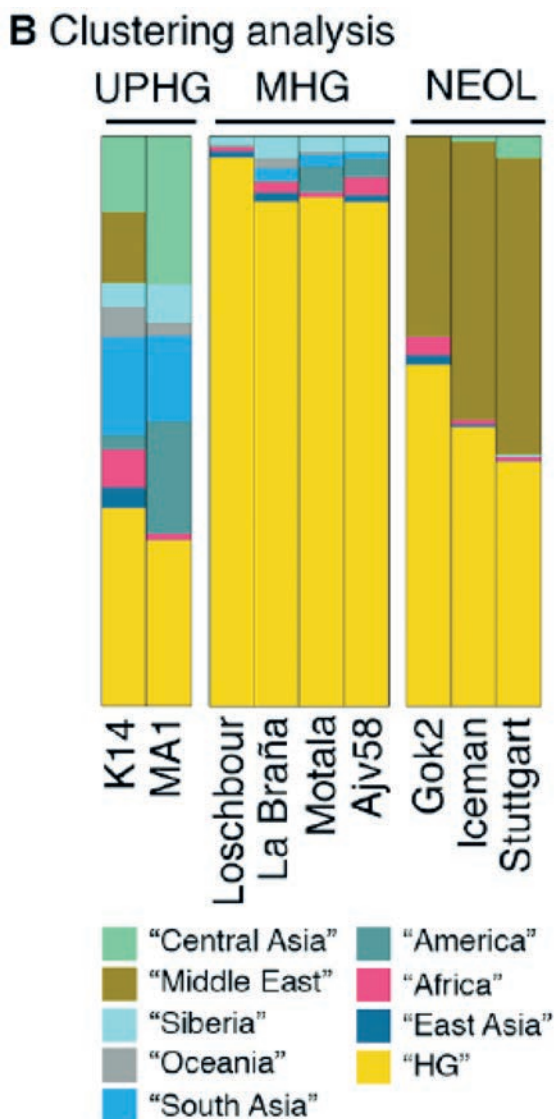


Рис. 8.5. Оценка доли в древних геномах генома охотников-собираателей и восьми компонентов современного населения разных географических регионов (метод кластерного анализа) [Seguin-Orlando et al., 2014]. UP – верхний палеолит, М – мезолит. HG – охотники-собираатели, NEOL – неолитические земледельцы. K14 – Костенки-14, MA1 – Мальта, Loschbour – Лошбор, La Braña – Ла Бранья, Motala – Мотала, Aiv58 – Альвиде, Gok2 – Гёкхем, Iceman – тирольский «ледяной человек», Stuttgart – Штуттгарт.

При этом обнаружена удивительная картина: геном из Воронежских Костенок так же, как и геном из Байкальской Мальты, содержит все восемь генетических компонентов из восьми крупных регионов современной ойкумены. Похоже, что в его геноме содержатся аллели, которые впоследствии разошлись по всем основным группам современного населения.

Из всех древних геномов образец из Костенок наиболее близок к геномам мезолитических охотников-собираателей (Лошбор, Ла Бранья), а также к геному верхнепалеолитического мальчика со стоянки Мальта в Восточной Сибири (MA1, 24 тыс. лет). На основании этого сходства в работе сделан вывод об интенсивных потоках генов, связывающих верхнепалеолитические популяции Сибири с предками мезолитических жителей Европы.

Эта работа несколько проясняет и загадочный «четвертый элемент» – тот «базальный» компонент генофонда Европы, который проявился при моделировании истории сложения генофонда Европы в работе [Lazaridis et al., 2014] (см. раздел 8.4, рис 8.20) – предковой евразийской группой, которая внесла свой большой вклад и в геном неолитических земледельцев. Из аналогичной по методам модели, созданной в рассматриваемой работе [Seguin-Orlando et al., 2014], следует (рис. 8.6), что в геном человека из Костенок эти таинственные «базальные евразийцы» внесли не менее важный вклад, чем и верхнепалеолитические западные евразийцы. Также из модели следует, что он имел и общих, хотя и более отдаленных, предков с древними северными евразийцами восточного ствола.

Из рассмотренных геномов 167 современных популяций геном из Костенок (K14) наиболее сходен с населением Европы, а в ее пределах – с геномами северных европейцев. Некоторое сходство отмечалось с современными популяциями Сибири (из изученных – особенно с хакасами), с популяциями Ближнего Востока и Кавказа, при этом практически отсутствует генетическое сходство с современным населением Восточной Азии (рис. 8.7).

Исходя из сравнительного анализа человека из Костенок и накопленных данных о геномах древнего и современного населения Евразии, сделан вывод [Seguin-Orlando et al., 2014] о раннем разделении популяций Восточной и Западной Евразии: верхнепалеолитический человек из Костенок принадлежал к западной евразийской линии микроэволюции (по антропологической классификации их, видимо, можно отнести к европеоидам), которая уже отделилась от населения Восточной Евразии (по антропологической классификации их, видимо, можно отнести к монголоидам). Исходя из датировки человека из Костенок (K14), разделение этих двух стволов генофонда Евразии (западноевразийского и восточноевразийского) произошло не позже, чем 36 тыс. лет назад. Поскольку у костен-

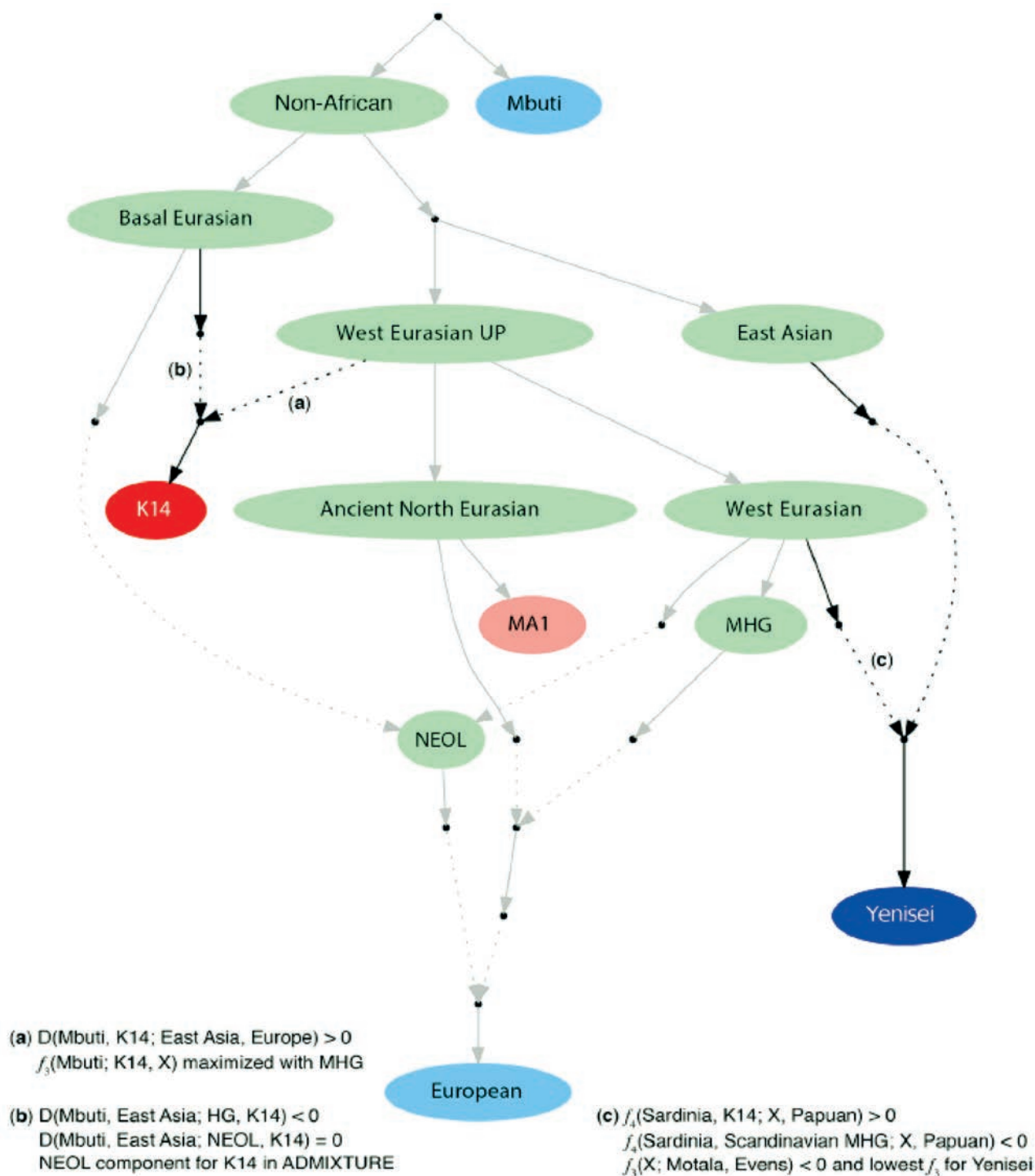


Рис. 8.6. Предполагаемые генетические связи между восточноевропейским образцом из Костенок (K14), сибирским образцом из Мальты (MA1) и генофондами древнего и современного населения [Sequin-Orlando et al., 2014].

Обозначения:

- Basal Eurasia – базальная евразийская группа популяций,
- West Eurasian UP – верхнепалеолитические западные евразийцы,
- Ancient North Eurasian – древние популяции Северной Евразии,
- MHG – мезолитические охотники-собиратели,
- NEOL – неолитические земледельцы,
- Yenisei – современные популяции на Енисее (хакасы),
- European – современные европейцы,
- Mbuti – пигмеи мбути.

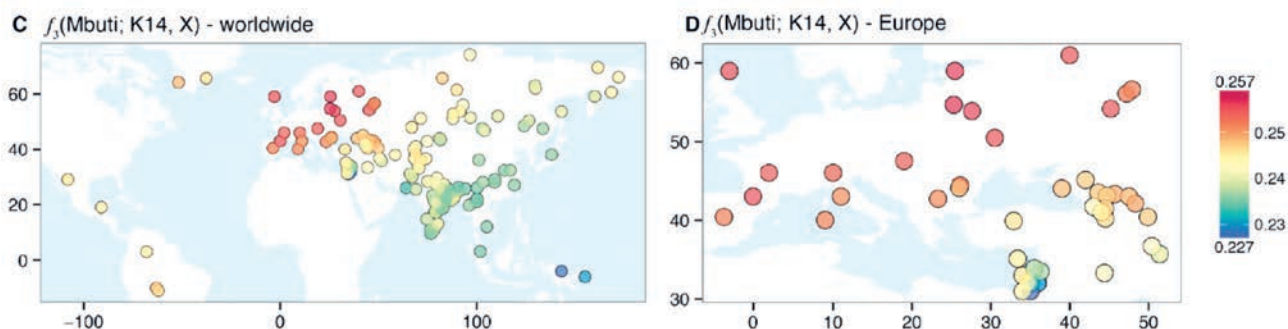


Рис. 8.7. Генетическое сходство между верхнепалеолитическим человеком из Костенок (K14) и современными популяциями мира (слева) и Европы (справа) [Seguin-Orlando et al., 2014].

Обозначения: более теплые цвета, в соответствии со шкалой (справа) указывают на большее генетическое сходство, холодные тона – на меньшее сходство.

ковского образца не обнаружена генетическая связь с современным населением Меланезии (напомним, что такая связь была выявлена с денисовцем) и Австралии, сделан вывод, что предки австралийцев и меланезийцев отделились от общего ствола еще раньше развилки западноевразийских и восточноевразийских генофондов.

ОСКОЛКИ МЕТА-ПОПУЛЯЦИИ ЕВРАЗИИ

В целом, обобщая накопленные данные по изучению древних геномов, можно сделать важ-

ный вывод, что в Европе, начиная с верхнего палеолита, существовала метапопуляция, неразрывно связанная с другими генофондами Евразии, и генетический след этой метапопуляции прослеживается и сегодня. Причем этот верхнепалеолитический генетический вклад у современных северных европейцев относительно выше, чем у южных. В целом генетическая картина колонизации Европы от верхнего палеолита до наших дней оказалась гораздо сложнее, чем представлялось ранее – до подробного изучения древней ДНК.

ЕВРОПЕОИДНЫЙ КОМПОНЕНТ В СИБИРИ (МАЛЬТА)

ДРЕВНИЙ ОБРАЗЕЦ ИЗ ЭРМИТАЖА

Еще один важный, уже не раз упоминавшийся выше и один из древнейших изученных на сегодня

образец геномов происходит с позднепалеолитической стоянки Мальта, близ озера Байкал (рис. 8.8), возрастом около 24 тысяч лет. Он был изучен под руководством Eske Willerslev между-



Рис. 8.8. Географическое положение верхнепалеолитических стоянок Мальта (красный треугольник) и Афонтова гора (красный кружок) в Сибири [Ragvalan et al., 2013].

Также указаны: местоположения палеолитических образцов с гаплогруппой U мтДНК (черные треугольники): 1. Оберкассель (Германия), 2. Холе Фельс (Германия), 3. Дольни Вестонице (Чехия), 4. Костёнки-14 (Россия); местоположения палеолитических образцов с гаплогруппой В мтДНК (черный квадрат): 5. Пещера Тяньюань (Китай); другие палеолитические стоянки с фигурками «палеолитических Венер» (коричневые кружки): 6. Лоссель (Франция), 7. Леспюг (Франция), 8. Гримальди (Италия), 9. Виллендорф (Австрия), 10. Гагарино (Россия); другие известные палеолитические стоянки (серые кружки): 11. Сунгирь (Россия), 12. Яна RHS (Россия).

народной командой, включившей и меня в числе многочисленных других коллег [Ragvalan et al., 2013].

Части скелета, принадлежащего мальчику стоянки Мальта (МА-1), находились в коллекции Эрмитажа. Ведь Мальта (ударение на последний слог) – это одна из самых известных стоянок человека позднего палеолита, место множества археологических находок. Например, именно там были найдены 30 знаменитых сибирских «палеолитических Венер» – скульптурных женских изображений верхнего палеолита (другие «палеолитические Венеры» обнаружены лишь в Европе, рис. 8.8). Благодаря полногеномному исследованию мальчика из Мальты и его неожиданным результатам, мировая известность этой сибирской стоянки еще больше возросла.

ЗАПАДНОЕВРАЗИЙСКИЙ КОМПОНЕНТ В ДРЕВНЕЙ СИБИРИ И АМЕРИКЕ

Как и в других исследованиях, геном из Мальты (МА-1) сравнивался с современным населением всего мира методом главных компонент (рис. 8.9). Результат оказался неожиданным: геном из верхнепалеолитической Сибири обнаружил сходство и с современным населением Западной Евразии, и с современными индейцами Америки, но оказался далек от своих географических соседей –

от современных популяций Восточной Азии. И многие современные популяции Сибири удаляются от Мальты на значительное генетическое расстояние. А те современные сибирские геномы, которые более близки к Мальте, все же находятся от нее примерно на том же расстоянии в генетическом пространстве 1 и 2 главных компонент, что и народы Европы, Америки и Центральной Азии (рис. 8.9). Эту картину уточняет рис. 8.10 – древо родства популяций, построенное методом TreeMix, допускающим не только разветвления, но некоторое число слияний ветвей дерева.

Подчеркнем, что не отдельные геномы, а представители всех 48 столь разных популяций современного коренного населения Америки оказались генетически сходны с верхнепалеолитической Сибирью (МА-1). Можно предполагать, что именно мета-популяция северной части Евразии верхнего палеолита, охватывающая ареал от Европы до Восточной Сибири, дала вклад 14–38% в генофонд современных индейцев Америки. Причем «западный» компонент должен был появиться в Сибири до того, как мигрировавшие через берингийский перешеек переселенцы из Сибири стали расселяться по Американскому континенту. Предполагается, что именно таким образом, коренное население обеих Америк вобрало в себя два потока генов – из Восточной и из Западной Евразии. Такая модель объясняет, например, каким образом у индей-

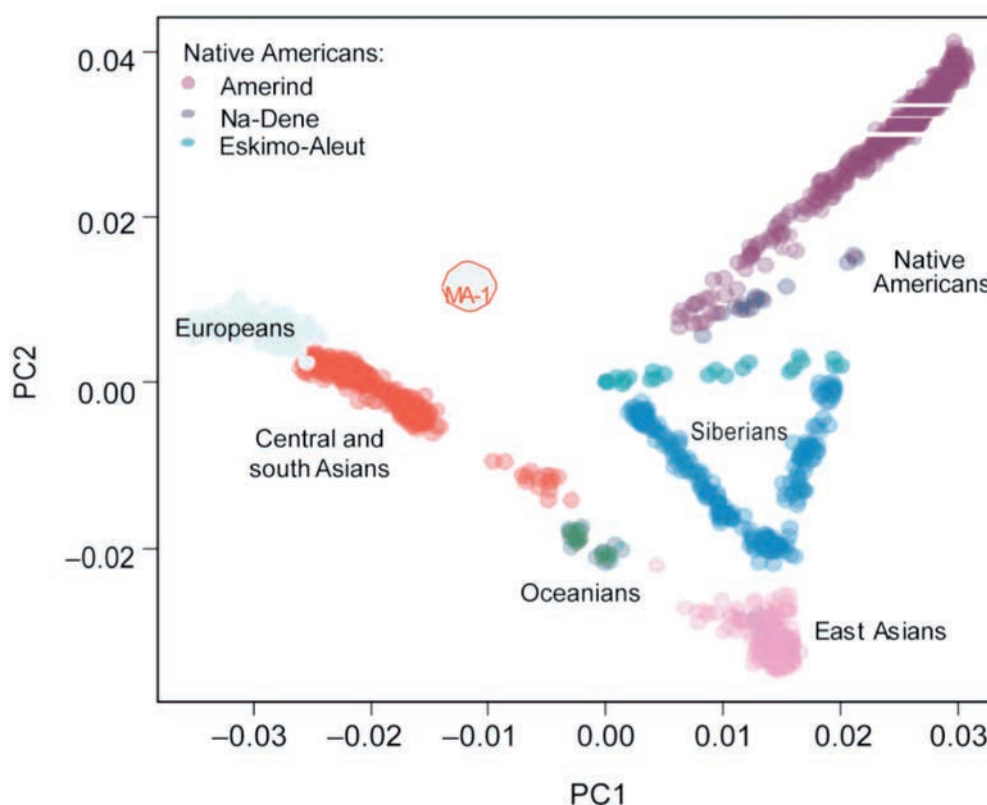


Рис. 8.9. Положение верхнепалеолитического генома Мальта (МА-1) из Восточной Сибири среди генофондов современного коренного населения Евразии и Америки (график главных компонент) [Ragvalan et al., 2013].

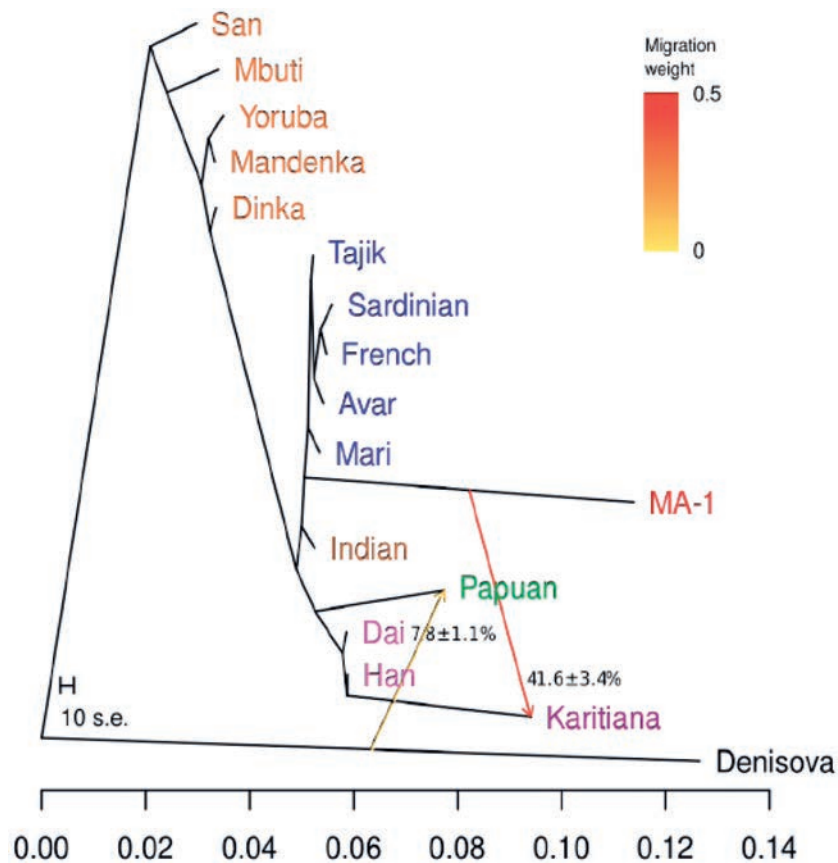


Рис. 8.10. Реконструкция вероятного родства 15 современных и 2 древних сибирских геномов (денисовец взят в качестве генетического масштаба) [Ragvalan et al., 2013].

Обозначения: MA-1 – Мальта, Karitiana – одна из популяций американских индейцев, Han и Dai – представители генофонда Восточной Азии. Полные геномы современного населения Сибири не были доступны авторам этого исследования и отсутствуют на графике, поэтому генофонд американских индейцев выводится из единственных присутствующих на дереве монголоидов (Han и Dai), но реконструируется и генетический вклад Мальты (красная стрелка). Сам геном Мальты по этим результатам выводится хотя и из западноевразийских генофондов, но почти так же близко к месту их развилки с популяциями Восточной Азии, как и геном из Индии.

цев Америки появилась митохондриальная гаплогруппа X, которая распространена в Европе, но отсутствует в Восточной Азии.

Этот западноевразийский компонент, свойственный верхнепалеолитическому населению Сибири и пришедший через Берингию из Сибири в Америку, может объяснить и «парадокс черепов», который долгое время является предметом дискуссий палеоантропологов. Парадокс заключается в том, что некоторые ископаемые черепа палеоамериканцев не обладают восточноевразийскими (монголоидными) чертами и не очень похожи на черепа современных индейцев.

Кроме того, и у современных индейцев есть некоторые европеоидные черты. Предполагалось, что эти антропологические черты у индейцев – результат позднего смешения в ходе послеколумбовой европейской колонизации. Но, в соответствии с генетическими данными, они могли прийти и из Берингии с самым древним населением Америки.

ЛЮДИ ЖИЛИ В СИБИРИ И НА МАКСИМУМЕ ОЛЕДЕНЕНИЯ

Геном мальчика из Мальты (MA-1, 24 тыс. лет назад) сравнили с другим древним сибирским геномом – из скелета с датировкой 17 тысяч лет назад, найденного на стоянке позднего палеолита близ Афонтовой горы (AG-2), на берегу Енисея (рис. 8.8). Оказалось [Ragvalan et al., 2013], что генетический профиль AG-2 сходен с MA-1, то есть за это время в населении Сибири не произошло больших генетических изменений. Но этих двух сибиряков разделяют не только семь тысяч лет – их разделяет сама природа. Дело в том, что время жизни мальчика из Мальты (MA-1) приходится на время последнего ледникового максимума (26-20 тыс. лет назад) – самого серьезного максимума оледенения за всю историю человечества. А сибиряк с Афонтовой горы (AG-2) жил уже после прохождения пика оледенения, когда климат стал значительно теплее. Из этого следует важный для ре-

конструкции истории населения Евразии вывод, что часть Сибири в течение последнего ледникового периода была заселена людьми постоянно, даже на пике оледенения. И что их генетический

след не исчез в этот самый суровый период жизни человечества, а они оставили потомство, которое продолжило цепь передачи генофонда и уже в приближении к эпохе мезолита.

МЕЗОЛИТ ВОСТОЧНОЙ ЕВРОПЫ (ЮЖНЫЙ ОЛЕНЬИЙ ОСТРОВ)

Генетическое исследование могильника Южного Оленьего Острова проведено нами в сотрудничестве с Центром древней ДНК университета Аделаиды (Австралия), музеем этнографии и антропологии имени Петра Великого (Кунсткамера) и Институтом археологии РАН и опубликовано в статье [Sarkissian et al., 2013].

Как известно, этот Оленеостровский могильник (см. табл. 1.7 в 1 главе), который оставило древнее население южной Карелии около 7,5 тысячелетий назад, является не только опорным памятником мезолита лесной зоны Восточной Европы, но и самым знаменитым. Он очень долго вызывал бурную дискуссию о степени и истоках монголоидности оставившего этот памятник населения, и соответственно о силе, направлении и географическом размахе миграций в Северной Евразии в ту эпоху. Недавнее антропологическое исследование, представляющее обобщающий анализ многочисленных древних и современных серий [Моисеев, Хартанович, 2012], склоняет чашу весов в пользу интерпретации этой популяции как раннего представителя недифференцированной уральской расы, которая является «третьей» расой Евразии, а не результатом метисации монголоидов и европеоидов (как предполагает альтернативная гипотеза). В этом контексте результаты независимого генетического исследования становятся особенно интересны и востребованы.

СПЕКТР ГАПЛОГРУПП

Проведен анализ митохондриальной ДНК нескольких десятков образцов от индивидуумов, захороненных на Южном Оленьем острове. Из них для 11 образцов удалось получить воспроизводимые и полные результаты по гаплогруппе и гаплотипу мтДНК (для увеличения выборки сюда включены и образцы из могильника Попово, антропологически и археологически близкого к Южному Оленьему острову). Обнаружены следующие гаплогруппы мтДНК (в порядке убывания частоты): **U4, C, U2, U5, J, H**.

Первая из этих гаплогрупп – **U4**, происходя изначально из южных регионов Евразии, в настоящее время распространена по всей Западной Евразии, достигая максимальных частот в Уральском регионе (причем по обе стороны хребта – как в Приуралье, так и в Западной Сибири). Поэтому обнаружение гаплогруппы **U4** в древних образцах

свидетельствует, во-первых, о том, что примерно та же география была свойственна этой гаплогруппе и семь тысяч лет назад. Во-вторых, эти результаты можно рассматривать как косвенное указание на генетическую связь мезолитического населения Северной Европы с современным населением Уральского региона.

Следующая по частоте обнаруженная на Южном Оленьем острове гаплогруппа **C** сейчас является типичной сибирской гаплогруппой. В случае, если гаплогруппу **C** обнаруживают с небольшими частотами в современных популяциях северо-востока Европы, исследователи рассматривают это как свидетельство потока генов в эти популяции из-за Уральского хребта. Естественно сохранить ту же интерпретацию и для наличия гаплогруппы **C** в древней мезолитической популяции. Но только если в современных популяциях Восточной Европы гаплогруппа **C** встречается у одного-двух людей на сто проанализированных (или обычно не встречается вовсе), то в популяции Южного Оленьего острова гаплогруппа **C** встречена у трех из двенадцати проанализированных образцов. Правда, пока нельзя исключить, что эти образцы происходят от родственных индивидуумов, и в таком случае они фактически являются копиями одной и той же митохондриальной ДНК. Но и в этом случае наличие гаплогруппы **C** является значимым индикатором генетических связей популяции Южного Оленьего острова с мезолитической или палеолитической Сибирью.

Остальные четыре гаплогруппы, встреченные в образцах из Южного Оленеостровского могильника (**U2, U5, J, H**), являются типичными западно-евразийскими гаплогруппами, преобладающими в современном населении как Восточной, так и Западной Европы.

В целом, обнаруженный спектр гаплогрупп мтДНК (самые частые **U4** и **C**, присутствуют также **U2e, U5a, J, H**) позволяет сделать два вывода. Во-первых, найдено сходство с другими мезолитическими популяциями Европы (изученными в работе [Bramanti et al., 2009]), в особенности по преобладанию субвариантов гаплогруппы **U**. Во-вторых, выявленный баланс западно-евразийских и восточно-евразийских гаплогрупп характерен для современных популяций Западной Сибири и отчасти Урала. Впрочем, более информативным является не описание генетических связей между древним и современным населением Евразии, а их картографирование.

КАРТА ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ

Карта генетических расстояний (рис. 8.11) обобщает данные по всем гаплогруппам и визуализирует степень близости реконструируемого древнего генофонда к современным популяциям Евразии. В основе такой карты лежит не только анализ палеодНК, но и обширная информация базы данных MURKA по изменчивости мтДНК в современном населении. Картографированные генетические расстояния были рассчитаны по частотам 31 гаплогруппы мтДНК: **A, B, C, D, F, G, H, HV, I, J, K, L, N1a, N1b, R, T, U1, U2, U3, U4, U5a, U5b, U6, U7, U8, V, W, X, Y, Z, other** (сумма прочих редких гаплогрупп). При построении такой карты важно использовать не только те гаплогруппы, что обнаружены в мезолитической популяции, но и всю возможную панель гаплогрупп, поскольку отсутствие какой-либо гаплогруппы – это не менее важная характеристика популяции, чем ее наличие. Конечно, всегда остается вопрос, что выборка древних образцов слишком мала, и образцы с иной гаплогруппой просто не попали в выборку. Но такие вопросы снимаются только в результате дальнейшего разворачивания исследований древней ДНК, а по меркам древней ДНК выборка 11 образцов для эпохи мезолита считается весьма представительной.

Карта для каждой области географического пространства показывает генетическое расстояние от современной популяции, населяющей эту область, до изучаемой древней популяции. При визуализации карты наиболее генетически непохо-

жие популяции (большие генетические расстояния) окрашиваются в коричневые цвета гор. Чем более генофонд современной популяции схож с изучаемым древним генофондом (чем меньше генетические расстояния), тем в более светлые тона окрашиваются ареалы таких популяций. При максимальном сходстве ареалы окрашиваются в зеленые цвета равнин (шкала расстояний от 0 до 1 приведена в легенде внизу карты, рис. 8.11). В результате мы собственными глазами можем видеть степень генетического сходства древней популяции с разными современными популяциями.

При интерпретации такой карты необходимо учитывать два момента.

Во-первых, ограниченность данных о древнем и современном генофондах. Поэтому для учета изученности современного населения на карте красными кружками обозначены генетически изученные современные популяции. Это позволяет осторожно относиться к тем ареалам генетического сходства, которые обеспечены информацией лишь о единственной современной популяции (как, например, популяция уйгуров в ареале северо-запада Китая на рис. 8.11). Более надежны оценки в тех случаях, когда генетическое сходство подтверждено несколькими современными популяциями (как например, бассейн Оби на рис. 8.11).

Второе ограничение связано с тем, что генофонды «не стоят на месте» – они меняются не только в современном пространстве (что показывает карта), но и во времени, поэтому экстраполяция современного генофонда на генофонд той же терри-

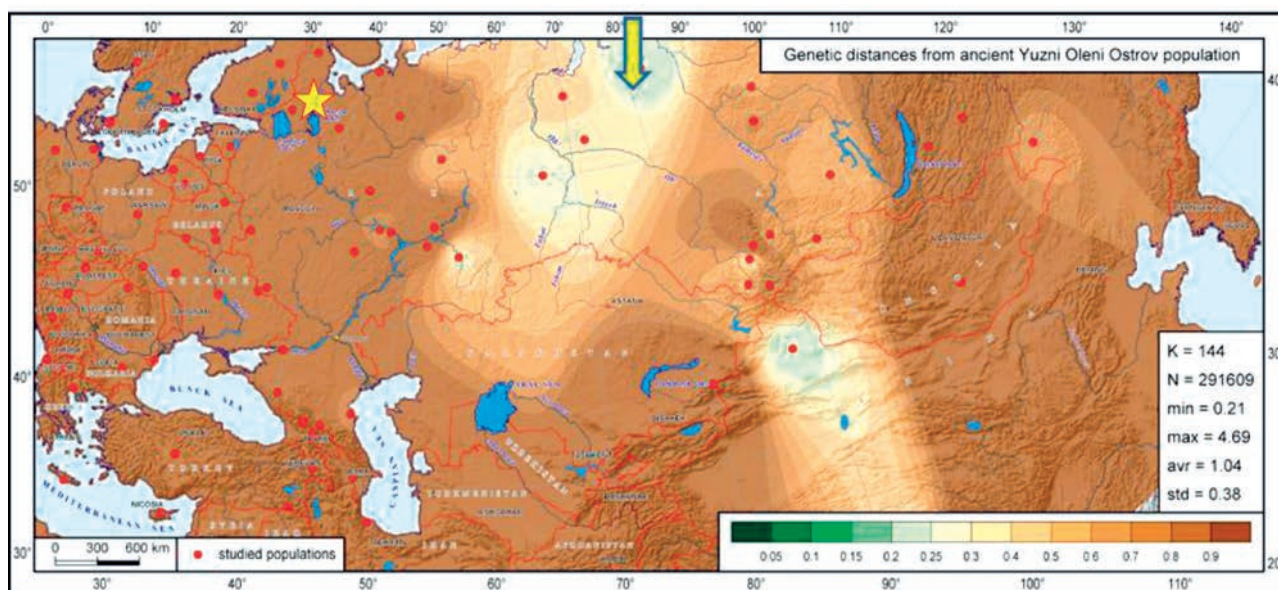


Рис. 8.11. Карта генетических расстояний от мезолитической популяции Южного Оленьего Острова до генофондов 144 современных популяций Евразии.

Обозначения: звездочкой отмечено положение изученной древней популяции; красными точками – положение генетически изученных современных популяций; стрелкой указана зона наибольшего сходства древней популяции с современным генофондом.

тории в древности является хотя и допустимой, но все же условностью. Снять это ограничение помогут лишь будущие обширные исследования древней ДНК, которые позволят постепенно проследить динамику генофондов Евразии во времени и в пространстве. Помня об этих ограничениях, попробуем с осторожностью приступить к интерпретации карты генетических расстояний.

Карта на рис. 8.11 демонстрирует, в какой степени современное население разных территорий Евразии является сходным с оленеостровской мезолитической популяцией. Легко заметить, что популяции, генетически сходные с оленеостровцами (светлые тона карты со значениями генетических расстояний от 0,20 до 0,40) наиболее распространены на Урале, в Западной Сибири и на северо-

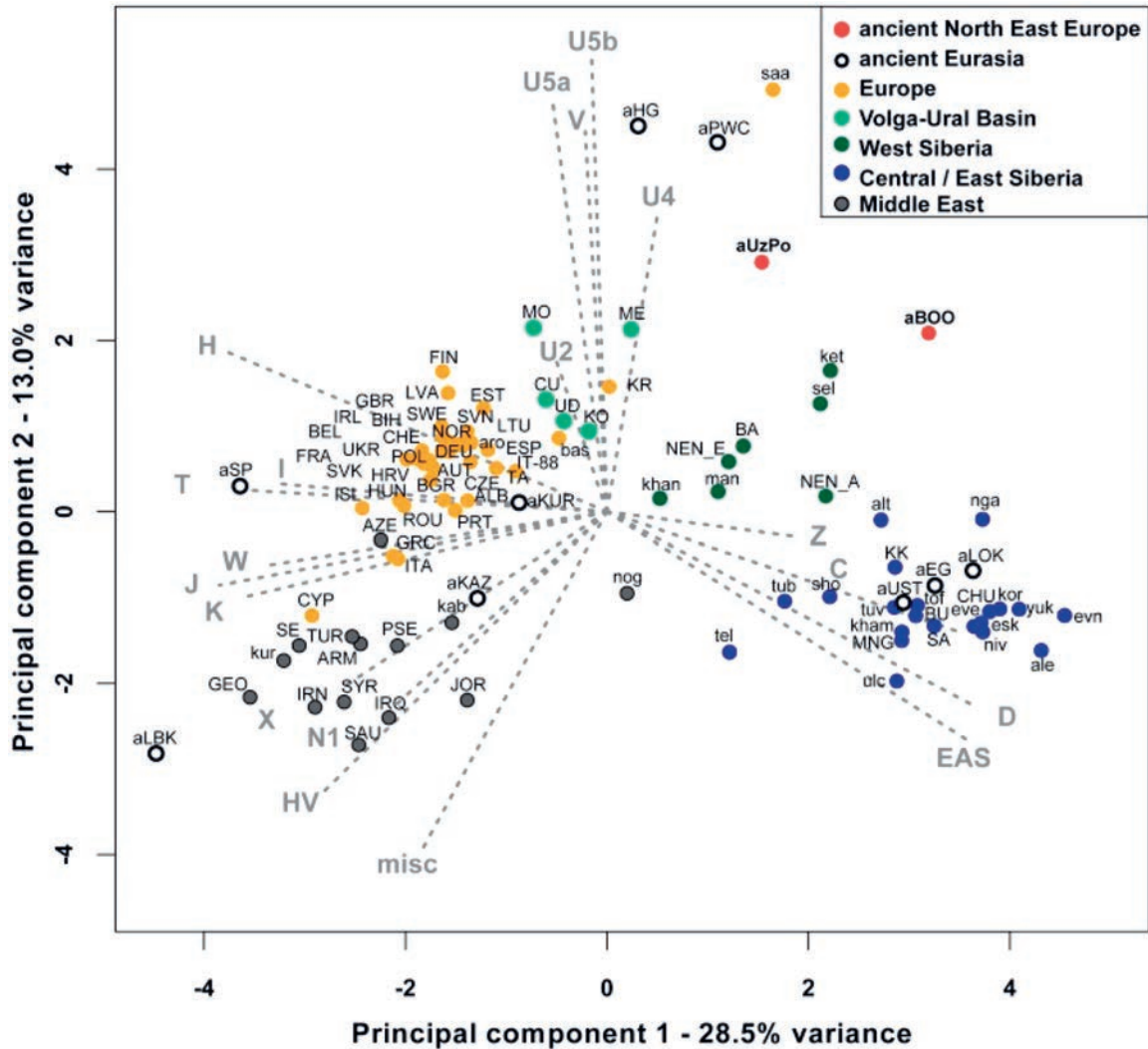


Рис. 8.12. Положение генофонда древних популяций северо-востока Европы относительно современного населения Европы и Ближнего Востока в пространстве 1 и 2 главных компонент [Sarkissian et al., 2013].

Обозначения: серым пунктиром показаны векторы гаплогрупп, показывающие вклад каждой гаплогруппы; красные кружки – древние популяции, анализируемые в данной статье; черно-белые кружки – другие древние популяции, использованные для сравнения; желтые кружки – современные популяции Европы; светло-зеленые кружки – современные популяции Волго-Уральского региона; темно-зеленые – современные популяции Западной Сибири; желтые кружки – современные популяции Центральной и Восточной Европы; серые кружки – современные популяции Передней Азии.

Древние популяции: aUzPo – Южный Олений остров и Попово (мезолит Северо-Восточной Европы, 7 500 лет назад); aBOO – Большого Олений остров (ранний металл Северо-Восточной Европы, 3 500 лет назад); aHG – охотники-собиратели палеолита и мезолита (Центральная и Восточная Европа, 30 000–4250 лет назад); aLBK – неолит Германии (7 500–7 100 лет назад); aLOK – могильник Локомотив китойской культуры (ранний неолит Прибайкалья, 7 140–6 130 лет назад); aUST – могильник Усть-Ида (развитый неолит Прибайкалья, 5 800–4 000 лет назад); aSP – неолит Испании (5 500–5 000 лет назад); aPWC – скандинавская культура ямочной керамики (5 300–4 500 лет назад); aEG – объединенная группа кочевников Ксионгну (Xiongnu) (азиатские гунны, 4 250–2 300 лет назад); aKUR – курганы Сибири (3 800–1 600 лет назад); aKAZ – кочевники Казахстана (3 400–2 100 лет назад).

Современные популяции:

ALB – албанцы, ALE – алеуты, ALT – алтайцы, ARM – армяне, ARO – аромунцы, AUT – австрийцы, AZE – азербайджанцы, BA – башкиры, BAS – баски, BEL – белорусы, BGR – болгары, BIN – боснийцы, BU – буряты, CHE – швейцарцы, CHU – чукчи, CU – чувашаи, CYP – Cypriots, CZE – чехи, DEU – немцы, ESK – эскимосы, ESP – испанцы, EST – эстонцы, EVE – эвенки, EVN – эвены, FIN – финны, FRA – французы, GBR – британцы, GEO – грузины, GRC – греки, HRV, Croatians HUN – венгры, ING - Ingrians IRL – ирландцы, IRN – иранцы, IRQ – иракцы, ISL – исландцы, IT-88 – сардинцы, ITA – итальянцы, JOR- иорданцы, KAB – кабардинцы, KET – кеты, KHAM – хамнигане, KHAN – ханты, KK – хакасы, KOM – коми, KOR – коряки, KR – карелы, KUR – курды, LTU – литовцы, LVA – латыши, MAN – манси, ME – марийцы, MNG – монголы, MO – мордва, NEN_A – восточные ненцы, NEN_E – западные ненцы, NGA – нганасаны, NIV – нивхи, NOG – ногайцы, NOR – норвежцы, POL – поляки, PRT – португальцы, PSE – палестинцы, ROU – румыны, RUS – русские, SA – якуты, SAA – саамы, SAU – арабы Саудовской Аравии, SE – осетины, SEL – селькупы, SHO – шорцы, SVK – словаки, SVN – словенцы, SWE – шведы, SYR – сирийцы, TA – татары, TEL – теленгиты, TOF – тофалары, TUB – тубалары, TUR – турки, TUV – тувинцы, UD – удмурты, UKR – украинцы, ULC – ульчи, VEP – вепсы, YUK – юкагиры.

западе Средней Сибири, причем ядром этой области является Западная Сибирь. Это позволяет – весьма гипотетично – говорить о Западной Сибири как о возможной прародине мезолитического населения Восточной Европы.

Более осторожная интерпретация карты такова: соотношение западноевразийского и восточноевразийского генофондов на Южном Оленьем острове было примерно таким, как в современном населении Западной Сибири. То есть (в сравнении с современным населением Восточной Европы, где доминируют западно-евразийские гаплогруппы) это соотношение было значительно сдвинуто в пользу восточного генетического ствола.

АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТ И АНАЛИЗ ОБЩИХ ГАПЛОТИПОВ

К сходным выводам приводит и анализ двумя другими независимыми методами.

При анализе методом главных компонент (первые две главные компоненты вбирают 41,5% общей генетической изменчивости) четко выделились европейский, ближневосточный, уральский и сибирский кластеры (рис. 8.12). На этом графике оленеостровцы расположены между уральским и сибирским кластерами, что подтверждает вывод об их принадлежности к «пограничному» генофонду.

При анализе методом общих гаплотипов мы переходим с уровня гаплогрупп мтДНК на уровень отдельных гаплотипов мтДНК (митотипов). В связи с их большим числом одним из наиболее информативных методов анализа оказывается подсчет доли гаплотипов, совпадающих между двумя популяциями. По смыслу, это аналог генетического расстояния между популяциями, но рассчитанный не на уровне гаплогрупп, а на уровне гаплотипов.

Как было показано нами ранее [Балановская, Балановский, 2007], при таком подсчете исследователя подстерегает опасность значительного завышения показателей сходства при использовании выборок малого объема. Причина в существова-

нии гаплотипов, распространенных практически повсеместно (как правило, это гаплотипы-основатели для частых гаплогрупп мтДНК, например, rCRS для гаплогруппы **H** или 16069–16126 для гаплогруппы **J**). Эти и подобные им гаплотипы являются как бы фоновым уровнем сходства почти для любой пары западноевразийских популяций. И чем меньше объем выборки, тем больше будет относительная доля таких гаплотипов. Чтобы избежать этого нежелательного влияния объема выборки на получаемые результаты, мы предложили [Балановская, Балановский, 2007] использовать выборки большого и одинакового объема для всех сравниваемых популяций. Этот подход был применен и к анализу общих гаплотипов в данном исследовании (табл. 8.1). На практике это делалось так. Сначала составлялись обширные сборные выборки путем объединения имеющихся выборок из географически и лингвистически близких популяций. Затем из этих выборок методом случайных чисел извлекалось одно и то же число образцов (по 500 образцов), а все прочие образцы не включались в анализ общих гаплотипов.

В таблице 8.1 представлены результаты проведенного таким образом сравнения митохондриального генофонда древней оленеостровской популяции и 34 крупных выборок из современных популяций Европы и Азии. Таблица показывает, что только семь современных популяций имеют по два общих гаплотипа с оленеостровской мезолитической популяцией, при этом число образцов с такими гаплотипами колеблется от 2 до 17. Особенно важно, что эти семь популяций примерно поровну распределены между Сибирью и Европой. То есть и этот вид анализа указывает на генетические связи оленеостровцев и с западноевразийскими, и с восточноевразийскими популяциями Северной Европы.

Таким образом, все методы анализа полученных данных по древней ДНК показывают промежуточное положение населения, оставившего Южный Оленеостровский могильник: в целом принадлежа к западноевразийскому генетическому

Таблица 8.1. Доля общих гаплотипов между древними популяциями (мезолитической ЮОО и эпохи раннего металла БОО) и современными популяциями Европы и Азии [Sarkissian et al., 2013].

Современные популяции	N	ЮОО		БОО	
		N _{HT}	N _S	N _{HT}	N _S
Балканы 1 (боснийцы, словенцы)	500	1	14	2	9
Балканы 2 (хорваты)	500	1	14	3	18
Ближний Восток 1 (Аравийский п-ов, Марокко)	500	1	8	1	1
Ближний Восток 2 (Турция)	489	1	2	2	4
Британские острова 1 (англичане и валлийцы, группа 1)	500	1	19	3	7
Британские острова 2 (англичане и валлийцы, группа 2)	500	1	19	2	2
Британские острова 3 (шотландцы)	500	1	27	2	8
Волго-Уральский регион (коми, мари, мордва, удмурты, татары)	500	1	19	4	25
Восточная Европа 1 (русские)	500	1	22	3	4
Восточная Европа 2 (украинцы)	500	2	12	3	9
Восточная Европа 3 (карелы, ингры, вепсы)	500	1	15	1	2
Восточная Европа 4 (башкиры, ногайцы, казахи, чувашы)	499	1	4	4	28
Западная Сибирь (ханты, манси, ненцы, селькупы)	500	0	0	5	29
Прибалтика 1 (латыши, литовцы)	500	1	11	3	4
Прибалтика 2 (эстонцы)	497	2	13	1	3
Северная Европа 1 (норвежцы)	500	1	25	5	13
Северная Европа 2 (шведы восточные)	500	1	14	4	7
Северная Европа 3 (шведы прочие)	500	1	14	3	4
Северная Европа 4 (саамы)	500	0	0	2	8
Северная Европа 5 (финны)	500	1	8	1	8
Северная Сибирь (нганасаны, кеты, эвенки, якуты)	500	2	5	5	48
Северо-Восточная Сибирь 1 (чукчи, эвены, коряки, юкагиры)	500	0	0	2	48
Северо-Восточная Сибирь 2 (эскимосы)	500	0	0	0	0
Средний Восток (Иран, Ирак, Сирия)	500	1	3	0	0
Центральная Азия (монголы, хамнегане)	500	2	2	5	26
Центральная Европа 1 (немцы, австрийцы, швейцарцы)	500	1	9	2	4
Центральная Европа 2 (чехи, словаки)	500	1	16	3	6
Центральная Европа 3 (поляки)	500	2	17	3	5
Южная Европа 1 (португальцы)	500	1	13	2	5
Южная Европа 2 (испанцы, французы)	500	2	7	1	3
Южная Европа 3 (итальянцы)	500	1	9	1	2
Южная Сибирь 1 (алтайцы, хакасы, шорцы, тофалары)	500	1	4	4	43
Южная Сибирь 2 (тувинцы)	500	0	0	2	61
Южная Сибирь 3 (буряты)	500	2	7	5	31

Примечания: приведены результаты анализа совпадающих гаплотипов мтДНК между современными митохондриальными генофондами и двумя древними популяциями: мезолитической популяцией Южного Оленьего острова и популяцией эпохи раннего металла Большого Оленьего Острова.

Обозначения:

ЮОО – Южный Олений Остров, Карелия, 7500 лет назад.

БОО – Большой Олений Остров, Кольский п-ов, 3500 лет назад.

N – объем выборки (все современные выборки приведены к единому объему 500 образцов).

N_{HT} – число общих гаплотипов (разных вариантов мтДНК) между современной и древней популяциями.

N_S – число образцов, несущих эти общие гаплотипы.

стволу, оно несло в своем генофонде и следы генетической связи с Уралом и Западной Сибирью.

Учитывая многолетнюю антропологическую дискуссию о частичной монголоидности оленеостровцев, можно заключить, что в мезолите (или раньше) взаимосвязи населения Восточной Европы, Урала и прилежащих районов Сибири были намного интенсивнее, чем в последующие эпохи. Вероятно, население востока Европы, Урала, Западной Сибири представляло единый континуум. Отметим, что это согласуется с зоной распространения уральских языков и теорией «уральской расы» [Бунак, 1980; Моисеев, 1999; Перевозчиков, 2003].

ПОЛНОЕ СЕКВЕНИРОВАНИЕ мтДНК ВЫЯВИЛО НОВУЮ ГАПЛОГРУППУ В МЕЗОЛИТЕ

Продолжением нашего исследования древней ДНК из Оленеостровского могильника стало секвенирование полного митохондриального генома мезолитического охотника-собирателя с Южного Оленьего острова [Sarkissian, 2014]. МтДНК этого образца относилась к уже упомянутой выше гаплогруппе С. Эта гаплогруппа чаще всего встречается в аборигенных популяциях Азии и Америки (рис. 8.13). На сегодня известно четыре основных варианта этой гаплогруппы: **C1**, **C4**, **C5** и **C7**. Сре-

ди них гаплогруппа **C1**, широко распространенная на Азиатском и Американском континентах, чрезвычайно редко встречается в современных европейских популяциях: всего несколько ее носителей найдены среди современных жителей Германии, Канарских островов, Исландии и Башкирии. Генетики предполагают, что гаплогруппа **C1** стала редкой в Европе из-за демографических событий: «бутылочного горлышка» и других разновидностей дрейфа генов, а также многочисленных миграций. Место ее происхождения, вероятно, находится в Восточной Сибири, о чем свидетельствует ее высокая частота у местного населения.

Гаплогруппу **C1** подразделяют на несколько субгаплогрупп: до сих пор были известны **C1a**, **C1b**, **C1c**, **C1d** и **C1e**, отличающиеся разным географическим распределением (рис. 8.14). Три из них (**C1b**, **C1c** и **C1d**) характерны для коренного населения Америки и широко распространены по всему Американскому континенту. Поскольку первое заселение Америки происходило в ходе миграции людей из Восточной Сибири через перешеек на месте нынешнего Берингова пролива, вероятно, эти субгаплогруппы были принесены в Америку из Азии. Субгаплогруппа **C1a** в небольшом количестве встречается у коренного населения Восточной Евразии. Последняя из известных до сих пор субгаплогруппа **C1e** недавно найдена в

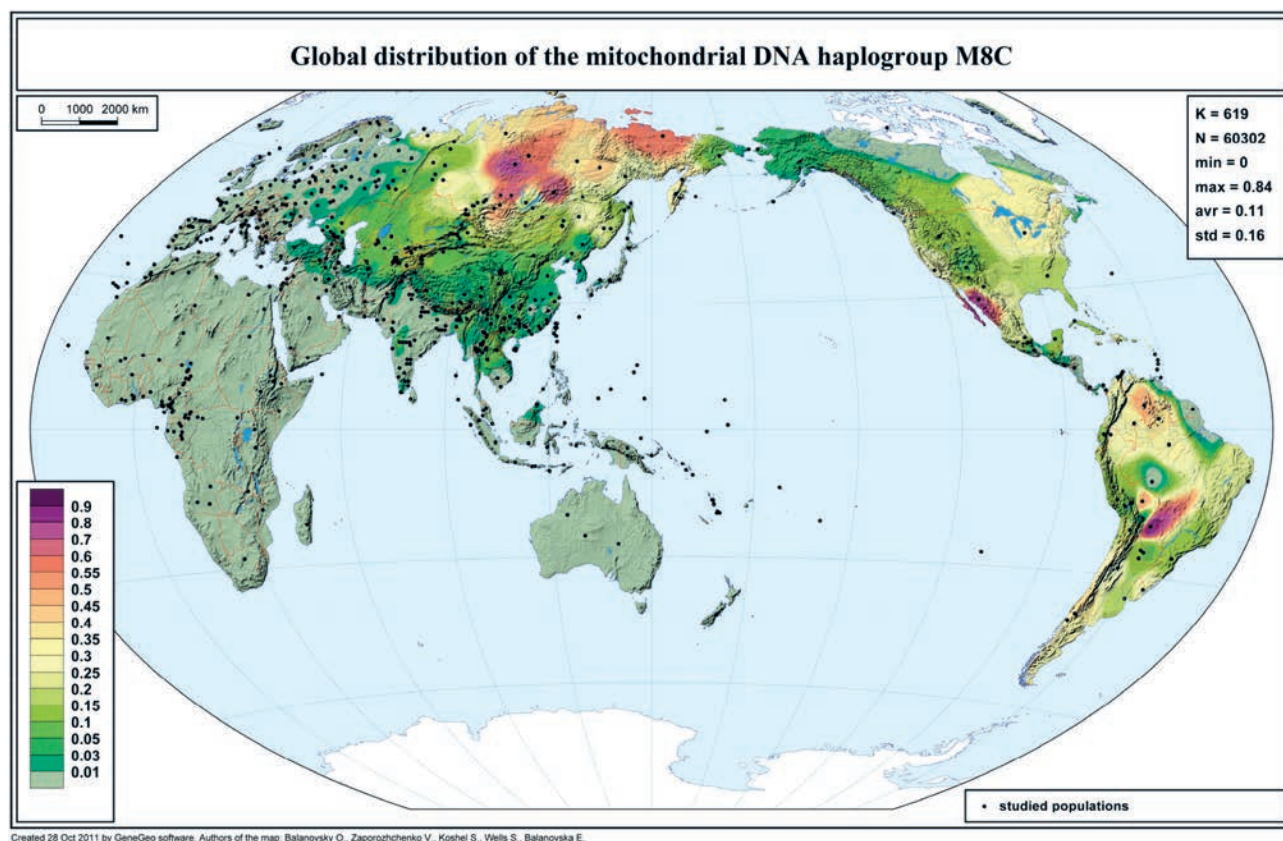


Рис. 8.13. Географическое распределение гаплогруппы С митохондриальной ДНК человека. Цветом обозначена частота встречаемости гаплогруппы (шкала слева внизу).

Исландии (рис. 8.14), причем всего у нескольких человек, и довольно сильно отличается от своих – американских и азиатских – сестринских вариантов.

Из трех образцов с Южного Оленьего острова мтДНК одного индивида (UZOO-74) была полностью секвенирована. Так же, как и в других работах по полному секвенированию древних митохондриальных геномов, использовали метод секвенирования нового поколения, в данном случае ионного полупроводникового секвенирования (Ion Torrent). В мтДНК обнаружили 58 нуклеотидных замен, отличающих ее от реконструированного исходного предкового варианта мтДНК человека («корня» митохондриального древа человечества). В том числе пять мутаций, отличающих ее от всех остальных вариантов C1, что и позволило отнести ее к новой субгаплогруппе C1f. Следуя по алфавиту, ее назвали C1f (рис. 8.15).

После того, как генотипировали другие два образца, UZOO-7 и UZOO-8 из Оленеостровского могильника, выяснилось, что в их мтДНК имеются те же пять новых мутаций, поэтому их отнесли

к той же субгаплогруппе C1f. Таким образом, мтДНК всех трех мезолитических охотников-собирателей оказалась очень сходной, что говорит об их родстве по материнской линии, все они были носителями субгаплогруппы C1f. Сравнение ее с базой данных подтвердило, что речь идет о действительно новой линии, которая не представлена в современных популяциях.

Относительно появления гаплогруппы C1 в Европе имеется три гипотезы. Первая отводит ведущую роль позднему потоку генов из Азии в Европу в историческое время. Вторая предполагает, что C1 появилась в Европе через Америку, в результате того, что шло смешение генов между коренными американцами и европейцами в период европейской колонизации Американского континента. Или же во время более ранней колонизации Америки, еще в доколумбовы времена она была принесена исландскими викингами, которые, как известно, создавали первые поселения на северо-западном побережье Америки в X веке. Ну, а третья гипотеза допускает, что гаплогруппа C1 была представлена в Европе с древнейших времен. В

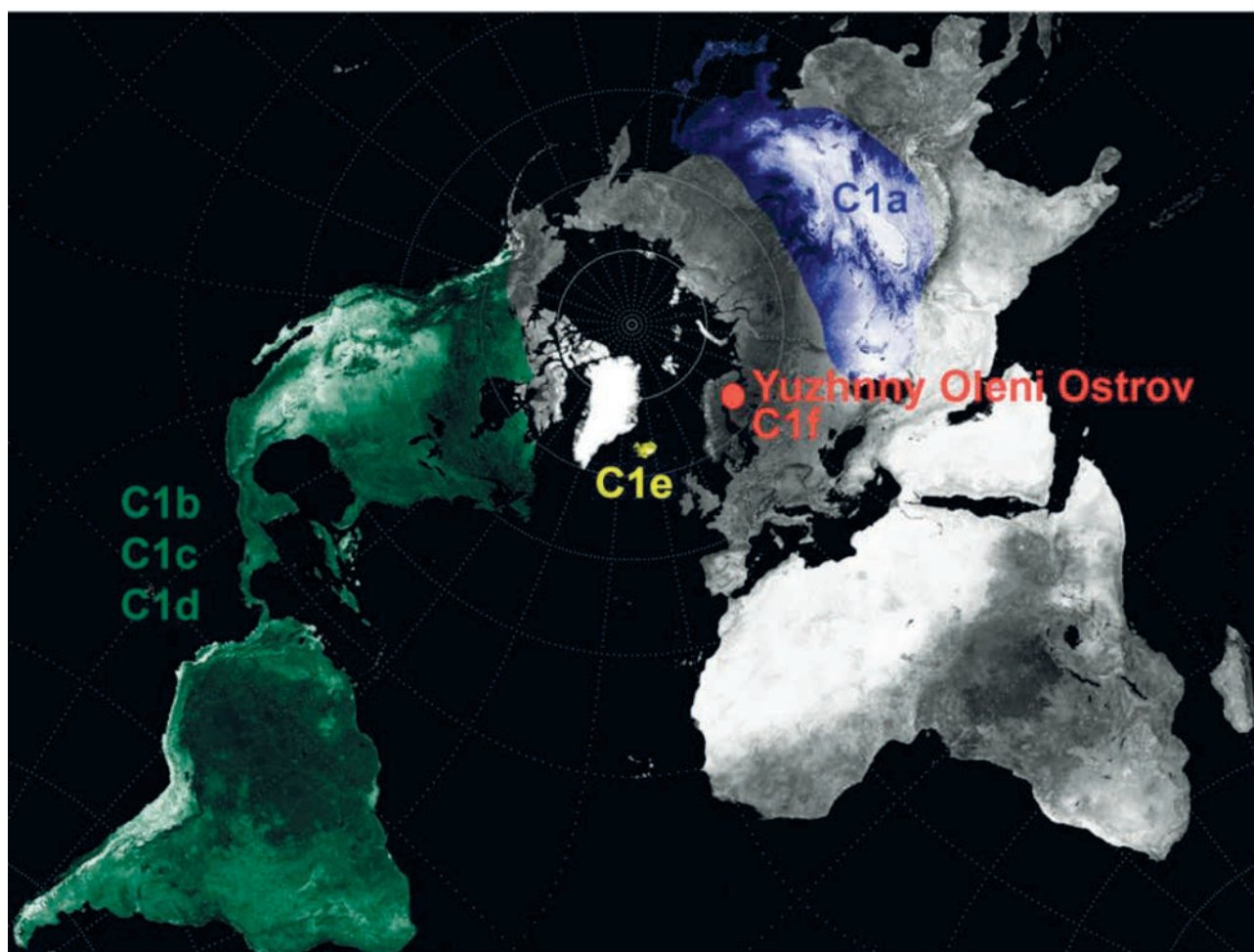


Рис. 8.14. Примерное географическое распределение субгаплогрупп C1a, C1b, C1c, C1e и C1f в современных популяциях и в мезолитической популяции Южного Оленьего острова [Sarkissian et al., 2014]. Локализация Южного Оленьего Острова (Yuzhny Oleni Ostrov) показана красной точкой.

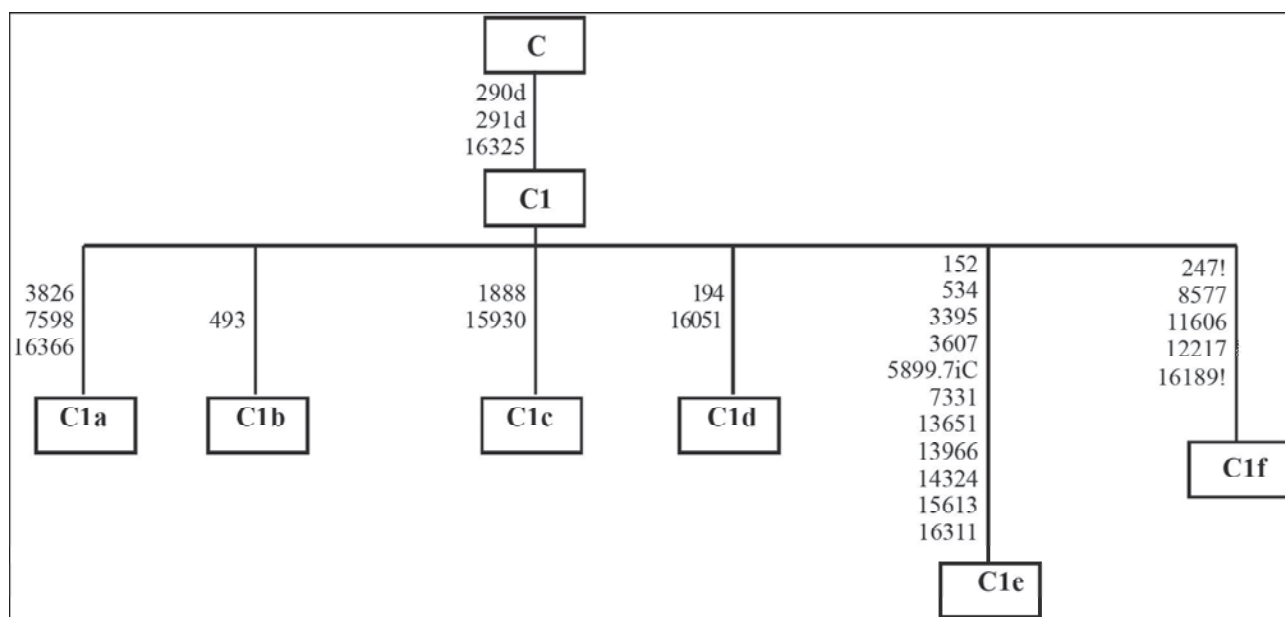


Рис. 8.15. Филогенетическое дерево гаплогруппы C1 [Sarkissian et al., 2014].

Показаны все известные к настоящему времени субгаплогруппы в пределах C1. Цифры вдоль ветвей дерева обозначают номера позиций в последовательности митохондриальной ДНК человека, в которых произошли мутации, специфичные для данной ветви (в качестве референтной последовательности использован RSRS (Behar et al., 2012)). “i” обозначает инсерцию, “d” – делецию, “!” – обратную мутацию.

пользу именно этой гипотезы говорит ее обнаружение в мтДНК мезолитических охотников-собирателей с Южного Оленьего острова.

Где же возникла субгаплогруппа C1f? Пока на этот вопрос нет однозначного ответа. По нашим предположениям, скорее всего, она возникла в Сибири, но довольно рано отделилась от других евразийских линий C1, и в дальнейшем они оказались

привязаны к разным региональным популяциям. Наибольшее родство C1f, по-видимому, имела с «исландской» субгаплогруппой C1e. Возможно, они разделились от общего предка где-то в Азии и позднее достигли Северной Европы разными путями независимых миграций. Субветвь C1f можно сравнить с мостом, соединяющим исландские (C1e) и восточноазиатские (C1a) линии гаплогруппы C1.

8.4. НЕОЛИТИЗАЦИЯ ЕВРОПЫ

По данным археологии, палеоантропологии и генетики первая волна заселения Европы связана с палеолитом. В качестве второй волны часто выделяют мезолитическую реколонизацию Европы после отступления ледника (один из наиболее северных памятников этой эпохи рассмотрен в предыду-

щем разделе). Но больше всего споров вызывает третья волна – неолитических земледельцев, поскольку именно с неолитом связан резкий демографический взрыв численности населения, и поэтому именно неолит может быть претендентом на определение основных черт европейского генофонда.

ПЕРВЫЕ ЗЕМЛЕДЕЛЬЦЫ (МИТОХОНДРИАЛЬНЫЙ ГЕНОФОНД)

ГИПОТЕЗЫ «ДЕМИЧЕСКОЙ» И «КУЛЬТУРНОЙ» ДИФФУЗИИ

В классическом труде археолога Аммермана и генетика Кавалли-Сфорца была сформулирована гипотеза «демической диффузии», постулирующая, что именно третья – неолитическая – волна сформировала основные черты европейского генофонда [Ammerman, Cavalli-Sforza, 1984]. Важным элементом гипотезы «демической диффузии» является демографический эффект перехода от при-

сваивающего хозяйства к производящему: экспоненциальный рост численности населения и постепенное расселение земледельческих популяций, замещающих или вбирающих в себя малочисленные аборигенные популяции охотников-собирателей – потомков палеолитического и мезолитического населения Европы.

Археологически земледельческие (неолитические) культуры прослеживаются с 7 тыс. лет до н.э. на Ближнем Востоке, откуда они затем проникают на Балканы и оттуда распространяются на осталь-

ную территорию Европы в 5–3 тысячелетиях до н.э. (7000–5000 лет назад). При этом по мере удаления от Ближнего Востока и Балкан доля мигрировавшего населения убывала, и соответственно, должна была убывать доля мигрантных генов в генофонде и возрастать доля генофонда автохтонного палеолитического населения Европы. И действительно, исследования классических генетических маркеров выявили географический градиент на карте первой главной компоненты изменчивости генофонда [Cavalli-Sforza et al., 1994], полностью соответствующий предполагаемой доле мигрантных генов неолитического населения и ее убывания по мере удаления от Ближнего Востока и Балкан.

Однако позднее данные по митохондриальной ДНК показали значительно более древний – палеолитический (порядка 20–30 тыс. лет назад) – возраст большинства гаплогрупп мтДНК, распространенных в населении Европы [Richards et al., 1996]. Из этого следовал вывод, что распространение неолита в Европе было «культурной» диффузией: культура неолита распространялась не ее непосредственными генетическими потомками – выходцами из популяций Ближнего Востока, а местным европейским населением, перенявшим эту культуру, но сохранившим свой генофонд.

Последовала длительная и бурная дискуссия сторонников неолитического и палеолитического времени формирования европейского генофонда [Barbujani et al., 1998; Barbujani, Bertorelle, 2001; Chikhi et al., 2002]. Точку в ней поставила работа [Richards et al., 2000], обобщившая все имевшиеся к тому времени данные о полиморфизме мтДНК в населении Европы и методологически корректно подошедшая к проблеме соотношения возраста популяции и возраста распространенных в ней гаплогрупп. В этой работе было показано, что большинство субгаплогрупп мтДНК, свойственных населению Европы и отсутствующих на Ближнем Востоке, имеют палеолитический возраст. Это противоречит гипотезе демической диффузии; ведь если гипотеза справедлива, то современный генофонд Европы имеет преимущественно ближневосточное происхождение, и тогда «старые» гаплогруппы, принесенные земледельцами с Ближнего Востока, должны быть общими между Европой и Ближним Востоком, а гаплогруппы, возникшие уже в Европе и потому ограниченные Европой, должны быть «молодыми».

В то же время, обе конкурирующие гипотезы основываются на реконструкции генофонда былых эпох по генетической структуре их современных популяций-потомков. Поэтому очевидно, что данные по древней ДНК (полученные в надежных лабораториях и получившие мировое признание) имеют большое, если не решающее, значение для выбора между этими двумя гипотезами, поскольку

дают прямую информацию о генофонде древних популяций.

МТДНК ПОПУЛЯЦИЙ НЕОЛИТИЧЕСКИХ ЗЕМЛЕДЕЛЬЦЕВ ЕВРОПЫ

Примерно в то же время, когда на севере Восточной Европы существовал мезолитический памятник Южного Оленьего острова, рассмотренный в предыдущем разделе, Центральная Европа уже была охвачена процессом неолитизации. На территории Центральной и отчасти Южной Европы (Балканы) возникла и распространилась первая обширная неолитическая культура Европы – археологическая культура линейно-ленточной керамики.

Данные по поздним неолитическим культурам Иберийского полуострова выявили сходство неолитического населения с современным митохондриальным генофондом [Sampietro et al., 2007]. Но данные по ранней неолитической культуре линейно-ленточной керамики [Naak et al., 2005], напротив, выявили ее резкие отличия от современного генофонда Европы. Эти отличия наиболее ярко проявились в высокой частоте гаплогруппы **N1a** мтДНК, почти отсутствующей у современных европейцев. Специально проведенное байесовское моделирование показало, что такая элиминация гаплогруппы в современном населении Европы не могла произойти за счет дрейфа генов [Naak et al., 2005].

Эта работа 2005 года ответила на вопрос “что произошло с генофондом первых земледельцев после их появления в Европе?”. Она показала, что генофонд первых неолитических земледельцев Европы резко отличался и от предшествующего, и от современного населения Европы, но постепенно растворился в окружающем населении и почти не дошел до нашего времени.

Чтобы ответить на следующий вопрос “откуда появились первые земледельцы”, в нашем исследовании [Naak, Balanovsky et al., 2010] был проведен анализ дополнительных 17 образцов древней ДНК и сопоставление неолитической популяции с современными популяциями из нашей базы данных MURKA. Все новые проанализированные образцы принадлежали к популяциям культуры линейно-ленточной керамики. Более того, они происходили из одного археологического сайта (Деренбург), что снимает часто возникающие у археологов вопросы о «целостности» изучаемой генетикой древней выборки и принадлежности ее к одной популяции и одной археологической культуре. Вместе с ранее проанализированными образцами, общий объем выборки для Деренбурга составил $N=22$, а суммарный объем выборки по культуре линейно-ленточной керамики (включая Деренбург и другие популяции, относившиеся к той

Таблица 8.2. Доля общих гаплотипов между популяциями неолита и современности [Haak, Balanovsky et al., 2010].

Современные популяции	Объем выборки	№ гаплотипов	№ информативных гаплотипов	№ образцов	№ информативных образцов	Доля информативных образцов
Ближний Восток 3 (Иран)	500	15	4	83	6	7%
Ближний Восток 2 (Сирия, Ирак, Палестина, Кипр)	500	13	3	88	4	5%
Центральная Европа 5 (чехи, словаки)	500	13	3	144	5	3%
Балканы 1 (хорваты, словенцы)	500	13	2	148	5	3%
Ближний Восток 1 (турки, курды, армяне)	500	12	3	97	3	3%
Кавказ	500	13	2	97	3	3%
Британские острова 1 (центр и север Англии)	500	16	5	164	5	3%
Восточная Европа 3 (южные русские)	500	14	3	130	3	2%
Балканы 2 (венгры, румыны)	500	14	3	138	3	2%
Британские острова 4 (шотландцы)	500	12	1	162	3	2%
Балканы 4 (греки)	500	11	2	109	2	2%
Центральная Европа 1 (австрийцы, швейцарцы, южные немцы)	500	14	3	164	3	2%
Северная Европа 1 (шведы, норвежцы)	500	13	2	180	3	2%
Восточная Европа 4 (северные русские)	500	13	2	121	2	2%
Центральная Европа 2 (немцы Нижней Саксонии)	500	11	1	137	2	1%
Прибалтика (литовцы, латыши, эстонцы)	500	11	2	149	2	1%
Центральная Европа 6 (поляки)	500	13	2	158	2	1%
Волго-Уральский регион 2 (башкиры, татары, чувашаи)	500	11	1	114	1	1%
Балканы 3 (сербы, албанцы, болгары, боснийцы, македонцы)	500	9	1	128	1	1%
Британские острова 6 (Оркнейские, северные)	500	10	1	132	1	1%
Британские острова 2 (запад и восток Англии)	500	12	1	145	1	1%
Восточная Европа 2 (белорусы)	499	12	1	148	1	1%
Южная Европа 4 (итальянцы)	500	12	1	154	1	1%
Южная Европа 1 (португальцы)	500	11	1	163	1	1%
Южная Европа 5 (сицилийцы, сардинцы, корсиканцы)	500	12	1	166	1	1%
Британские острова 3 (валлийцы, ирландцы)	500	11	0	199	0	0%
Центральная Европа 3 (немцы прочие)	500	11	0	166	0	0%
Восточная Европа 1 (украинцы)	500	11	0	136	0	0%
Центральная Европа 4 (северные французы)	500	11	0	159	0	0%
Южная Европа 2 (баски, северные испанцы)	500	10	0	169	0	0%
Южная Европа 3 (прочие испанцы)	500	10	0	153	0	0%
Британские острова 5 (Шетландские о-ва)	499	9	0	142	0	0%
Северная Африка	498	9	0	103	0	0%
Волго-Уральский регион 1 (коми, мари, мордва, удмурты)	500	9	0	169	0	0%
Восточная Европа 5 (финны, карелы)	500	9	0	136	0	0%
Северная Европа 2 (исландцы)	500	9	0	146	0	0%

Примечание: Современные популяции расположены по степени убывания генетической близости с неолитическим населением Европы.

же культуре) составил $N=42$. Отметим, что по меркам анализа древней ДНК эти объемы выборок велики и позволяют проводить полноценный популяционный анализ.

АНАЛИЗ ОБЩИХ ГАПЛОТИПОВ

Анализ на уровне гаплотипов мтДНК (совпадения гаплотипов между неолитической популяцией и широким кругом современных популяций) проведен точно так же, как и аналогичный анализ для мезолитической популяции в предыдущем разделе. Среди 42 образцов неолитической популяции были выявлены 25 различных гаплотипов. Из них 11 оказались широко распространенными среди почти всех изученных современных популяций, 10 имели ограниченное географическое распространение в современном населении, а оставшиеся 4 были уникальны для древней популяции (отсутствовали среди включенных в анализ современных

популяций). Одиннадцать «повсеместных» гаплотипов представляли собой преимущественно гаплотипы-основатели для частых западно-евразийских гаплогрупп **H**, **HV**, **V**, **K**, **T**, **W**. В силу своей повсеместности они малоинформативны для выявления того, какие из современных популяций в наибольшей степени сходны с неолитическим генофондом.

Поэтому более информативны те десять гаплотипов, которые обнаружены в древней популяции и приурочены лишь к некоторым их современных популяций. Из результатов анализа общих гаплотипов, представленных в табл. 8.2, следует, что семь современных популяций характеризуются наибольшим сходством с неолитической популяцией. Это следующие популяции (условно перечислим входящие в их состав отдельные народы, страны или регионы): Иран, Сирия, Ирак, Палестина, Кипр, чехи, словаки, хорваты, словенцы, турки, курды, армяне, Кавказ, Англия. Из этих попу-

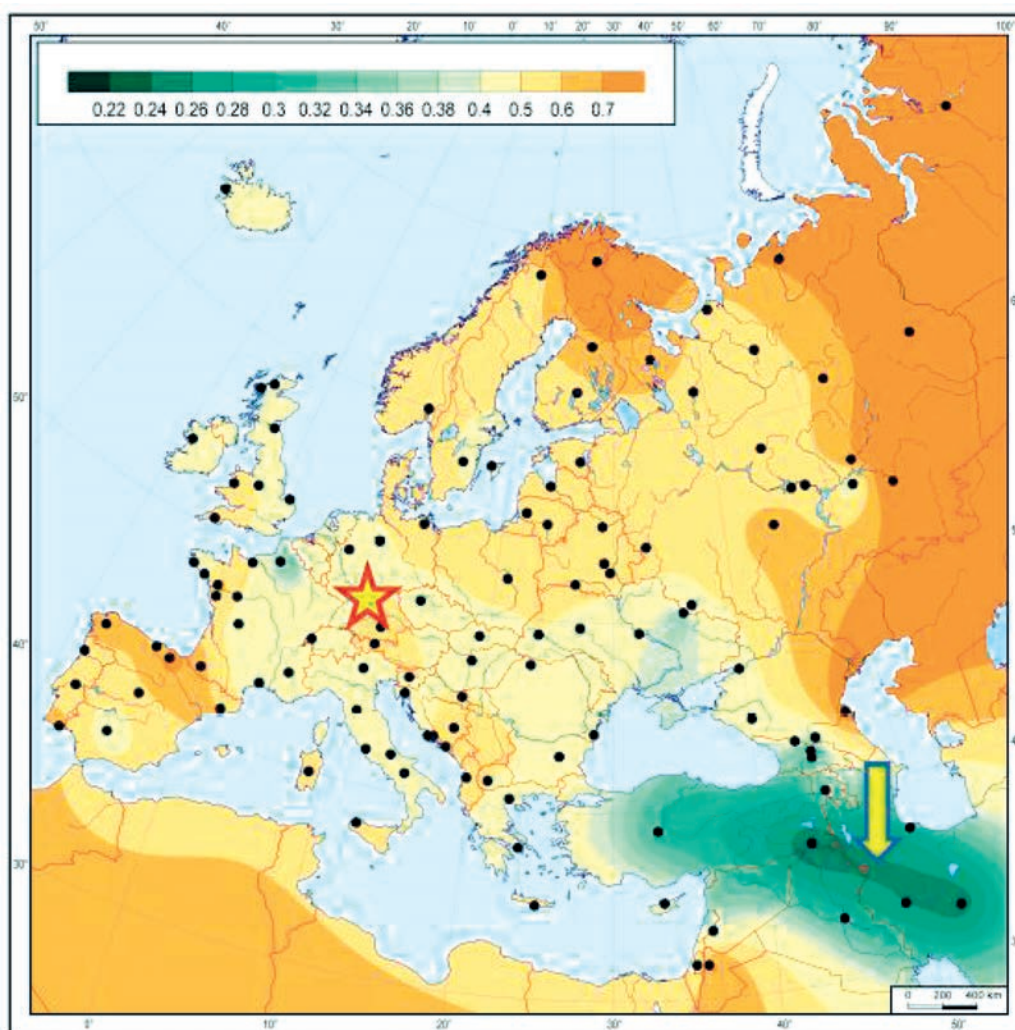


Рис. 8.16. Карта генетических расстояний от неолитической популяции Центральной Европы до генофондов 118 современных популяций Европы и Азии [Naak, Balanovsky et al., 2010].

Звездочкой отмечено положение изученной древней популяции; черными точками – положение генетически изученных современных популяций; стрелкой – зона наибольшего сходства древней популяции с современным генофондом.

ляций часть (чехи, словаки, хорваты, словенцы) расположены в предполагаемом географическом центре культуры линейно-ленточной керамики в Центральной Европе и, вероятно, именно они частично сохранили генофонд эпохи неолита. Однако все прочие популяции (за исключением Англии) происходят с Ближнего Востока (или зон его мощного влияния), то есть из региона, откуда, по археологическим данным, неолитическая волна заселения пришла в Европу.

КАРТА ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ

Объективное представление о степени близости современных популяций к неолитическому генофонду дают геногеографические карты. При анализе на уровне гаплогрупп мтДНК были рассчитаны и картографированы генетические расстояния (рис. 8.16) от популяции европейского неолита (ее географическое положение показано звездочкой) до современных популяций (их географи-

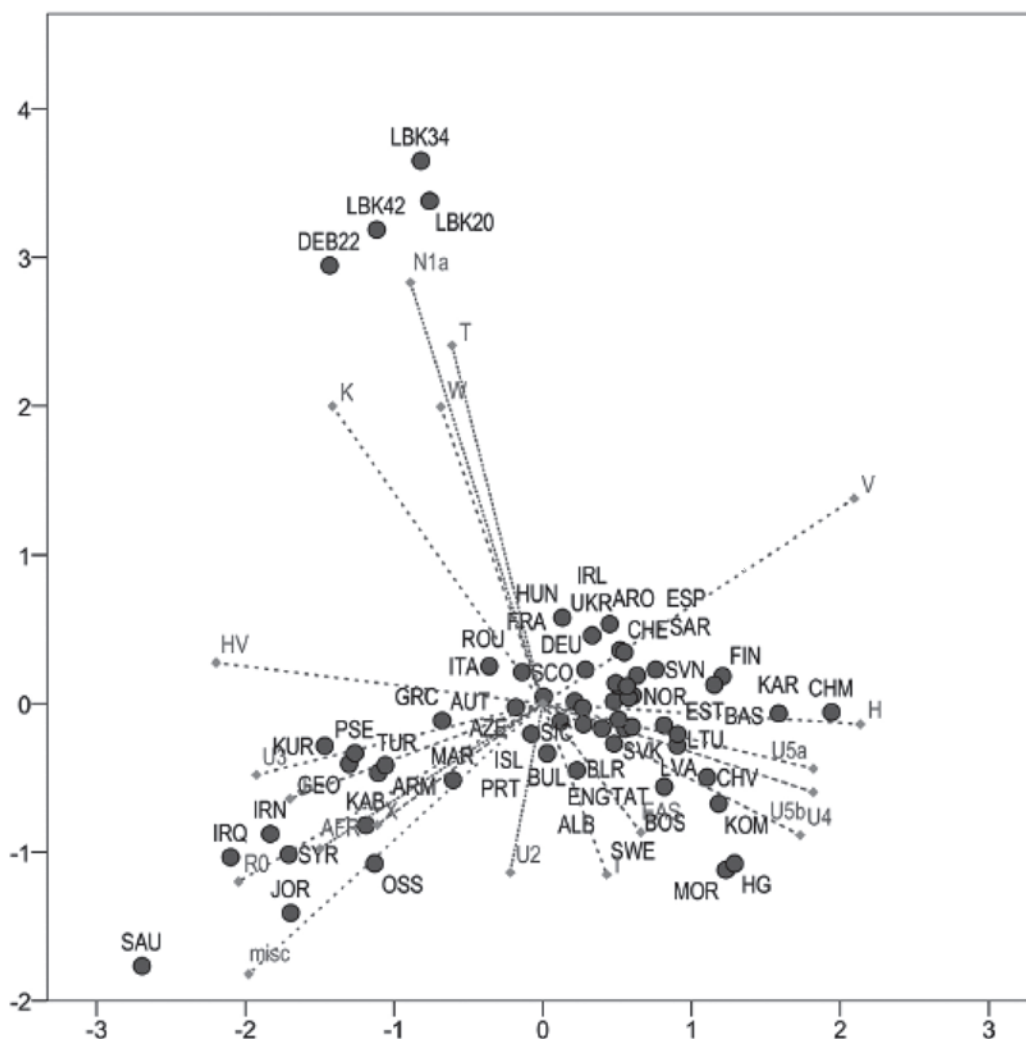


Рис. 8.17. Положение генофонда неолитических популяций относительно современного населения Европы и Ближнего Востока [Naak, Balanovsky et al., 2010].

Древние популяции: DEB22 – популяция «Деренбург» культуры линейно-ленточной керамики; LBK20 – популяции культуры линейно-ленточной керамики, кроме Деренбурга; LBK42 – все популяции культуры линейно-ленточной керамики; LBK34 – все популяции культуры линейно-ленточной керамики после исключения образцов возможных родственников; HG – мезолитические охотники-собиратели.

Современные популяции: ALB – албанцы, ARM – армяне, ARO – аромунцы, AUT – австрийцы, AZE – азербайджанцы, BAS – баски, BLR – белорусы, BOS – боснийцы, BUL – болгары, CHE – швейцарцы, CHM – марийцы, CHV – чуваша, CRO – хорваты, CZE – чехи, DEU – немцы, ENG – англичане, ESP – испанцы, EST – эстонцы, FIN – финны, FRA – французы, GEO – грузины, GRC – греки, HUN – венгры, IRL – ирландцы, IRN – иранцы, IRQ – иракцы, ISL – исландцы, ITA – итальянцы, JOR – иорданцы, KAB – кабардинцы, KAR – карелы, KOM – коми, KUR – курды, LTU – литовцы, LVA – латыши, MAR – марокканцы, MOR – мордва, NOR – норвежцы, OSS – осетины, POL – поляки, PRT – португальцы, PSE – палестинцы, ROU – румыны, SAR – сардинцы, SAU – арабы Саудовской Аравии, SCO – шотландцы, SIC – сицилийцы, SVK – словаки, SVN – словенцы, SWE – шведы, SYR – сирийцы, TAT – татары, TUR – турки, UKR – украинцы.

ческое положение показано черными точками). Шкала приведена в верхней части карты. Наименьшие генетические расстояния (т.е. наибольшее генетическое сходство) окрашены в зеленые тона низменностей. По мере роста величин генетических различий между генофондами современных популяций и неолитической, по аналогии с картой физической географии, тона становятся сначала песочных оттенков, а затем красноватых тонов возвышенностей. Карта (рис. 8.16) четко показывает, что наименьшие расстояния (со значениями от 0,22 до 0,38) локализируются в северной части Ближнего Востока (северная Месопотамия, Закавказье, восточная Анатолия). Хотя генофонд Ближнего Востока, конечно же, мог с неолита несколько измениться, все же все имеющиеся данные указывают, что именно эта зона была наиболее вероятным источником миграции для неолитического населения культуры линейно-ленточной керамики Европы.

АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТ

Еще одним независимым методом анализа на уровне частот гаплогрупп является анализ главных компонент. На графике (рис. 8.17) показаны генетические взаимоотношения неолитической популяции (представленной в виде четырех точек, несколько различающихся способами расчета частот гаплогрупп), 55 современных популяций Европы и Ближнего Востока, а также мезолитической популяции охотников-собирателей. На графике первая главная компонента (горизонтальная ось), объясняющая 25% общей изменчивости, разделяет ближневосточные и европейские популяции, при этом неолитическая популяция присоединяется к популяциям Ближнего Востока и Закавказья.

Однако вторая главная компонента (вертикальная ось, 18% общей дисперсии) четко отделяет

неолитическую популяцию от всех современных.

На графике обозначен и вклад каждой из гаплогрупп (пунктирные линии). Сходство неолитической популяции с Ближним Востоком по первой главной компоненте определяется повышенными частотами гаплогрупп мтДНК **HV**, **J**, **U3** в неолитической популяции, тогда как ее своеобразие (по второй компоненте) определяется, главным образом, гаплогруппами **T**, **N1a**, **K** и **W**.

ОБЩИЙ ВЫВОД

Таким образом, все три вида независимого анализа – и общих гаплотипов, и главных компонент, и в особенности карта генетических расстояний – указывают, что генофонд одной из первых в Европе неолитических популяций (культуры линейно-ленточной керамики) наиболее сходен с генофондом популяций Ближнего Востока.

Выше указывалось, что более раннее исследование палеоДНК [Haak et al., 2005] показало, что современный генофонд Европы значительно отличается от неолитического. Это означало, что первое земледельческое население Европы не оставило многочисленных потомков, что согласуется с гипотезой «культурной диффузии» земледелия. Новые данные, представленные в этом разделе, позволяют уточнить этот вывод: они показали, что эта неолитическая популяция генетически близка к популяциям Ближнего Востока.

В результате проведенного исследования древней ДНК теперь можно утверждать, что хотя *распространение* земледелия внутри Европы носило характер «культурной диффузии», но *появление* земледелия в Европе связано с далекой миграцией первых земледельцев. Этот вывод стал своеобразным синтезом «демической» и «культурной» гипотез распространения земледелия по Европе.

ТРИ ИСТОЧНИКА ГЕНОФОНДА ЕВРОПЕЙЦЕВ

ГЕНОФОНДЫ ОХОТНИКОВ-СОБИРАТЕЛЕЙ И ЗЕМЛЕДЕЛЬЦЕВ ПО ПОЛНОГЕНОМНЫМ ДАННЫМ

«Полногеномный» ответ на вечный вопрос: «Из каких предковых популяций сформировался генофонд Европы в неолите?» – был получен в исследовании большого международного коллектива под руководством Дэвида Райха (David Reich) и Иоханнеса Краузе (Johannes Krause) с участием многих соавторов, включая и меня самого. Результаты этой работы опубликованы в журнале Nature [Lazaridis et al., 2014]. Это исследование основано не на митохондриальном геноме неолитического населения (на изучении которого основано исследование, описанное в предыдущем подразделе) и который в современном населении Европы, как уже не раз

описано в этой книге, крайне гомогенен. Новое исследование сделано уже на новом витке развития генетических технологий и использует много больший объем данных о геноме – результаты полногеномного секвенирования, включающего, кроме аутосомного генома, также анализ митохондриальной ДНК и Y- хромосомы.

Были секвенированы девять древних геномов из трех популяций (рис. 8.18). Названия образцов, как всегда, даются по названиям археологических сайтов, из которых извлечен изученный палеоантропологический материал.

1) «Образец Штуттгарт» - от скелета женщины из Центральной Европы с датировкой 7 тыс. лет назад, найденного близ Штуттгарта, Германия, который относят к популяции первых европейских

земледельцев культуры линейно-ленточной керамики.

2) «Образец Лошбор» - от скелета мужчины из Центральной Европы с датировкой 8 тыс. лет назад, найденный в горах Лошбор в Люксембурге, принадлежащий к одной из археологических культур охотников-собирателей и относимый к мезолиту.

3) «Образцы Мотала12» – от семи скелетов из Северной Европы возрастом 8 тыс. лет из захоронения в северной части ареала охотников-собирателей, найденного в городе Мотала в Швеции и относимом к мезолиту.

В работе также использовали и результаты исследования некоторых других древних геномов, секвенированных и опубликованных ранее, например, верхнепалеолитический образец со стоянки Мальта близ озера Байкал (МА-1, с датировкой около 24 тыс. лет) и мезолитический «образец Ла Бранья» из Иберии (с датировкой около 8 тыс. лет).

Хотя главные выводы исследования основаны на изучении аутосомных маркеров, у всех образцов проанализированы также гаплогруппы митохондриальной ДНК, а у принадлежащих мужчинам – гаплогруппы Y-хромосомы. Выяснилось, что «земледелец» из Штуттгарта обладал гаплогруппой T2 мтДНК, типичной для неолитических европейцев, а «охотники-собиратели» из Лошбор и все семь образцов из Мотала принадлежат к гаплогруппам U5 или U2, типичным для митохондриального генофонда мезолитических охотников-собирателей Европы. Что касается Y-хромосомы, то все шесть образцов мужского пола из двух популяций (образец из Лошбор и пять образцов из Мотала) обладали гаплогруппой I, так что, по-видимому, эта гаплогруппа была характерна для Y-хро-

мосомного генофонда мезолитических охотников-собирателей Европы.

При исследовании аутосомного генома показано, что уровень средней гетерозиготности «земледелец» из Штуттгарта располагается у верхнего предела гетерозиготности современных европейцев, а у «охотника-собирателя» из Лошбора гетерозиготность ниже, чем у всех существующих европейских популяций. Это позволило предположить, что предки «охотника-собирателя» из Лошбора прошли через узкое популяционное «бутылочное горлышко» (резкое снижение генетического разнообразия внутри популяции). Такие события могли быть нередкими в малочисленных популяциях охотников-собирателей. Геном «земледелец» из Штуттгарта отличался от геномов «охотников-собирателей» большим числом копий гена амилазы слюны (AMY1), что связывают с высоким содержанием крахмала в рационе – признак перехода к сельскому хозяйству.

Удалось реконструировать некоторые черты внешности древних европейцев по вариантам аутосомных генов. Жители Центральной Европы – и «земледелец» из Штуттгарта, и «охотник-собиратель» из Лошбора – с высокой вероятностью (>99%) были темноволосыми. При этом «охотники-собиратели» и Южной Европы (Ла Бранья), и Центральной (Лошбор) и Северной (Мотала12) с вероятностью >75% имели голубые или иные светлые глаза, а «земледелец» из Штуттгарта с вероятностью >99% – карие глаза. Представитель первых земледельцев, так же, как и современные европейцы, обладал аллелем светлой кожи в гене SLC24A5 в гомозиготном состоянии, в отличие от охотников-собирателей. В то же время, у «охотни-

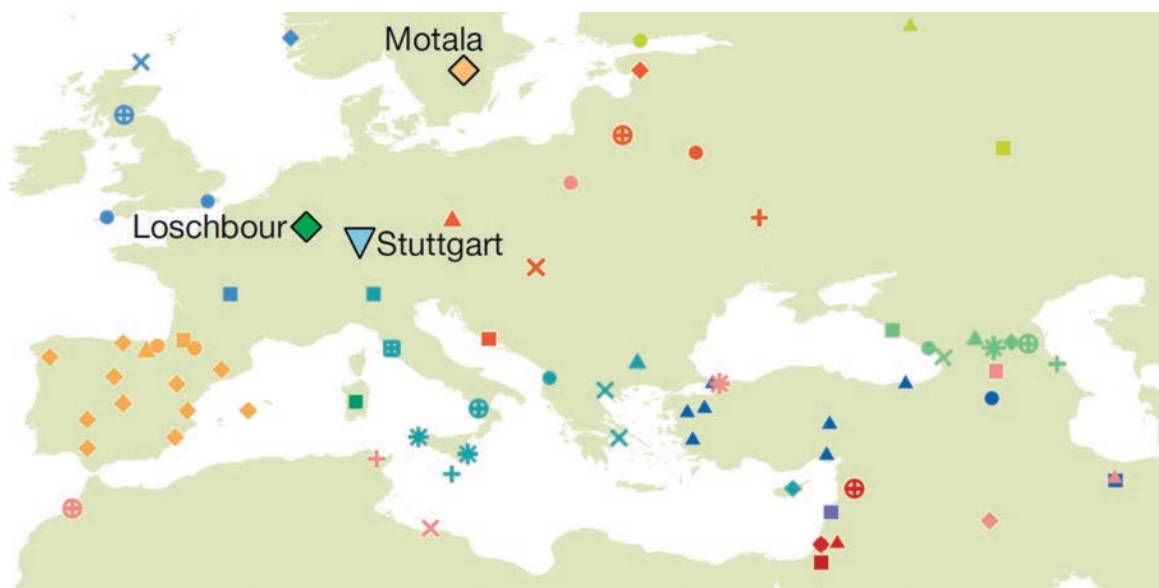


Рис. 8.18. Карта изученных западноевразийских геномов [Lazaridis et al., 2014].

Географическое местоположение изученных современных образцов обозначено мелкими цветными значками. Местоположение древних образцов Loschbour, Stuttgart, Motala обозначено крупными значками и подписано.

ка-собирателя» Мотала12 присутствовала, по крайней мере, одна копия этого аллеля, а значит, он был в наличии у европейцев уже в мезолите, до появления земледелия.

ТРИ ПРЕДКОВЫХ ПОПУЛЯЦИИ

Древнюю ДНК исследователи сравнили с ДНК более чем двух тысяч ныне живущих людей из 203 популяций. Для анализа использовали генотипирование по панели 500 тысяч однонуклеотидных полиморфизмов (SNP) с помощью панели Human

Origins (охарактеризована в главе 4). Для выделения кластеров популяций использовали программу ADMIXTURE и метод анализа главных компонент.

Эти результаты анализа главных компонент мы уже рассматривали в главе 4 (рис. 4.3), но тогда сосредоточились лишь на генетических взаимоотношениях современных популяций, теперь же обратим внимание в первую очередь на положение древних образцов.

На графике главных компонент все геномы сгруппировались в четыре кластера (рис. 8.19):

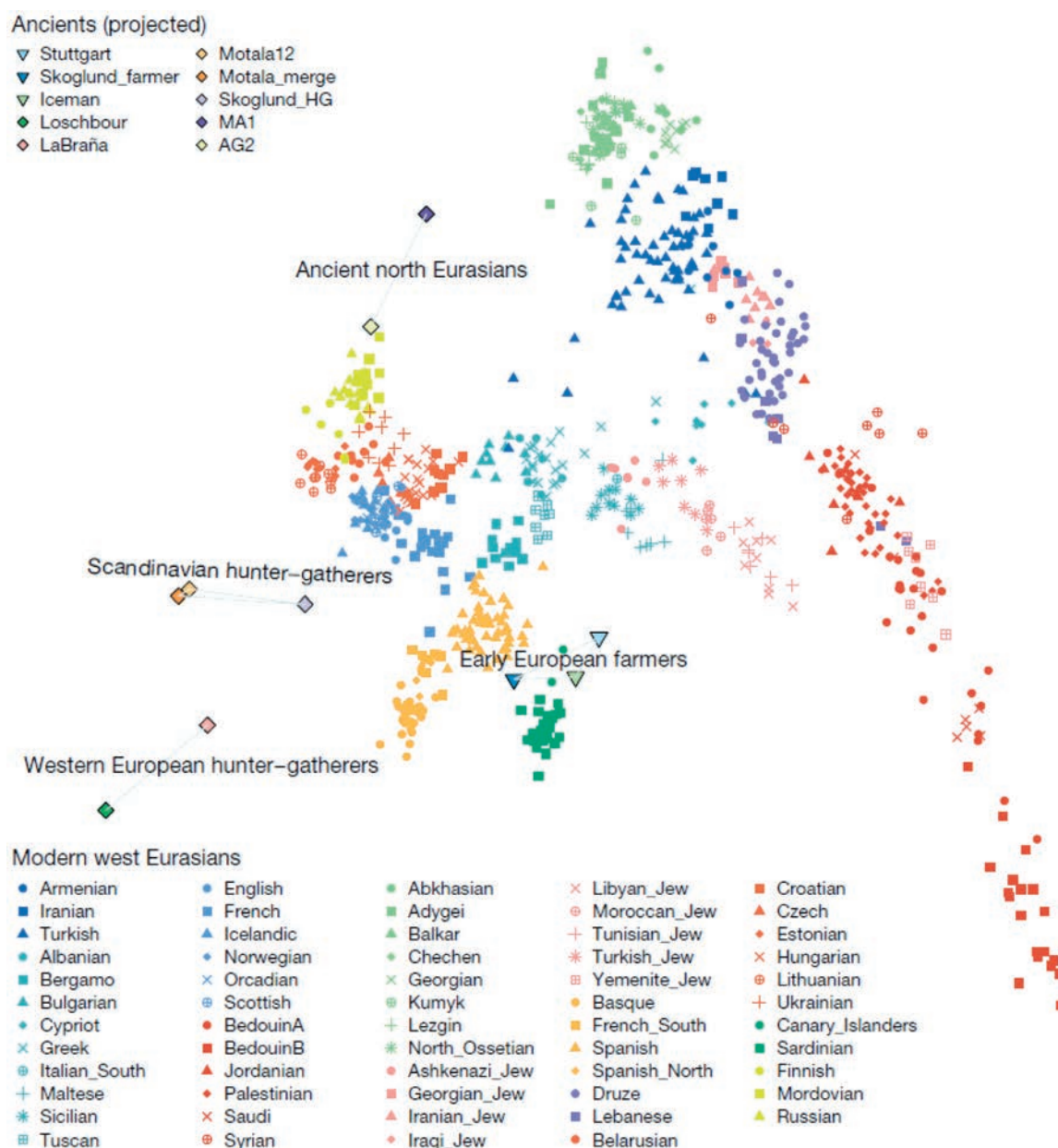


Рис. 8.19. Структура мезолитического, неолитического и современного генофонда Европы - анализ главных компонент по широкогеномной панели [Lazaridis et al., 2014].

Каждая точка соответствует одному образцу современной или древней ДНК. Тип и цвет значка каждой популяции указаны в легенде карты. *Древние образцы:* Stuttgart (Штуттгарт), Skoglund farmer (Скогланд, земледельцы), Iceman (тирольский «ледяной человек»), Loschbour (Лощбор), LaBvana (Ла Бранья), Motala (Мотала), Skoglund_HG (Скогланд, охотники-собиратели), MA1 (образец с верхнепалеолитической стоянки Мальта), AG2 (образец с верхнепалеолитической стоянки Афонтова гора).

– Западноевропейские охотники-собиратели (WHG – Western Hunter-Gatherers), к которым относятся европейские охотники-собиратели из Испании, Люксембурга и Швеции (в том числе Лошбор и Ла Бранья), а также скандинавские охотники-собиратели (в том числе Мотала).

– Охотники-собиратели верхнего палеолита из Сибири, в том числе индивидуум со стоянки Мальта (МА-1), предположительно входящие в метапопуляцию древних северных евразийцев (ANE – Ancient North Eurasians).

– Земледельцы Ближнего Востока (сюда вошли геномы современного населения и, предполо-

жительно, сюда же относились геномы древних жителей Ближнего Востока).

– Посередине между этими тремя находятся современные европейцы, составляющие четвертый кластер, в который также вошли ранние европейские земледельцы (EEF); к нему относятся земледельцы из Штуттгарта, тирольский «ледяной человек» (датировка – около 5300 лет) и шведские земледельцы с датировкой около 5 тыс. лет.

Европейские охотники-собиратели не только находятся вне кластера современных европейцев, но и смещены относительно него в сторону большего удаления от Ближнего Востока. Земледельцы из

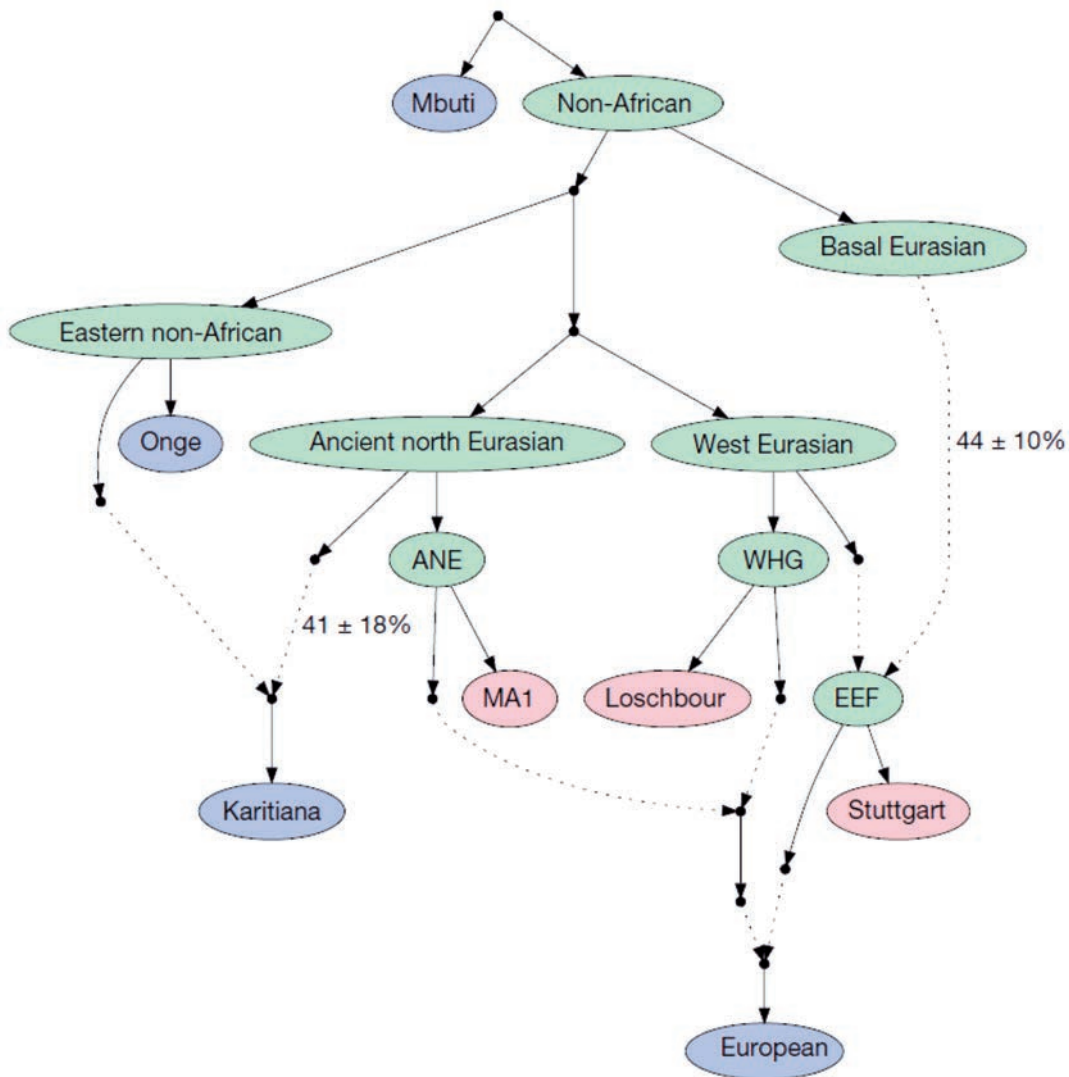


Рис. 8.20. Филогенетическое дерево популяций западной Евразии [Lazaridis et al., 2014].

Обозначения: Современные популяции обозначены фиолетовым цветом, изученные в данной работе древние популяции – розовым, статистически реконструированные древние популяции – зеленым. Сплошные линии обозначают переход от предковой популяции без генетического смешения, пунктирные линии обозначают смешение. Модель смешения популяций из трех источников удовлетворяет имеющимся данным для большинства популяций.

Изученные образцы: Mbuti (пигмеи), Karitiana (индейцы каритиана), Onge (этническая группа в Индии), ANE (древние северные евразийцы), WHG (западно-евразийские охотники-собиратели), EEF (ранние европейские земледельцы), Basal Eurasian (базальная евразийская линия), MA1 (образец со стоянки Мальта), Louschbour (Лошбор), Stuttgart (Штуттгарт).

Штуттгарта попадает в кластер с другими неолитическими европейцами. Образец с сибирской палеолитической стоянки Мальта (МА-1, около 24 тыс. лет) отклоняется от современных западных евразийцев в третьем направлении, далеко и от ближневосточных, и от западноевропейских популяций.

Были рассмотрены две гипотезы-модели о популяциях, послуживших источниками формирования генофонда Европы. В соответствии с первой моделью, популяций было две: западноевропейские охотники-собиратели (WHG) и ранние европейские земледельцы (EEF). В соответствии со второй гипотезой – три популяции сформировали генофонд современных европейцев: к первым двум добавляются еще и древние северные евразийцы (ANE). Гены ранних европейских земледельцев у современных европейцев составляют от 30% в балтийском регионе до 90% в средиземноморском. Гены древних северных евразийцев предположительно присутствуют почти у всех современных европейцев (составляя максимум 20%).

Основной вывод этого исследования – трехкомпонентное происхождение генофонда Европы – был признан столь важным, что попал на обложку Nature. Впервые было показано, что генофонд современного населения Европы сформировался в результате смешения не двух, а трех групп населения:

- 1) первоначального населения Европы, сохранившегося еще с древнекаменного века (WHG);
- 2) мигрантов с Ближнего Востока, принесших в Европу технологии земледелия (EEF);
- 3) населения Северной Евразии, в том числе Европейской части России и Сибири (ANE).

ЗАГАДКА «БАЗАЛЬНЫХ» ЕВРАЗИЙЦЕВ

Для детализации вопроса о путях сложения генофонда, был применен и третий метод – построено особое древо родства древних популяций (рис. 8.20), для которого, в отличие от традиционных методов построения деревьев, ветвям разрешили не только разделяться, но и сливаться, что лучше соответствует причудливым реалиям популяционной истории человечества. Большая часть дерева отражала уже известные события древнейшей истории. Так, от вышедшего из Африки чело-

вечества отделилась ветвь восточных евразийцев, которые внесли свой генетический вклад в формирование коренного населения Америки (представленные индейцами каритиана). В формирование генофонда коренного населения Америки внесли большой вклад и древние северные евразийцы (ANE), популяция которых широко расселилась на восток ойкумены (как мы выше рассматривали в подразделе по Мальте). Древние западные евразийцы внесли вклад как в генофонд западноевропейских охотников-собирателей (Лошбор), так и ранних европейских земледельцев (Штуттгарт).

Но в этой уже известной картине было две новости. Первая – выявленная и всеми предыдущими методами – важный вклад населения Северной Евразии (ANE) в формирование не только генофонда населения самого севера Европы, но всего европейского генофонда. Вторая новость обнаружилась на смоделированном дереве – на нем возникла и совершенно новая, неизвестная до сих пор ветвь – это «базальная линия евразийцев» (basal Eurasian на рис. 8.20). Эта популяция выделилась уже после выхода из Африки, но до разделения на западных евразийцев (европеоидов) и восточных евразийцев (монголоидов) – возможно, в период раннего заселения Ближнего Востока. Что и говорить, популяция загадочная, и ей сложно найти соответствия в известной картине древних групп человечества, но ее существование в прошлом упрямо вылезало из статистического анализа. В соответствии с реконструированной историей потомки этой (пока что виртуальной) популяции составили почти половину (44±10%) генофонда ранних европейских земледельцев.

В заключение отметим, что в окончательную модель происхождения европейского генофонда хорошо вписывается большинство современных европейских популяций за двумя исключениями. Первое исключение составляют сицилийцы, жители средиземноморского острова Мальта (не путать с верхнепалеолитическим образцом со стоянки Мальта из Сибири) и евреи ашкенази, которые имеют большую долю ближневосточных предков, чем остальные. Второе исключение – финны, мордва и северо-западные русские, в генофонде которых велик вклад «восточноазиатских» генов.

НЕОЛИТИЗАЦИЯ В ПИРЕНЕЯХ

ГЕНЕТИЧЕСКИЕ ОСОБЕННОСТИ ИБЕРИЙСКОГО ЭНЕОЛИТА

Секвенирование полных геномов древних образцов с Пиренейского полуострова [Günther et al., 2015] добавило важный элемент в мозаику формирования генофонда европейцев и, по-видимому, стало ключом к разгадке происхождения народа басков, которое уже давно интригует археологов и лингвистов.

Шведские и испанские исследователи секвенировали полные геномы восьми останков людей из пещеры Эль Порталон в Сьерра-де-Атапуэрка в Испании (рис. 8.21). Пиренейский полуостров известен как место длительного обитания человека с палеолита до исторических времен. Останки, из которых была извлечена ДНК (четыре мужских и четыре женских индивида) относятся к энеолиту (медно-каменному веку – конец эпохи неолита, с

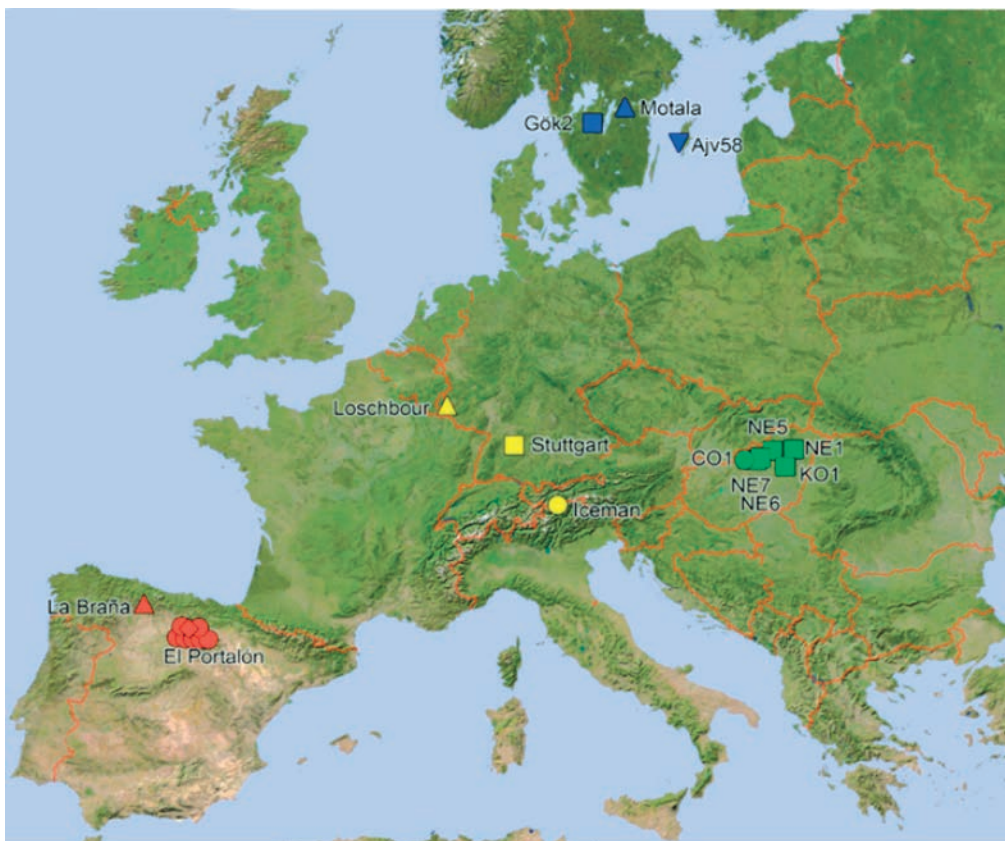


Рис. 8.21. Местоположение анализируемых образцов древней ДНК [Günthera et al., 2015].

Красным цветом обозначены образцы с Иберийского полуострова: треугольник – мезолит (Ла-Бранья), кружки – энеолит (медно-каменный век), ромбик – бронзовый век (Эль Порталон). Желтым цветом обозначены образцы из Центральной Европы: треугольник – мезолит (Лосшбор), кружок – энеолит (медно-каменный век, Тирольский ледяной человек), квадрат – неолит (Штуттгарт). Зеленым цветом обозначены образцы из Венгрии: кружок – энеолит (медно-каменный век), квадраты – неолит. Голубым цветом обозначены образцы из Скандинавии: треугольник вершиной вниз – поздние охотники-собиратели, треугольники вершиной вверх – мезолит, квадрат – неолит.

датировкой около 5500 лет назад) и бронзовому веку (с датировкой около 3500 лет назад). Датировки всех древних образцов, анализируемых в этой работе приведены на шкале рис. 8.22. Из восьми индивидов, изученных в Пиренеях, семь были представлены фрагментарными останками, но один был почти полным скелетом мальчика. Ядерную ДНК из этих образцов секвенировали с покрытием от 0,01x до 4,08x.

В ДНК восьми древних обитателей Пиренеев определили гаплогруппы митохондриальной ДНК (мтДНК) и Y-хромосомы. Материнские линии митохондриальной ДНК оказались характерными для ранних европейских земледельцев (**К**, **Ж** и **Х**), охотников-собирателей (**U5**) или и тех, и других (**Н**). Отцовские линии Y-хромосомы принадлежали к гаплогруппам **H2** и **I2a2a**. Таким образом, эти генетические маркеры демонстрируют смешение генофондов ранних европейских земледельцев, пришедших на Пиренейский полуостров, с местными охотниками-собирающими, причем с участием обоих полов.

По данным секвенирования полных геномов обитателей пещеры Эль Порталон в Атапуэрке,

другим древним геномам из разных частей Европы и геномам современных европейцев провели анализ главных компонент (рис. 8.23).

На графике видно, что в пространстве 1 и 2 главных компонент генофонды охотников-собирающих, так же, как и ранних земледельцев, четко группируются независимо от их географической локализации. Причем наиболее близко к охотникам-собирающим и Южной Европы (Пиренейского полуострова), и Центральной Европы, и Северной (Скандинавия) находятся современные народы Северной Европы. А к ранним земледельцам и Южной Европы, и Центральной, и Северной, напротив, более близки современные южные европейцы.

Образцы из Атапуэрки, несмотря на то, что они относятся к энеолиту (переходная эпоха от финала неолита к бронзе), по результатам разных видов анализа несут значительный генетический след охотников-собирающих. В этом они сходны со скандинавскими неолитическими земледельцами и отличаются от земледельцев Центральной Европы (образцы NE1 и Штуттгарт). Это показывает, что когда ранние земледельцы распространялись по

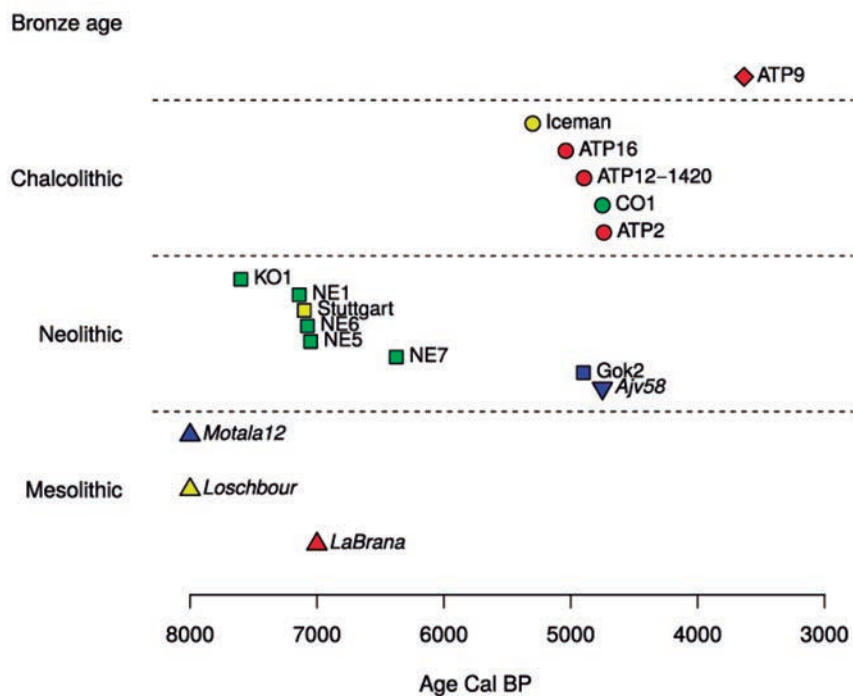


Рис. 8.22. Датировка анализируемых образцов древней ДНК [Günthera et al., 2015].

Красным цветом обозначены образцы с Иберийского полуострова: треугольник – мезолит (Ла-Бранья), кружки – энеолит (медно-каменный век), ромбик – бронзовый век (Эль Порталон). Желтым цветом обозначены образцы из Центральной Европы: треугольник – мезолит (Лошбор), кружок – энеолит (медно-каменный век, Тирольский ледяной человек), квадрат – неолит (Штуттгарт). Зеленым цветом обозначены образцы из Венгрии: кружок – энеолит (медно-каменный век), квадраты – неолит. Голубым цветом обозначены образцы из Скандинавии: треугольник вершиной вниз – поздние охотники-собиратели, треугольники вершиной вверх – мезолит, квадрат – неолит.

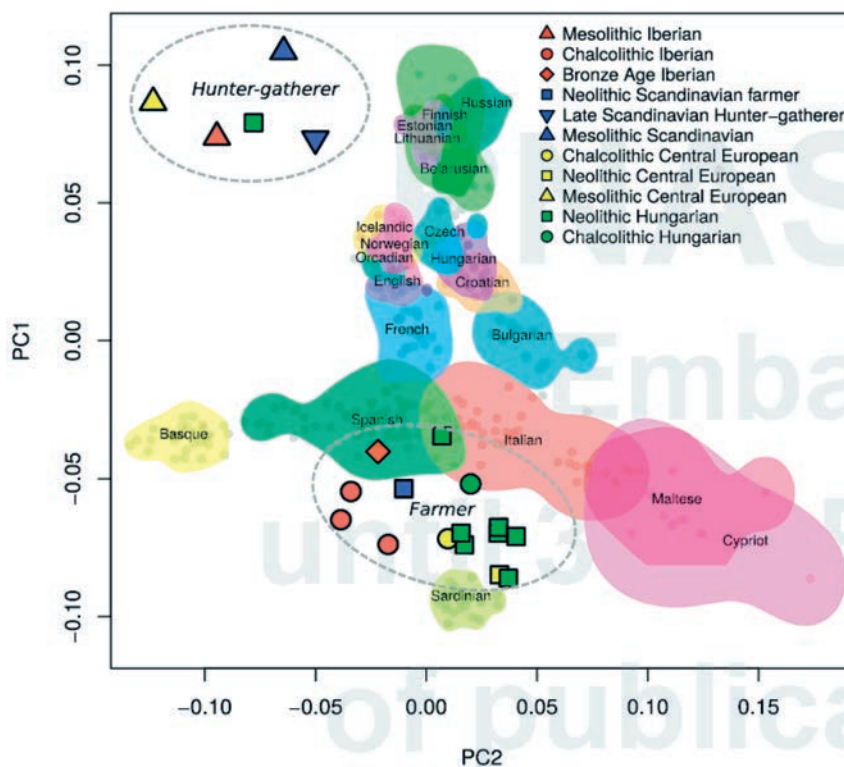


Рис. 8.23. Генетические взаимоотношение неолитической популяции Иберийского полуострова с другими древними и современными популяциями (анализ главных компонент по полногеномным данным) [Günthera et al., 2015].

Пунктирные линии обозначают группы охотников-собирателей и земледельцев. Обозначения образцов указаны в легенде. Цветным фоном залиты представители каждого изученного этноса современного населения Европы.

Европе и ассимилировали местные популяции охотников-собирателей, то в разных регионах доля ассимилированного местного генетического компонента резко различалась.

АНАЛИЗ ПРЕДКОВЫХ КОМПОНЕНТОВ ПОДТВЕРДИЛ ИЗОЛЯЦИЮ

Этот вывод был подтвержден и анализом методов ADMIXTURE. В анализ включены 16 древних популяций и множество современных – из Евразии и Северной Африки (рис. 8.24). У всех современных популяций Пиренейского полуострова выявились два основных предковых компонента – это европейские земледельцы (розовый цвет) и европейские охотники-собиратели (голубой).

Удивительным оказалось то, что баски (как предполагается, говорящие на древнем языке населения «до-индоевропейской» эпохи) в этом отношении не отличаются от остальных популяций полуострова. Отличаются они в другом. У всех популяций, кроме басков и популяций Сардинии, имеются еще два предковых компонента – один маркирует современные популяции Кавказа/Центральной Азии (черный), а другой – современные

популяции Северной Африки (желтый). А у басков и у населения Сардинии их нет (рис. 8.24). Эти две южноевропейские популяции в наиболее полном виде сохранили генетическое наследие ранних европейских земледельцев и местных охотников-собирателей. Кроме того, именно баски оказались генетически ближе всего к людям из пещеры Эль Порталон в Атапуэрке. Эти новые данные значительно «омолодили» генофонд басков, показав, что немалая его часть получена от популяций ранних земледельцев. При этом изоляция басков подтверждается – отсутствием кавказско-азиатского и африканского предковых компонентов, присутствующих у большинства народов Европы. Видимо, к популяциям басков не докатились волны миграций, как времен Римской империи, так и вторжений мусульман из Северной Африки, начавшихся в 711 году.

Происхождение же языка басков пока так и остается неясным. Можно только предположить, что их язык сохранился либо от первых земледельцев (если они говорили не на индоевропейских языках), либо от еще более древних охотников-собирателей и избежал влияния распространившихся по Европе индоевропейских языков.

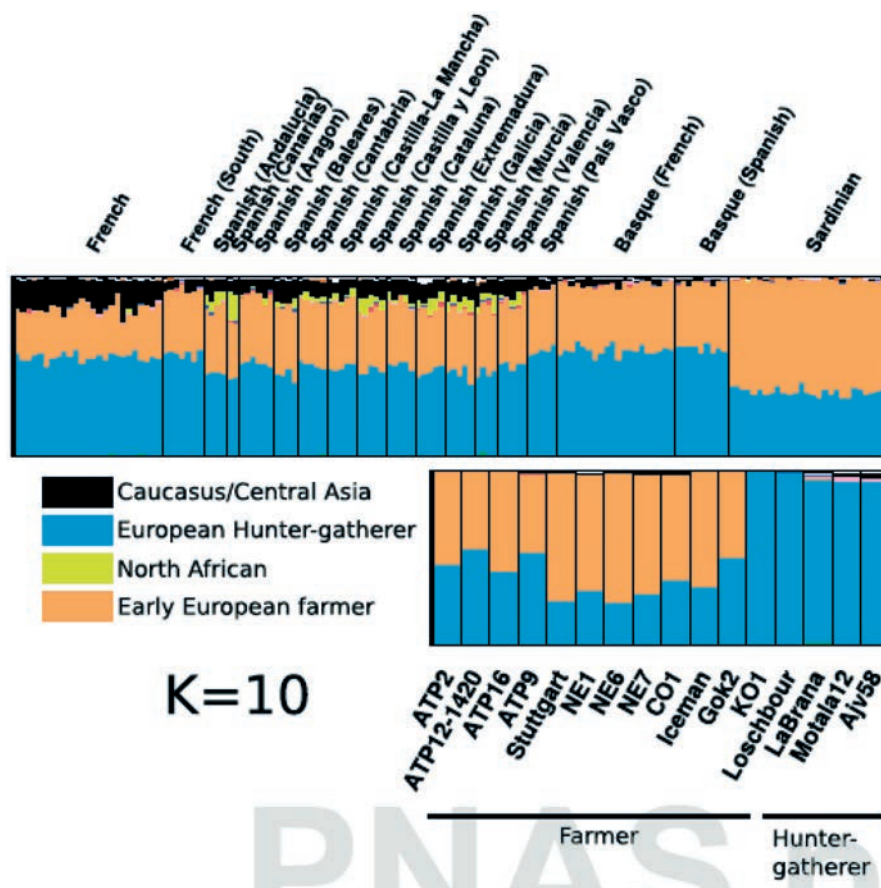


Рис. 8.24. Результаты анализа полных геномов методом ADMIXTURE (показаны только популяции Юго-западной Европы) [Günthera et al., 2015].

Предковые компоненты обозначены на цветной панели слева: Кавказ/Центральная Азия (черный), европейские охотники-собиратели (голубой), Северная Африка (желтый), ранние европейские земледельцы (розовый). Общее число предковых компонентов, заложенных в данный вариант модели: $k=10$.

8.5. ЭПОХА МЕТАЛЛА

РАННИЙ МЕТАЛЛ НА СЕВЕРЕ ЕВРОПЫ

Археологический памятник Могильник Большого Оленьего острова – с названием, столь похожим на знаменитый мезолитический Южный Олений Остров, – расположен примерно в том же географическом регионе, но не так широко известен. Он датируется примерно 3,5 тыс. лет назад (табл. 1.7) и относится к эпохе раннего металла (следующей после неолита). Географически он расположен на севере Кольского полуострова, на небольшом острове рядом с берегом.

Анализ антропологических данных из этого могильника, рассматриваемых с учетом нового материала, полученного в раскопках последнего десятилетия, указывает промежуточный антропологический облик населения, оставившего могильник Большого Оленьего острова, между европеоидными и монголоидными популяциями [Моисеев, Хартанович, 2011].

СПЕКТР ГАПЛОГРУПП мтДНК

Благодаря хорошей сохранности материала, для Большого Оленьего острова удалось для 23 образцов провести анализ палеоДНК, то есть в нашем распоряжении оказалась выборка в два раза большая, чем для мезолитического памятника Южного Оленьего острова. Обнаруженные гаплогруппы мтДНК, выстроенные по порядку убывания их ча-

стоты, рисуют своеобразный генетический профиль пост-неолитического населения Большого Оленьего острова: **C, U5, D, Z, U4, T**. Как и в населении мезолитического Южного Оленьего острова, в этом перечне мирно соседствуют типичные европейские и сибирские варианты. Но в генофонде населения эпохи раннего металла именно сибирским гаплогруппам (**C, D, Z**) принадлежат лидирующие места, причем появляются (не встреченные в мезолите) восточноевразийские гаплогруппы **D** и **Z**. Все это заставляет сделать вывод, что в популяции, оставившей могильник на Большом Оленьем острове, сибирский генетический компонент преобладал над европейским. А это заставляет предполагать миграционную волну с востока – причем, вероятно, не из ближайших районов за Уралом, а из более отдаленных внутренних частей Сибири. Отметим, что ни в одной из современных популяций Европы (за исключением тундровых ненцев, являющихся совсем недавними пришельцами из Сибири) нет столь высоких частот сибирских гаплогрупп.

КАРТА ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ

Картографированные генетические расстояния были рассчитаны по частотам 31 гаплогруппы мтДНК: **A, B, C, D, F, G, H, HV, I, J, K, L, N1a, N1b,**

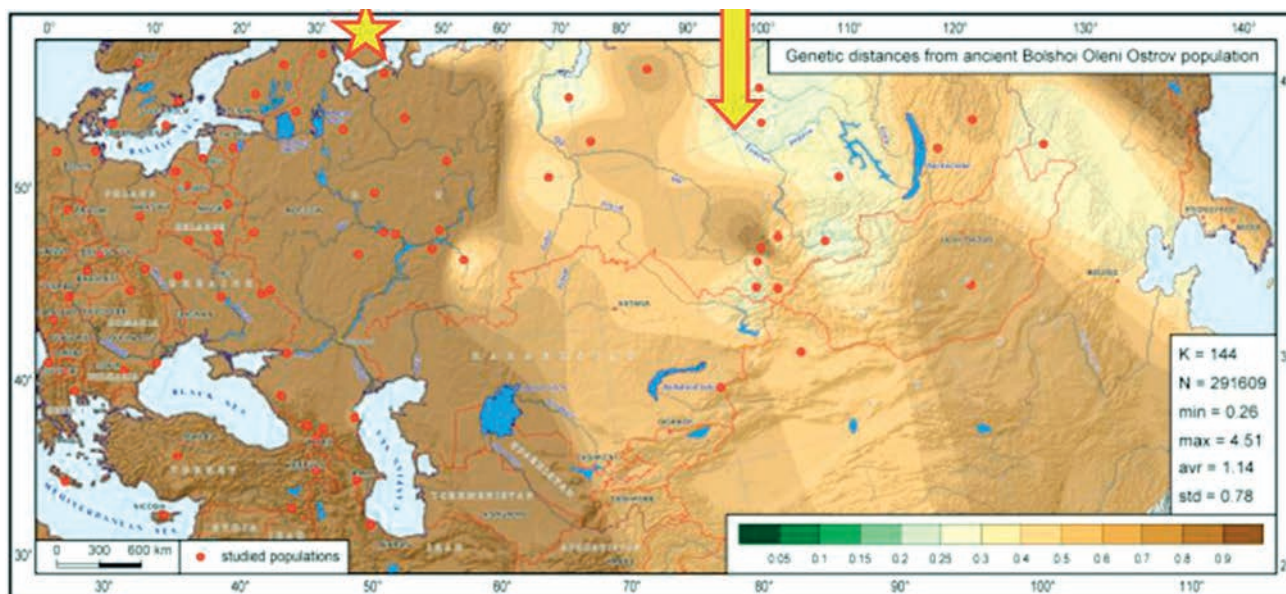


Рис. 8.25. Карта генетических расстояний от популяции эпохи раннего металла Большого Оленьего Острова (Кольский п-ов) до генофондов 144 современных популяций Евразии.

Звездочкой – отмечено положение изученной древней популяции, красными точками – положение генетически изученных современных популяций, стрелкой – зона наибольшего сходства древней популяции с современным генофондом.

R, T, U1, U2, U3, U4, U5a, U5b, U6, U7, U8, V, W, X, Y, Z, other (сумма прочих редких гаплогрупп). Карта на рис. 8.25 демонстрирует, в какой степени сходным с древней популяцией является современное население разных территорий Евразии. Шкала генетических расстояний приведена внизу карты и читается как обычно: наименьшие обнаруженные расстояния (наибольшая близость генофондов) на этой карте отражены как светлые области песочных тонов, а генетически далекие генофонды – коричневыми тонами гор и возвышенностей.

Популяции, генетически сходные с оленеостровцами Кольского полуострова (светлые тона карты) отсутствуют в Европе, но характерны для современного населения Сибири. Этот ареал популяций, близких к генофонду древнего населения Кольского полуострова, начинается от бассейна Оби, захватывает не только Среднюю, но и Восточную Сибирь, а на юге простирается до Казахстана, Монголии и северо-запада Китая. Наиболее генетически близкая к древним пост-неолитическим оленеостровцам географическая область, представленная не одной, а рядом современных популяций, находится в сердце этого ареала – в бассейне Енисея, в Южной Сибири и Прибайкалье.

В целом карта генетических расстояний показывает, что современное население отнюдь не Восточной Европы, а многих частей Сибири, и главным образом бассейна Енисея, проявляет наибольшее генетическое сходство с древним генофондом Большого Оленьего острова. Это позволяет выдвигать гипотезу о Средней Сибири как о возможной прародине населения Кольского полуострова 3500 лет назад. При более осторожной интерпретации карты можно утверждать, что соотношение западноевразийского и восточноевразийского генофондов в населении Кольского полуострова 3500 лет назад было примерно таким, как и в современном населении Средней и Южной Сибири. То есть (в сравнении с современным населением Восточной Европы, где доминируют западноевразийские гаплогруппы) это соотношение еще более, чем для мезолитического населения (Южный Олений Остров), было сдвинуто в пользу восточного генетического ствола Северной Евразии.

АНАЛИЗ ОБЩИХ ГАПЛОТИПОВ

Похожие результаты дал и анализ общих гаплотипов. Этот анализ выполнен аналогично рассмотренному выше анализу для мезолитической и неолитической популяций, поэтому, не останавливаясь на технических подробностях, перечислим группы популяций, обладающих наибольшим сходством с оленеостровцами по спектру гаплотипов. Это популяции Северной, Западной, Южной Сибири, Центральной Азии, а также Северной Европы (норвежцы). Преобладание в этом списке си-

бирских популяций (при том, что в общем списке включенных в анализ популяций резко преобладают европейские) указывает, что и на уровне гаплотипов митохондриальный генофонд Большого Оленьего Острова тяготеет к Сибири.

СРАВНЕНИЕ С ГЕНОФОНДОМ ЭПОХИ НОВОГО ВРЕМЕНИ

Для более точного прослеживания динамики генофонда во времени была изучена еще одна древняя популяция, происходящая с той же территории – с Кольского полуострова. Но эта популяция (Чальмны-Варрэ) была уже далеко не столь древней и датируется Новым временем (около XVIII века, табл. 1.7. в 1 главе). По антропологическим данным можно полагать, что эта популяция представляет саамов. К тем же выводам привело и генетическое исследование. В этой популяции обнаружены только три гаплогруппы мтДНК, перечисляемые в порядке убывания частоты: **V, U5b, U5a**. Такой спектр гаплогрупп (среди которых почти полностью преобладают **U5b** и **V**) среди всех современных популяций мира встречается только у саамов. К тому же в этой популяции Нового времени с высокой частотой обнаружены гаплотипы мтДНК с так называемым «саамским» мотивом. Хорошая сохранность образцов позволила проанализировать обширную выборку (N=42), что снимает вопросы о статистической достоверности этих результатов.

Из этого можно сделать два вывода. Первый вывод относится к саамам и указывает на стабильность их генофонда: по крайней мере, на протяжении трех-четырёх столетий заметных изменений он не претерпел. Второй вывод относится уже к популяции Большого Оленьего Острова эпохи раннего металла, древнюю территорию которой сейчас занимают саамы: кардинальное различие их генофондов указывает, что популяция эпохи раннего металла не оставила следов в современном населении региона. Более подробный анализ на уровне отдельных гаплотипов показал, что даже в тех немногих случаях, когда одна и та же гаплогруппа встречена и в популяции Большого Оленьего Острова, и у современных популяций саамов, то конкретные гаплотипы не совпадают. Вырисовывается картина практически стопроцентного замещения генофонда (полной смены населения) на этой территории.

ОБЩИЙ ВЫВОД

Таким образом, популяция северной Европы эпохи раннего металла (Большой Олений Остров, 3500 лет назад), так же, как и мезолитический Южный Олений остров (7 500 лет назад), несет в своем генофонде четкие свидетельства сибирского влияния. Это выявляется и по спектру гаплог-

рупп мтДНК, и картой генетических расстояний, и при анализе совпадающих гаплотипов мтДНК. Однако, в отличие от мезолита (7500 лет назад), в этом более позднем населении восточноевропейские гаплогруппы преобладают: вероятным источником миграции была уже не Западная, а Средняя Сибирь.

Можно заключить, что популяция эпохи раннего металла (Большого Оленьего острова) сформировалась в результате миграции из Сибири (вероятно, вдоль арктического побережья по тундровой зоне). Впоследствии эта миграция из Сибири, вероятно, угасла, не оставив значимого следа в современном генофонде.

БРОНЗОВЫЙ ВЕК В ЕВРОПЕ

Полногеномное исследование большого числа образцов древней ДНК, недавно проведенное командой Дэвида Райха (David Reich), привело к открытию миграции в центральную Европу из причерноморских степей около 4,5 тысяч лет назад [Naak et al., 2015], связанную со степными популяциями ямной культуры. В контексте двух гипотез о появлении в Европе индоевропейских языков – анатолийской и степной (курганной), эти новые генетические данные хорошо согласуются именно со степной гипотезой. Неудивительно, что после появления этих генетических данных споры о происхождении индоевропейцев разгорелись с новой силой. Рассмотрим эти важные результаты подробнее.

БОЛЬШОЕ ЧИСЛО ДРЕВНИХ ОБРАЗЦОВ

В исследование вовлечены новые данные о геномах 69 древних европейцев, живших от 8000 до 3000 лет назад, изученных по полногеномной панели, включающей 400 тысяч однонуклеотидных полиморфизмов (SNP). А при добавлении к ним 25 секвенированных ранее образцов ДНК древних европейцев, изучаемая выборка достигла знаковой отметки – почти сто (94) древних геномов (рис. 8.26). Среди носителей изученной древней ДНК: 19 охотников-собирателей (>5500 лет до н.э.); 28 ранне-неолитических земледельцев (6000–4000 лет до н.э.); 11 средне-неолитических земледельцев (4000–3000 лет до н.э.), включая тирольского «ледяного человека»; 9 представителей позднего медного/раннего бронзового века (ямная культура 3300–2700 лет до н.э.); 15 представителей позднего неолита (2500–2200 до н.э.); 9 представителей раннего бронзового века (2200–1500 лет до н.э.), два – позднего бронзового века (1200–1100 лет до н.э.), и один – железного века (около 900 лет до н.э.). Образцы происходят из археологических сайтов (рис. 8.26) Германии (41), Испании (10), России (14), Швеции (12), Венгрии (15), Италии (1) и Люксембурга (1).

ПОТОК ГЕНОВ В ПОЗДНЕМ НЕОЛИТЕ

Сравнение всех образцов древней ДНК и 777 образцов из современных популяций Западной Европы традиционно проведено с помощью анализа

главных компонент (РСА). Важно подчеркнуть, что основные черты этого итогового графика (рис. 8.27) во многом вторят аналогичному графику из работы по трем источникам генофонда Европы (рис. 8.19), что подчеркивает устойчивость обнаруживаемых закономерностей. Генофонд современных европейцев (множество серых точек в центре графика) оказался расположен между генофондами Ближнего Востока (множество серых точек в правой части графика) и европейских охотников-собирателей. Первые европейские земледельцы смещены от большинства современных европейцев и совпадают лишь с современными жителями Сардинии. Это не удивительно, поскольку население Сардинии традиционно считается сохранившим свой генофонд со времени первого появления земледельцев в Европе.

Но благодаря большому массиву новым образцов древней ДНК на графике появились новые черты, которые обнаруживают доселе неизвестные детали генетической истории европейцев.

Во-первых, если до сих пор охотники-собиратели были представлены по их популяциям в Западной Европе, Северной Европе (Скандинавия) и в Сибири (стоянка Мальта), то теперь добавились охотники-собиратели из Восточной Европы (из Карелии и Самарской области). В полном согласии с их географическим расположением они и на графике расположились между североевропейскими и сибирскими охотниками-собирателями.

Во-вторых, новые изученные образцы ДНК среднего неолита несколько отличаются от ранее изученных популяций раннего неолита. Если сравнить на графике образцы раннего неолита (треугольники вершиной вниз) и среднего неолита (треугольники вершиной вверх), то видно, что с ходом времени неолитическая популяция Европы генетически вполне ощутимо сдвинулась в сторону генофонда европейских охотников-собирателей (ромбики слева), что указывает на продолжение ассимиляции автохтонного населения Европы и перехода его на неолитические технологии и отношения.

В-третьих, новые изученные образцы европейцев позднего неолита и бронзового века (кружки на графике) попали в тот же кластер, что и современные европейцы (серые точки) – довольно далеко отойдя от популяций раннего и среднего неолита.

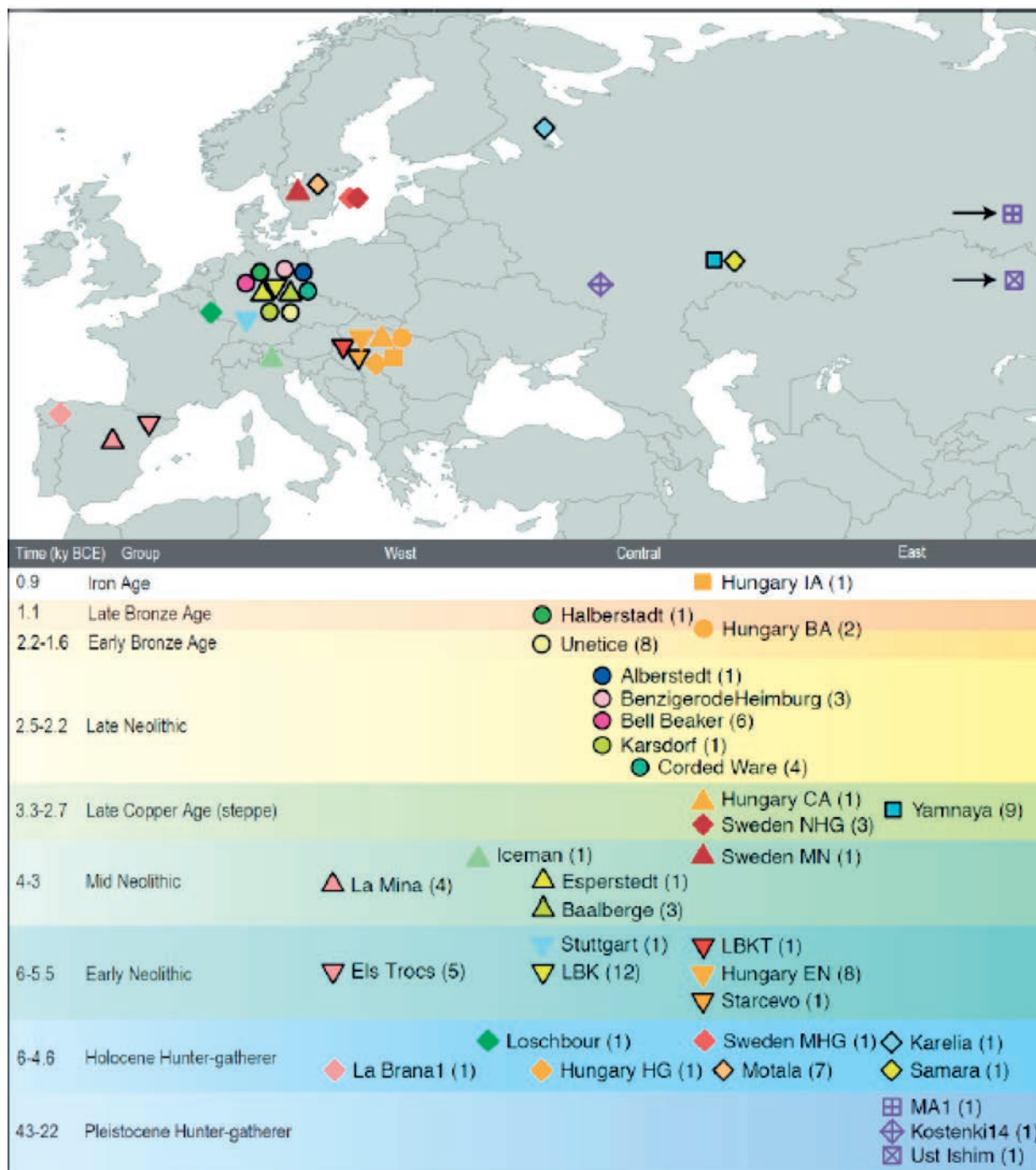


Рис. 8.26. Географическое местоположение находок образцов древней ДНК (69 новых и 25 из литературы) и их хронология от палеолита до железного века [Haak et al., 2015].

Стрелочками обозначено то, что местоположения образцов MA1 (Мальта) и Усть-Ишим находятся восточнее, за пределами карты. Iron Age – железный век, Late Bronze Age – поздний бронзовый век, Early Bronze Age – ранний бронзовый век, Late Neolithic – поздний неолит, Late Copper Age – поздний медный век, Mid Neolithic – средний неолит, Early Neolithic – ранний неолит, Holocene Hunter-gatherer – охотники-собиратели голоцена, Pleistocene Hunter-gatherer – охотники-собиратели плейстоцена.

лита. А значит, именно на этом отрезке времени – при переходе от среднего неолита к финальному – в генофонде Европы последний раз произошли существенные изменения (последний – потому что после этого он практически совпал с современным генофондом, а значит, более уже кардинально не менялся).

В-четвертых – что стало ключом к пониманию происходивших процессов – появились данные по степному населению Восточной Европы (ямная культура, голубые квадраты на графике). Их геномы расположились между охотниками-собираателями Восточной Европы и некоторыми популяциями из круга Юго-Западной Азии и смежных регионов

(в частности, Предкавказья). Напомню, что ареал археологической ямной культуры – «степняков-кочевников» – занимал всю южную часть Восточно-Европейской равнины от Приуралья до Подунавья с востока на запад и от Предкавказья до среднего Поволжья с юга на север, а датируется эпохой позднего неолита-ранней бронзы (период около 3000–

2200 гг. до н.э., то есть 5000–4200 лет назад). На основании новых генетических данных выдвинута гипотеза, что именно эта популяция ямной культуры могла стать той решающей «добавкой» к неолитическому генофонду Европы, которая изменила генофонд раннего и среднего неолита Европы на этапе его перехода к позднему неолиту/бронзе.

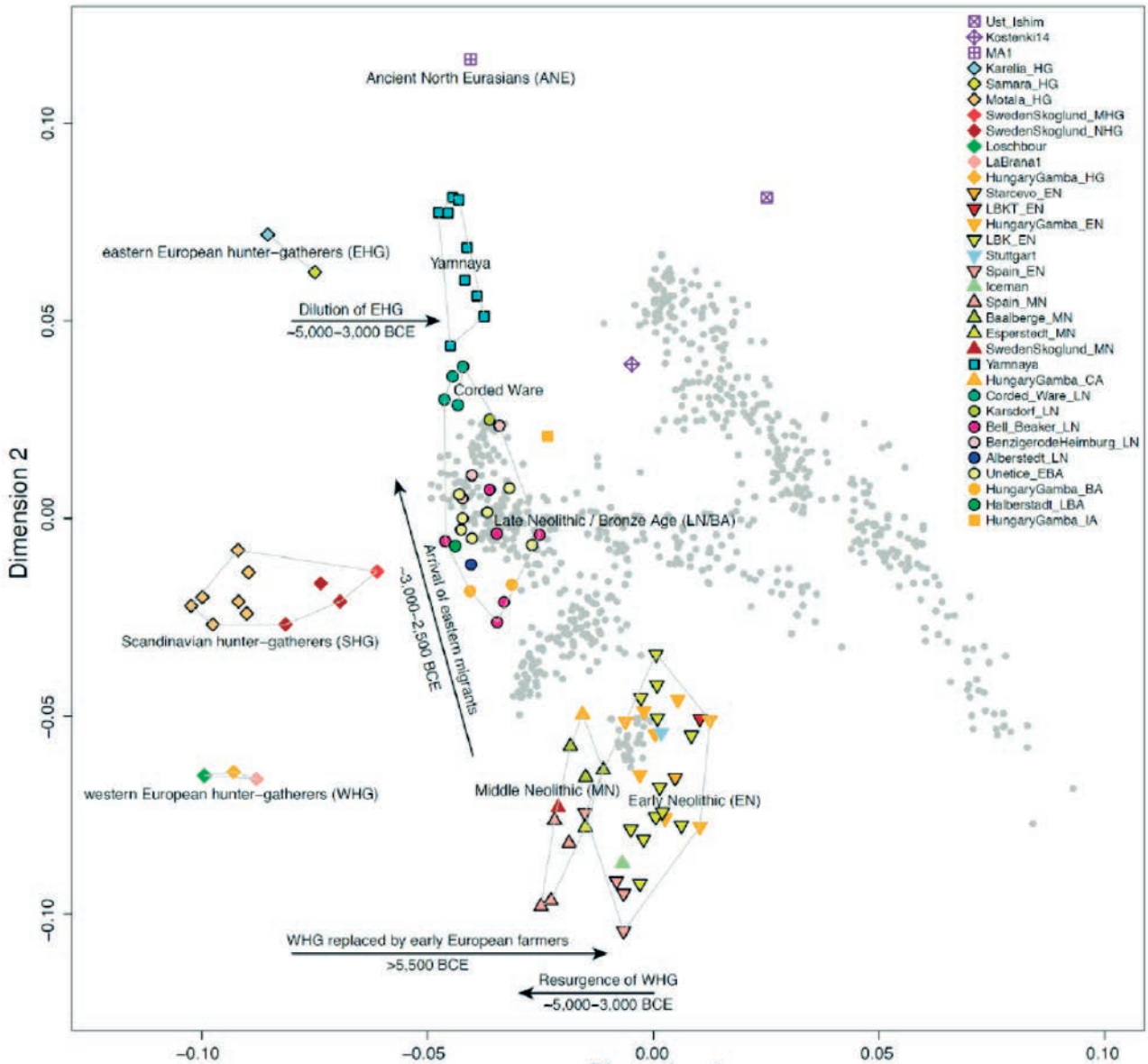
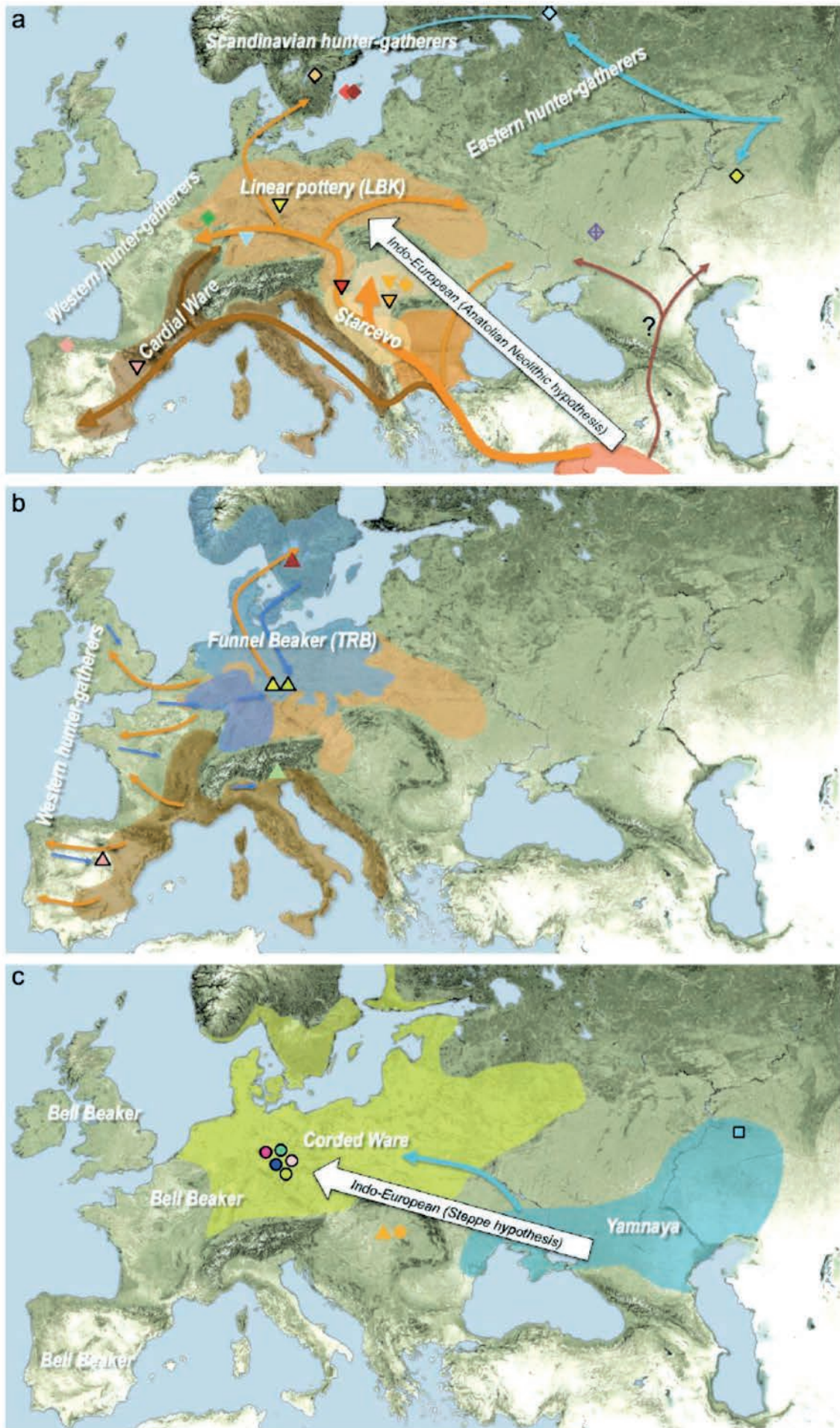


Рис. 8.27. Анализ главных компонент 94 древних и 777 современных представителей населения Европы [Haak et al., 2015].

Обозначения: Ancient North Eurasians (ANE) – древние северные евразийцы; Eastern European hunter-gatherers (EHG) – восточноевропейские охотники-собиратели; «Dilution of EHG» – «разбавление» генофонда восточноевропейских охотников-собирателей генофондом, родственным ближневосточному, и формирование генофонда ямной культуры; Corded Ware – культура шнуровой керамики; Late Neolithic/Bronze Age (LN/BA) – поздний неолит/бронзовый век; Scandinavian hunter-gatherers (SHG) – скандинавские охотники-собиратели; Western European hunter-gatherers (WHG) – западноевропейские охотники-собиратели; «Arrival of eastern migrants» – миграция из восточноевропейских степей; Middle Neolithic (MN) – средний неолит; Early Neolithic (EN) – ранний неолит; «WHG replaced by early European farmers» – западноевропейские охотники-собиратели сменились ранними европейскими земледельцами; Resurgence of WHG – возвратное усиление доли генетического компонента западноевропейских охотников-собирателей.



РЕКОНСТРУКЦИЯ СТЕПНОЙ МИГРАЦИИ

Реконструкция миграций групп населения и генетических потоков между ними показывает, что популяции Германии и Испании в среднем неолите имели на 18–34% больше генетического вклада западноевропейских охотников-собирателей (WHG), чем в раннем неолите. Это говорит о постепенном увеличении доли генетического компонента охотников-собирателей в популяциях, перешедших на неолитическое ведение хозяйства (рис. 8.27). Но в позднем неолите и начале бронзового века население этой же территории уже обнаруживает довольно резкий генетический сдвиг: на 22–39% больше сходства с генофондом популяций Восточной Европы (EHG-компонент), чем их предки в среднем неолите. Именно этот генетический сдвиг позволяет выдвинуть гипотезу о важном влиянии миграции с востока популяций «степняков», носителей ямной культуры (или генетически сходных с ней популяций).

Древнеямная культурно-историческая общность занимала все степные территории Восточной Европы, ограничиваясь Причерноморьем и Прикаспием на юге, Южным Приуральем на востоке, а на западе доходила до Подунавья и Балкан, заканчиваясь в Венгерских степях. Это кочевая и полукочевая культура скотоводов с элементами мотыжного земледелия. Они хоронили своих умерших в ямах под курганами (от чего культура и получила название), и что важно – у них были колесные повозки, запряженные волами (были и кони, но их использовали как мясо).

По представленным в статье результатам генетический вклад популяций ямной культуры в генофонд Центральной Европы оказался гораздо значительнее, чем предполагалось ранее. Изучение ДНК представителей более поздней культуры шнуровой керамики показало, что их генофонд на три четверти (73%) родственен генофонду ямной культуры. Поэтому была выдвинута логичная с точки зрения генетики гипотеза генетической преемственности от популяций ямной культуры к популяциям культуры шнуровой керамики.

Однако следует учесть, что не все так однозначно. Во-первых, культура шнуровой керамики (в на-

стоящее время период ее существования оценивается как 3000–2300 гг. до н.э., то есть 5000–4300 лет назад) практически синхронна ямной культуре (период около 5000–4200 лет назад, а при более строгих оценках – 4600–4200 лет назад). А во-вторых, ее ареал расположен к северу и северо-западу от ямной. Скорее речь может идти о двух синхронных смежных культурах, что позволяет думать не об их «вертикальной» по времени генетической преемственности, а об их «горизонтальном» генетическом родстве, то есть о том, что они восходят к некоему общему генетическому субстрату, поиск которого еще предстоит провести. Или же необходимо доказывать, что ямная культура существовала уже в IV тыс. до н.э., и выяснять генофонд и ареал этой ранней «протоямной» культуры.

В целом, благодаря этому исследованию добавлены три новых фрагмента в пазл общей картины формирования генофонда Европы. Основа этой картины уже была очерчена в работе [Lazaridis et al., 2014], выявившей три основных источника генофонда Европы (раздел 8.4): западноевропейские охотники-собиратели, древние северные евразийцы и ранние европейские земледельцы. Три новых фрагмента в этой картине: во-первых, равноправное включение генофонда охотников-собирателей Восточной Европы (EHG) в общую картину генофонда Европы; во-вторых – значительная роль генофонда популяции ямной культуры в формировании современного облика генофонда Европы; в-третьих – гипотеза о формировании генофонда популяций культуры шнуровой керамики в результате смешения популяций ямной культуры и потомков европейцев эпохи среднего неолита.

Полученная в результате картина сложения европейского генофонда основана на двух больших миграционных волнах, сыгравших ведущую роль в предыстории Европы. Первая – в раннем неолите – принесла на территорию Европы культуру и генофонд первых земледельцев с Ближнего Востока. Вторая – в позднем неолите – привела к появлению кочевников-скотоводов ямной культуры, пришедших из степей Восточной Европы (рис. 8.28). Причем, как показали генетические данные, за обеими миграциями последовало частичное воз-

Рис. 8.28. Географическое местоположение археологических культур и иллюстрация предположительных миграций [Naak et al., 2015].

а) вероятный путь миграции ранних земледельцев в Европу 9000–7000 лет назад; б) возвращение охотников-собирателей в течение среднего неолита 7000–5000 лет назад; в) появление степных мигрантов в центральной Европе в позднем неолите, около 4500 лет назад.

Белыми стрелками обозначены две альтернативные гипотезы появления в Европе индоевропейских языков:

а) анатолийская неолитическая гипотеза; б) степная гипотеза.

Обозначения: Linear pottery (LBK) – культура линейно-ленточной керамики; Cardial Ware – культура импрессо, Starcevo – Старчево-Кришская культура; Funnel Beaker (TRB) – культура воронковидных кубков; Bell Beaker – культура колоколовидных кубков; Corded Ware – культура шнуровой керамики; Yamnaya – ямная культура.

вращение генофонда к предыдущему облику вследствие ассимиляции пришельцами предшествующего населения Европы.

ОДНОРОДИТЕЛЬСКИЕ МАРКЕРЫ У ЯМНИКОВ

Описанные выше результаты основаны на анализе аутосомных маркеров из полногеномных панелей. Но получены и данные по одnorodительским маркерам – мтДНК и Y-хромосомы. Гаплогруппы Y-хромосомы определены для 34 древних образцов. Самыми распространенными, как и в наши дни, оказались гаплогруппы **R1a** и **R1b** – они встречены у 60% европейцев от позднего неолита до бронзового века. Причем среди образцов, найденных на территории России, все 100% принадлежали к носителям этих гаплогрупп, и на всем временном интервале – от 9700 до 4700 лет назад. Самые ранние изученные носители этих гаплогрупп – два охотника-собирателя с территории России из Карелии (**R1a**) и Самары (**R1b**). Характерно, что их Y-хромосомные варианты не принадлежали к линиям **R1a-M417** и **R1b-M269**, которые сегодня доминируют у европейцев. В то же время, более поздние образцы – все семь мужчин ямной культуры были носителями линии гаплогруппы **R1b-M269**.

В целом, данные по одnorodительским генетическим маркерам – Y-хромосоме и митохондриальной ДНК – не противоречат предложенной модели миграций. Считается, что Y-хромосомная гаплогруппа **G2a** и митохондриальная гаплогруппа **N1a**, обычные у ранних земледельцев Центральной Европы, почти исчезают в течение позднего неолита и бронзового века, когда они по большей части замещаются Y-хромосомными гаплогруппами **R1a** и **R1b** и мтДНК гаплогруппами **I**, **T1**, **U2**, **U4**, **U5a**, **W** и **H**. Это не только подтверждает вклад степных популяций в генофонд неолитической Европы, но также говорит о том, что в миграциях принимали участие оба пола. А по пути в Центральную Европу степные мигранты-кочевники, вероятно, смешивались с земледельцами Восточной Европы. В то же время, среди изученных мужчин ямной культуры все обладали только гаплогруппой **R1b** Y-хромосомы, а у мужчин культуры шнуровой керамики преобладала гаплогруппа **R1a**: это плохо согласуется с результатами по аутосомному геному о наследовании «шнуровиками» трех четвертей генофонда «ямников». Но и не дает повода отвергнуть такую гипотезу – поскольку пока изучено слишком мало образцов в каждой из выборок.

О СВЯЗИ «ЯМНИКОВ» И «ШНУРОВИКОВ»

В статье [Naak et al., 2015] упоминается некоторое сходство между обеими рассматриваемыми

археологическими культурами – ямной и шнуровой керамики. Такие параллели действительно упоминаются некоторыми археологами, но большинство из них все же подчеркивают, напротив, кардинальные и многочисленные различия между этими культурами, а не их немногочисленные сходные элементы. Поэтому гипотеза, в которой миграция из степей приравнивается к миграции ямной культуры, не нашла поддержки у многих археологов, в том числе Л.С. Клейна. Он подчеркивает, что представления об археологическом сходстве ямной культуры с культурами шнуровой керамики и боевого топора не выдерживают критики. Тем более, что генетики прослеживают лишь миграции населения, а культура далеко не всегда двигалась вместе с населением – у нее свои пути. Это дает генетикам право говорить о преемственности (или родстве) населения, но не о преемственности этих двух археологических культур.

На рис. 8.29 представлена модель вклада в генофонды древних и современных популяций Европы трех предковых компонентов – ранних неолитических популяций (Early Neolithic), западноевропейских охотников-собирателей (Western European Hunter-Gatherers) и представителей ямной культуры (Yamnaya). Она показывает, что в Северной Европе, как ни парадоксально, генетический вклад представителей ямной культуры гораздо выше, чем в южной. А в Южной Европе, в свою очередь, гораздо больше генетический вклад ранних неолитических популяций. Итак, за счет древних внутриевропейских миграций (которые пока еще практически не изучены) картина получается более сложной, чем простая модель двух миграций – «земледельческой» и «степной».

ИНДОЕВРОПЕЙЦЫ ЛИ?

Как известно, существует множество гипотез прародин и путей распространения индоевропейцев, в том числе две основные: анатолийская утверждает, что индоевропейцы расселились из Анатолии (Малой Азии) около 8500 лет назад, а курганная постулирует миграцию индоевропейцев из причерноморских степей. И та, и другая гипотеза отражены в статье [Naak et al., 2015] белыми стрелками на карте миграций (рис. 8.28), хотя изображены, на мой взгляд, чрезмерно схематично с точки зрения географии и археологии.

Дэвид Райх и его соавторы [Naak et al., 2015] считают, что полученные ими результаты вносят решающий вклад в дебаты о путях экспансии индоевропейских языков в Европу. В связи с этим мне вспомнилась одна из работ Майта Метспалу, посвященная сравнению митохондриальных генофондов Крита и Кипра. (Интересная, кстати, ситуация – два острова в Восточном Средиземноморье изначально имели сходное население, но потом их

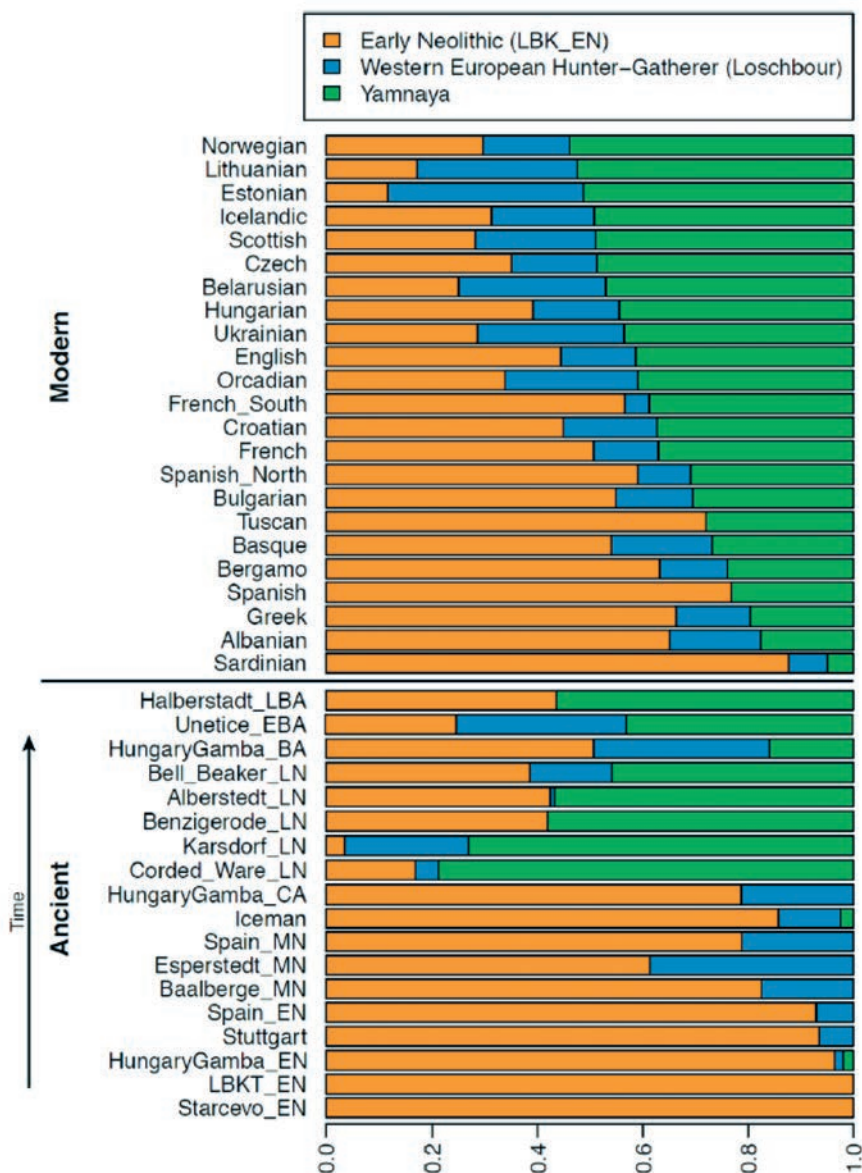


Рис. 8.29. Пропорция генетического вклада ранних неолитических популяций (Early Neolithic), западноевропейских охотников-собирателей (Western European hunter-gatherers) и представителей ямной культуры (Yamnaya) в генофонды различных современных и древних популяций [Haak et al., 2015].

генофонды разошлись). И один из выводов статьи был шутивным: «Ни один из результатов генетического исследования не противоречит широко известному утверждению, что Афродита родилась на Кипре». Вот и здесь – ни один из результатов генетического исследования не противоречит утверждению, что индоевропейцы расселились из степей.

Говоря серьезно, результаты статьи [Haak et al., 2015], конечно, к вопросам миграций индоевропейцев отношение имеют. Ведь хотя древняя ДНК не может ответить на вопрос, на каком языке говорили люди, она выявляет миграции самих этих людей. Разные гипотезы распространения индоевропейских языков предполагают определенные миграции, а генетика может их подтвердить, опровергнуть или же обнаружить новые миграции, не пред-

полагавшиеся ранее. И обнаруженная миграция, действительно, очень похожа на ту, которая должна была иметь место согласно степной гипотезе распространения индоевропейцев. Но похожесть, корреляция – не доказательство, а лишь дополнительный аргумент в пользу правдоподобности. Основной аргумент, который приводят сами авторы, – не «за степную гипотезу», а «против анатолийской». Аргумент состоит в том, что обнаружены массовые миграции *после* распространения земледелия в Европе (около 4500 лет назад популяция, генетически родственная носителям ямной культуры, переместилась в Центральную Европу). А анатолийская гипотеза во многом опиралась на то, что после неолитизации очень больших подвижек населения уже не было, а неолитизация-то явно шла через Анатолию.

Кроме того, авторы статьи [Haak et al., 2015] осторожны в своих лингвистических выводах. Они не утверждают, что все индоевропейские языки в Европе могли появиться в результате первых земледельческих миграций. И подчеркивают, что их данные ничего не говорят о *прародине* индоевропейских языков, а лишь о том, каким путем часть этих языков могла достигнуть Центральной Европы. А если прародину индоевропейцев допустить в восточной Анатолии, то с тем, что одна ветвь этих языков достигла северного Причерноморья и оттуда со степными кочевниками разошлась по большей части Евразии, не спорит, насколько я знаю, почти никто из сторонников анатолийской гипотезы. Так что большого спора гипотез — если ограничивать его событиями внутри Европы — возможно, и нет.

МИГРАНТЫ, ПОХОЖИЕ НА «ЯМНИКОВ»

Хоть похоже на Россию,
Только все же не Россия...
Александр Гордницкий

Размышляя о результатах этой ключевой работы [Haak et al., 2015], я выделил в ее внутренней логике два главных силлогизма:

1а) генофонд раннего неолита и среднего неолита Европы отличается от генофонда позднего неолита Европы;

1б) генофонд позднего неолита Европы практически совпадает с генофондом современной Европы.

Вывод 1: при переходе от среднего к позднему неолиту Европа (имеется в виду Западная и Центральная Европа) приняла какую-то массовую миграцию.

С этим выводом трудно поспорить: из четырех факторов микроэволюции ни дрейф, ни отбор, ни тем более мутации так подействовать не могли — остается миграция! Но далее следует второй силлогизм:

2а) была массовая миграция при переходе к позднему неолиту;

2б) генофонд носителей ямной культуры хорошо подходит на роль мигрантов (если добавить к генофонду среднего неолита Европы генофонд «ямников», то получится как раз генофонд позднего неолита Европы).

Вывод 2: «ямники» и мигрировали.

А вот этот вывод, строго говоря, логически не обоснован, потому что вместо «ямников» могли мигрировать и носители других археологических культур — генетически похожие на них. Правда, точное следование аристотелевской логике в биологии большая редкость. И было бы несправедливо предъявлять к этой статье требования на порядок более строгие, чем ко всем прочим статьям.

Но все же надо задуматься о том, что если была миграция из Восточной Европы (больше, похоже, неоткуда), то по Восточной Европе в распоряжении пока данные только по трем популяциям:

– мезолитическим охотникам-собираателям (всего два генома!);

– носителям ямной культуры;

– современному населению.

И больше ничего!

То, что охотники-собиратели хуже подходят на роль добавки в средненеолитический генофонд, чем «ямники», авторы статьи [Haak et al., 2015] проверили. Но от охотников-собираателей до современного населения простирается десяток тысячелетий. И хотя то, что миграция должна была произойти не позднее финала неолита, сокращает временной интервал в два раза, но и на этом временном интервале и в пространстве Восточной Европы кто только не жил!

Лично я считаю, что авторы статьи выбрали вполне вероятного кандидата. Например, правдоподобнее предполагать миграцию не из лесной, а действительно из степной полосы Восточной Европы (в лесах и плотность населения меньше, и доля генофонда охотников-собираателей должна быть больше, а вклад, как упоминалось, был не «охотничий»). Но все же «вероятно» и «правдоподобнее» — это предположения, а не доказательства. И учитывая, что большинство археологов отрицают возможность прямого мощного влияния «ямников» на «шнуровиков», стоит предполагать, что реальный поток генов был более сложным, хотя и действительно вел во 2–4 тысячелетии до н.э. из Восточной Европы в Европу Центральную.

КАРТЫ ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ ОТ ДРЕВНИХ ПОПУЛЯЦИЙ

Одним из дополнительных аргументов сходства «шнуровиков» и «ямников» в работе [Haak et al., 2015] является то, что «шнуровики» генетически более всего похожи именно на «ямников» (минимальное Fst расстояние). Ну, судя по таблице в статье (Extended Data Table 3 [Haak et al., 2015]), популяция унетичской культуры (Unetice) еще более похожа на шнуровиков. А среди современного населения на шнуровиков похожи самые разные популяции — и венгры, и англичане, и скандинавы, и русские с мордвой.

Поэтому по данным этой таблицы из статьи я построил карты генетических расстояний (рис. 8.30–8.33) от основных древних популяций до современных популяций. Как обычно на картах расстояний, области, которые генетически далеки от генофонда реперной популяции, окрашиваются в красно-коричневые тона возвышенностей, а те области, до которых генетическое расстояние мало, — в интенсивно зеленые тона низменностей. Для

всех четырех карт использована одна и та же шкала (приведенная в верхней части каждой карты), что позволяет сравнивать их напрямую.

Вот карта генетических расстояний от «шнуровиков» (рис. 8.30): темно-зеленые пятна генетического сходства с современными популяциями разбросаны по всей Европе.

А вот карта генетических расстояний от популяции ямной культуры (рис. 8.30): очень большого сходства (темно-зеленые тона) нет нигде. Умеренное генетическое сходство (ярко-зеленые тона) — охватывает всю Восточную Европу (изучены русские, мордва, лезгины). А прочая Европа генетически еще меньше похожа на «ямников». Это не очень-то согласуется с идеей, что именно ямники во многом сформировали генофонд Европы (хотя если говорить о сравнительно небольшом влиянии, то это карте не противоречит). Особенно непохож на «ямников» генофонд современной южной Европы — что мы видели и в результатах ADMIXTURE (рис. 8.29).

Взглянем еще на карту расстояний от восточно-европейских охотников-собирателей (рис. 8.32). Тональность карты — красноватые тона очень больших расстояний (генофонд Европы сейчас совсем иной, чем в мезолите), но все же вклад этого компонента прослеживается на «своей» территории — Восточной Европе — явственнее, чем в Западной

Европе. Ну, это ожидаемый результат. Показательна, конечно, близость генофонда восточно-европейских охотников-собирателей к Северной Европе и северу Центральной Европы, хотя она и меньше, чем к Восточной.

А вот культуры позднего неолита/бронзы Центральной Европы (приведена карта от культуры колоколовидных кубков, рис. 8.32) кажутся родными почти всей современной Европе, окрасившейся в темно-зеленые тона минимальных расстояний. Что это? Результат того, что они — «общие дети» обеих половинок Европы, соединившихся на излете неолита, как предполагают авторы статьи [Haak et al., 2015]? Или напротив, результат распространения потомков этих культур по всей Европе уже в более поздние времена?

Правда, на картах бросается в глаза очень небольшое число опорных точек — современных популяций, изученных к моменту выхода статьи [Haak et al., 2015] по панели Human Origin. Поэтому выявляемые как картах географические закономерности наверняка еще уточнятся и скорректируются по мере более подробного изучения современного генофонда.

Будем надеяться, что Давид Райх и его команда, ежегодно выпускающие статьи в Nature, не заставят нас долго ждать следующей серии их увле-

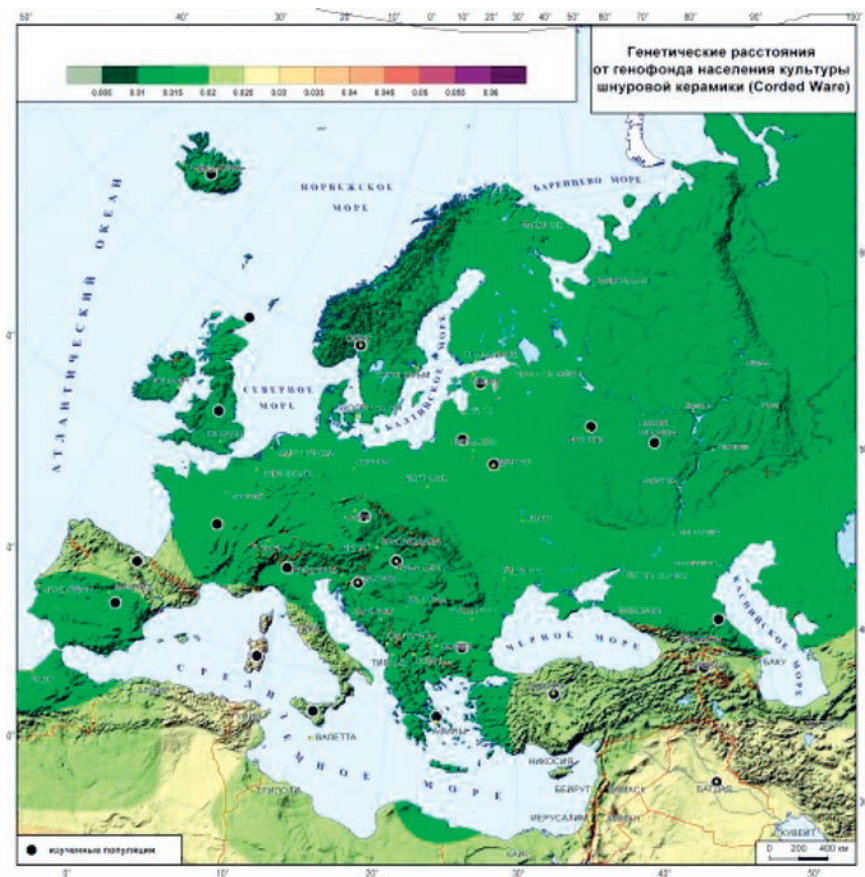


Рис. 8.30. Карта генетических расстояний от генофонда населения культуры шнуровой керамики (Corded Ware).

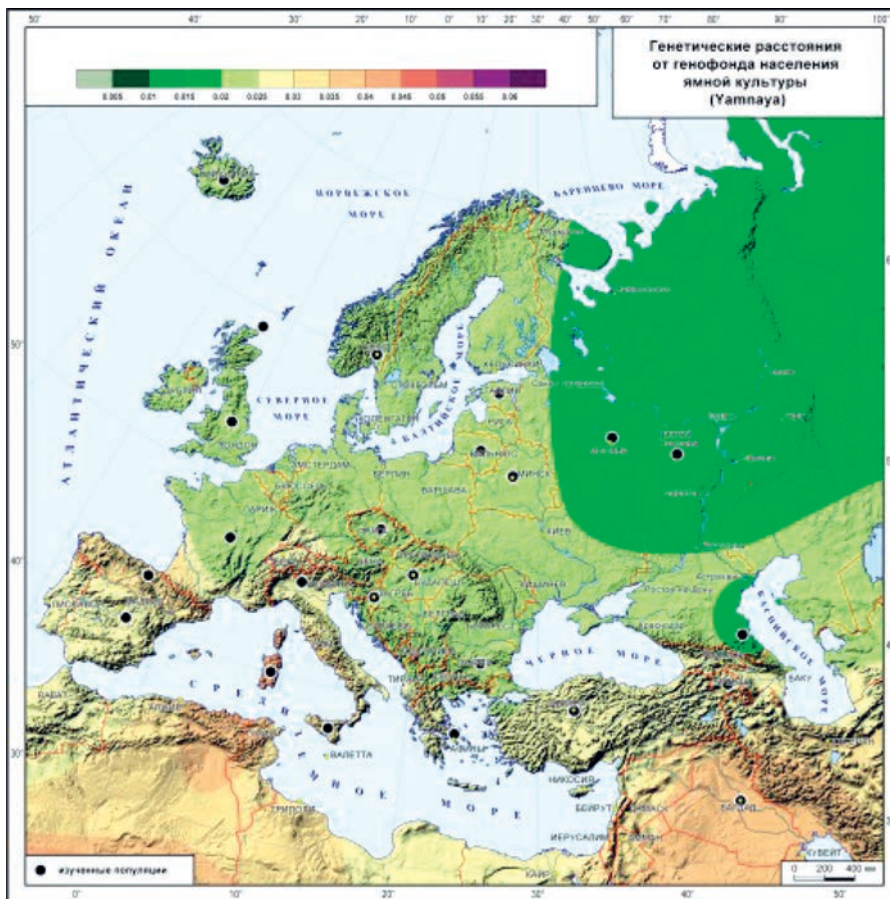


Рис. 8.31. Карта генетических расстояний от генофонда населения ямной культуры (Yamna).

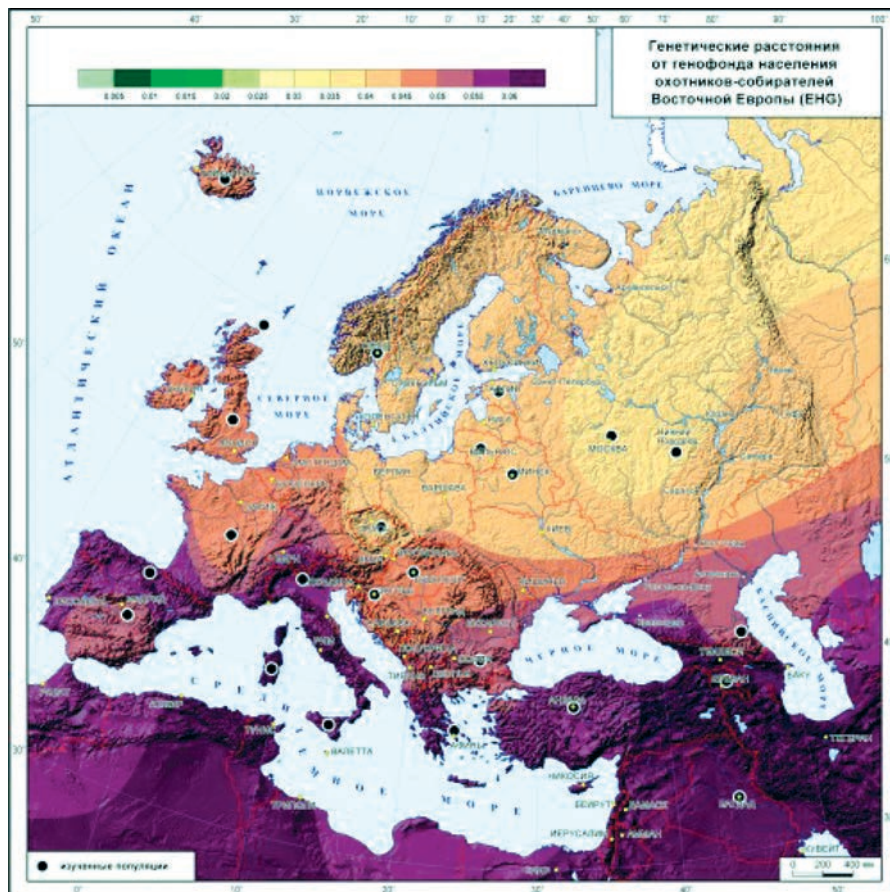


Рис. 8.32. Карта генетических расстояний от генофонда охотников-сборателей Восточной Европы (ЕНГ).

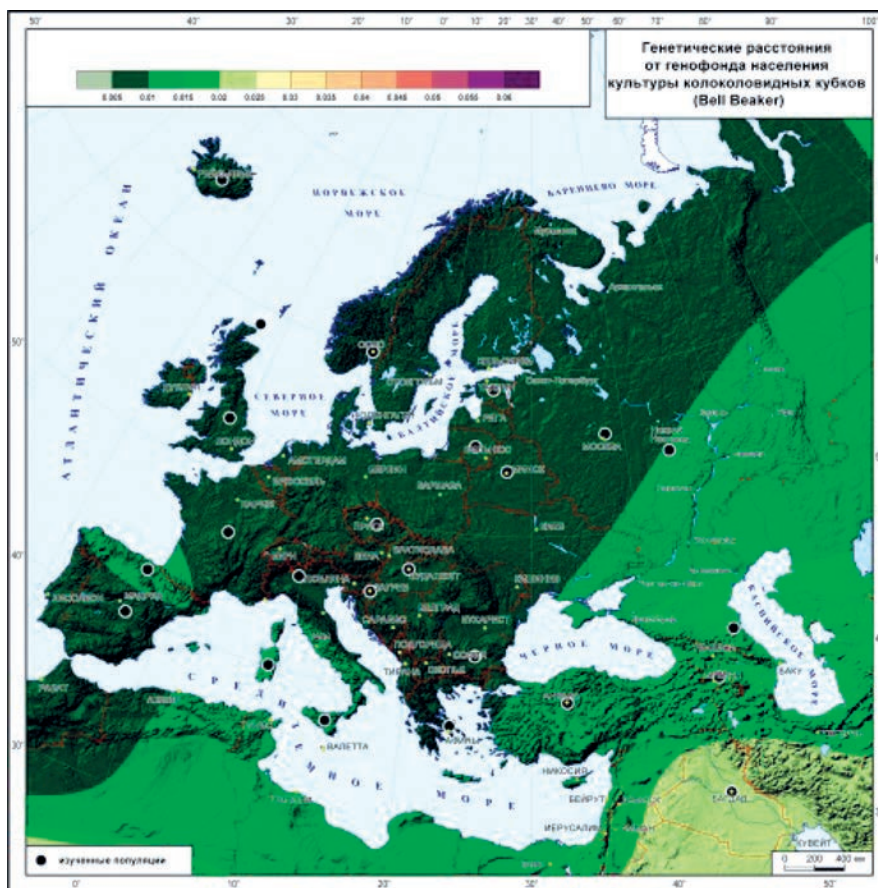


Рис. 8.33. Карта генетических расстояний от генофонда населения культуры колоколовидных кубков (Bell Beaker).

кательного «бостонского сериала», потому что, похоже, только полногеномный анализ других – и древних, и современных – популяций может луч-

ше всего распутать пути миграций. Похоже, они окажутся сложнее, чем цепочка «ямники» – «шнуровики» – «европейцы».

БРОНЗОВЫЙ ВЕК В ЕВРАЗИИ

ГЕНЕТИЧЕСКАЯ МОЗАИКА БРОНЗОВОГО ВЕКА

Бронзовый век, как становится ясно, был периодом важных миграционных событий, во многом сформировавших современный генофонд Евразии. Наиболее полно генетическая мозаика бронзового века выявлена в исследовании команды Eske Willerslev [Allentoft et al., 2015]. Эта статья по своим выводам близка к только что описанному исследованию [Naak et al., 2015], но рассматривает популяции бронзового века не только Европы, но и значительной части Азии. Также эта работа [Allentoft et al., 2015] важна тем, что в ней представлен анализ самого большого до сих пор числа древних геномов представителей разных археологических культур (рис. 8.34). Как видно из этого рисунка, в исследовании были отобраны образцы в основном бронзового века, а также несколько образцов до и после – эпох позднего неолита и железного века. Географически они охватывают Скандинавию, Цен-

тральную, Южную и Восточную Европу, Южный Урал и Центральную Сибирь (рис. 8.35).

Полногеномное секвенирование проведено для 101 образца, но покрытие варьировало очень сильно – от 0,01x до 7,4x (цифра показывает, сколько раз был прочитан каждый нуклеотид). С относительно высоким (по меркам древней ДНК) покрытием от 1,1x до 7,4x секвенированы 19 геномов, а также использовали ранее опубликованные данные.

На втором этапе из каждой древней популяции выбрали образцы, в которых успешно генотипировано не менее 300 тысяч SNP маркеров. В результате выборка составила 17 индивидуумов, представляющих 16 популяций. Анализ главных компонент выявил в этих данных 11 генетических кластеров, соответствующих географической и археологической принадлежности образцов. Вслед за авторами статьи, отдельно рассмотрим их результаты по Европе и по Азии.

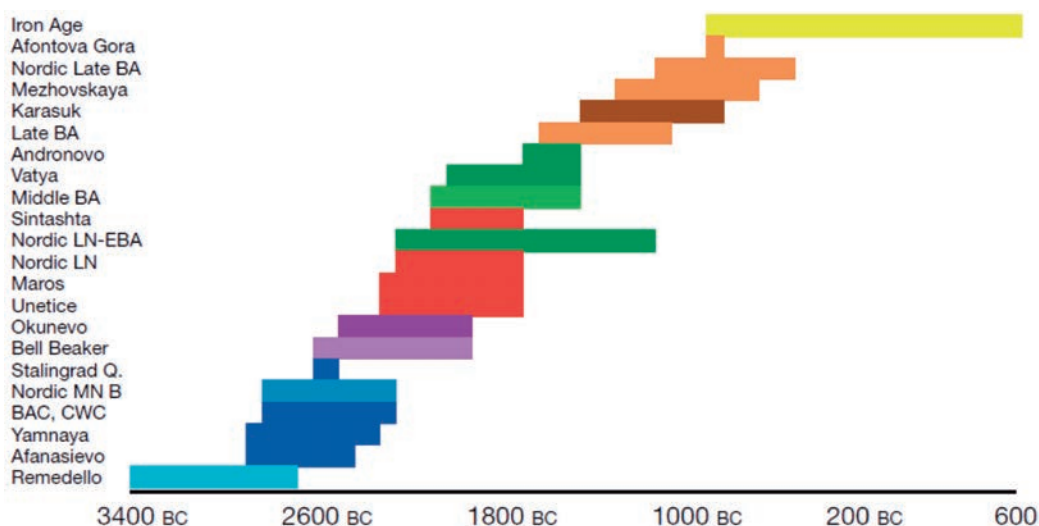


Рис. 8.34. Название (или местоположение) и датировка изучаемых культур [Allentoft et al., 2015].

По горизонтальной оси: датировки в годах относительно н.э. (BC – до н.э.)

По вертикальной оси (сверху вниз): железный век, Афонтова гора, северный поздний бронзовый век, межовская, карасукская, поздний бронзовый век, андроновская, Ватя, средний бронзовый век, синташтинская, северный поздний неолит – ранний бронзовый век, северный поздний неолит, Марос, унетичская, окуневская, колоковидных кубков, сталинградский карьер, северный средний неолит, культура шнуровой керамики или боевых топоров (BAC, CWC), ямная, афанасьевская, Ремеделло.

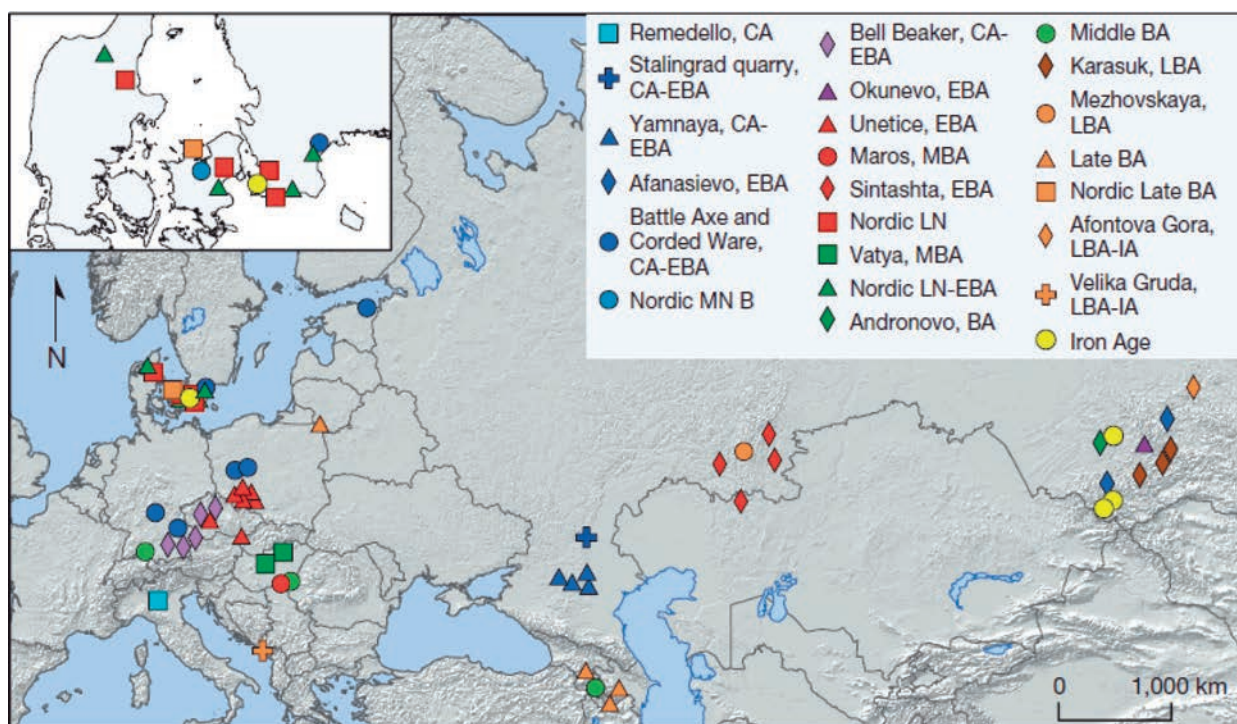


Рис. 8.35. Местоположение изученных образцов древней ДНК, слева вверху перечислены названия культур и их обозначения [Allentoft et al., 2015].

ЕВРОПА ЭПОХИ БРОНЗЫ

Основной результат состоит в том, что популяции в Северной и Южной Европе были сформированы смешением ранних охотников-собирателей и неолитических земледельцев, а в начале бронзового века получили поток генов из степей Предкавказья,

представленной ямной культурой (рис. 8.36). По сути, этот вывод [Allentoft et al., 2015] идентичен подробно разобранному выше выводу [Haak et al., 2015]. Похожи – с небольшими уточнениями – и выводы о том, что европейские культуры позднего неолита и бронзового века – шнуровой керамики, колоковидных кубков, унетичская и скандинавс-

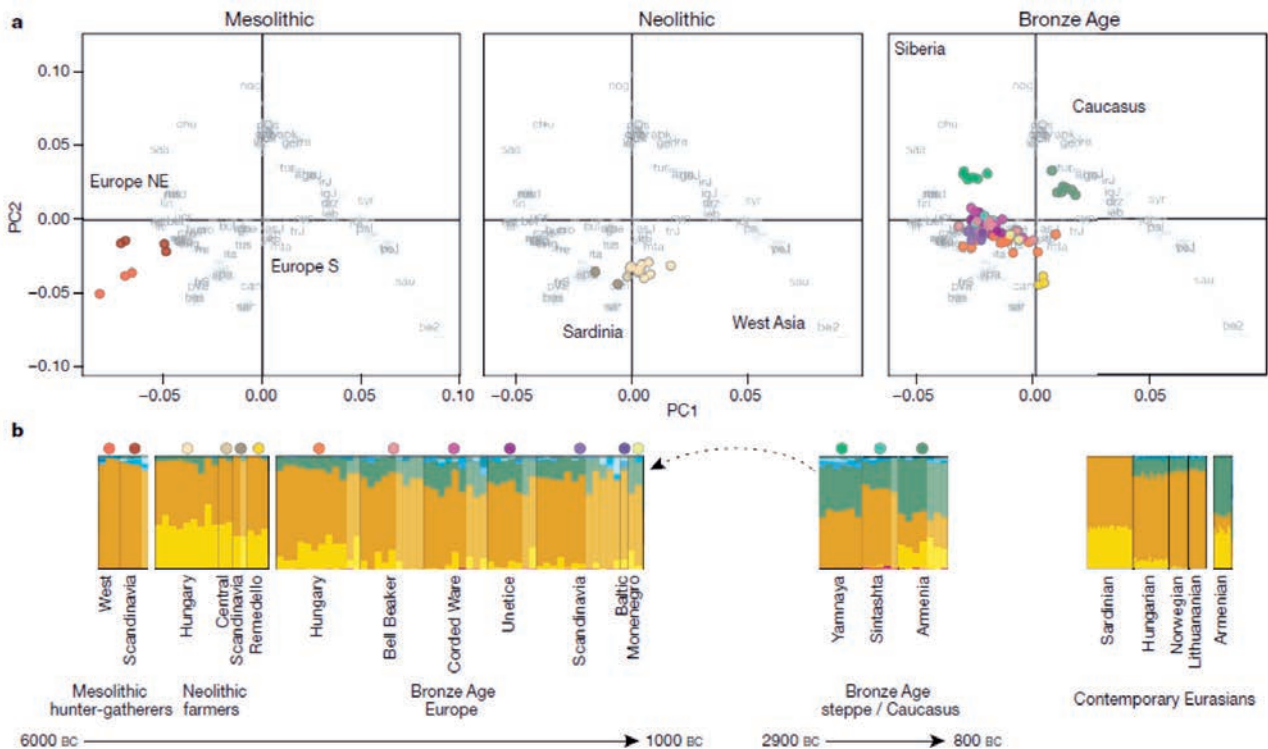


Рис. 8.36. Генетическая структура древней Европы и причерноморских степей [Allentoft et al., 2015].

а) Анализ главных компонент (PCA) древних образцов ($n=93$) из разных археологических эпох, спроецированный на современный генофонд Европы, Западной Азии и Кавказа. Серым представлены положение в генетическом пространстве 1 и 2 главных компонент отдельных индивидов (мелкий шрифт) и популяций (крупный шрифт). Цветные кружки обозначают древние образцы.

б) Анализ ADMIXTURE предковых компонентов ($K=16$) для геномов древних ($n=93$) и современных индивидов. Цветные кружки обозначают те же древние образцы, что и на рис. 8.36а (на графике PCA). Пунктирной линией обозначен предполагаемый генетический поток от ямной культуры.

кая – оказались генетически близки друг другу и проявляют сходство также с ямной культурой. Наиболее близка к ней культура шнуровой керамики, наименее – культуры Венгрии, промежуточное положение занимает культура колоколовидных кубков Центральной Европы. Доля этого «ямного» компонента, по данным [Allentoft et al., 2015], как и по данным [Haak et al., 2015], различается в разных регионах Европы. Так, генофонд популяции культуры медного века Ремеделло в Италии не имеет его вовсе, но он широко представлен в культурах поздней бронзы в Черногории.

Есть, правда, и одно – кажущееся – расхождение. В статье [Allentoft et al., 2015] часто упоминается «кавказский компонент», но это лишь расхождение в географических терминах: под «Caucasus» в работе [Allentoft et al., 2015] понимается никак не Южный Кавказ и даже не Северный, а те же самые степные области Предкавказья и Причерноморья, которые в статье [Haak et al., 2015] называются “steppe”.

АЗИЯ В ЭПОХУ БРОНЗЫ

Бронзовый век в Азии был столь же динамичен – он характеризуется обширными миграция-

ми и замещением популяций (рис. 8.37). Популяция раннего бронзового века – афанасьевская культура в Алтае-Саянском регионе – генетически не отличима от ямной. Это свидетельствует о том, что популяция ямной культуры из степей распространялась сразу в двух направлениях – как на запад в Центральную Европу, так и на восток в азиатские степи (рис. 8.38). Таким образом, потоки генов на большие расстояния в раннем бронзовом веке соединили такие отдаленные регионы, как Алтай и Скандинавию.

При этом синташтинская культура – географически промежуточная между ямной и афанасьевской – имеет несколько меньшую долю «ямного» генетического компонента и приближается к культуре шнуровой керамики из географически далекой от нее Центральной Европы. Пока можно только гадать, означает ли это далекую восточную миграцию шнуровиков или в синташтинскую культуру влился какой-то местный субстрат, имеющий древние общие корни с центральноевропейским субстратом «шнуровиков».

Следующая, культура, андроновская, возникшая в Центральной Азии в эпоху поздней бронзы, генетически близка к синташтинской, но четко отличается и от ямной, и от афанасьевской. Из чего

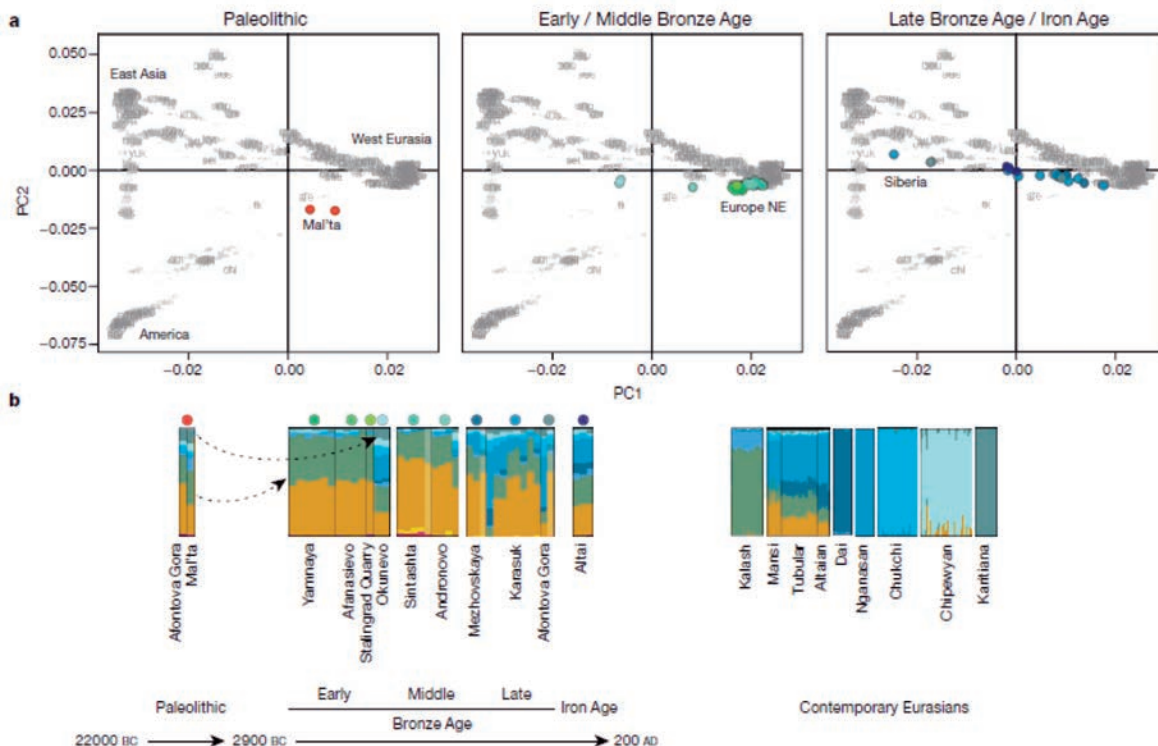


Рис. 8.37. Генетическая структура Азии бронзового века [Allentoft et al., 2015].

Обозначения те же, что и на предыдущем рисунке (8.36). Пунктирная линия обозначает предполагаемую генетическую связь ямной и окуневской культуры с палеолитическим образцом со стоянки Мальта.

вытекает гипотеза, что андроновская культура представляет собой продолжение генофонда синташтинской культуры во времени и пространстве (рис. 8.39).

К концу бронзового века в Алтае-Саянском регионе андроновская культура замещается карасукской и межовской культурами и культурами железного века. Эти культуры, вероятно, появились в результате миграции в Южную Сибирь (Алтае-Саяны) популяций из Восточной Азии или других регионов Сибири, принесшими более «азиатский» генофонд. Действительно, в результатах ADMIXTURE в генофондах этих популяций резко возрастает доля «синих» компонентов, свойственных современному населению Сибири (рис. 8.37).

Наиболее интригующие результаты получены для популяции окуневской культуры бронзового века из Саяно-Алтайского региона, которая демонстрирует генетическое сходство с современными коренными американцами. Этот факт означает, что окуневская культура может представлять часть популяции, генетически родственной верхнепалеолитическому охотнику-собирателю со стоянки Мальта, который, в свою очередь, был родственен популяции, отправившейся заселять Америку.

Реконструированные миграции третьего и второго тысячелетий до нашей эры схематически показаны, соответственно, на рис. 8.38 и 8.39. Например, мигрирующие на восток «ямники», предположительно говорившие на индоевропейских язы-

ке, дали начало афанасьевской культуре близ Алтая около 3000 лет до н.э. А представители афанасьевской культуры, в свою очередь, могли продвигаться немного южнее, и этим можно объяснить загадочное присутствие одного из старейших индоевропейских языков, тохарского, в бассейне реки Тарим.

В целом, совокупность описанных результатов, и в особенности своеобразное положение окуневской культуры, хорошо согласуется с предыдущими исследованиями палеоантропологов. Это не только мое мнение: на конференции в Йене (Linguistics, archaeology, and genetics, 2015) Мортен Аллентофт в докладе привел цитату из рецензента своей статьи в Nature. Смысл цитаты сводился к тому, что многие из обнаруженных в его статье генетических взаимоотношений популяций были описаны в русскоязычной научной литературе по физической антропологии, и кто бы мог подумать, что русские антропологи были так близки к истине. Пожалуй, это явное указание, что такие исследования генетикам стоит проводить в содружестве с представителями более старых и опытных наук.

КОГДА ВЗРОСЛЫЕ НАУЧИЛИСЬ ПИТЬ МОЛОКО

В работе [Allentoft et al., 2015] проанализирована также динамика в европейском генофонде отдельных генетических вариантов (всего 104), которые связаны с важными фенотипическими при-

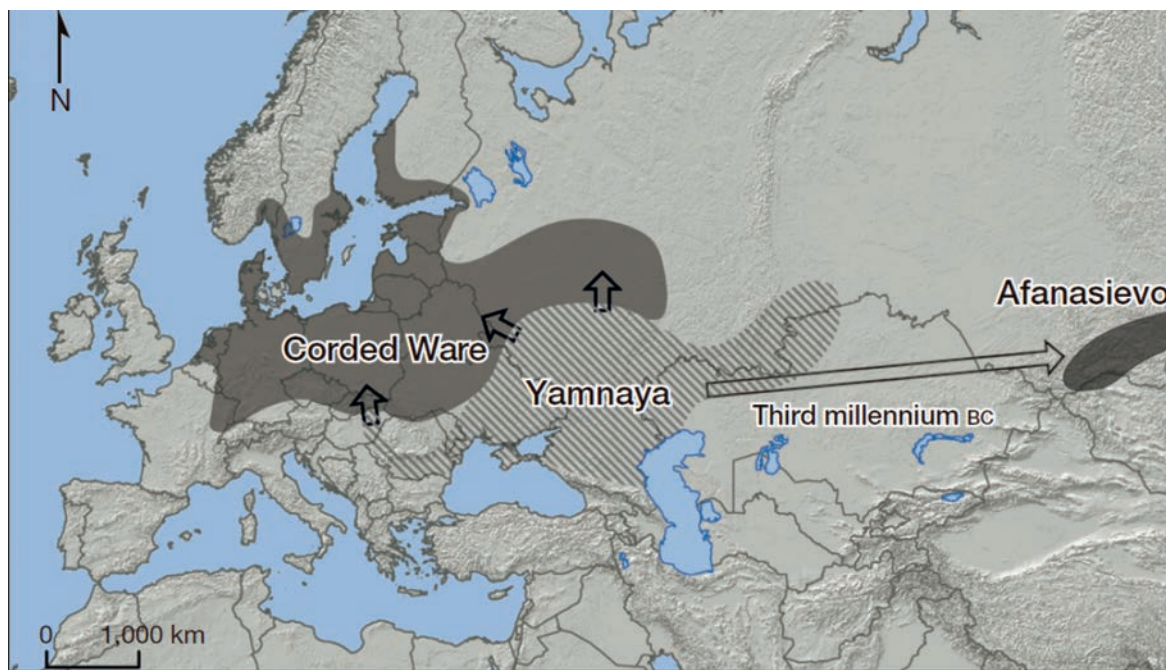


Рис. 8.38. Местоположение культур раннего бронзового века – ямной, шнуровой керамики и афанасьевской [Allentoft et al., 2015].

Стрелками показана экспансия ямной культуры.



Рис. 8.39. Культуры среднего и позднего бронзового века: синташтинская, андроновская, окуневская и карасукская, возникшие из миграции на восток [Allentoft et al., 2015].

Черные ромбы обозначают места находок колесных повозок и лошадей как доказательство распространения кочевых культур. Пунктиром обозначено место распространения тохарского языка, второго по древности среди индоевропейских языков.

знаками, что позволяет предположить, что некоторые из этих генетических вариантов поддерживались естественным отбором.

Например, выяснилось, что два аллеля, ассоциированных со светлой пигментацией кожи, у европейцев с ходом времени стремительно увеличи-

ваются в частоте. Аллель, обеспечивающий голубые глаза, хотя возникает еще у мезолитических охотников-собирателей, но отсутствует в черноморско-каспийских степных популяциях. Однако у европейцев бронзового века достигает заметной частоты.

Генетический вариант, связанный с толерантностью к лактозе (rs4988235), дает возможность взрослым людям употреблять в пищу молоко. Распространение этого аллеля традиционно связывали с появлением животноводства – он обеспечивал дополнительный источник питания. Полученные результаты оказались неожиданными: в бронзовом веке частота данного аллеля была довольно низкой – в

среднем 10%. Особенно он был характерен для популяции ямной культуры, поэтому авторы полагают, что именно от «ямников» он распространился по Европе. Однако своих высоких частот (среди современных европейцев его частота в среднем 70%) этот вариант достиг в более позднее время. Видимо, молоко вошло в обычный рацион далеко не сразу после того, как люди одомашнили скот.

СКИФЫ ДОНА

Античное время Европы (около 2500 лет назад, VI–III вв. до н.э.) пока представлено в исследованиях древней ДНК не так подробно, как более ранняя эпоха бронзового века. Но и изучая этот период времени, большое внимание вновь уделялось популяциям евразийской степи. В том числе нами совместно с Австралийским центром древней ДНК был изучен митохондриальный генофонд европейских скифов (низовья Дона, табл. 1.7. в 1 главе).

генофонде скифов Причерноморья западноевразийских гаплогрупп.

Другие авторы [Lalueza-Fox et al., 2003] обнаружили сходный генофонд у родственных скифам культур на территории Казахстана. Это указывает, что не только в эпоху бронзы, но и в скифскую эпоху «западноевразийский» митохондриальный генофонд распространялся много дальше на восток, чем сейчас.

СПЕКТР ГАПЛОГРУПП

Была исследована митохондриальная ДНК 16 образцов из скифских могильников на территории Ростовской области и обнаружены следующие гаплогруппы, перечисленные в порядке убывания: **T, U5a, H, I, D, A, C, F, U2e, U7**. Столь обширный перечень гаплогрупп при сравнительно небольшом числе проанализированных образцов указывает на большое генетическое разнообразие популяции скифов. Также можно отметить преобладание в

КАРТА ГЕНЕТИЧЕСКИХ РАССТОЯНИЙ

Этот вывод подтверждает и карта генетических расстояний (рис. 8.40). Картографированные генетические расстояния были рассчитаны по частотам 31 гаплогруппы мтДНК: **A, B, C, D, F, G, H, HV, I, J, K, L, N1a, N1b, R, T, U1, U2, U3, U4, U5a, U5b, U6, U7, U8, V, W, X, Y, Z, other** (сумма прочих редких гаплогрупп). Карта демонстрирует, в какой мере сходным с древним генофондом донских скифов является современное население разных территорий Евразии (144 современных популяции).

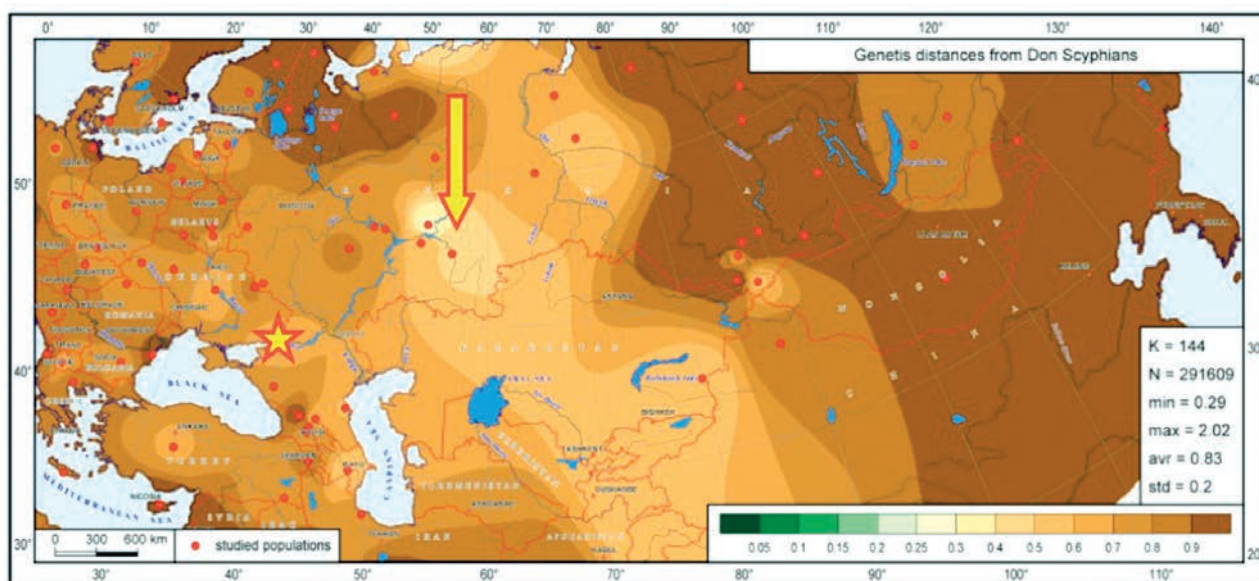


Рис. 8.40. Карта генетических расстояний от популяций скифов Причерноморья до генофондов 144 современных популяций Евразии.

Звездочкой отмечено положение изученной древней популяции, красными точками – положение генетически изученных современных популяций, стрелкой – зона наибольшего сходства древней популяции с современным генофондом.

Популяции, генетически сходные со скифами Причерноморья (более светлые тона карты) обнаруживаются как в Европе, так и в Азии. Однако в Европе лишь три области обнаруживают значения генетических расстояний в интервале от 0,6 до 0,7 (т.е. генетически довольно «далекие» популяции) – это низовья Днепра, Дона и Волги. Вся остальная территория Европы еще более генетически далека от генофонда скифов (расстояния в интервале от 0,7 до 0,9). А ареал популяций, близких к генофонду скифов (расстояния в интервале от 0,3 до 0,6) охватывает Южный Урал, юг Западной Сибири, Казахстан и отчасти Южную Сибирь. И наиболее генетически близкие к скифам современные популяции обнаруживаются в низовьях Камы и Вятки.

В целом карта генетических расстояний показывает, что к генофонду донских скифов наиболее близко современное население не западного, а центрального сегмента степного пояса Евразии (включающего юг Урала, Западной Сибири, Алтай и Казахстан). Это позволяет говорить, что западно-евразийский генофонд доминировал в генетическом портрете скифов Дона, но что сейчас область сходных генофондов находится не в Европе, а несколько восточнее – в центральном сегменте евразийской степи.

ОБЩИЙ ВЫВОД

Таким образом, рассмотрение и спектра гаплогрупп, и в особенности карты генетических расстояний указывает, во-первых, на то, что популяция скифов Дона, в отличие от других изученных нами

древних популяций, относилась почти полностью к «европейскому», а не к «азиатскому» генофонду. Во-вторых, результаты указывают, что в то время этот «европейский» генофонд географически далеко выходил за пределы Европы на восток.

В этом плане большой интерес представляют результаты, полученные другими авторами [Keyser et al., 2009], по древней популяции носителей родственной скифам курганной культуры в Южной Сибири. Эти результаты приводят, по сути, к тем же выводам о «европеоидности» генофонда скифов и об их широком распространении на восток, что не является удивительным, поскольку хорошо согласуется и с антропологическими, и с археологическими данными.

Скорее вызывает удивление сравнение с нашими собственными данными по Большому Оленьему Острову (см. рис. 8.25): две популяции, обе представляющие Восточную Европу и не так уж сильно различающиеся по датировкам (3500 лет назад на севере Восточной Европы и 2500 лет назад на юге Восточной Европы), выявляют противоположно направленные миграции: северная популяция Большого Оленьего Острова выявляет миграции из Сибири в Европу, а южная популяция скифов Дона – распространение «европейского» генофонда в Сибирь. Это является указанием на то, что результаты, полученные по отдельным древним популяциям, следует с большой осторожностью экстраполировать на крупные регионы, но вместе с тем и вселяет надежду, что по мере генетического изучения все новых древних популяций сложная картина разнонаправленных миграций в населении Евразии будет все больше проясняться.

ЗАКЛЮЧЕНИЕ

...И вот она – подобная кораллу,
Приросшая к Кавказу и к Уралу,
Земля морей и полуостровов,
Здесь вздутая, там сдавленная узко,
В парче лесов и в панцире хребтов,
Жемчужница огромного моллюска,
Атлантикой рожденная из пен –
Опаснейшая из морских сирен.
Страстей ее горючие сплетенья
Мерцают звездами на токах вод –
Извилистых и сложных, как растенья.
Она водами дышит и живет.
Ее провидели в лучистой сфере
Блудницею, сидящею на звере,
На водах многих с чашею в руке,
И девушкой, лежащей на быке...

Максимилиан Волошин. «Европа» (1918)

Как средиземный краб или звезда морская,
Был выброшен последний материк.
К широкой Азии, к Америке привык,
Слабеет океан, Европу омывая.

Изрезаны ее живые берега,
И полуостровов воздушны изваянья;
Немного женственны заливы очертанья:
Бискайи, Генуи ленивая дуга.
(...)

Европа цезарей! С тех пор, как в Бонапарта
Гусиное перо направил Меттерних, –
Впервые за сто лет и на глазах моих
Меняется твоя таинственная карта!

Осип Мандельштам. «Европа» (1914)

Я пишу эти строки еще сто лет спустя. Но на моих глазах карта Европы показывала не только таинственные перемены, но и удивительную неизменность. Неизменность была в картах ее современного генофонда – как ни менялась научная мода, как ни расширялись технологии анализа и не сужались подходы к синтезу, генофонд народонаселения Европы являл все те же черты – однородность большую, чем другие регионы мира, а его изменчивость была мала и следовала очертаниям «морей и полуостровов». Перемены же случаются не раз в сто лет, а ежегодно – но касаются они генофондов былых эпох, реконструкций генетической карты древней Европы. Перемены разительны: среди специалистов одно время даже ходила шутка, что различия между палеолитическим, мезолитическим, неолитическим и современным генофондами Европы лучше всего объясняются внезапными колонизациями с Марса. Но в последние годы анализ древней ДНК набрал силу и набрал объем выборок. И калейдоскоп непохожих кадров – моментальных снимков генофондов разных эпох – начал складываться в захватывающий фильм истории генофонда Европы. И уже не нашествия марсиан, а земные миграции начинают объяснять изменения геногеографической карты Европы. И очередность этих изменений наглядна, если рассмотреть временные срезы генофондов соседних эпох.

Как примирить эти перемены прошлых генофондов Европы и неизменность ее современного генофонда? Легко – взглянув на шкалу времени. В классический период популяционной генетики мы верили, что последний раз генофонд Европы ме-

нялся в неолите. Сейчас, спустя десятилетия, анализ древней ДНК уверил нас, что последняя масштабная перемена случилась в бронзовом веке. Но и то, и другое было очень давно. Тысячелетиями измеряются времена, проходящие от одного изменения до другого. А изменения, происходящие раз в сто лет, важны для истории стран – но не для истории генофондов.

Потому желающих найти в этой книге подтверждения своих исторических или тем более политических взглядов ждет разочарование: не замечает генофонд Европы тех потрясений, о которых писали поэты и пишут блогеры. Похоже, что не замечает он и «глобализаций». Например, считалось, что европейские генофонды перемешались из-за второй мировой войны. Можно ли это проверить? Можно, так как к этому времени уже широко изучались группы крови АВ0. И были сопоставлены все европейские популяции, изученные по АВ0 до второй мировой, и изученные после нее, причем так, чтобы их разделяло не меньше, чем одно поколение – 25 лет. Сравнение показало, что средние различия как в пределах одного народа, так и между европейскими народами не изменилось.

Зато какой простор такая стабильность открывает тем, кто хочет понять единство и многообразие населения Европы, а заодно и его историю. Можно век за веком изучать народонаселение Европы – и не опасаться, что объект изучения изменится. И действительно, такие исследования биологической основы народонаселения Европы ведутся уже не одно столетие – сперва по признакам физической антропологии, затем по классическим мар-

керам генетики, а сейчас уже и по полным геномам. Как дерево, растут наши знания о генетической изменчивости народонаселения Европы – о ее генофонде. И в этой книге предпринята попытка представить срез этого дерева знаний, каков он есть на 10е годы XXI века. Это дерево будет расти и дальше. Но я честно попробовал подвести итоги и обобщить то, что известно о генофонде Европы сейчас.

ЗАДАЧА ЗАКЛЮЧЕНИЯ

Заклучение книги нередко является единственной частью, которую прочитывает большинство читателей, не имеющих нужды в подробном изучении всего исследования. Поэтому я попы-

тался представить тут не один-два самых интересных мне вывода, а подытожить все разделы монографии, проведя перед мысленным взором читателя многочисленные результаты исследования в их тезисном виде. Неизбежным недостатком такого подхода является большой объем заключения, и, несмотря на это, все же отсутствие в нем многих существенных частных. Возможным достоинством – краткое (по сравнению с основным текстом) изложение главных мыслей и выводов, за обоснованием которых можно при желании обратиться к соответствующим главам. Судить об эффективности такого подхода предоставляю читателю в надежде на его снисходительность.

ДААННЫЕ И МЕТОДЫ ИХ АНАЛИЗА

РАЗНЫЕ МАРКЕРЫ – ОДНА ЕВРОПА

В последние четверть века популяционная генетика переживает новый подъем, связанный с использованием нерекombинирующих систем (митохондриальной ДНК и Y-хромосомы), а в последние пять-семь лет – еще и с использованием широкогеномного и полногеномного анализа. На рисунке 9.1 приведены графики – как менялось число статей, ежегодно публикуемых в мире по разным генетическим системам. На них хорошо видно постепенное появление в арсенале популяционной генетики все новых систем. Видна и меняющаяся мода на маркеры. В этой книге объединены резуль-

таты исследования всех трех современных генетических систем – Y-хромосомы, митохондриальной ДНК и полногеномных панелей маркеров.

БАЗА ДАННЫХ ПО Y-ХРОМОСОМЕ «Y-BASE»

Чтобы изучить генофонд всей Европы, надо сначала собрать всю информацию о нем. Сделать это легко с собственными данными, хотя их у нашего коллектива много – каждая пятая популяция, изученная в Европе по маркерам Y-хромосомы, изучена нами. Но чтобы узнать эту долю – и, главное, чтобы проанализировать все имеющиеся данные – потребовалось собрать и все данные, опуб-



Рис. 9.1. Использование разных генетических маркеров за сто лет генетических исследований.

По оси абсцисс указаны годы, по оси ординат – число статей, опубликованных в данном году по данной генетической системе. Приведены данные по пяти наиболее широко использовавшимся типам генетических маркеров. Публикации по классическим маркерам подсчитаны по библиографическому приложению в [Cavalli-Sforza et al., 1994], по остальным типам маркеров – по данным Pubmed.

ликованные другими. База содержит 144 тысячи - образцов, генотипированных по SNP-маркерам (по гаплогруппам), из них для 32 тысяч есть и данные по STR-маркерам. А в главном зарубежном аналоге – базе данных криминалистов YHRD – все наоборот: 143 тысячи по STR-маркерам (для криминалистики они важнее), из них только для 17 тысяч есть данные по SNP-гаплогруппам.

Но собрать данные – этого еще мало. Ведь разные авторы часто пользуются разными маркерами: один определяет в популяции какую-то гаплогруппу, а другой определяет не ее, а какую-то ее часть. Конечно, все гаплогруппы Y-хромосомы вложены одна в другую, точнее, вырастают одна из другой, как ветви на дереве. Но чтобы воспользоваться этим свойством, пришлось придумать и реализовать в программной оболочке принцип использования данных не о том, к какой гаплогруппе принадлежит образец, а самих первичных данных о его генотипе. Это обеспечивает *Y-base* долгую жизнь, независимо от того, сколько сотен (или сотен тысяч) новых маркеров будет открыто на Y-хромосоме, и как изменятся имена гаплогрупп. Потому что разработанные нами программные модули автоматически преобразуют первичную информацию о статусе SNP маркеров в образце в информацию о том, к какой гаплогруппе (по сегодняшнему дереву) принадлежит образец, а следующий модуль рассчитывает частоты гаплогрупп в популяциях.

В наступившую эпоху полного секвенирования Y-хромосомы стремительно растет объем данных о дереве гаплогрупп – какие гаплогруппы существуют и как друг с другом связаны. Но вот данные о частотах этих гаплогрупп в популяциях растут куда медленнее – ведь для этого множество популяционных образцов надо типировать на специфические маркеры новооткрытых ветвей. И *Y-base* аккумулирует не столько данные по самому дереву (с этим прекрасно справляются ISOGG и YFull), сколько данные по распространенности гаплогрупп в популяциях.

Все эти технические детали нужны лишь для одного: иметь под руками полную базу данных по частотам гаплогрупп Y-хромосомы во множестве популяций всего мира. Для этой книги, понятно, использовалась только часть данных – по Европе и прилегающим регионам. И то использовались не все, а только надежные данные – например, малые выборки исключались или объединялись, а для картографирования использовались только выборки, для которых совершенно ясно, в какой точке Европы они собраны.

БАЗА ДАННЫХ ПО мтДНК (MURKA)

Эту базу мы создали и ведем уже больше 10 лет, и объем ее рос стремительно, вбирая данные

из лавины статей по изменчивости мтДНК в популяциях, сошедшей в мировой науке в конце прошлого – начале нынешнего века. В ней сейчас данные по двум тысячам популяций мира – суммарно 132 тысячи образцов, для каждого из которых секвенирован хотя бы ГВС1, а для многих – и полный мт-геном. Это данные из пятисот публикаций по населению всех континентов, а также наши (в том числе еще неопубликованные) данные по народам Северной Евразии.

Как и база по Y-хромосоме, MURKA отличается тщательным описанием популяций: не только их название и объем выборки, но и принадлежность к стране и провинции, этническая и лингвистическая принадлежность, географические координаты, ссылки на источник данных и многое другое.

БАЗЫ ДАННЫХ ПО АУТОСОМНЫМ МАРКЕРАМ

Этими базами я занимаюсь со своего студенчества – первой моей большой работой было как раз создание (и анализ) базы данных по аутосомным микросателлитам на основе еще ранее созданной нашим коллективом базы по классическим маркерам «GENEPOOL». Сейчас классические маркеры почти не используются (к сожалению, потому что за десятилетия их использования накопились данные по великому множеству популяций). Зато используются данные по STR-маркерам ДНК-идентификации, практическая важность которых помогла им быстро обогнать по числу изученных популяций почти все другие аутосомные ДНК-маркеры. И по этим STR-маркерам ДНК-идентификации нами тоже создана база данных (18 тысяч образцов из 80 популяций Евразии).

Впрочем, для исследования генофонда Европы пригодились не эти базы аутосомных маркеров, а данные по аутосомным широкогеномным маркерам. Этот массив данных включает хотя и сотни тысяч маркеров, но пока лишь считанные десятки популяций и единицы статей по Европе. Поэтому острой необходимости оформлять их в сложную инфосистему пока нет – все статьи хорошо известны специалистам, и данные из них иногда объединяются для совместного анализа.

СОБСТВЕННЫЕ ДАННЫЕ

Чтобы как можно полнее охарактеризовать генофонд народонаселения Европы, в созданные базы данных наравне включались и собственные, и литературные данные. В то же время собственные данные в ряде видов анализа составляли основную часть, и тогда нередко анализировались отдельно, чтобы обеспечить методологическое единообразие.

Это единообразие состоит в подходах к формированию выборок – в какой мере они отражают

именно коренное население. Ведь хорошо известно, что выборки популяционных генетиков (которые на каждой территории стремятся изучить коренное население) и выборки криминалистов (которым важно общее население региона) нередко оказываются генетически контрастны. Но и популяционные генетики разнятся строгостью своих требований к отбираемым для изучения индивидам, поэтому рискованно смешивать в одном анализе выборки, собранные при разном уровне строгости. Наш коллектив на протяжении уже почти четверти века следует шести строгим правилам.

1) «Три поколения» – только индивиды, у которых все бабушки и дедушки родились в данной популяции и относили себя к данной этнической группе. Если же более далекие предки пришли издалека, то, раз уж они оставили в этой популяции потомство на протяжении более трех поколений, их миграция уже влилась в генофонд, и поэтому их современные потомки включаются в анализируемую выборку.

2) «Не родственники». Родственники (до третьей степени) исключались, поскольку их наличие не только уменьшает эффективный размер выборки, но и создает иллюзию миграции, если в выборку попали потомки единичного мигранта.

3) «Информированное согласие» – письменное согласие на предоставление образца крови для популяционно-генетического анализа и контроль Этической комиссией.

4) «Только мужчины». В большинстве выборов мужчины составляют более 95%, что позволяет изучать не только мтДНК и аутосомы, но и Y-хромосому, а также облегчает анализ X-хромосомы.

5) «Образцы крови» – ведь нужно много ДНК. Сбор слюны – редкое, а буккального эпителия – редчайшее исключение.

6) «Субпопуляции» – большинство популяций исследовалось не в одной, а в нескольких географических или субэтнических локальностях, чтобы охватить весь генофонд, а не одну его своеобразную часть.

Созданный по этим правилам биобанк «Коренное население Северной Евразии, Е.В. Балановская, О.П. Балановский» включает 24 тысячи образцов от представителей 90 народов России и сопредельных стран – это 250 популяций, равномерно покрывающих всю Северную Евразию. Для анализа генофонда Европы использованы не все эти образцы, а «только» четыре тысячи образцов из 70 популяций по Y-хромосоме, три тысячи образцов по мтДНК, и чуть менее сотни – по широкогеномным маркерам.

ГЕНОГЕОГРАФИЯ И GENE GEO

Большинство методов, использованных в этой книге, традиционны для популяционной генетики,

и нет нужды говорить о них в Заклучении. Но картографические методы геногеографии – нечто особенное.

Теория и практика картографического изучения генофондов стала лицом российской геногеографии еще в 90^е годы прошлого века. Целью картографического анализа является «проявление» структуры генофонда. Поэтому карта – это не иллюстрация, а *инструмент* исследования. Не картинка, а научная *модель*. Не фотография, а *образ* генофонда. Как всякий хороший инструмент, карта требует правильного обращения. Как всякая модель, может существовать во множестве вариантов – в зависимости от задаваемых параметров. Как всякий образ, карта может быть прокомментирована, но не передана словами. Карта нередко красива, и всегда логична, проверяема, математически точна.

На первом этапе, согласно логике картографического анализа генофонда, строятся карты распространения отдельных генетических маркеров – аллелей, гаплогрупп и т.д. Эти карты содержат значения частоты данного маркера, рассчитанные для каждого из тысяч узлов карты, равномерной сетью покрывающих территорию. Это цифровая модель – поэтому нужно подобрать оптимальные параметры картографического моделирования.

На втором этапе эти рассчитанные значения подвергаются статистическому анализу – от простого усреднения между соседними узлами до корреляций с неким другим признаком.

На третьем этапе карты разных признаков комбинируются друг с другом – создаются обобщенные карты. Поскольку карты – это математические матрицы значений, их можно хоть складывать, хоть брать производные. Можно рассчитать по картам (и тем самым картографировать) любой популяционно-генетический параметр. В этой книге я чаще всего пользуюсь картами генетических расстояний. Они просто и наглядно показывают, население каких территорий генетически сходно с интересующей нас популяцией, а каких – генетически удалено от нее, тем самым очерчивая для популяции «родную» для нее территорию.

На всех этих этапах карта представляет собой файл с данными, не воспринимаемый человеческим глазом. Но этот файл всегда можно визуализировать, показав разные интервалы значений разными цветами. И тогда статистическая модель предстает в обличье привычной географической карты, на которую можно наносить любую дополнительную информацию – от береговой линии до мест археологических раскопок.

В итоге можно получить карту распространения любого аллеля (например, гаплогруппы) или карту любого параметра генофонда, провести ее последующий анализ для уточнения закономерностей в географической изменчивости,

связать с другими картами и создать в итоге точный и выразительный геногеографический образ генофонда.

Кроме того, картографический способ представления данных о генофонде удобопонятен и универсален: он допускает использование генетических результатов не только генетиками, но и специалистами в смежных областях.

Чтобы воплощать всю эту теоретическую красоту в жизнь, нужна картографическая программа, заточенная под генетическую специфику. В конце прошлого века под руководством Е.В. Балановской и Ю.Г. Рычкова создавался целый ряд картографических программ, лучшей из которых стал GGMAg. Но в начале текущего столетия назрела нужда в новой, современной программе. Им стал GeneGeo. Как видно из названия, программа предназначена специально для геногеографии, а разрабатывается она под руководством автора и совместно со старшим научным сотрудником кафедры картографии МГУ к.г.н. С.М. Кошелем (вместе с которым в свое время разрабатывался и GGMAg). Хвастаться нехорошо, но не сказать, что программа удалась, я тоже не могу: она не только реализует

еще широкий спектр методов (который ежегодно расширяется), но и весьма удобна в использовании, а также соответствует современным мировым стандартам картографических программ, в том числе в разделе графического отображения карт и совместимости с другими картографическими пакетами (Surfer, ArcGIS).

ДАТИРОВКИ ВОЗРАСТА ГАПЛОГРУПП

Это еще одна группа методов, которые хотя и используются широко, вызывают дискуссии в популяционной генетике и вокруг нее. Но я не в силах втиснуть в Заключение то, что я выяснил, систематизировал, несложными словами описал и еле уместил в большом разделе про датировки и споры о них. Поэтому отсылаю интересующихся к разделу 1.4. Здесь же только напомним, что, на мой взгляд, не стоит путать *скорость* мутирования и *метод* расчета возраста, не стоит полностью доверять ни «генеалогической», ни особенно «эволюционной» скорости мутирования STR-маркеров, зато очень даже стоит пользоваться датировками гаплогрупп Y-хромосомы по данным ее полного секвенирования.

СТРУКТУРА ГЕНОФОНДА ЕВРОПЫ

ЧТО БЫЛО ИЗВЕСТНО

Из всех регионов мира европейский генофонд является наиболее изученным по обеим «однородительским» генетическим системам (мтДНК и Y-хромосома). Из предшествовавших работ известно, что главными чертами митохондриального генофонда Европы является его гомогенность [Simoni et al., 2000] и предположительно палеолитический возраст формирования [Richards et al., 1996; 2000]. Изменчивость же гаплогрупп Y-хромосомы демонстрирует четкие географические градиенты, а формирование генофонда тоже обычно возводится к верхнему палеолиту [Semino et al., 2000; Battaglia et al., 2008].

И для митохондриальной ДНК, и для Y-хромосомы эти основные черты изменчивости были установлены еще в 2000 году [Richards et al., 2000; Simoni et al., 2000; Rosser et al., 2000; Semino et al., 2000]. После этого накапливались и анализировались подробные данные по отдельным странам, народам, гаплогруппам, но анализ в масштабе всей Европы не проводился. Попытка обобщающего анализа на новом витке наших знаний предпринята в этой книге. Объем данных стал на порядок больше: анализ изменчивости Y-хромосомы в Европе основывается на данных о 585 популяциях (суммарный объем выборки N=38 000 образцов), что в 10 раз превышает объем выборки (N=3 616) в предыдущем обобщающем исследовании [Rosser

et al., 2000]. Аналогично и для мтДНК – наш анализ основывается на данных о 118 популяциях (суммарный объем выборки N=25 500 образцов, тщательно отобранный из общего массива данных N=31 935 образцов), что тоже в 10 раз превышает объем данных (N=3 113), использованный в предыдущем обобщающем анализе [Richards et al., 2002]. Что касается широкогеномных маркеров, то они сейчас находятся на стадии «первичного» обобщающего анализа (как мтДНК и Y-хромосома в 2000 году), и в книге я разбираю полученные результаты, дополняя их нашим собственным анализом балто-славянских популяций, в том числе и по широкогеномным маркерам. На нескольких последующих страницах я попробую суммировать главное, что получилось в новом обобщающем анализе генофонда Европы.

ГОМОГЕННОСТЬ

В шеренге регионов Евразии – Европа, Сибирь, Восточная Азия, Передняя Азия, Индостан (можно назвать еще Юго-Восточную Азию и Центральную Азию) – Европа выделяется гомогенностью своего генофонда. Это значит, что все популяции Европы похожи друг на друга – похожи больше, чем сходны друг с другом разные популяции Сибири или разные популяции Индии. Вообще говоря, степень гетерогенности генофонда задается характером демографической истории популяций,

балансом между изоляцией и миграциями: если преобладает изоляция, то дрейф генов увеличивает межпопуляционные различия (гетерогенность), если преобладают миграции, то все наоборот – они различия между популяциями сглаживают.

Поскольку дрейф и миграции действуют на все гены одинаково, то степень гетерогенности является чертой не одного какого-то маркера, а всего генофонда. В полном согласии с этим, относительная (по сравнению с другими регионами Евразии) гомогенность Европы была показана и по классическим маркерам, и по аутосомным ДНК-маркерам, и по мтДНК, и по Y-хромосоме, и можно держать пари, что будет показана по любым другим маркерам, которые популяционные генетики облюбовуют для себя в будущем.

Но что же это такое?! – скажет читатель. Гомогенность Европы по Y-хромосоме! Да как же так, если она гетерогенна донельзя – чуть ли не каждая гаплогруппа сидит в своем регионе, а в другие почти не заходит! Так-то оно так, но только если измерить количественно степень «сидения в своем регионе» и отсутствия в других регионах, то эта величина для Y-хромосомы Европы будет хотя и велика, но куда меньше, чем для Y-хромосомы в коренном населении Сибири.

Но вот изменчивость мтДНК в той же Сибири будет поменьше этой величины. В том-то и дело, что можно сравнивать регионы, но только по одной и той же системе маркеров. Или разные системы маркеров, но для одного и того же региона. Ведь мы начали с того, что «дрейф и миграции действуют на все гены одинаково». Но гены-то при этом разные, им свойственна разная «фоновая» степень межпопуляционной изменчивости. Например, подразделяя гаплогруппы, можно переходить к маркерам со все большей географической приуроченностью и потому большими межпопуляционными различиями (аналогично и среди аутосомных маркеров можно находить более или менее изменчивые в географическом пространстве). Еще важнее, что для любых популяций с патрилокальной традицией браков (то есть для большинства народов планеты, начиная с неолита, а то и раньше) изменчивость Y-хромосомы будет особенно высока, изменчивость мтДНК особенно низка, а аутосомные маркеры будут в промежутке, но ближе к мтДНК. Вот эта-то фоновая гетерогенность всех трех типов маркеров в одних регионах согласованно поднимется, а в других согласованно опустится в зависимости от баланса дрейфа и миграций.

В Европе она опустилась. Опустилась настолько, что межпопуляционная изменчивость мтДНК (наименее изменчивой еще на «фоновом» уровне) стала если не нулевой, то трудноуловимой – лишь полное секвенирование позволяет выделять географически специфичные варианты, да специальные методы анализа (последовательного увеличения

«генетического разрешения») позволяют разглядеть остов генофонда. Изменчивость аутосомных маркеров на «фоновом» уровне промежуточная, и потому она опустилась не так сильно, так что всегда позволяла выявлять общие тренды в европейском генофонде, а сейчас при использовании обширных широкогеномных панелей стала выявлять и детали. Изменчивость Y-хромосомы на «фоновом» уровне высочайшая. И в Европе она осталась высокой, так что даже этот один ген (ведь для популяционной генетики Y-хромосома – самый длинный, но все же один локус) замечательно выявляет и общую структуру генофонда, и многие ее детали.

ГЕНОФОНД СТРУКТУРИРОВАН ПО ГЕОГРАФИИ

Y-хромосома. Потому с Y-хромосомы и начнем обзор структуры европейского генофонда. Мы картографировали все 34 гаплогруппы, для которых известно не просто их существование, но и частоты встречаемости во многих популяциях Европы. Получившийся атлас Европы хорошо выявляет и общие паттерны, и многие детали распространения вариантов Y-хромосомы. Глядя на эти карты отдельных гаплогрупп бросается в глаза приуроченность каждой гаплогруппы к своему географическому домену. Можно сказать, что генофонд Европы по Y-хромосоме составлен из этих доменов, как из паззлов: в каждом регионе хотя и присутствует множество гаплогрупп, но особый облик ее генофонда формируют лишь одна или две.

Если же сравнивать популяции Европы по всему спектру гаплогрупп, получится генетическая картина, очень напоминающая географическую карту. Все популяции Европы распределятся по четырем группам, занимающим четыре зоны на географической карте и четыре квадранта на генетическом графике. Одна группа (северо-западная) включит популяции Западной Европы. Вторая (северо-восточная) – популяции Восточной Европы. Третья (юго-западная) – популяции европейского Средиземноморья. Четвертая (юго-восточная) – популяции Передней Азии (формально уже не Европа) и Северного Кавказа, который формально, географически – Европа, но генетически от нее резко отделен, и является частью переднеазиатского генофонда.

Внутри каждой группы популяции кластеризуются по географическим регионам. Для Средиземноморья и Западной Европы кластеры задаются очертаниями «морей и полуостровов»: народы Иберийского, Аппенинского, Балканского полуостровов, Британских островов и континентальной Европы собираются в соответствующие кластеры, и даже расположение кластеров относительно друг друга на генетическом графике вторит расположению этих регионов на географической карте. Для

Восточной Европы, не подразделенной морями, домены ее генофонда разделяет Волга, а также различия «лес – степь», то есть опять-таки география. На Северном Кавказе подразделение на кластеры популяций особенно четкое, но тут оно следует не только географической (Западный Кавказ, Центральный и Восточный), но еще в большей мере лингвистической классификации (абхазо-адыгская группа, иранская, нахская, дагестанская).

Все эти закономерности выявлены при использовании разных наборов гаплогрупп и разных методов анализа, и такая устойчивость картины «географической» структурированности генофонда Европы свидетельствует об ее объективности.

Широкогеномные маркеры. Еще больше в ее объективности убеждает анализ широкогеномных (аутосомных) данных. При включении большого числа популяций генетический график начинает с занудной методичностью воспроизводить географическую карту: на графике главных компонент, показывающем генетические кластеры, можно найти «Бискайский залив», «Ирландское море», «Португалию» и так далее. Правда, включение дополнительных популяций может заметно трансформировать паттерн, видимый на графике. Дело в том, что генетические расстояния не прямо пропорциональны географическим, а лишь высоко коррелируют с ними. Поэтому очертания популяций на генетическом графике могут быть не пропорциональны своим географическим прототипам. Особенно это касается популяций севера Восточной Европы: за счет усиленного дрейфа генов они генетически куда более своеобразны, чем должны бы быть, судя по географическому расстоянию до них. Поэтому включение генетически своеобразных финнов или коми меняет пропорции генетического графика Европы: широкое облако остальных европейских популяций сжимается в узкую полоску, а своеобразные популяции северо-востока Европы располагаются сбоку от нее. Это та же географическая карта – но настолько «сдавленная узко» в западной части и настолько «вспученная» в восточной, что узнать ее нелегко. Похожим образом, если добавить в анализ Европы самый генетически близкий ей регион мира – Переднюю Азию, то европейское облако опять сжимается в полоску, только рядом с ней, вместо северного «финского острова» появляется южный материк переднеазиатско-кавказских популяций. Но и в этом случае сохраняется географический принцип: популяции Средиземноморья будут находиться между европейским и переднеазиатским материками, а внутри европейского материка географически согласованно расположатся популяции Западной, Центральной и Восточной Европы.

Другой метод – выявление предковых компонентов ADMIXTURE – тоже обнаруживает геогра-

фически точные закономерности. Даже в названиях основных предковых компонентов, выделяемых в масштабе Европы – западноевропейский, восточноевропейский, средиземноморский, кавказский – видны все те же основные домены генофонда Европы. (Не удержусь подчеркнуть, что анализ в масштабе мира подчеркивает всю надуманность идеи, что расы человека – это миф. Генетическая реальность рас как нельзя более наглядно выпирает из карт предковых генетических компонентов – четко выделяются негроидный, европеоидный, монголоидный, американоидный, австралоидный и другие компоненты).

мтДНК. Неудивительно, что и по мтДНК выявляются похожие закономерности в генофонде Европы. Неудивительно и то, что уловить их в митохондриальной гомогенности нелегко. Это удалось лишь при использовании особого приема многомерной статистики: последовательного изменения масштаба анализа.

При обычном анализе – включении всех популяций Европы и прилегающих регионов – выделяются 4 кластера: саамы (вспомним своеобразие севера Восточной Европы при широкогеномном анализе и классические работы L.L. Cavalli-Sforza про обособленность генофонда саамов); Передняя Азия и Кавказ (совпадение с широкогеномной и Y-хромосомной картинами); популяции восточных рубежей Европы (это уже специфика мтДНК – следствие четкого деления на западноевразийские и восточноевразийские гаплогруппы); единый кластер всей Европы – популяции и Западной Европы, и почти всей Восточной, и Средиземноморья попали в один компактный кластер. Вот она, митохондриальная гомогенность Европы!

Сменим масштаб – проанализируем только популяции, вошедшие в этот четвертый «пан-европейский» кластер. И тут увидим, что внутри него все же существуют два разных кластера. Первый – генетически обособленные популяции северо-востока Европы (снова напомним их своеобразие по широкогеномным маркерам и подчеркну, что это вызвано тем, что баланс дрейфа генов и миграций в этой части Европы сдвинут в сторону дрейфа, хотя и не так сильно, как в соседней Сибири). Но подавляющее большинство европейских популяций попадает в другой кластер – по-прежнему единый, генетически гомогенный, но географически обширный.

Еще раз сменим масштаб и проанализируем этот кластер отдельно. И внутри него, наконец, увидим генетическую структуру: популяции германской, славянской, романской и балтской языковых групп расположены в разных зонах на графике. Этот паттерн лингвистический – но вместе с тем и географический, ведь большинство популяций одной зоны не только лингвистические родственники, но и географические соседи.

ИЛИ ВСЕ ЖЕ ПО ЛИНГВИСТИКЕ?

Когда мы только что рассматривали структуру генофонда Европы, то наткнулись на яркие совпадения не только с географическими, но и с лингвистическими закономерностями. Это вечный вопрос для народонаселения любого региона мира: что определяет структуру его генофонда – географическое соседство (а значит, поток генов) или же лингвистическое родство языков (а значит, родство предковых популяций)? Оценить связь структуры генофонда с географией и лингвистикой позволяют коэффициенты *парной корреляции*. Но зачастую географически соседние народы и по языку родственны, поэтому значимые корреляции получаются с обоими возможными факторами. Как тогда различить, который из факторов повлиял? Для этого можно посмотреть, будут ли сходными популяции, говорящие на родственных языках, но географически далекие, или же чаще будут сходными популяции, географически соседние, но говорящие на совсем разных языках. Математическая статистика реализует этот подход с помощью *частных корреляций*.

Этот подход прочно вошел в популяционную генетику в виде теста Мантеля. Рассчитывается матрица генетических расстояний (от каждой популяции до каждой), рассчитывается аналогичная матрица географических расстояний, и нужна также матрица лингвистических расстояний между языками, на которых говорят эти популяции. Используемые в книге матрицы лингвистических расстояний являются одними из многочисленных результатов, полученных в ходе многолетних исследований московской школы глоттохронологии, и были любезно предоставлены мне коллегами-лингвистами – член-корр. РАН А.В. Дыбо, проф. О.А. Мудраком, д.б.н. А.С. Касьяном.

Вся Европа. При анализе всех индоевропейских народов Европы парные корреляции по Y-хромосоме достигали 0,7 с географией и 0,5 с лингвистикой. Частные корреляции тоже были достоверны и высоки: 0,7 с географией и 0,4 с лингвистикой. В предыдущем обобщающем исследовании Россера с коллегами аналогичные коэффициенты корреляции были намного ниже (самая высокая – 0,4 с географией). Это показывает, что по мере накопления данных географическая структурированность генофонда Европы становится все более и более явной. В целом же, оба фактора – и география, и лингвистика – играют каждый свою роль, совместно формируя генетический ландшафт Европы по маркерам Y-хромосомы.

И по мтДНК нами впервые показана достоверная связь генофонда и с географией (парная корреляция 0,4) и с лингвистикой (0,3), хотя эти коэффициенты и поменьше, чем в случае Y-хромосомы, а частные корреляции еще немного ниже (0,3

и 0,2, соответственно). То есть опять видим, что мтДНК выявляет те же самые закономерности, просто вновь менее рельефно.

Балто-славяне. Для большей части европейского генофонда – народов балто-славянской группы – проведен еще более подробный анализ. Все народы этой группы изучены по всем трем генетическим системам – Y-хромосоме, мтДНК и широкогеномным маркерам. И снова – чем подробнее изучаем и чем лучше формируем выборки, тем более четкие вырисовываются закономерности. Парные корреляции генофонда балто-славян с географией по Y-хромосоме и широкогеномным маркерам достигают невиданной величины – 0,9, а по мтДНК – 0,8. Парные корреляции с лингвистикой тоже очень велики: по Y-хромосоме и широкогеномным – 0,8, а по мтДНК – 0,7. Но вот частные корреляции еще более рельефно, чем при анализе всей Европы, показывают *примат географии*: корреляции с ней остаются очень велики (0,8 для ведущих систем и 0,5 для мтДНК), а вот корреляции с лингвистикой у балто-славян становятся недостоверными (от 0,1 до 0,3 для разных систем).

Обратим внимание, что важнейшие корреляции (Y-хромосомы и широкогеномных маркеров с географией) превышают 0,9. Выше уже некуда – стопроцентная связь (корреляция, равная 1) вряд ли возможна, да и была бы скучна.

Поскольку для балто-славян были подробные и надежные данные по пяти системам (трех генетическим, лингвистике и географии), я проанализировал, как все они связаны друг с другом. Оказалось, что Y-хромосомный генофонд, аутомсомный генофонд (широкогеномная панель) и географическое положение популяций формируют триаду, причем очень тесную (корреляции выше 0,9), а мтДНК и лингвистика рисуют хотя и похожие, но слегка отличающиеся картины.

Тогда что же получается? Структура генофонда задается географическим пространством, а с лингвистическим родством народов связана лишь в той мере, в какой само расселение лингвистических групп зависит от географии? Для балто-славянских народов – да. А вот в других регионах мира – вовсе необязательно.

Кавказ. Например, при изучении генофонда Северного Кавказа – столь же подробном, как и для балто-славян – мы получили обратный результат. Структура кавказского генофонда хотя и коррелирует с географией, но еще лучше коррелирует с лингвистической структурой. Это показано пятью методами:

- 1) генетическая и лингвистическая дендрограммы являются отражениями друг друга;
- 2) генетические границы на Кавказе совпадают с лингвистическими;

3) и парная, и частная корреляции генетики с лингвистикой выше, чем с географией;

4) лингвистическая классификация точнее воспроизводит структуру генофонда, чем географическая;

5) датировки ветвей лингвистического древа хорошо согласуются с генетическими датировками кластеров гаплотипов, специфичных для этих ветвей (групп популяций).

Все это позволяет считать, что генетическая структура Кавказа развивалась параллельно с многообразием языков Северного Кавказа. И связующим звеном между генетической и языковой эволюцией, их общей причиной являлась демографическая история популяций, главным образом, разделение популяций на дочерние группы. Горный рельеф способствовал именно разделению, а не смешению популяций. И эта изоляция дочерних популяций друг от друга приводила к тому, что кластеры гаплотипов, возникавшие в одной из них, не распространялись к соседям. По той же причине шла дивергенция и языков этих популяций. Когда же эта дочерняя популяция, в свою очередь, дробилась на внучатые, они не только сохраняли (генетическую и лингвистическую) память об общем происхождении, но и приобретали каждая свои (генетические и лингвистические) особенности. Тем самым одна и та же (демографическая) история популяций отражалась в двух зеркалах – генофонде и языке. Поэтому совпадение генетических и лингвистических реконструкций не является чем-то исключительным: оно лишь свидетельствует, что оба вида анализа проведены корректно, что оба зеркала были незамутненными.

Модель. Тогда почему же такое совпадение не найдено у балто-славян? Да и в других регионах мира - почему генофонд чаще оказывается скоррелирован с географией, чем с лингвистикой? Дело в том, что характер связи генетических и лингвистических данных зависит от того, по какой модели шло формирование народонаселения. На Кавказе это была модель саморазвития единой предковой популяции – ее постепенное дробление или, лучше сказать, ветвление. И в других регионах мира, где история народонаселения следует этой модели, совпадение генетики и лингвистики является естественным, а связь с географией будет вторична.

Но в регионах, где народонаселение сформировалось в результате смешения одной группы населения с другой, фактор лингвистического родства оказывается менее значимым, чем фактор географического соседства. Ведь география в какой-то мере работает всегда (ни одна миграция не может игнорировать пространство, да и интенсивность дрейфа генов часто имеет географические причины). А лингвистика как предиктор генофонда хорошо работает лишь тогда, когда родство язы-

ков связано - через демографию - с общностью происхождения *всего генофонда*. Если же группы, родственные по языку, смешались с другими популяциями, то сформировавшиеся генофонды будут лишь *частично* иметь общий компонент (маркируемый языком), и корреляция с лингвистикой станет меньше, чем с географией.

Именно это и произошло со славянами: славянские группы, расселившиеся по Восточно-Европейской равнине, вобрали в себя один субстратный генофонд, а славяне, расселившиеся по Балканам, вобрали в себя другой; поэтому одни славянские генофонды стали весьма отличаться от других славянских же генофондов. Как результат, структура генофонда стала лучше коррелировать не с лингвистикой (общеславянским родством), а с географией (которая аппроксимирует генетическое сходство субстратных генофондов друг с другом).

Тюрки. Еще более ярко, чем у славян, преобладание субстрата проявляется у тюркоязычных народов. В их демографической истории часто работала модель доминирования элиты: группа населения переходит на новый язык – язык элиты (чаще всего завоевателей), при этом численность элиты и ее генетический вклад ничтожны по сравнению со всей массой народа.

Получается, что модель доминирования элиты противоположна модели саморазвития, а модель смешения популяций находится посередине.

Любопытно, что и в случае модели доминирования элиты ее генетический компонент, пусть он и очень мал, иногда можно уловить, если ее субстрат различается столь драматически, как у тюрков. Это показано Б. Юнусбаевым с соавторами для тюрков, генофонды которых, хотя и кардинально разнятся, скажем, в Сибири, в Европе и в Закавказье, но все же несут общий компонент. Но конечно, этот компонент слишком мал, чтобы сформировать значимые корреляции генетики и лингвистики при анализе всей совокупности тюркоязычных народов.

Завершу свои рассуждения о связи генофонда и языка тем, что в случае любой модели – от саморазвития до доминирования элиты – связь генетики и лингвистики (соответственно, сильная, слабая, или отсутствующая) должна проявляться в разной степени, но во всех генетических системах – что мы и видели для балто-славян.

ИЛИ СТРУКТУРИРОВАН ПО ЭТНОСАМ?

Кроме географического соседства и лингвистического родства, важнейших в масштабе всей Европы, в масштабе ее отдельных регионов стоит посмотреть и на роль этнического фактора. Под этим непростым словом я имею в виду только одну и очень простую вещь: поскольку большинство этно-

сов являются по совместительству популяциями (более половины браков заключается с представителями того же этноса), то принадлежность разных локальных популяций к одному народу (вложенность их в популяцию более высокого, этнического уровня), возможно, скажется и на большем *генетическом* сходстве этих популяций друг с другом, чем с популяциями, относящимися к другому народу. Этот этнический фактор структурирования генофондов, по сути, является тем же лингвистическим (большинство народов отличаются от других по языку), но только на очень дробном уровне лингвистического древа – его листьях-языках.

Анализ генофондов более сотни популяций Европы, относящихся к двадцати народам, дей-

ствительно, выявил удивительно четкую картину. Почти во всех случаях популяции, относящиеся к одному народу, кластеризуются совместно, причем образующиеся на генетическом графике «этнические облака», за редким исключением, почти не перекрываются друг с другом.

При этом размеры этнических облаков очень разные. Есть народы гомогенные (их географически разные популяции генетически очень похожи друг на друга), например, поляки и бельгийцы. А есть очень гетерогенные (например, башкиры и финны) – в этом случае разные популяции одного и того же народа генетически весьма разнообразны, но все же обычно ближе друг к другу, чем к популяциям других народов.

ИСТОРИЯ ГЕНОФОНДА ЕВРОПЫ

Эта история пишется на наших глазах. Точнее, записывалась она в нашей ДНК в течение десятков тысячелетий, но вот расшифровывать эти письмены – расшифровывать массово и убедительно – смогли лишь в самые последние годы. Я говорю о древней ДНК.

Анализировать ее не только сложно (уж очень ее мало сохраняется), но и рискованно – слишком велик риск ошибиться, проанализировав вместо древней современную ДНК, случайно залетевшую в пробирку. Эта проблема контаминации если и не решается полностью, то сводится до минимума лишь в специализированных центрах древней ДНК. В России, к сожалению, еще не создан такой центр, который получил бы широкое международное признание. Поэтому наши исследования древней ДНК проводятся совместно с признанными центрами в Австралии и Германии, и полученные результаты вписались в общую картину истории генофонда Европы.

Несомненно, что картина эта не окончательная, и в ближайшем будущем нас ждет еще немало новостей о нашем прошлом. Но я все же рискну описать картину, какой она представляется сегодня. Предполагаю, что будущие исследования внесут и уточнения, и дополнения, но не перерисуют ее полностью.

СРЕДНИЙ ПАЛЕОЛИТ

Первыми людьми в Европе были неандертальцы. Жили они и в Азии, и даже, похоже, были там более генетически разнообразны. Вообще говоря, они не столько наши предки, сколько братья (параллельная линия эволюции гоминид). Но когда человечество современного вида вышло из Африки на просторы Евразии и повстречало неандертальцев, то небольшая метисация между ними происходила. Поэтому неандертальцы отчасти нам и предки – но лишь в небольшой части. Метисация

была не только в месте первой встречи – на Ближнем Востоке, но продолжалась и в других частях Азии, и в Европе. В результате доля «неандертальского» компонента в генофонде европейцев составляет один-два процента.

ВЕРХНИЙ ПАЛЕОЛИТ И МЕЗОЛИТ

Расселяясь по просторам Евразии, человечество быстро поделилось на два ствола – западноевразийский (европеоидный) и восточноевразийский (монголоидный). Народонаселение Европы является частью западного ствола. В северной половине Евразии он отнюдь не был ограничен Уралом, но заходил и далеко на восток в Сибирь. Впрочем, в эту эпоху – верхнего палеолита, включая и период максимума последнего оледенения, – генетические различия между западноевразийским и восточноевразийским стволами еще не были так выражены, как сегодня (подобно тому, как и антропологические различия между протоевропеоидами и протомонголоидами тогда не были резкими, а наряду с ними могли существовать еще и протоуралоиды). Поэтому верхнепалеолитическое население и Европы, и соседних с ней более восточных частей Евразии представляло континуум популяций. За счет дрейфа генов эти популяции были генетически весьма различными, – но эти различия были географически постепенными. На генетических графиках хорошо видно, как верхнепалеолитические генофонды разных регионов (Западной Европы, Скандинавии, Восточной Европы, Южной Сибири) по своему взаимному сходству друг с другом выстраиваются в цепочку, воспроизводящую такую же цепь этих регионов на географической карте.

После таяния ледника, в мезолите, генетические взаимоотношения региональных генофондов оставались примерно такими же. Точнее сказать, пока мало данных, чтобы проследить, как забола-

чивание одних территорий и освобождение от ольда других сказалоcь на демографической и генетической структуре. Впрочем, выявляется вероятная миграция населения из Западной Сибири в северные области Восточной Европы. По крайней мере, мезолитический генофонд Южного Оленьего острова наиболее сходен с современными популяциями именно Зауралья – если это и не прямая миграция, то свидетельство значительного генетического сходства мезолитического населения на огромных территориях по обе стороны Урала.

НЕОЛИТ

В неолите же многое изменилось. В Передней Азии произошла неолитическая революция (в том числе переход от охоты и собирательства к земледелию и скотоводству), которая потом докатилась до Европы. Докатилась она вместе с потомками «революционеров». Поэтому первые неолитические культуры Европы – от Испании до Венгрии, от Балтийского моря до Мраморного – генетически похожи, как близнецы. Очевидно, это была единая волна земледельцев, быстро расселившихся по Европе. И генетические данные находятся в хорошем согласии с археологической картиной, разве что не дифференцируют «прибрежно-средиземноморский» и «дунайско-континентальный» пути миграций земледельцев, которые так четко выделяет археология.

Генетически эти первые земледельцы Европы были промежуточными между генофондом автохтонного населения Европы (тех самых палеолитических и мезолитических популяций, о которых шла речь выше и которые жили охотой и собирательством) и современных популяций Передней Азии. Но вот непонятно – то ли эти первые земледельцы были промежуточными потому, что смешались с автохтонами Европы, то ли современные генофонды Передней Азии теперь уже не те, что в неолите. Как бы то ни было, по прошествии пары тысячелетий смешение земледельцев и охотников-собирателей в Европе возобновляется (или начинается), и генофонды земледельческих популяций еще немного (но именно что немного) приближаются к автохтонному верхнепалеолитическому генофонду Европы.

Напомню, что эти верхнепалеолитические популяции, хотя и формировали географически непрерывный ряд генофондов, но различия между ними были очень велики. В частности, весьма различался генофонд Западной и Восточной Европы. Поэтому сложение современного генофонда Европы хорошо аппроксимируется не двумя источниками (пришлый компонент переднеазиатских земледельцев и автохтонный компонент европейских охотников-собирателей), а тремя: пришлый компонент переднеазиатских земледельцев, автохтонный

компонент западноевропейских охотников-собирателей, автохтонный компонент восточноевропейских охотников-собирателей.

БРОНЗОВЫЙ ВЕК

В бронзовом веке в причерноморско-каспийских степях сформировались яркие культуры скотоводов – прообраз более поздних степных кочевников. Генетически наиболее подробно изучена ямная культура, хотя есть данные и по другим, в том числе более поздним культурам этого круга – срубной, синташтинской, афанасьевской, андроновской. Все они оказались родственными друг другу по своему генофонду, хотя, конечно, у каждой обнаружены и любопытные особенности.

Генофонд ямной культуры, с которой все началось (по крайней мере, с нее началось генетическое изучение и от нее сейчас танцуют генетические реконструкции), был двусоставным. Первый компонент – все те же восточноевропейские охотники-собиратели (автохтоны Восточной Европы со времен палеолита). Второй компонент прибыл откуда-то с юга, из Передней Азии, и пока можно лишь гадать, обогнул ли он Каспий с востока, или Черное море с запада, или каким-то образом перепрыгнул Кавказ (хотя массовая миграция из Армянского нагорья в причерноморские степи лично мне кажется наименее вероятной). Этот пришедший в Восточную Европу второй компонент, хотя и был переднеазиатским, но имел мало общего с тем другим переднеазиатским компонентом, который земледельцы принесли в Западную Европу.

Образовавшийся в итоге слияния «иногo перднеазиатского» и «местного восточноевропейского» компонентов генофонд носителей ямной культуры, похоже, повлиял на генофонд всей Европы в бронзовом веке не меньше, чем миграция земледельцев из Передней Азии повлияла на него в неолите. Носители (или потомки?) ямной и родственных ей культур широко расселились и на восток, и на запад. На восток они дошли по степям до Алтая, принеся туда свой генофонд, который с перипетиями и модификациями, но в целом сохранялся вплоть до тюркской экспансии на рубеже нашей эры. Все это хорошо согласуется с археологическими и палеоантропологическими данными.

Генетические результаты показывают, что в бронзовом веке генофонд «ямников» двинулся и на запад – в Центральную Европу, а оттуда в Западную и Северную Европу, и лишь до Средиземноморья почти не дошел. Но вот этой западной миграции «ямников» археологические данные, мягко говоря, не поддерживают. Конечно, мигрировать могли носители не самой ямной культуры, а какие-то генетически родственные им популяции. Но в любом случае, тут серьезная научная проблема, и в ближайшие годы мы можем стать свидетелем

лями – или участниками – любопытного процесса поиска ее решения.

Эта миграция так важна потому, что была последним крупным мазком, завершившим композицию генофонда Европы. В результате этой миграции генофонд Европы приобрел те частоты генов, которые и сейчас являются среднеевропейскими. А изменчивость этих частот в разных европейских популяциях – помним про гомогенность генофонда Европы – невелика.

НАСТОЯЩЕЕ И БУДУЩЕЕ

Конечно, история генофонда Европы на бронзовом веке не закончилась. И древняя ДНК пока-

зывает ряд последующих изменений в отдельных регионах (один из многих примеров – миграция из Сибири в Европу по тундровой полосе 3,5–2 тысячи лет назад, представленная Большим Оленьим островом на Кольском полуострове, но впоследствии угасшая). И анализ современной ДНК реконструирует многие процессы в генофонде – не обязательно только миграции. Сам я возлагаю большие надежды на полное секвенирование Y-хромосомы – первые опыты таких исследований тоже описаны в этой книге и оказались весьма вдохновляющими, – посмотрим, чего нам удастся добиться в ближайшем будущем.

О СЛАВЯНАХ И ПРОЧИХ ЧАСТНОСТЯХ

В предыдущих разделах Заключения я кратко описал основные черты генофонда Европы и основные вехи его истории. Но это касалось всей Европы – ее общей истории. При этом народонаселение Европы состоит из множества конкретных популяций, и у каждой из них есть и своя частная история. Конечно, общая история генофонда Европы является и их историей тоже, и в общей структуре генофонда Европы находится свое место для каждой из ее частей. Но все же мне хотелось уравновесить описание общих закономерностей генофонда Европы характеристикой частных закономерностей – особенностей отдельных популяций. Сделать это для каждого народа в рамках одной книги невозможно – например, наша монография об одном только русском генофонде была еще объемнее этой книги о генофонде всей Европы. Но все же и в этой книге нашлось место подробному изучению генофонда славян, Кавказа, Русского Севера, картам генетических расстояний от многих народов Европы и прочим частным, но важным закономерностям. Поэтому в заключении Заключения кратко перечислю важнейшие из них.

СЛАВЯНЕ

Генофонд народов, говорящих на балто-славянских языках, удалось изучить подробно – практически каждый народ был изучен по большим выборкам и по всем трем современным генетическим системам: Y-хромосоме (6 тысяч образцов из шестидесяти популяций), мтДНК (7 тысяч образцов из пятидесяти популяций), широкогеномным маркерам (300 образцов из двадцати популяций).

Общий паттерн. По всем трем системам получена примерно одна и та же картина трех кластеров, только по Y-хромосоме и широкогеномным маркерам она четкая, а по мтДНК, как обычно, смазанная.

Первый кластер – восточные и западные славяне. Среди них генофонды поляков, русских и украинцев очень близки, но не идентичны – их «этнические облака» почти соприкасаются, но не перекрываются друг с другом. Белорусы подразделились – большинство их популяций перекрывается с русским генофондом, а полесские популяции – с украинским. Балты (латыши и литовцы) формируют свой небольшой кластер в непосредственной близости от кластера западных и восточных славян, тогда как генофонды словаков и особенно чехов отдаляются от него, приближаясь к народам Западной Европы.

Второй кластер – северные русские. Эти популяции генетически далеки от остальных восточнославянских популяций.

В третьем кластере перемешаны славянские и неславянские популяции Балкан.

Аутосомные гаплотипы. По числу общих фрагментов генома популяции первого кластера (западные и восточные славяне) больше похожи на неславянские народы северо-восточной Европы (балты и западные финно-угры), на южных славян и неславянские генофонды Балкан они похожи меньше, а на народы Волго-Уральского региона, Кавказа и Западной Европы – еще меньше. Если поставить в центр рассмотрения южных славян, они примерно одинаково похожи и на остальных славян, и на неславянские народы Балкан (кроме греков, сходство с которыми невелико). Эти результаты подчеркивают в первую очередь сходство каждой группы славян со своими географическими соседями, но во вторую выявляют и большее сходство разных групп славян друг с другом, чем это можно было бы ожидать просто из географического положения популяций. Подчеркну, что результаты одновременно свидетельствуют об этих обеих сторонах одного процесса – как выяснилось при обсуждении этих результатов в СМИ, многие, в

ладу со своими предпочтениями и предубеждениями, видят только одну из сторон, напрочь «забывая» о второй.

Преобладание субстрата. Где бы ни находилась прародина славян, это был лишь кусочек Европы, выйдя из которого, славяне расселились по половине Европы. Поэтому на каждой территории славяне не могли не ассимилировать местные популяции. Вопрос лишь в доле, какую этот субстрат составил в генофонде расселившихся славян. Генетически выделяются два разных субстрата. «Центрально-восточноевропейский субстрат» приняли в себя западные и восточные славяне. Другой, «южно-восточноевропейский субстрат», впитали в себя южные славяне. И в обоих случаях доля этого субстрата была преобладающей. В пользу вывода о преобладании субстрата в формировании генофонда славян говорят три аргумента.

Во-первых, тот факт, что западные и восточные славяне имеют меньше общих фрагментов генома с южными славянами, чем с популяциями северо-восточной Европы (родственников популяций, ассимилированных славянами).

Во-вторых, тест АМОВА также указывает на важную роль субстрата, поскольку генетическое разнообразие между разными ветвями славян намного превышает разнообразие внутри ветвей; такая картина и должна была сформироваться, если восточные и южные ветви славян ассимилировали генетически различные популяции.

В-третьих, преобладающая роль географии, а не лингвистики в формировании генофонда славян говорит о том же. А лингвистика была реконструирована хорошо – специально для нашего совместного исследования коллеги-лингвисты обновили базы данных, уточнили древо родства балтославянских языков и диалектов, и мы надеялись найти сходство этого уточненного древа с генофондом. Но не тут-то было. Примат же географии говорит именно о субстрате. Ведь если бы включе-

ния субстрата не было, то общее происхождение, фиксированное в языке, не могло не сказаться и на сходстве генофондов, даже когда какие-то группы славян мигрировали на далекое расстояние от своих родственников. Но такая роль лингвистического родства не выявилась. И напротив: генетическое сходство между дославянскими популяциями, жившими на территории половины Европы, должно было быть примерно пропорционально географическим расстояниям между ними. Тогда, если в современных славянских генофондах преобладает субстрат, то и сходство этих генофондов должно следовать географическим расстояниям. Что и было выявлено.

Славянские пироги. Любой генофонд можно сравнить со слоистым пирогом – каждую эпоху в нем появляются новые напластования, новые слои. И если смотреть на пирог генофонда южных славян, то в нем дославянский слой будет самым толстым, преобладающим, а славянский слой – лишь тонкой глазурью на поверхности. То же самое с западными и восточными славянами – у них преобладает восточноевропейский дославянский субстратный слой, а славянство – опять лишь тонкая глазурь (рис. 9.2). Так что внешний вид этих пирогов похож – глазурь-то одинаковая, и те, и другие – славяне. Но вот внутренние слои у пирогов не одинаковые, потому что их «генетический вкус» разнится.

Балтский пирог? Если, как мы выяснили, генофонд западных и восточных славян сформирован в основном автохтонным восточноевропейским субстратом, то эти славяне должны походить на всех своих соседей по Восточной Европе, а друг от друга разные славянские популяции тоже должны отличаться примерно так же, как разные неславянские популяции Восточной Европы, живущие на примерно таком же расстоянии друг от друга. Но это не так! Разные популяции западных и восточных славян удивительно похожи друг на



Балканский пирог



Восточноевропейский пирог

Рис. 9.2. Многослойность генофонда: одинаковая славянская глазурь на двух разных субстратах.

друга генетически, и выражено отличаются от некоторых своих соседей – германоязычных на западе и народов Волго-Уральского региона на востоке. А вот от других соседей – балтов – они отличаются мало. Поэтому можно – с осторожностью – выдвинуть гипотезу, что ассимилированный субстрат был по преимуществу балтским. Ведь балты были расселены очень широко. Балтский субстрат у славян (правда, наряду с финно-угорским) выявляли и антропологи. Лингвистически, балты – ближайшие родственники славян. И генофонды их весьма похожи. Поэтому если предположить, что расселяющиеся на восток славяне ассимилировали по преимуществу ранее расселившиеся там балтов, то это объяснит и сходство современных славянских и балтских генофондов друг с другом, и их отличия от окружающих их не балто-славянских групп Европы.

Северные русские. Генетическое своеобразие северных русских твердо установлено. Его причины – нет. Чаще всего это своеобразие объясняется ассимиляцией финно-угорских популяций, но специально проведенный подробный анализ заставил меня усомниться в справедливости такой трактовки.

По данным об Y-хромосоме оказалось, что северные русские похожи не столько на финно-угров, сколько на географически отдаленных балтов. По митохондриальной ДНК Русский Север тоже обнаруживает близость вовсе не с финно-уграми, а с популяциями Западной и Центральной Европы, тогда как остальные русские популяции тяготеют к генофондам Восточной Европы. По данным о фонде фамилий своеобразие северных русских хотя и четко выражено, но не выходит за орбиту основного массива русских популяций (например, ниже, чем у кубанских казаков). Такое вот «двойное подданство» – принадлежность к широкому кругу североευропейских популяций и одновременно к русскому генофонду.

Сходство генофонда Русского Севера с целым рядом генофондов северной половины Европы позволяет выдвинуть гипотезу о сохранении на этих территориях (в том числе и на Русском Севере) древнего генофонда Европы, восходящего, возможно, еще к эпохе мезолита. Частицы этого палеоевропейского генофонда унаследованы как северными русскими, так и частью финноязычных племен, расселившихся здесь еще до них, а также балтоязычными и германоязычными популяциями. В то же время традиционную гипотезу именно «финно-угорского» субстрата у северных русских отвергнуть нельзя. Возможно, дальнейшие исследования смогут дифференцировать эти две гипотезы, но на данном этапе нужно рассматривать обе.

ФИННО-УГРЫ

Про генофонд финно-угорских популяций можно написать не меньше интересного, чем про славян. Но я ограничусь лишь одной чертой. На языке популяционной генетики она описывается очень кратко – высокое межпопуляционное разнообразие при низком внутривидовом. За этими словами стоит очень много.

Высокое межпопуляционное – да ведь оно уникально высоко, по меркам Европы! Вспомним как на митохондриальном графике все прочие европейские популяции сбивались в маленькую кучку, и только финно-угры вольготно располагались в своем обширном кластере. Или как на широкогеномных графиках одно лишь появление финнов с эстонцами сжало европейскую генетическую карту в узкую полоску, а появление коми добавило новый «генетический полюс»; да и анализ предковых компонентов выделяет два разных специфически уральских компонента. Или как на картах генетических расстояний по Y-хромосоме разные финно-угорские популяции следуют самым различным паттернам. Все это следствие того, что финно-угорские популяции – генетически очень разные, хотя происхождение у них, похоже, общее.

Низкое внутривидовое – еще как низкое! Если межпопуляционное разнообразие измеряет, насколько отличаются друг от друга разные популяции, то внутривидовое – насколько различаются друг от друга разные индивиды в пределах одной и той же популяции. Мера внутривидового разнообразия – это гетерозиготность (насколько отличаются объединенные в одном человеке геномы папы и мамы – разных индивидов из одной и той же популяции). И аналог гетерозиготности для митохондриальной ДНК – гаплотипическое разнообразие – закономерно убывает в Восточной Европе с юга на север, причем север – это как раз финно-угорские популяции. Эта карта, кстати, хорошо коррелирует с отягощенностью наследственной патологией, как и должно быть – там, где низка гетерозиготность, высока вероятность выхода в гомозиготу рецессивных аллелей заболеваний.

Две стороны одной медали. И высокое межпопуляционное, и низкое внутривидовое разнообразие – это две стороны одной медали. И называется эта медаль дрейфом генов, который у финно-угров сработал куда сильнее, чем в остальных частях Европы. Причина же дрейфа, в свою очередь, отнюдь не в финно-уграх как таковых, а лишь в особенностях той территории, где им дове-

лось жить. Сниженная (по сравнению с остальной Европой) продуктивность территории обусловила сниженную плотность населения и меньшую численность элементарных популяций (то есть прямую причину дрейфа генов), а косвенно еще и усложнила дальние брачные миграции, то есть дополнительно сдвинула баланс дрейфа и миграций в пользу генетического дрейфа.

ОТДЕЛЬНЫЕ НАРОДЫ

В этой книге никак не уместились бы подробные характеристики генофонда каждого народа Европы. Но все же большинству из них уделена хотя бы одна карта – карта генетических расстояний от этого народа. Она очерчивает зону генетического сходства – показывает, население каких территорий генетически сходно с этим народом, каких – менее похоже, а каких – разительно отличается. Такие карты построены для большинства народов Восточной Европы и Балкан. Понятное дело, что на каждой карте народ похож на самого себя – ареал того народа, от которого построена карта, залит зеленым цветом сходства. Понятно и то, что соседние народы обычно более или менее похожи, а находящиеся на другом конце Европы – генетически далеки. Но при этом карты генетических расстояний оказались поразительно разнообразны. На одних область сходства занимает огромную часть Европы, на других – сужается до маленького глазка, да и очертания области сходства часто оказываются неожиданными.

Еще более поразительно, что это разнообразие сводится к нескольким устойчивым типам. Например, карты расстояний от поляков, украинцев, русских (кроме северных), мордвы-эрзи – все очерчивают одну и ту же область генетического сходства в Центральной и Восточной Европе. Эти карты различаются совсем немного и лишь тем, какие территории внутри этой единой области высокого сходства обладают «самым сходным» генофондом с тем народом, от генетических характеристик которого построена карта.

Выделяются четыре основных типа карт расстояний:

- «восточноевропейский» (поляки, кашубы, словаки, сорбы, русские, украинцы, белорусы, эрзя);
- «северо-восточный» (карелы, вепсы, эстонцы, коми ижемские и прилузские, латыши, литовцы, северные русские);
- «северо-балканский» (венгры, румыны, молдаване, гагаузы, словенцы);
- «южно-балканский» (сербы, хорваты, боснийцы, герцеговинцы, македонцы).

Каждый из этих четырех типов, похоже, выявляет устойчивый генофонд, эдакую литосферную плиту, из которых составлен генофонд восточной половины Европы. Есть, конечно, и народы, карты

расстояний от которых своеобразны и не относятся ни к одному из типов. Особенно много таких народов в Поволжье – казанские татары, мишари, мордва-мокша, чуваша – все они генетически настолько своеобразны, что карты генетических расстояний от них не похожи ни на основные типы, ни друг на друга. Это хорошо согласуется с тем, что я только что написал о большом генетическом разнообразии финно-угров, к которым тут добавились и тюркоязычные народы того же региона.

«ПОЛНЫЙ ИГРЕК»

Генетика – модная наука. И сами генетики следят за научной модой, причем нередко – и не без оснований – называют новую моду новой эрой. Сначала в популяционную генетику пришла «эра ДНК», сменившая эру классических маркеров, потом появились мтДНК и Y-хромосома и основанная на них филогеография, потом «эра чипов», а сейчас наступает и «эра полных геномов». Поэтому не будет слишком высокопарным, если я и набирающее силу полное секвенирование Y-хромосомой назову еще одной эрой. Эрой, когда Y-хромосома – самый информативный генетический маркер – становится сверхинформативна, потому что выделяются уже не десятки и сотни, а десятки тысяч гаплогрупп, многие из которых маркируют отдельные популяции или отдельные события в их истории. Более того, гаплогруппы Y-хромосомы наконец-то стало возможным точно датировать. Раньше для точности и числа выявляемых мутаций не хватало, и данные о скорости их мутирования были противоречивы. Но при полном секвенировании Y-хромосомы (точнее, суб-полном, поскольку около половины хромосомы состоит из повторов и надежному секвенированию не поддается) выявляется множество мутаций. Еще важнее, что «полнохромосомную» скорость мутирования удастся определить с приемлемой точностью. Исследования, основанные на полном секвенировании Y-хромосомы, можно условно разделить на три направления.

1) Определение скорости мутирования. Они определялись, прежде всего, классическим методом калибровок – по заселению Америки и Сардинии. То есть это аналог «эволюционной» скорости, только теперь не для Y-STR маркеров, а для SNP маркеров в данных полного секвенирования. Вторым методом был анализ числа мутаций в известных родословных (исландских, английских), соответствующий понятию «генеалогической» скорости. Нами реализован и промежуточный способ – «клановая» оценка: хотя члены клана восходят к общему предку (как при «генеалогической» оценке), клан представляет собой эволюционирующую популяцию (как при «эволюционной» оценке). Четвертый подход основан на древней ДНК образ-

цов, датированных радиоуглеродом, и оценке генетического разнообразия их современных потомков. Скорости определены шестью различными коллективами на основе этих четырех различных подходов, и консенсусная оценка – около $0,8 \cdot 10^{-9}$ мутаций на нуклеотид в год.

2) Исследования отдельных гаплогрупп. Это направление принципиально важно, потому что анализируется не только филогенетика (какие гаплогруппы существуют), но и филогеография (в каких популяциях они распространены), и для этого проводится скрининг тысяч популяционных образцов на их принадлежность к ключевым новооткрытым гаплогруппам. Несомненно, что число таких работ будет быстро расти. Уже изучена гаплогруппа **R1a** у евреев-ашкенази, и показано ее ближневосточное (а не восточно-европейское, как думали раньше) происхождение у евреев Европы. В нашей работе по гаплогруппе **G1** выделены ее субварианты (армянский, казахский, башкирский и индийский) каждый из которых характеризуется полной специфичностью для соответствующих популяций. Сейчас завершаются еще по меньшей мере четыре исследования – итальянских генети-

ков по средиземноморской гаплогруппы **E-M35**, эстонских – по североевразийской **N1c** и наши работы по центральноазиатской гаплогруппе **C3-M217** и западноевразийской **R1b**.

3) Реконструкция демографической истории. В превосходном исследовании Y-хромосомной изменчивости человечества, проведенном под руководством Тоомаса Кивисилда, обнаружено «бутылочное горлышко», через которое в конце неолита прошла мужская часть популяции (именно мужская, потому что общий эффективный размер в это время только возрастал). Также дано подробнейшее филогенетическое дерево как в целом, так и по отдельным гаплогруппам, оценены параметры разнообразия для отдельных регионов мира, применены «эволюционная» и «генеалогическая» скорости мутирования по Y-STR маркерам и многое другое.

Это направление, а также анализ древней ДНК и разработка статистических методов расшифровки популяционной истории из полногеномных данных, пожалуй, еще дорисуют написанную картину генофонда Европы и расцветят ее новыми красками.

СПИСОК ЛИТЕРАТУРЫ

1. Абдушлишвили М.Г. Антропология древнего и современного населения Грузии. Тб.: Мецниереба, 1964. 208 с.
2. Абрамова М.П. Центральный Кавказ в сарматскую эпоху // Степи европейской части СССР в скифо-сарматское время. М.: Наука. 1989. С. 268-281. (Археология СССР).
3. Агеева Р. А. Какого мы роду-племени? Народы России: имена и судьбы: Словарь-справочник. М.: Academia, 2000. 422 с.
4. Алексеев В.П. География человеческих рас. М.: Мысль, 1974. 350 с.
5. Алексеев В.П. Историческая антропология и антропогенез. М.: Наука, 1989. 446 с.
6. Алексеева Т.И. Этногенез восточных славян по данным антропологии. М.: МГУ, 1973. 331 с.
7. Алексеева Т.И. Этногенез и этническая история восточных славян по данным антропологии // Восточные славяне. Антропология и этническая история. М.: Научный Мир, 1999. С. 307-315.
8. Алексеева Т.Н., Дяченко В.Д. Глава 5. Антропологический облик // «Украинцы», серия «Народы и культуры» Отв. ред. - Н.С. Полищук, А.П. Пономарев. М. 2000. 535 с.
9. Алтухов Ю. П., Рычков Ю. Г. Популяционные системы и их структурные компоненты. Генетическая стабильность и изменчивость // Журн. общ. биол. 1970. Т. 31. № 5. С. 507-526.
10. Алтухов Ю. П. Внутривидовое генетическое разнообразие: мониторинг и принципы сохранения // Генетика. 1995. Т. 31. № 10. С. 1333 - 1357.
11. Алтухов Ю. П. Гетерозиготность генома, скорость полового созревания и продолжительность жизни // Доклады РАН. 1996. Т. 348. № 6. С. 842-845.
12. Алтухов Ю. П. Гетерозиготность генома, интенсивность метаболизма и продолжительность жизни // Доклады РАН. 1999. Т. 369. № 5. С. 704-707.
13. Ахметова В.Л., Хусаинова Р.И., Юрьев Е.Б., Туктарова И.А., Петрова Н.В., Макаров С.В., Кравчук О.И., Пай Г.В., Балановская Е.В., Гинтер Е.К., Хуснутдинова Э.К.. Анализ полиморфизма девяти ДНК-локусов ядерного генома в популяции марийцев // Генетика. 2006.
14. Багашев А.Н. Палеоантропология Западной Сибири: лесостепь в эпоху раннего железа. Новосибирск: Наука, 2000. 374 с.
15. Бадер Н.О., Церетели Л.Д. Мезолит Кавказа // Мезолит СССР. М.: Наука, 1989. С. 93-105 (Археология СССР).
16. Балаганская О.А., Балановский О.П., Лавряшина М.Б., Кузнецова М.А., Романов А.Г., Дибирова Х.Д., Фролова С.А., Захарова Т.А., Баранова Е.Е., Сабитов Ж., Нимадава П., Балановская Е.В. Генетическая структура по маркерам Y хромосомы народов Алтая (России, Казахстана, Монголии) // Вестник Московского университета. Серия XXIII. Антропология. 2011. № 2. С. 25-36.
17. Балаганская О.А., Балановская Е.В., Лавряшина М.Б., Исакова Ж.Т., Сабитов Ж.М., Фролова С.А., Романов А.Г., Дибирова Х.Д., Кузнецова М.А., Захарова Т.А., Pitchappan R., Урасин В.М., Балаганский А.Г., Баранова Е.Г., Балановский О.П. Полиморфизм Y хромосомы у тюркоязычного населения Алтае-Саян, Тянь-Шаня и Памира в контексте взаимодействия генофондов западной и восточной Евразии // Медицинская генетика. 2011. Т. 10. № 3. С. 12-22.
18. Балановская Е. В., Рычков Ю. Г. Этническая генетика: Этногеографическое разнообразие народов мира // Генетика. 1990. Т. 26. № 1. С. 114-121.
19. Балановская Е.В., Рычков Ю.Г. Этническая генетика: Адаптивная структура генофонда народов мира по данным о полиморфных генетических маркерах человека // Генетика. 1990. Т. 26. № 4. С. 739-748.
20. Балановская Е. В., Нурбаев С. Д. Компьютерная технология геногеографического изучения генофонда. III. Вычленение трендовых поверхностей // Генетика. 1995. Т. 31. № 4. С. 536-559.
21. Балановская Е. В. Генофонд народов Северной Евразии: внутренняя структура и положение в мировом генофонде // Антропология на пороге III тысячелетия / Под ред. Т. И. Алексеевой. М.: Старый сад, 2004. Т. 1. С. 358-389.
22. Балановская Е.В., Балановский О.П. Русский генофонд на Русской равнине. М.: Луч. 2007.
23. Балановская Е.В., Пежемский Д.В., Романов А.Г., Баранова Е.Е., Ромашкина М.В., Агджоян А.Т., Балаганский А.Г., Евсеева И.В., Виллемс Р., Балановский О.П. Генофонд Русского Севера: славяне? Финны? Палеоевропейцы? // Вестник Московского Университета. Серия XXIII "Антропология". 2011. №3. С. 27-58.
24. Балановский О.П., Бужилова А.П., Балановская Е.В. Русский генофонд. Геногеография фамилий // Генетика. 2001. Т. 37. № 7. С. 974-990.
25. Балановский О.П. Генофонд высоких широт Евразии или откуда пришли саамы? // Путь на север: окружающая среда и самые ранние обитатели Арктики и Субарктики (материалы международной конференции) / Под ред. А.А. Величко, С.А. Васильева. Москва. Институт географии РАН. 2008. С. 277-282.
26. Балановский О.П., Тегако О.В. Генофонд белорусов по данным о трех типах генетических маркеров – аутосомных, митохондриальных, Y хромосомы. // Актуальные вопросы антропологии. Минск 2008. Т. 2. С. 53-65.

27. Балановский О.П., Пшеничнов А.С., Фролова С.А., Чурносов М.И., Атраментова Л.А., Тегако О.В., Микулич А.И., Кузнецова М.А., Васинская О.А. Изоляция северных русских популяций на фоне восточнославянского единства: рассказывает митохондриальная ДНК // VIII Конгресс этнографов и антропологов России, г. Оренбург, 1–5 июля 2009 г.
28. Балановский О.П., Пшеничнов А.С., Фролова С.А., Васинская О.А., Дибирова Х.Д., Кузнецова М.А., Кошель С.М., Запорожченко В., Чурносов М.И., Атраментова Л.А., Утевская О., Тегако О.В., Почешхова Э.А., Микулич А.И., Виллемс Р., Балановская Е.В. Основные черты митохондриального генофонда восточных славян // Медицинская генетика. 2010. Т9. №1. С 29-37.
29. Балановский О.П., Кошель С.М., Запорожченко В.В., Пшеничнов А.С., Фролова С.А., Кузнецова М.А., Баранова Е.Е., Теучеж И.Э., Кузнецова А.А., Ромашкина М.В., Утевская О.М., Чурносов М.И., Виллемс Р., Балановская Е.В. Эколого-генетический мониторинг в популяциях человека: гетерозиготность, гаплотипическое разнообразие мтДНК и генетический груз // Генетика. 2011. Т. 47. № 11. С. 1523-1535.
30. Баранов В.С. Персонализированная медицина: ожидания, разочарования и надежды // Вестник Российской академии медицинских наук. 2011. № 9. С. 27-35.
31. Беневоленская Ю.Д., Давыдова Г.М. Псковские поозеры // Антропология современного и древнего населения Европейской части СССР. Л., 1986. С. 3-52.
32. Бермишева М. А., Тамбетс К., Виллемс Р., Хуснутдинова Э. К. Разнообразие гаплогрупп митохондриальной ДНК у народов Волго-Уральского региона // Молекулярная биология. 2002. Т. 36, № 6. С. 990-1001.
33. Бермишева М.А., Кутуев И.А., Коршунова Т.Ю., Дубова Н.А., Виллемс Р., Хуснутдинова Э.К. Филогенетический анализ мтДНК ногайцев: высокий уровень смешения материнских линий из восточной и западной Евразии. // Молекулярная биология. 2004. Т.38. №4. С. 617-624.
34. Бермишева М.А., Кутуев И.А., Спицын В.А., Виллемс Р., Батырова А.З., Коршунова Т.Ю., Хуснутдинова Э.К. Анализ изменчивости митохондриальной ДНК в популяции ороков // Генетика. 2005. Т. 41. № 1. С. 78-84.
35. Бетров Р. Происхождение и этнокультурные связи адыгов. Нальчик: Нарт, 1991. 168 с.
36. Бжания В.В. Неолит Юга. Кавказ // Неолит Северной Евразии. М: Наука, 1996. с. 73-86.
37. Бокарев Е. А. Сравнительно-историческая фонетика восточно-кавказских языков. М.: Наука, 1981. 140 с.
38. Боровкова Н.П., Шереметьева В.А., Евсюков А.Н., Спицын В.А. Закономерности распределения аллелей аполипопротеина Е (АроЕ) среди мирового народонаселения // Вестник Московского университета. Серия 23: Антропология. 2010. № 2. С. 21-35.
39. В.В. Бунак. Род Ното, его возникновение и последующая эволюция. М.: Наука, 1980. 328 с.
40. Витов М.В. Антропологические данные как источник по истории колонизации Русского Севера. М.: ИЭА РАН, 1997. 201 с.
41. Воевода М.И., Ромащенко А.Г., Ситникова В.В., Шульгина Е.О., Кобзев В.Ф. Сравнение полиморфизма митохондриальной ДНК пазырыкцев и современного населения Евразии // Археология, этнография и антропология Евразии. Новосибирск, 2000. №4. С.88-94.
42. Воевода М.И., Степанов В.А., Ромащенко А.Г., Максимов В.Н. Этногенетические особенности подверженности атеросклерозу в этнических группах Сибири (на примере гена аполипопротеина Е) // Бюллетень Сибирского отделения Российской академии медицинских наук. 2006. № 2. С. 63-72.
43. Генофонд и геногеография народонаселения. Под ред. Ю.Г. Рычкова. Т.1. Генофонд населения России и сопредельных стран. Отв. ред. Ю.Г. Рычков, Ю.П. Алтухов. СПб: Наука, 2000. 611 с.
44. Герасимова М.М., Рудь Н.М., Яблонский Л.Т. Антропология античного и средневекового населения Восточной Европы. М.: Наука, 1987. 252 с.
45. Гигинейншвили Б.К. Сравнительная фонетика дагестанских языков. Тбилиси, Изд-во ун-та, 1977. 165 с.
46. Гинтер Е. К. Влияние генетической структуры на груз наследственных болезней в русских популяциях // Вестн. РАМН. 1993. N 9. С. 23-26.
47. Гольцова Т.В., Осипова Л.П., Жаданов С.В., Виллемс Р. Влияние брачной миграции на генетическую структуру популяции нганасан Таймыра: генеалогический анализ по маркерам митохондриальной ДНК // Генетика. 2005. Т.41. №7. С.954-965.
48. Гравере Р.У. Одонтологический аспект этногенеза и этнической истории восточнославянских народов // Восточные славяне: Антропология и этническая история. М.: Науч. мир, 1999. С. 205-218.
49. Грехова Л.В., Балановская Е.В., Рычков Ю.Г. Разработка технологии создания компьютерных региональных археологических атласов: поздний палеолит Северной Евразии // Гуманитарная наука в России: Соросовские лауреаты (История. Археология. Культурная антропология и этнография). М.: 1996. С. 286-304.
50. Григоренко А.П., Боринская С.А., Янковский Н.К., Рогаев Е.И. Достижения и особенности в работе с древней ДНК и ДНК из сложных криминалистических образцов Acta Naturae (русскоязычная версия). 2009. № 3. С. 64-76.
51. Гриценко П.Е. Украинский язык и говоры // «Украинцы», серия «Народы и культуры» Отв. ред. - Н.С. Полищук, А.П. Пономарев. М. 2000. 535 с.
52. Гусар В.А. Мінливість та поліморфізм мітохондріальної ДНК української популяції // Журнал АМН України. – 2006. Т.12, №4. С.739-749.

53. Деренко М.В., Малярчук Б.А., Захаров И.А. О происхождении европеоидного компонента митохондриальных генофондов этнических групп Алтае-Саянского нагорья // *Генетика*. 2002. Т. 38. №9. С. 1292-1297.
54. Деренко М.В., Лункина А.В., Малярчук Б.А., Захаров И.А., Цэдэв Ц., Парк К.С., Чо Я.М., Ли Х.К., Чу Ч.Х. Рестрикционный полиморфизм митохондриальной ДНК у корейцев и монголов // *Генетика*. 2004. Т. 40. № 11. С. 1562-1570.
55. Деренко М.В., Чарный Я., Малярчук Б.А., Возняк М., Дамбуева И.К., Гржибовский Т., Захаров И.А. 2007. Изменчивость 15 аутосомных микросателлитных локусов ДНК в пяти популяциях коренного населения Южной Сибири // *Молекулярная биология*. 2007. Т. 41. №. 4. С. 593-600.
56. Деренко М.В., Малярчук Б.А., Возняк М., Денисова Г.А., Дамбуева И.К., Доржу Ч.М., Гржибовский Т., Захаров И.А. Распространенность мужских линий “чингизидов” в популяциях Северной Евразии // *Генетика*. 2007. Т. 43. № 3. С. 422-426.
57. Дибирова Х.Д. Роль географической подразделенности и лингвистического родства в формировании генетического разнообразия населения Кавказа (по данным об Y хромосоме). Диссертация на соискание ученой степени кандидата биологических наук. Москва, 2011. 150 с.
58. Дибирова Х.Д., Балановская Е.В., Кузнецова М.А., Фролова С.А., Васинская О.А., Почешхова Э.А., Запорожченко В.В., Дружинина Е.Г., Пшеничных А.С., Раджабов М.О., Теучеж И.Э., Схалыхо Р.А., Захарова Т.А., Евсеева И.В., Дубинецкая Е., Балановский О.П. Генетический рельеф Кавказа: четыре лингвистико-географических региона по данным о полиморфизме Y хромосомы // *Медицинская генетика*. 2010. Т.9. №10. С. 9-18.
59. Динамика популяционных генофондов при антропогенных воздействиях / Под ред. Ю. П. Алтухова. М.: Наука, 2004. 619 с.
60. Сравнительно-историческая грамматика тюркских языков. Прагюрский язык - основа. Картина мира прагюрского этноса по данным языка / Под ред. Э. Р. Тенишева и А. В. Дыбо. М.: Наука, 2006. 908 с.
61. Ефимова С.Г. Население средневековой Европы: соотношение антропологических и этнокультурных общностей // *На путях биологической истории человечества*. Т. I. М., 2002. С. 157-178.
62. Захаров И.А. 16 миллионов потомков Чингисхана // *Вавиловский журнал генетики и селекции*. 2003. № 27. С. 4.
63. Захаров И.А. Поиски “хромосомы Чингиз-хана” *Генетика*. 2010. Т. 46. № 9. С. 1276-1277.
64. Зинченко Р.А., Гинтер Е.К. Особенности медико-генетического консультирования в различных популяциях и этнических группах // *Медицинская генетика*. 2008. Т.7, №10(76). С.20-29.
65. Зинченко Р.А., Ельчинова Г.И., Балановская Е.В., Нурбаев С.Д., Гинтер Е.К. 2000. Влияние генетической структуры популяций на размеры груза моногенных наследственных болезней, в российских популяциях. *Вестник РАМН*. 5, 5-10.
66. Исмагулов О. Этническая антропология Казахстана : (Соматологическое исслед.) / О. Исмагулов. Алма-Ата: Наука, 1982. 228 с.
67. Кашибадзе В.Ф. Кавказ в антропоисторическом пространстве Евразии. Одонтологическое исследование. Ростов-на-Дону: Изд-во ЮНЦ РАН, 2006. 312 с.
68. Ковалевская В.Б. Северокавказские древности // *Степи Евразии в эпоху средневековья*. Серия “Археология СССР с древнейших времён до средневековья”, том 18 (Отв. ред. С.А. Плетнева). М.: Наука, 1981. с. 83–188.
69. Козенкова В.И. Кобанская культура Кавказа. // *Степи европейской части СССР в скифо-сарматское время*. М.: Наука, 1989. с. 252–267.
70. Козлов А.И., Балановская Е.В., Нурбаев С.Д., Балановский О.П. Геногеография первичной гиполактазии в популяциях Старого Света // *Генетика*. 1998. Т.34. № 4. С.551-561.
71. Кожекбаева Ж. М., Бородин Т.А., Боринская С. А., Гусар В. А., Фещенко С.П., Ахметова В.Л., Хусаинова Р.И., Гупало Е.Ю., Спицын В. А., Гречанина Е.Я., Хуснутдинова Э.К., Янковский Н. К. Распределение ВИЧ-протективных аллелей (CCR5 Δ 32, CCR2-64I и SDF1-3'A) в выборках русских, украинцев и белорусов. *Генетика*, 2004, Т.40. №10. С. 1394-1401.
72. Крупнов Е.И. Древняя история Северного Кавказа. М.: Изд-во Акад. наук СССР, 1960. 520 с.
73. Крупнов Е.И. О времени формирования основного ядра нартского эпоса у народов Кавказа//Сказания о нартах — эпос народов Кавказа. М.: Наука. 1969. С.15-29.
74. Кутуев И.А. Генетическая структура и молекулярная филогенетика народов Кавказа. Автореферат диссертации на соискание ученой степени доктора биологических наук. Уфа, 2010. 44 с.
75. Лимборская С.А., Хуснутдинова Э.К., Балановская Е.В. Этногеомика и геногеография народов Восточной Европы. М. Наука. 2002. 261 с.
76. Лобов А.С. Структура генофонда субпопуляций башкир. Автореферат диссертации на соискание ученой степени доктора биологических наук. Уфа, 2009. 23 с.
77. В.М. Массон, Н.Я. Мерперт, Р.М. Мунчаев, Е.К. Черныш. Энеолит СССР / отв. ред. В.М. Массон, Н.Я. Мерперт; М.: Наука, 1982 . 360 с. (Археология СССР. Свод археологических источников).
78. Марусин А.В., Степанов В.А., Спиридонова М.Г., Бохан Н.А., Мандель А.И., Харьков В.Н., Пельс Я.Р., Пузырев В.П. Полиморфизм генов этанол-метаболизирующих ферментов ADH1B, ADH7 И CYP2E1, связь с риском развития алкоголизма, коморбидного с туберкулезом, в русской популяции Западно-Сибирского региона // *Медицинская генетика*. 2006. Т. 5. № 7. С. 51-56.

79. Марусин А.В., Степанов В.А., Спиридонова М.Г., Харьков В.А., Пельс Я.Р., Пузырёв В.П. Анализ ассоциаций генетического полиморфизма ферментов, метаболизирующих этанол, с риском формирования коронарного атеросклероза // Генетика. 2007. Т. 43. № 3. С. 409-416.
80. Степи европейской части СССР в скифо-сарматское время. / Под ред. А. И. Мелюковой. М.: Наука, 1989. 464 с. (Археология СССР).
81. Моисеев В.Г. Происхождение уралоязычных народов по данным краниологии / Отв. ред. А.Г. Козинцев; РАН. МАЭ им. Петра Великого (Кунсткамера). СПб.: Наука, 1999. 132 с.
82. Моисеев В.Г., Хартанович В.И. Краниологические материалы из могильника эпохи раннего металла на Большом Оленьем острове Баренцева моря // Археология, этнография и антропология Евразии. 2012. №1 (49). С. 145-154.
83. Молодин В.И. Пазырыкская культура: проблемы этногенеза, этнической истории и исторических судеб // Археология этнография и антропология Евразии. Новосибирск, 2000. № 4. С. 131-142.
84. Мудрак О.А. Чукотско-камчатская и эскимосская плоттохронология и некоторые алтайско-эскимосские этимологии, обнаруживаемые в списке Суодеша // Аспекты компаративистики 3. М., изд. РГГУ, 2008, стр. 297-336.
85. Мунчаев Р.И. Майкопская культура // Археология. Эпоха бронзы Кавказа и Средней Азии. Ранняя и средняя бронза Кавказа. М.: Наука, 1994. С. 158-225.
86. Наследственные болезни в популяциях человека / Под ред. Е.К. Гинтера. М.: Медицина, 2002. 303 с.
87. Наумова О.Ю., Рычков С.Ю., Морозова И.Ю., Хаят С.Ш., Семиков А.В., Жукова О.В. Разнообразие митохондриальной ДНК у тоболо-иртышских сибирских татар // Генетика. 2008. Т. 44. № 2. С. 257-268.
88. Нидерле Л. Славянские древности. М.: Изд-во иностранной литературы, 1956. 453 с.
89. Орехов В.А. Характеристика митотипов представителей трех этнических групп европейской части России: Автореферат диссертации на соискание ученой степени кандидата биологических наук. М.: ИОГен РАН, 2002. 22 с.
90. Перевозчиков И.В. Проблема “третьей” расы // Горизонты антропологии: Труды международной научной конференции памяти академика В.П. Алексеева. М.: Наука, 2003. С. 97.
91. Пономарев А.П. Историко-этнографическое районирование // «Украинцы», серия «Народы и культуры» Отв. ред. – Н.С. Полищук, А.П. Пономарев. М. 2000. 535 с.
92. Постановление Правительства РФ от 24 марта 2000 г. N 255 “О Едином перечне коренных малочисленных народов Российской Федерации” (с изменениями от 30 сентября 2000 г., 13 октября 2008 г., 18 мая, 17 июня, 2 сентября 2010 г., 26 декабря 2011 г.).
93. Почешхова Э.А. Геногеографическое изучение народов Западного Кавказа. Диссертация на соискание ученой степени доктора медицинских наук. Москва, 2008. 298 с.
94. Пузырев В.П., Назаренко С.А., Одинокова О.Н., Степанов В.А. Геномная медицина: подходы и достижения // Бюллетень Сибирского отделения Российской академии медицинских наук. 2000. № 2. С. 107.
95. Пузырёв В.П., Степанов В.А., Голубенко М.В., Пузырёв К.В., Максимова Н.Р., Харьков В.Н., Спиридонова М.Г., Ноговицына А.Н. Линии мтДНК и Y-хромосомы в популяции якутов // Генетика. 2003. Т. 39. № 7. С. 975.
96. Пузырев В.П., Степанов В.А., Назаренко С.А. Геномные исследования наследственной патологии и генетического разнообразия сибирских популяций // Молекулярная биология. 2004. Т. 38. № 1. С. 129-138.
97. Пшеничных А.С. Структура генофонда украинцев по данным о полиморфизме митохондриальной ДНК и Y хромосомы. Диссертация на соискание ученой степени кандидата биологических наук. Москва, 2007. 155 с.
98. Ревазов А.А., Парадеева Г.М., Русакова Г.И. 1986. Пригодность русских фамилий в качестве «квазигенетического» маркера. Генетика. Т. 22. 4. 699-703.
99. Рычков Ю. Г., Ящук (Балановская) Е. В. Генетика и этногенез // Вопросы антропологии. 1980. № 64. С. 23-39.
100. Рычков Ю.Г., Ящук (Балановская) Е.В. Генетика и этногенез: Состояние и тенденции генетического процесса в связи с особенностями развития народонаселения Европы (зарубежной) // Вопросы антропологии. 1983. Вып. 72. С.3-17.
101. Рычков Ю.Г., Ящук Е.В. Генетика и этногенез. Историческая упорядоченность генетической дифференциации популяций человека (модель и реальность) // Вопросы антропологии. Вып. 75. М., 1985.
102. Рычков Ю. Г., Ящук Е. В. Генетическая память об этногенезе // Этнические связи народов Севера Азии и Америки по данным антропологии / Под ред. М.С. Великановой, И.М. Золотаревой : сборник. М.: Наука, 1986. С. 149-156.
103. Рычков Ю.Г. Генохронология исторических событий // Вопросы антропологии. 1986. Вып. 77. С.3-18.
104. Рычков Ю.Г., Балановская Е.В. Концепция эколого-генетического мониторинга населения России // Успехи современной генетики. Вып. 20. М.: Наука, 1996. С.3-38.
105. Рычков Ю.Г., Балановская Е.В., Нурбаев С.Д. Историческая геногеография Восточной Европы // Тр. Междунар. симпозиума памяти В.П.Алексеева “Горизонты антропологии”. М.: Русский мир, 1997. С.152-162.
106. Рычков Ю. Г., Балановская Е. В., Нурбаев С. Д., Шнейдер Ю. В. Историческая геногеография восточных славян // Восточные славяне и их соседи / Отв. ред. Т. И. Алексеева. М.: Научный мир, 1999. С. 109-135. (Изд. 2. М.: Научный мир, 2002. С. 109-134).
107. Сломинский П.А., Шадрин М.И., Спицын В.А. и др. Простой и быстрый способ определения делеции 32 пн в гене рецептора хемокинов CCR5 // Генетика. 1997. Т. 33. № 11. С. 1596-1598.

108. Соловьева Д.С., Балановская Е.В., Кузнецова М.А., Васинская О.А., Фролова С.А., Почешхова Э.А., Евсеева И.В., Болдырева М.Н., Балановский О.П. Русский генофонд: геногеография Alu-инсерций (АСЕ, АРОА1, В65, PV92, ТРА25) // Молекулярная биология. 2010. Т.44. №3. С. 447-455.
109. Сорокина И.Н., Балановская Е.В., Чурносос М.И. Генофонд населения Белгородской области. I. Дифференциация всех районных популяций по данным антропоники // Генетика. 2007. Т.43. №6. С. 841-849.
110. Спицын В. А., Макаров С. В., Пай Г. В., Кузьмина Л. П., Бычкова Л. С., Кравчук О. И. Генетический полиморфизм и профессиональные заболевания: итоги 10-летних исследований // Вестник Российской академии медицинских наук. 2000. №5. С. 27-32.
111. Спицын В.А., Макаров С.В., Пай Г.В., Бычкова Л.С. Полиморфизм в генах человека, ассоциирующихся с биотрансформацией ксенобиотиков // Вавиловский журнал генетики и селекции. 2006. Т. 10. № 1. С. 97-105.
112. Спицын В.А. Экологическая генетика человека. М.: Наука, 2008. 503 с.
113. Спицын В.А., Бацевич В.А., Ельчинова Г.И., Кобылянский Е.Д. Генетическое положение чувашей в системе финно-угорских и тюркоязычных народов // Генетика. 2009. Т. 45. № 9. С. 1270-1276.
114. Суверник Р. И., Володько Н. В., Мазунин И. О., Ельцов Н. П., Стариковская Е. Б. Генетическая история русских старожилков полярного севера Восточной Сибири по результатам анализа изменчивости мтДНК // Генетика. 2010. Т. 46. № 11. С. 1571-1579.
115. Старостин С. А. Сравнительно-историческое языкознание и лексикостатистика // Лингвистическая реконструкция и древнейшая история Востока. М.: Наука, 1989, стр. 3-39.
116. Степанов В.А., Пузырев В.П. Гаплотипы двух диаллельных локусов Y-хромосомы у коренного и пришлого населения Сибири // Генетика, 2000. Т.36. № 1. С.87-92.
117. Степанов В.А., Пузырев В.П. Анализ аллельных частот семи микросателлитных локусов Y хромосомы в трех популяциях тувинцев // Генетика. 2000. Т. 36. №2. С. 241-248.
118. Степанов В.А., Пузырев В.П. Микросателлитные гаплотипы Y-хромосомы демонстрируют отсутствие подразделенности и наличие нескольких компонентов в мужском генофонде тувинцев // Генетика. 2000. Т. 36. №3. С. 377-384.
119. Степанов В.А., Харьков В.Н., Пузырев В.П., Солтобаева Ж.О., Стегний В.Н. Гаплотипы Y-хромосомы в популяциях Средней Азии // Генетика. 2001. Т. 37. №2. С. 256-259.
120. Степанов В.А. Этногеномика и наследственные основы широко распространенных болезней // Вестник Российской академии медицинских наук. 2003. № 12. С. 85.
121. Степанов В.А., Спиридонова М.Г., Тадинова В.Н., Пузырёв В.П. Анализ генетического разнообразия популяций Северной Евразии по аутосомным микросателлитным локусам // Генетика. 2003. Т. 39. № 10. С. 1381.
122. Степанов В.А., Спиридонова М.Г., Пузырёв В.П. Сравнительное филогенетическое исследование коренных этносов Северной Евразии по панели аутосомных микросателлитных локусов // Генетика. 2003. Т. 39. № 11. С. 1564.
123. Степанов В.А., Харьков В.Н., Пузырев В.П. Эволюция и филогеография линий Y-хромосомы человека // Вавиловский журнал генетики и селекции. 2006. Т. 10. № 1. С. 57-73.
124. Султанаева З.М., Викторова Т.В., Асеев М.В., Баранов В.С., Хуснутдинова Э.К. Молекулярно-генетический анализ полиморфизма локуса DXS52 у народов Волго-Уральского региона // Генетика. 2001. Т.37, №5, С.661-667.
125. Схалыхо Р.А., Почешхова Э.А., Теучеж И.Э., Дибирова Х.Д., Агджоян А.Т., Утевская О.М., Юсупов Ю.М., Дамба Л.Д., Исакова Ж.Т., Момыналиев К.Т., Тагрили Ш.Г., Кузнецова М.А., Коньков А.С., Фролова С.А., Балановская Е.В., Балановский О.П. Тюрки Кавказа сравнительный анализ генофондов по данным о Y-хромосоме // Вестник МГУ. 2013. № 2. С. 34-48.
126. Схалыхо Р.А. Геногеография тюркоязычных народов Кавказа: анализ изменчивости Y-хромосомы. Диссертация на соискание ученой степени кандидата биологических наук. Москва, 2013. 140 с.
127. Талибов Б. Б. Сравнительная фонетика лезгинских языков. М.: Наука. 1980. 350 с.
128. Федоров Я. А. Историческая этнография Северного Кавказа. М.: Изд-во МГУ. 1983. 125с.
129. Федоров Я. А., Федоров Г. С. Ранние тюрки на Северном Кавказе. М.: Изд-во Моск. ун-та, 1978. 296 с.
130. Федорова С.А., Бермишева М.А., Виллемс Р., Максимова Н.Р., Хуснутдинова Э.К. Анализ линий митохондриальной ДНК в популяции якутов // Молекулярная биология. 2003. Т.37. №4. С.643-653.
131. Харьков В.Н., Степанов В.А., Боринская С.А., Кожекбаева Ж.М., Гусар В.А., Гречанина Е.Я., Пузырев В.П., Хуснутдинова Э.К., Янковский Н.К. Структура генофонда восточных украинцев по гаплогруппам Y-хромосомы // Генетика. 2004. Т. 40. № 3. С. 415-421.
132. Харьков В.Н., Степанов В.А., Фещенко С.П., Боринская С.А., Янковский Н.К., Пузырев В.П. Частоты диаллельных гаплогрупп Y-хромосомы у белорусов // Генетика. 2005. Т. 41. № 8. С. 1132-1136.
133. Харьков В.Н., Степанов В.А. Формирование генофонда коренного населения Сибири по данным: Y-хромосомы // Медицинская генетика. 2005. Т. 4. № 6. С. 282-283.
134. Харьков В.Н., Степанов В.А., Медведева О.Ф., Спиридонова М.Г., Воевода М.И., Тадинова В.Н., Пузырев В.П. Различия структуры генофондов северных и южных алтайцев по гаплогруппам Y-хромосомы // Генетика. 2007. Т. 43. № 5. С. 675-687.

135. Харьков В.Н., Медведева О.Ф., Степанов В.А. Разнообразие тувинского генофонда по данным Y-хромосомы // Генетика человека и патология. Томск, 2007. Вып. 8. С. 36-40.
136. Харьков В.Н., Степанов В.А., Медведева О.Ф., Спиридонова М.Г., Максимова Н.Р., Ноговицына А.Н., Пузырев В.П. Происхождение якутов: анализ гаплотипов Y-хромосомы // Молекулярная биология. 2008. Т. 42. № 2. С. 226-237.
137. Харьков В.Н., Медведева О.Ф., Лузина Ф.А., Колбаско А.В., Гафаров Н.И., Пузырев В.П., Степанов В.А. Сравнительная характеристика генофонда телеутов по данным маркеров Y-хромосомы // Генетика. 2009. Т. 45. № 8. С. 1132-1142.
138. Харьков В.Н., Хамина К.В., Медведева О.Ф., Штыгашева О.В., Степанов В.А. Разнообразие генофонда хакасов: внутриэтническая дифференциация и структура гаплогрупп Y-хромосомы // Молекулярная биология, 2011. Т. 45. № 3. С. 446-458.
139. Харьков В. Н. Структура и филогеография генофонда коренного населения Сибири по маркерам Y-хромосомы. Автореферат диссертации на соискание ученой степени доктора биологических наук. Томск, 2012. 45 с.
140. Хуснутдинова Э.К., Викторова Т.В., Ахметова В.Л., Мустафина О.Е., Фатхлисламова Р.И., Балановская Е.В., Петрова Н.В., Макаров С.В., Кравчук О.И., Пай, Гинтер Г.В. Популяционно-генетическая структура чувашей (по данным о восьми ДНК-локусах ядерного генома) // Генетика. 2003. Т. 39. № 11. С. 1550-1563.
141. Хуснутдинова Э.К., Литвинов С.С., Кутуев И.А., Юнусбаев Б.Б., Хусаинова Р.И., Ахметова В.Л., Ахатова Ф.С., Метспалу Э., Роотси С., Виллемс Р. Генофонд этнических групп Кавказа по данным комплексного исследования Y-хромосомы, митохондриальной ДНК и полногеномного анализа // Генетика, 2012, Т.48, N 6 С. 750-61.
142. Чикишева Т.А., Поздняков Д.В. Антропология населения Горного Алтая в гунно-сарматское время // Археология, этнография и антропология Евразии. Новосибирск, 2000. № 3. С. 116-131.
143. Чурносов М.И., Костоглодова И.Н., Сорокина И.Н., Жерлицына М.С., Балановская Е.В. Анализ популяционной структуры Белгородской области по комплексу биохимических и квазигенетических маркеров // Антропология на рубеже тысячелетий. Москва, 2004. Т.1. С. 443-453.
144. Чурносов М.И., Сорокина М.И., Балановская Е.В. Генофонд населения Белгородской области. Динамика индекса эндогамии в районных популяциях // Генетика. 2008. Т. 44. № 8. С. 1117-1125.
145. Шагилов А. К. Этимологический словарь адыгских (черкесских) языков. А-Н. М.: Наука, 1977. 290 с.
146. Шорохова Д.А., Степанов В.А., Удовенко Ю.Д., Новоселов В.П., Пузырев В.П. Генетическая вариабельность и дискриминирующий потенциал четырех микросателлитных локусов ДНК в русской популяции // Молекулярная биология. 2005. Т. 39. № 6. С. 965-970.
147. Янковский Н.К., Иванов Д.В., Шестопалов В.И. Функциональная геномика: зачем она медицине? Медицинская генетика. 2005. Т. 4. № 6. С. 295b-295.
148. Ярхо А.И. Алтае-Саянские тюрки. Абакан, 1949. 148 с.
149. Achilli A, Olivieri A, Pala M, Metspalu E, Fornarino S, Battaglia V, Accetturo M, Kutuev I, Khusnutdinova E, Pennarun E, Cerutti N, Di Gaetano C, Crobu F, Palli D, Matullo G, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Semino O, Villems R, Bandelt HJ, Piazza A, Torroni A. Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans. *Am J Hum Genet.* 2007 Apr; 80(4):759-68. Epub 2007 Feb 6.
150. Adams SM, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, Arroyo E, López-Parra AM, Aler M, Grifo MS, Brion M, Carracedo A, Lavinha J, Martínez-Jarreta B, Quintana-Murci L, Picornell A, Ramon M, Skorecki K, Behar DM, Calafell F, Jobling MA. The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet.* 2008 Dec; 83(6):725-36.
151. Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas AS, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Della Casa P, Dąbrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolář J, Křiška A, Lasak I, Longhi C, McGlynn G, Merkevičius A, Merkyte I, Metspalu M, Mkrtychyan R, Moiseyev V, Paja L, Pálfi G, Pokutta D, Pospieszny Ł, Price TD, Saag L, Sablin M, Shishlina N, Smrčka V, Soenov VI, Szeverényi V, Tóth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sichert-Pontén T, Brunak S, Nielsen R, Kristiansen K, Willerslev E. Population genomics of Bronze Age Eurasia. *Nature.* 2015 Jun 11;522(7555):167-72. doi: 10.1038/nature14507.
152. Alonso S, Flores C, Cabrera V, Alonso A, Marty'n P, Albarra'n C, Izagirre N, de la Ru'a C and Garcí'a O. The place of the Basques in the European Y-chromosome diversity landscape. *Eur J Hum Genet.* 2005 Dec;13(12):1293-302.
153. Alvarez JC, Johnson DL, Lorente JA, Martinez-Espin E, Martinez-Gonzalez LJ, Allard M, Wilson MR, Budowle B. Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. *Leg Med (Tokyo).* 2007 Nov;9(6):293-304. Epub 2007 Jul 5.
154. Alvarez L, Santos C, Ramos A, Pradesaba R, Francalacci P, Aluja MP. Mitochondrial DNA patterns in the Iberian Northern plateau: population dynamics and substructure of the Zamora province. *Am J Phys Anthropol.* 2010 Aug;142(4): 531-9.
155. Alvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia MT, Cuscó I, Lareu MV, García O, Pérez-Jurado L, Carracedo A, Salas A. New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS One.* 2009;4(4):e5112. Epub 2009 Apr 2.

156. Ammerman A.J., Cavalli-Sforza L.L. Neolithic Transition and the Genetics of Populations in Europe, Princeton. N.J.: Princeton University Press. 1984. 176 p.
157. Anderson B., Bankier A.T., Barrel B.G. Sequence and organization of the human mitochondrial genome. *Nature*. 1981 Apr; 290(5806):457-65.
158. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999 Oct;23(2):147.
159. Anikovich MV, Sinitsyn AA, Hoffecker JF, Holliday VT, Popov VV, Lisitsyn SN, Forman SL, Levkovskaya GM, Pospelova GA, Kuzmina IE, et al (2007) Early Upper Paleolithic in Eastern Europe and implications for the dispersal of modern humans *Science*. 2007 Jan 12;315(5809):223-6.
160. Babalini C, Martínez-Labarga C, Tolk HV, Kivisild T, Giampaolo R, Tarsi T, Contini I, Barac L, Janičević B, Martinović Klarić I, Perićić M, Sujoldzić A, Villems R, Biondi G, Rudan P, Rickards O. The population history of the Croatian linguistic minority of Molise (southern Italy): a maternal view. *Eur J Hum Genet*. 2005 Aug;13(8):902-12.
161. Balanovsky O, Zhabagin M, Agdzhoyan A, Chukhryaeva M, Zaporozhchenko V, Utevska O, Highnam G, Sabitov Z, Greenspan E, Dibirova K, Skhalyakho R, Kuznetsova M, Koshel S, Yusupov Y, Nymadawa P, Zhumadilov Z, Pocheshkhova E, Haber M, A Zalloua P, Yepiskoposyan L, Dybo A, Tyler-Smith C, Balanovska E. Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y-chromosome and reveals migrations of Iranic speakers. *PLoS One*. 2015 Apr 7;10(4):e0122968. doi: 10.1371/journal.pone.0122968. eCollection 2015.
162. Balanovsky O., Dibirova Kh., Dybo A., Mudrak O., Frolova S., Pocheshkhova E., Haber M., Platt D., Schurr T., Haak W., Kuznetsova M., Radzhabov M., Balaganskaya O., Druzhinina E., Zakharova T., Hernanz D., Zalloua P., Koshel S., Ruhlen M., Renfrew C., Wells R.S., Tyler-Smith C., Balanovska E., The Genographic Consortium. Parallel Evolution of Genes and Languages in the Caucasus Region. *Mol Biol Evol*. 2011 Oct;28(10):2905-20. Epub 2011 May 13.
163. Balanovsky O, Rootsi S, Pshenichnov A, Kivisild T, Churnosov M, Evseeva I, Pocheshkhova E, Boldyreva M, Yankovsky N, Balanovska E, Villems R. Two sources of the Russian patrilineal heritage in their Eurasian context. *Am J Hum Genet*. 2008 Jan;82(1):236-50.
164. Balanovsky O, Pocheshkhova E, Pshenichnov A, Solovieva D, Kuznetsova M, Voronko O, Churnosov M, Tegako O, Atramentova L, Lavryashina M, Evseeva I, Borinska S, Boldyreva M, Dubova N, Balanovska E. Is spatial distribution of the HIV-1-resistant CCR5Delta32 allele formed by ecological factors? *J Physiol Anthropol Appl Human Sci*. 2005 Jul;24(4):375-82.
165. Balanovsky O.P., Human genetics and Neolithic dispersal. In: The East European Plain on the Eve of Agriculture, eds. P. M. Dolukhanov, G. R. Sarson & A. M. Shukurov, British Archaeological Reports, International Series 1964, 2009, pp. 235-46.
166. Balaesque P, Bowden GR, Adams SM, Leung HY, King TE, Rosser ZH, Goodwin J, Moisan JP, Richard C, Millward A, Demaine AG, Barbujani G, Previderč C, Wilson IJ, Tyler-Smith C, Jobling MA. A predominantly neolithic origin for European paternal lineages. *PLoS Biol*. 2010 Jan 19;8(1):e1000285.
167. Bandelt HJ, Forster P, Sykes BC, Richards MB. Mitochondrial portraits of human populations using median networks. *Genetics*. 1995 Oct;141(2):743-53.
168. Bandelt HJ, Kong QP, Parson W, Salas A. More evidence for non-maternal inheritance of mitochondrial DNA? *J Med Genet*. 2005 Dec;42(12):957-60. Epub 2005 May 27.
169. Barbujani G, Bertorelle G. Genetics and the population history of Europe. *Proc Natl Acad Sci USA*. 2001 Jan 2; 98(1): 22-5.
170. Barbujani G, Bertorelle G, Chikhi L. Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet*. 1998 Feb;62(2):488-92.
171. Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, King RJ, Rootsi S, Marjanovic D, Primorac D, Hadziselimovic R, Vidovic S, Drobic K, Durmishi N, Torroni A, Santachiara-Benerecetti AS, Underhill PA, Semino O. Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. *Eur J Hum Genet*. 2009 Jun;17(6):820-30. Epub 2008 Dec 24.
172. Behar DM, Thomas MG, Skorecki K, Hammer MF, Buluygina E, Rosengarten D, Jones AL, Held K, Moses V, Goldstein D, Bradman N, Weale ME. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. *Am J Hum Genet*. 2003 Oct;73(4):768-79. Epub 2003 Sep 17.
173. Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, Herrnstadt C, Howell N, Balanovsky O, Kutuev I, Pshenichnov A, Gurwitz D, Bonne-Tamir B, Torroni A, Villems R, Skorecki K. The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. *Am J Hum Genet*. 2006 Mar;78(3):487-97. Epub 2006 Jan 11.
174. Behar DM, Rosset S, Blue-Smith J, Balanovsky O, Tzur S, Comas D, Mitchell RJ, Quintana-Murci L, Tyler-Smith C, Wells RS; Genographic Consortium. The Genographic Project public participation mitochondrial DNA database. *PLoS Genet*. 2007 Jun;3(6):e104.
175. Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S, Schurr TG, Santos FR, Quintana-Murci L, Bertranpetit J, Comas D, Tyler-Smith C, Balanovska E, Balanovsky O, Behar DM, Mitchell RJ, Soodyall H, Pitchappan

- R, Cooper A, Royyuru AK, Rosset S, Parida L, Blue-Smith J, Hernanz DS, Wells RS. The dawn of human matrilineal diversity. *Am J Hum Genet.* 2008 May;82(5):1130-40. Epub 2008 Apr 24.
176. Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, Chaubey G, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parfitt T, Hammer MF, Skorecki K, Vilems R. The genome-wide structure of the Jewish people. *Nature.* 2010 Jul 8;466(7303):238-42.
177. Behar DM, Metspalu M, Baran Y, Kopelman NM, Yunusbayev B, Gladstein A, Tzur S, Sahakyan H, Bahmanimehr A, Yepiskoposyan L, Tambets K, Khusnutdinova EK, Kushniarevich A, Balanovsky O, Balanovsky E, Kovacevic L, Marjanovic D, Mihailov E, Kouvatsi A, Triantaphyllidis C, King RJ, Semino O, Torroni A, Hammer MF, Metspalu E, Skorecki K, Rosset S, Halperin E, Vilems R, Rosenberg NA. No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Hum Biol.* 2013 Dec;85(6):859-900.
178. Belyaeva O, Bermisheva M, Khrunin A, Slominsky P, Bebyakova N, Khusnutdinova E, Mikulich A, Limborska S. Mitochondrial DNA variations in Russian and Belorussian populations. *Hum Biol.* 2003 Oct;75(5):647-60.
179. Bertoncini S, Bulayeva K, Ferri G, Pagani L, Caciagli L, Taglioli L, Semyonov I, Bulayev O, Paoli G, Tofanelli S. The dual origin of Tati-speakers from Dagestan as written in the genealogy of uniparental variants. *Am J Hum Biol.* 2012 Jul-Aug;24(4):391-9.
180. Bíró AZ, Zalán A, Völgyi A, Pamjav H. A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *Am J Phys Anthropol.* 2009 Jul;139(3):305-10.
181. Birnbaum Henrik. Common Slavic: progress and problems in its reconstruction. Cambridge Mass.: Slavica; 1975
182. Blazhek V, Novotna P. Retoromanske jazyky: prehled a klasifikace. *Linguistica Brunensia.* In Sbornik praci filozoficke fakulty brnenske univerzity 2008. A 56: 15-32.
183. Blažek V. From August Schleicher to Sergei Starostin. On the development of the tree-diagram models of the Indo-European languages. 2007;35. Available: <http://www.muni.cz/research/publications/725608> Yunusbaev et al., 2015
184. Borinskaya S, Kal'ina N, Marusin A, Faskhutdinova G, Morozova I, Kutuev I, Koshechkin V, Khusnutdinova E, Stepanov V, Puzyrev V, Yankovsky N, Rogaev E. Distribution of the alcohol dehydrogenase ADH1B*47His allele in Eurasia. *Am J Hum Genet.* 2009 Jan;84(1):89-92.
185. Bosch E, Calafell F, Gonzalez-Neira A, Flaiz C, Mateu E, Scheil HG, Huckenbeck W, Efremovska L, Mikerezi I, Xiritoris N, Grasa C, Schmidt H, Comas D. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet.* 2006 Jul;70(Pt 4):459-87.
186. Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ, Lueth F, Terberger T, Hiller J, Matsumura S, Forster P, Burger J. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science.* 2009 Oct 2;326(5949):137-40. Epub 2009 Sep 3.
187. Brandstätter A, Niederstätter H, Pavlic M, Grubwieser P, Parson W. Generating population data for the EMPOP database – an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. *Forensic Sci Int.* 2007 Mar 2;166(2-3):164-75. Epub 2006 Jul 7.
188. Brandstätter A, Egyed B, Zimmermann B, Tordai A, Padar Z, Parson W. Mitochondrial DNA control region variation in Ashkenazi Jews from Hungary. *Forensic Sci Int Genet.* 2008 Jan;2(1):e4-6. Epub 2007 Sep 24.
189. Brion M, Dupuy BM, Heinrich M, Hohoff C, Hoste B, Ludes B, Mevag B, Morling N, Niederstätter H, Parson W, Sanchez J, Bender K, Siebert N, Thacker C, Vide C, Carracedo A. A collaborative study of the EDNAP group regarding Y-chromosome binary polymorphism analysis. *Forensic Sci Int.* 2005 Oct 29;153(2-3):103-8.
190. Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* 2007;35(17):5717-28. Epub 2007 Aug 22.
191. Bulayeva K, Jorde LB, Ostler C, Watkins S, Bulayev O, Harpending H. Genetics and population history of Caucasus populations. *Hum Biol.* 2003 Dec;75(6):837-53.
192. Bulayeva KB, Jorde L, Watkins S, Ostler C, Pavlova TA, Bulayev OA, Tofanelli S, Paoli G, Harpending H. Ethnogenomic diversity of Caucasus, Daghestan. *Am J Hum Biol.* 2006 Sep-Oct;18(5):610-20.
193. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PL, Xuan Z, Rooks M, Bhattacharjee A, Brizuela L, Albert FW, de la Rasilla M, Fortea J, Rosas A, Lachmann M, Hannon GJ, Pääbo S. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science.* 2010 May 7; 328(5979): 723-5.
194. Capelli C, Redhead N, Romano V, Calč F, Lefranc G, Delague V, Megarbane A, Felice AE, Pascali VL, Neophytou PI, Poulli Z, Novelletto A, Malaspina P, Terrenato L, Berebbi A, Fellous M, Thomas MG, Goldstein DB. Population structure in the Mediterranean basin: a Y chromosome perspective. *Ann Hum Genet.* 2006 Mar;70(Pt 2):207-25.
195. Capelli C, Brisighelli F, Scarnicci F, Arredi B, Caglia' A, Vetrugno G, Tofanelli S, Onofri V, Tagliabracci A, Paoli G, Pascali VL. Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. *Mol Phylogenet Evol.* 2007 Jul;44(1):228-39. Epub 2006 Dec 13.
196. Cavalli-Sforza L.L., Menozzi P. and Piazza A. *History and Geography of Human Genes.* Princeton: Princeton University Press. 1994. 1059 p.
197. Chikhi L, Nichols RA, Barbujani G, Beaumont MA. Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A.* 2002 Aug 20;99(17):11008-13. Epub 2002 Aug 7.

198. Chirikba V.A. Common West Caucasian. The Reconstruction of its Phonological System and Parts of its Lexicon and Morphology. Leiden: Research School CNWS. 1996. 452 p.
199. Cinniodlu C, King R, Kivisild T, Kalfodlu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet.* 2004 Jan;114(2):127-48. Epub 2003 Oct 29.
200. Comrie B. *The World's Major Languages.* New York: Oxford University Press. 1987.
201. Cooper A, Poinar HN. Ancient DNA: do it right or not at all. *Science.* 2000 Aug 18;289(5482):1139.
202. Cooper A, Rambaut A, Macaulay V, Willerslev E, Hansen AJ, Stringer C. Human origins and ancient human DNA. *Science.* 2001 Jun 1;292(5522):1655-6.
203. Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet.* 2004 May;74(5):1014-22. Epub 2004 Mar 24.
204. Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R. Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. *Hum Mutat.* 2006 Aug;27(8):831-2.
205. Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon JM, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradsnik B, Assum G, Brdicka R, Kozlov AI, Efremov GD, Coppa A, Novelletto A, Scozzari R. Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol Biol Evol.* 2007 Jun;24(6):1300-11. Epub 2007 Mar 10.
206. Cruciani F, Trombetta B, Sellitto D, Massaia A, Destro-Bisol G, Watson E, Beraud Colomb E, Dugoujon JM, Moral P, Scozzari R. Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur J Hum Genet.* 2010 Jul;18(7):800-7. Epub 2010 Jan 6.
207. Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet.* 2002 Apr;70(4):1009-14. Epub 2002 Feb 13.
208. Derbeneva OA, Sukernik RI, Volodko NV, Hosseini SH, Lott MT, Wallace DC. Analysis of mitochondrial DNA diversity in the Aleuts of the Commander Islands and its implications for the genetic history of Beringia. *Am J Hum Genet.* 2002 Aug;71(2):415-21. Epub 2002 Jun 25.
209. Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet.* 2002 Apr;70(4):1009-14. Epub 2002 Feb 13.
210. Derenko MV, Malyarchuk BA, Dambueva IK, Shaikhaev GO, Dorzhu CM, Nimaev DD, Zakharov IA. Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of North Asia. *Hum Biol.* 2000 Dec;72(6):945-73.
211. Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miścicka-Sliwka D, Zakharov IA. The presence of mitochondrial haplogroup X in Altaians from South Siberia. *Am J Hum Genet.* 2001 Jul;69(1):237-41.
212. Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miścicka-Sliwka D, Woźniak M, Zakharov IA. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet.* 2003 Sep;67(Pt 5):391-411.
213. Derenko M, Malyarchuk B, Denisova GA, Wozniak M, Dambueva I, Dorzhu C, Luzina F, Miścicka-Sliwka D, Zakharov I. Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet.* 2006 Jan;118(5):591-604. Epub 2005 Oct 27.
214. Derenko M, Malyarchuk B, Denisova G, Wozniak M, Grzybowski T, Dambueva I, Zakharov I. Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J Hum Genet.* 2007;52(9):763-70. Epub 2007 Aug 17.
215. Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, Dambueva I, Zakharov I. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLoS One.* 2010 Dec 21;5(12):e15214.
216. Derenko M, Malyarchuk B, Denisova G, Perkova M, Rogalla U, Grzybowski T, Khusnutdinova E, Dambueva I, Zakharov I. Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of northern Asia and eastern Europe. *PLoS One.* 2012;7(2):e32179.
217. Der Sarkissian C, Brotherton P, Balanovsky O, Templeton JE, Llamas B, Soubrier J, Moiseyev V, Khartanovich V, Cooper A, Haak W; Genographic Consortium. Mitochondrial genome sequencing in Mesolithic North East Europe Unearths a new sub-clade within the broadly distributed human haplogroup C1. *PLoS One.* 2014 Feb 4;9(2):e87612. doi: 10.1371/journal.pone.0087612. eCollection 2014.
218. Der Sarkissian C., Balanovsky O., Brandt G., Khartanovich V., Buzhilova A., Koshel S., Zaporozhchenko V., Gronenborn D., Moiseyev V., Kolpakov E., Shumkin V., Alt K.W., Balanovska E., Cooper A., Haak W., The Genographic Consortium. Ancient DNA Reveals Prehistoric Gene-Flow from Siberia in the Complex Human Population History of North East Europe // *PLoS Genet.* 2013. 9(2): e1003296
219. Di Giacomo F, Luca F, Anagnou N, Ciavarella G, Corbo RM, Cresta M, Cucci F, Di Stasi L, Agostiano V, Giparaki M, Loutradis A, Mammi C, Michalodimitrakis EN, Papola F, Pedicini G, Plata E, Terrenato L, Tofanelli S, Malaspina P,

- Novelletto A. Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Mol Phylogenet Evol.* 2003 Sep;28(3):387-95.
220. Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrila L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarino V, Mammi C, Michalodimitrakis EN, Paoli G, Pappa KI, Pedicini G, Terrenato L, Tofanelli S, Malaspina P, Novelletto A. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet.* 2004 Oct;115(5):357-71. Epub 2004 Aug 21.
221. Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P. Forensic evaluation of mtDNA in a population from south west Switzerland. *Int J Legal Med.* 2000;113(2):89-97.
222. Dolukhanov PM. "Prehistoric revolutions" and languages in Europe. In: *The roots of peoples and languages of Northern Eurasia II and III* (A. Künnap, ed.). University of Tartu, Tartu, 2000. pp. 71-78.
223. Dubut V, Chollet L, Murail P, Cartault F, Béraud-Colomb E, Serre M, Mogentale-Profizi N. mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur J Hum Genet.* 2004 Apr;12(4):293-300.
224. Dupuy BM, Olaisen B. mtDNA sequences in the Norwegian Saami and main populations. In: Carracedo A, Brinkmann B, Bär W, editors. *Advances in forensic haemogenetics*. Berlin: Springer. 1996. pp. 23-25.
225. Embleton S. Lexicostatistics. Glottochronology: from Swadesh to Sankoff to Starostin to future horizons. In *Time Depth in Historical Linguistics*, C. Renfrew, A. McMahon & L. Trask, ed. Cambridge: The McDonald Institute for Archaeological Research Press. 2000. p. 143-165.
226. Esko T, Mezzavilla M, Nelis M, Borel C, Debniak T, Jakkula E, Julia A, Karachanak S, Khrunin A, Kisfali P, Krulisova V, Aušrelė Kučinskienė Z, Rehnström K, Traglia M, Nikitina-Zake L, Zimprich F, Antonarakis SE, Estivill X, Glavač D, Gut I, Klovins J, Krawczak M, Kučinskas V, Lathrop M, Macek M, Marsal S, Meitinger T, Melegh B, Limborska S, Lubinski J, Paolotie A, Schreiber S, Toncheva D, Toniolo D, Wichmann HE, Zimprich A, Metspalu M, Gasparini P, Metspalu A, D'Adamo P. Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. *Eur J Hum Genet.* 2013 Jun;21(6):659-65.
227. Falchi A, Giovannoni L, Calo CM, Piras IS, Moral P, Paoli G, Vona G, Varesi L. Genetic history of some western Mediterranean human isolates through mtDNA HVR1 polymorphisms. *Hum Genet.* 2006;51(1):9-14. Epub 2005 Nov 24.
228. Fechner A, Quinque D, Rychkov S, Morozowa I, Naumova O, Schneider Y, Willuweit S, Zhukova O, Roewer L, Stoneking M, Nasidze I. Boundaries and clines in the West Eurasian Y-chromosome landscape: insights from the European part of Russia. *Am J Phys Anthropol.* 2008 Sep;137(1):41-7.
229. Fedorova SA, Bermisheva MA, VILLEMS R, Maksimova NR, Khusnutdinova EK. Analysis of mitochondrial DNA haplotypes in yakut population. *Mol Biol (Mosk).* 2003 Jul-Aug;37(4):643-53.
230. Fedorova SA, Reidla M, Metspalu E, Metspalu M, Rootsi S, Tambets K, Trofimova N, Zhadanov SI, Hooshiar Kashani B, Olivieri A, Voevoda MI, Osipova LP, Platonov FA, Tomsky MI, Khusnutdinova EK, Torroni A, VILLEMS R. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol Biol.* 2013. 19;13:127.
231. Fenner JN. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol.* 2005. 128: 415-423.
232. Filippova IN, Khrunin AV, Limborska SA. Analysis of DNA variations in GSTA and GSTM gene clusters based on the results of genome-wide data from three Russian populations taken as an example. *BMC Genet.* 2012 Oct 22;13:89. doi: 10.1186/1471-2156-13-89.
233. Flores C, Maca-Meyer N, Pérez JA, González AM, Larruga JM, Cabrera VM. A predominant European ancestry of paternal lineages from Canary Islanders. *Ann Hum Genet.* 2003 Mar;67(Pt 2):138-52.
234. Flores C, Maca-Meyer N, González AM, Oefner PJ, Shen P, Pérez JA, Rojas A, Larruga JM, Underhill PA. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet.* 2004 Oct;12(10):855-63.
235. Forster P, Harding R, Torroni A, Bandelt HJ. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet.* 1996 Oct;59(4):935-45.
236. Fortson, Benjamin W. IV. *Indo-European Language and Culture: An Introduction*. Oxford: Blackwell; 2004
237. Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, Pilu R, Busonero F, Maschio A, Zara I, Sanna D, Useli A, Urru MF, Marcelli M, Cusano R, Oppo M, Zoledziewska M, Pitzalis M, Deidda F, Porcu E, Poddie F, Kang HM, Lyons R, Tarrrier B, Gresham JB, Li B, Tofanelli S, Alonso S, Dei M, Lai S, Mulas A, Whalen MB, Uzzau S, Jones C, Schlessinger D, Abecasis GR, Sanna S, Sidore C, Cucca F. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science.* 2013 Aug 2;341(6145):565-9. doi: 10.1126/science.1237947.
238. Caciagli L, Bulayeva K, Bulayev O, Bertoncini S, Taglioli L, Pagani L, Paoli G, Tofanelli S. The key role of patrilineal inheritance in shaping the genetic variation of Dagestan highlanders. *J Hum Genet.* 2009 Dec;54(12):689-94. Epub 2009 Nov 13.
239. García O, Fregel R, Larruga JM, Álvarez V, Yurrebaso I, Cabrera VM, González AM. Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. *Heredity (Edinb).* 2011 Jan;106(1):37-45. Epub 2010 Apr 21.

240. Ge J, Budowle B, Aranda XG, Planz JV, Eisenberg AJ, Chakraborty R. Mutation rates at Y chromosome short tandem repeats in Texas populations. *Forensic Sci Int Genet.* 2009 Jun;3(3):179-84. Epub 2009 Feb 14.
241. Gell-Mann M, Peiros I, Starostin SA. Lexicostatistics Compared with Shared Innovations: the Polynesian Case. In *Aspects of Comparative Linguistics*, 2008, v.3. Moscow: RSUH Publishers. p. 13-44.
242. González AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM. Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am J Phys Anthropol.* 2003 Apr;120(4):391-404.
243. Gonçalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, Underhill PA, Kivisild T, Brehm A. Y-chromosome lineages from Portugal, Madeira and Açores record elements of Sephardim and Berber ancestry. *Ann Hum Genet.* 2005 Jul;69(Pt 4):443-54.
244. Goodacre S, Helgason A, Nicholson J, Southam L, Ferguson L, Hickey E, Vega E, Stefánsson K, Ward R, Sykes B. Genetic evidence for a family-based Scandinavian settlement of Shetland and Orkney during the Viking periods. *Heredity (Edinb).* 2005 Aug;95(2):129-35.
245. Gokcumen O, Dulik MC, Pai AA, Zhadanov SI, Rubinstein S, Osipova LP, Andreenkov OV, Tabikhanova LE, Gubina MA, Labuda D, Schurr TG. Genetic variation in the enigmatic Altaian Kazakhs of South-Central Russia: insights into Turkic population history. *Am J Phys Anthropol.* 2008 Jul;136(3):278-93. doi: 10.1002/ajpa.20802.
246. Cox MP. 2008. Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol.* 80: 335-357.
247. Dulik MC, Osipova LP, Schurr TG. Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One.* 2011 Mar 11;6(3):e17548. doi: 10.1371/journal.pone.0017548.
248. Gray RD, Atkinson QD. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature.* 2003 Nov 27;426(6965):435-9.
249. Greenhill SJ, Atkinson QD, Meade A, Gray RD. The shape and tempo of language evolution. *Proc Biol Sci.* 2010 Aug 22;277(1693):2443-50. Epub 2010 Apr 7.
250. Grzybowski T, Malyarchuk BA, Derenko MV, Perkova MA, Bednarek J, Woźniak M. Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *Forensic Sci Int Genet.* 2007 Jun;1(2):141-7. Epub 2007 Mar 7.
251. Günther T, Valdiosera C, Malmström H, Ureña I, Rodríguez-Varela R, Sverrisdóttir ÓO, Daskalaki EA, Skoglund P, Naidoo T, Svensson EM, Bermúdez de Castro JM, Carbonell E, Dunn M, Storl J, Iriarte E, Arsuaga JL, Carretero JM, Götherström A, Jakobsson M. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci USA.* 2015 Sep 22;112(38):11917-22. doi: 10.1073/pnas.1509851112. Epub 2015 Sep 8.
252. Gusmão L, Sánchez-Diz P, Calafell F, Martín P, Alonso CA, Alvarez-Fernández F, Alves C, Borjas-Fajardo L, Bozzo WR, Bravo ML, Builes JJ, Capilla J, Carvalho M, Castillo C, Catanesi CI, Corach D, Di Lonardo AM, Espinheira R, Fagundes de Carvalho E, Farfán MJ, Figueiredo HP, Gomes I, Lojo MM, Marino M, Pinheiro MF, Pontes ML, Prieto V, Ramos-Luis E, Riancho JA, Souza Góes AC, Santapa OA, Sumita DR, Vallejo G, Vidal Rioja L, Vide MC, Vieira da Silva CI, Whittle MR, Zabala W, Zarrabeitia MT, Alonso A, Carracedo A, Amorim A. Mutation rates at Y chromosome specific microsatellites. *Hum Mutat.* 2005 Dec;26(6):520-8.
253. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature.* 2015 Jun 11;522(7555):207-11. doi: 10.1038/nature14317. Epub 2015 Mar 2.
254. Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CS, Brandt G, Schwarz C, Nicklisch N, Dresely V, Fritsch B, Balanovska E, Vilems R, Meller H, Alt KW, Cooper A; Members of the Genographic Consortium. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol.* 2010 Nov 9;8(11):e1000536.
255. Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, Tänzer M, Vilems R, Renfrew C, Gronenborn D, Alt KW, Burger J. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science.* 2005 Nov 11;310(5750):1016-8.
256. Haber M, Youhanna SC, Balanovsky O, Saade S, Martínez-Cruz B, Ghassibe-Sabbagh M, Shasha N, Osman R, el Bayeh H, Koshel S, Zaporozhchenko V, Balanovska E, Soria-Hernanz DF, Platt DE, Zalloua PA. mtDNA lineages reveal coronary artery disease-associated structures in the Lebanese population. *Ann Hum Genet.* 2012 Jan;76(1):1-8. doi: 10.1111/j.1469-1809.2011.00682.x. Epub 2011 Oct 20.
257. Haber M, Platt DE, Badro DA, Xue Y, El-Sibai M, Bonab MA, Youhanna SC, Saade S, Soria-Hernanz DF, Royyuru A, Wells RS, Tyler-Smith C, Zalloua PA; Genographic Consortium. Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. *Eur J Hum Genet.* 2011 Mar;19(3):334-40. Epub 2010 Dec 1.
258. Handt O, Krings M, Ward RH, Pääbo S. The retrieval of ancient human DNA sequences. *Am J Hum Genet.* 1996 Aug;59(2):368-76.

259. Hedman M, Brandstätter A, Pimenoff V, Sistonen P, Palo JU, Parson W, Sajantila A. Finnish mitochondrial DNA HVSI and HVSI-II population data. *Forensic Sci Int*. 2007 Oct 25;172(2-3):171-8. Epub 2007 Mar 2.
260. Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B. mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet*. 2001 Mar;68(3):723-37. Epub 2001 Feb 1.
261. Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefánsson K. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am J Hum Genet*. 2003 Jun;72(6):1370-88. Epub 2003 Apr 29.
262. Helgason A, Einarsson AW, Gudmundsdóttir VB, Sigurðsson Á, Gunnarsdóttir ED, Jagadeesan A, Ebenesersdóttir SS, Kong A, Stefánsson K. The Y-chromosome point mutation rate in humans. *Nat Genet*. 2015 May;47(5):453-7. doi: 10.1038/ng.3171. Epub 2015 Mar 25.
263. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. *Nat Rev Genet*. 2001 May;2(5):353-9.
264. Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A*. 1995 Jan 17;92(2):532-6.
265. Irwin J, Egyed B, Saunier J, Szamosi G, O'Callaghan J, Padar Z, Parsons TJ. Hungarian mtDNA population databases from Budapest and the Baranya county Roma. *Int J Legal Med*. 2007 Sep;121(5):377-83. Epub 2006 Dec 22.
266. Irwin J, Saunier J, Strouss K, Paintner C, Diegoli T, Sturk K, Kovatsi L, Brandstätter A, Cariolou MA, Parson W, Parsons TJ. Mitochondrial control region sequences from northern Greece and Greek Cypriots. *Int J Legal Med*. 2008 Jan;122(1):87-9. Epub 2007 May 11.
267. Qiaomei Fu, Mateja Hajdinjak, Oana Teodora Moldovan, Silviu Constantin, Swapan Mallick, Pontus Skoglund, Nick Patterson, Nadin Rohland, Iosif Lazaridis, Birgit Nickel Bence Viola, Kay Prüfer, Matthias Meyer, Janet Kelso, David Reich & Svante Pääbo. An early modern human from Romania with a recent Neanderthal ancestor // *Nature* (2015) doi:10.1038/nature14558. <http://www.nature.com/nature/journal/vaop/ncurrent/full/nature14558.html>
268. Karachanak S, Carossa V, Nesheva D, Olivieri A, Pala M, Hooshiar Kashani B, Grugni V, Battaglia V, Achilli A, Yordanov Y, Galabov AS, Semino O, Toncheva D, Torroni A. Bulgarians vs the other European populations: a mitochondrial DNA perspective. *Int J Legal Med*. 2012 Jul;126(4):497-503. Epub 2011 Jun 15.
269. Karafet TM, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F, Dashnyam B, Gerelsaikhan T, Omoto K, Hammer MF. Y chromosome markers and Trans-Bering Strait dispersals. *Am J Phys Anthropol*. 1997 Mar;102(3):301-14.
270. Karafet TM, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet*. 2001 Sep;69(3):615-28. Epub 2001 Jul 30.
271. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res*. 2008 May;18(5):830-8. Epub 2008 Apr 2.
272. Karafet TM, Bulayeva KB, Bulayev OA, Gurganova F, Omarova J, Yepiskoposyan L, Savina OV, Veeramah KR, Hammer MF. Extensive genome-wide autozygosity in the population isolates of Dagestan. *Eur J Hum Genet*. 2015. Jan. [Epub ahead of print]
273. Karlsson AO, Wallerström T, Götherström A, Holmlund G. Y-chromosome diversity in Sweden - a long-time perspective. *Eur J Hum Genet*. 2006 Aug;14(8):963-70. Epub 2006 May 24.
274. Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Talas UG, Rootsi S, Ilumäe AM, Mägi R, Mitt M, Pagani L, Puurand T, Faltyskova Z, Clemente F, Cardona A, Metspalu E, Sahakyan H, Yunusbayev B, Hudjashov G, DeGiorgio M, Loogväli EL, Eichstaedt C, Eelmets M, Chaubey G, Tambets K, Litvinov S, Mormina M, Xue Y, Ayub Q, Zoraqi G, Korneliusson TS, Akhatova F, Lachance J, Tishkoff S, Momyaliev K, Ricaut FX, Kusuma P, Razafindrazaka H, Pierron D, Cox MP, Sultana GN, Willerslev R, Muller C, Westaway M, Lambert D, Skaro V, Kovačević L, Turdikulova S, Dalimova D, Khusainova R, Trofimova N, Akhmetova V, Khidiyatova I, Lichman DV, Isakova J, Pocheshkhova E, Sabitov Z, Barashkov NA, Nymadawa P, Mihailov E, Seng JW, Evseeva I, Migliano AB, Abdullah S, Andriadze G, Primorac D, Atramantova L, Utevska O, Yepiskoposyan L, Marjanovic D, Kushniarevich A, Behar DM, Gilissen C, Vissers L, Veltman JA, Balanovska E, Derenko M, Malyarchuk B, Metspalu A, Fedorova S, Eriksson A, Manica A, Mendez FL, Karafet TM, Veeramah KR, Bradman N, Hammer MF, Osipova LP, Balanovsky O, Khusnutdinova EK, Johnsen K, Remm M, Thomas MG, Tyler-Smith C, Underhill PA, Willerslev E, Nielsen R, Metspalu M, Vilems R, Kivisild T. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res*. 2015 Apr;25(4):459-66. doi: 10.1101/gr.186684.114. Epub 2015 Mar 13.
275. Kasperaviciute D, Kucinkas V, Stoneking M. Y chromosome and mitochondrial DNA variation in Lithuanians. *Ann Hum Genet*. 2004 Sep;68(Pt 5):438-52.
276. Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edlmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, Jonkisz A, Kuzniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM, Pawłowski R, Poetsch M, Schmid D, Schmidt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziewska M, Roewer L, Dobosz T, Ploski R. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet*. 2005 Sep;117(5):428-43. Epub 2005 Jun 16.

277. Keyser C, Bouakaze C, Crubézy E, Nikolaev VG, Montagnon D, Reis T, Ludes B. Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet.* 2009 Sep;126(3):395-410. Epub 2009 May 16.
278. Khrunin AV, Khokhrin DV, Filippova IN, Esko T, Nelis M, Bebyakova NA, Bolotova NL, Klovins J, Nikitina-Zake L, Rehnström K, Ripatti S, Schreiber S, Franke A, Macek M, Krulišová V, Lubinski J, Metspalu A, Limborska SA. A genome-wide analysis of populations from European Russia reveals a new pole of genetic diversity in northern Europe. *PLoS One.* 2013;8(3):e58552.
279. King RJ, Di Cristofaro J, Kouvatsi A, Triantaphyllidis C, Scheidel W, Myres NM, Lin AA, Eissautier A, Mitchell M, Binder D, Semino O, Novelletto A, Underhill PA, Chiaroni J. The coming of the Greeks to Provence and Corsica: Y-chromosome models of archaic Greek colonization of the western Mediterranean. *BMC Evol Biol.* 2011 Mar 14;11:69.
280. King TE, Jobling MA. Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. *Mol Biol Evol.* 2009 May;26(5):1093-102. Epub 2009 Feb 9.
281. Kitchen A, Ehret C, Assefa S, Mulligan CJ. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proc Biol Sci.* 2009 Aug 7;276(1668):2703-10. Epub 2009 Apr 29.
282. Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol.* 2002 Oct;19(10):1737-51.
283. Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol.* 1999 Nov 18;9(22):1331-4.
284. Kornienko I.V., Ivanov P.L. Hypervariable region I-II data from the Russian Federation // Data deposited in Genbank. / URL: www.ncbi.nlm.nih.gov.
285. Kornienko I.V, Vodolazhski? D.I, Mikhalkovich L.S, Pavlichenko G.N, Ivanov P.L. Polymorphism of the gene for subunit 6 of the NADh dehydrogenase complex (ND6) in ethnic russian population in Russia. *Mol Biol (Mosk).* 2003 Jul-Aug;37(4):595-600. (В оригинале: Корниенко И.В., Водолажский Д.И., Михалкович Л.С., Павличенко Г.Н., Иванов П.Л. Полиморфизм гена 6-ой субъединицы комплекса НАДН-дегидрогеназы в российской популяции // Молекулярная Биология, 2003.- Т. 37.-№4,-С. 595-600.)
286. Kuipers AN. Caucasian. In: *Current Trends in Linguistics. Soviet and East European Linguistics*, T. Sebeok, ed. The Hague: Mouton. 1963. p. 315-344.
287. Kushniarevich A, Utevska O, Chuhryaeva M, Agdzhoyan A, Dibirova K, Uktveryte I, Möls M, Mulahasanovic L, Pshenichnov A, Frolova S, Shanko A, Metspalu E, Reidla M, Tambets K, Tamm E, Koshel S, Zaporozhchenko V, Atramentova L, Kučinskas V, Davydenko O, Goncharova O, Evseeva I, Churnosov M, Pocheshchova E, Yunusbayev B, Khusnutdinova E, Marjanović D, Rudan P, Rootsi S, Yankovsky N, Endicott P, Kassian A, Dybo A; Genographic Consortium, Tyler-Smith C, Balanovska E, Metspalu M, Kivisild T, Villems R, Balanovsky O. Genetic Heritage of the Balto-Slavic Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. *PLoS One.* 2015 Sep 2;10(9):e0135820. doi: 10.1371/journal.pone.0135820. eCollection 2015.
288. Lalueza-Fox C, Sampietro ML, Gilbert MT, Castri L, Facchini F, Pettener D, Bertranpetit J. Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc Biol Sci.* 2004 May 7;271(1542):941-7.
289. Lappalainen T, Laitinen V, Salmela E, Andersen P, Huoponen K, Savontaus ML, Lahermo P. Migration waves to the Baltic Sea region. *Ann Hum Genet.* 2008 May;72(Pt 3):337-48. Epub 2008 Feb 19.
290. Lappalainen T, Koivumäki S, Salmela E, Huoponen K, Sistonen P, Savontaus ML, Lahermo P. Regional differences among the Finns: a Y-chromosomal perspective. *Gene.* 2006 Jul 19;376(2):207-15. Epub 2006 Mar 18.
291. Lari M, Di Vincenzo F, Borsato A, Ghirotto S, Micheli M, Balsamo C, Collina C, De Bellis G, Frisia S, Giacobini G, Gigli E, Hellstrom JC, Lannino A, Modi A, Pietrelli A, Pilli E, Profico A, Ramirez O, Rizzi E, Vai S, Ventura D, Piperno M, Lalueza-Fox C, Barbujani G, Caramelli D, Manzi G. The Neanderthal in the karst: First dating, morphometric, and paleogenetic data on the fossil skeleton from Altamura (Italy). *J Hum Evol.* 2015 May;82:88-94. doi: 10.1016/j.jhevol.2015.02.007. Epub 2015 Mar 21.
292. Larruga JM, Diez F, Pinto FM, Flores C, González AM. Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. *Eur J Hum Genet.* 2001 Sep;9(9):708-16.
293. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet JM, Wahl J, Ayodo G, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GB, Cali F, Churnosov M, Cole DE, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon JM, Fedorova SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J, Näkkäläjärvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekgn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler

- EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, Krause J. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014. 18;513(7518):409-13.
294. Lehocý I, Baldovic M, Kádasi L, Metspalu E. A database of mitochondrial DNA hypervariable regions I and II sequences of individuals from Slovakia. *Forensic Sci Int Genet*. 2008 Sep;2(4):e53-9. Epub 2008 Feb 14.
295. Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet*. 2002 Jan; 70(1):192-206. Epub 2001 Nov 30.
296. Libert F, Cochaux P, Beckman G, Samson M, Aksenova M, Cao A, Czeizel A, Claustres M, de la Rúa C, Ferrari M, Ferrec C, Glover G, Grinde B, Güran S, Kucinskis V, Lavinha J, Mercier B, Ogur G, Peltonen L, Rosatelli C, Schwartz M, Spitsyn V, Timar L, Beckman L, Parmentier M, Vassart G. The *deltaccr5* mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Hum Mol Genet*. 1998 Mar;7(3):399-406.
297. Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L, Pericic M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzryev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventré A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Gölge M, Mikerezi I, Usanga E, Villems R. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol*. 2004 Nov;21(11):2012-21. Epub 2004 Jul 14.
298. Luca F, Di Giacomo F, Benincasa T, Popa LO, Banyko J, Kracmarova A, Malaspina P, Novelletto A, Brdicka R. Y-chromosomal variation in the Czech Republic. *Am J Phys Anthropol*. 2007 Jan;132(1):132-9.
299. Lutz S, Weisser HJ, Heizmann J, Pollak S. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med*. 1998;111(2):67-77.
300. Maca-Meyer N, Sánchez-Velasco P, Flores C, Larruga JM, González AM, Oterino A, Leyva-Cobián F. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann Hum Genet*. 2003 Jul;67(Pt 4):329-39.
301. Mallory JP, Adams DQ. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford: Oxford University Press; 2006.
302. Malyarchuk BA, Derenko MV. Mitochondrial DNA variability in Russians and Ukrainians: implication to the origin of the Eastern Slavs. *Ann Hum Genet*. 2001 Jan;65(Pt 1):63-78.
303. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Woźniak M, Miścicka-Sliwka D. Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet*. 2002 Jul;66(Pt 4):261-83.
304. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Drobnic K, Miścicka-Sliwka D. Mitochondrial DNA variability in Bosnians and Slovenians. *Ann Hum Genet*. 2003 Sep;67(Pt 5):412-25.
305. Malyarchuk B, Derenko M, Grzybowski T, Lunkina A, Czarny J, Rychkov S, Morozova I, Denisova G, Miścicka-Sliwka D. Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. *Hum Biol*. 2004 Dec;76(6):877-900.
306. Malyarchuk BA, Vanecek T, Perkova MA, Derenko MV, Sip M. Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. *Hum Biol*. 2006 Dec;78(6):681-96.
307. Malyarchuk B, Grzybowski T, Derenko M, Perkova M, Vanecek T, Lazur J, Gomolcak P, Tsybovsky I. Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Mol Biol Evol*. 2008 Aug;25(8):1651-8. Epub 2008 May 13.
308. Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, Tsybovsky I. The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One*. 2010 Apr 21;5(4):e10285.
309. Malyarchuk B, Derenko M, Denisova G, Kravtsova O. Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. *Mol Biol Evol*. 2010 Oct;27(10):2220-6. Epub 2010 May 10.
310. Manni F, Guerard E. *Barrier vs. 2.2*. (computer program). Population genetics team, Museum of Mankind (Musée de l'Homme), Paris, 2004.
311. Marchani EE, Watkins WS, Bulayeva K, Harpending HC, Jorde LB. Culture creates genetic structure in the Caucasus: autosomal, mitochondrial, and Y-chromosomal variation in Dagestan. *BMC Genet*. 2008. 17;9:47.
312. Martinez L, Mirabal S, Luis JR, Herrera RJ. Middle Eastern and European mtDNA lineages characterize populations from eastern Crete. *Am J Phys Anthropol*. 2008 Oct;137(2):213-23.
313. McEvoy B, Richards M, Forster P, Bradley DG. The Longue Durée of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. *Am J Hum Genet*. 2004 Oct;75(4):693-702. Epub 2004 Aug 12.
314. Meinilä M, Finnilä S, Majamaa K. Evidence for mtDNA admixture between the Finns and the Saami. *Hum Hered*. 2001;52(3):160-70.
315. Mendez FL, Karafet TM, Krahn T, Ostrer H, Soodyall H, Hammer MF. Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Hum Biol*. 2011 Feb;83(1):39-53.
316. Mielnik-Sikorska M1, Daca P, Malyarchuk B, Derenko M, Skonieczna K, Perkova M, Dobosz T, Grzybowski T. The history of Slavs inferred from complete mitochondrial genome sequences. *PLoS One*. 2013;8(1):e54360. doi: 10.1371/journal.pone.0054360. Epub 2013 Jan 14.

317. Mikkelsen M, Sřrensen E, Rasmussen EM, Morling N. Mitochondrial DNA HV1 and HV2 variation in Danes. *Forensic Sci Int Genet.* 2010 Jul;4(4):e87-8. Epub 2009 Aug 26.
318. Mirabal S, Regueiro M, Cadenas AM, Cavalli-Sforza LL, Underhill PA, Verbenko DA, Limborska SA, Herrera RJ. Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. *Eur J Hum Genet.* 2009. 17(10):1260-73.
319. Morozova I, Evsyukov A, Kon'kov A, Grosheva A, Zhukova O, Rychkov S. Russian ethnic history inferred from mitochondrial DNA diversity. *Am J Phys Anthropol.* 2012 Mar;147(3):341-51. doi: 10.1002/ajpa.21649. Epub 2011 Dec 20.
320. Mudrak O. A. Kamchukchee and Eskimo Glottochronology and Some Altaic Etymologies Found in the Swadesh List // *Aspects of comparative linguistics 3.* Moscow: RSUH Publ., 2008. pp. 297-336.
321. Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, Balanovsky O, Balanovska E, Rudan P, Baldovic M, Herrera RJ, Chiaroni J, Di Cristofaro J, VILLEMS R, Kivisild T, Underhill PA. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet.* 2011 Jan;19(1):95-101. Epub 2010 Aug 25.
322. Nakhleh L, Warnow T, Ringe D, Evans SN. A comparison of phylogenetic reconstruction methods on an Indo-European dataset. *Trans Philol Soc.* 2005;103: 171–192. doi:10.1111/j.1467-968X.2005.00149.x
323. Nasidze I, Stoneking M. Mitochondrial DNA variation and language replacements in the Caucasus. *Proc Biol Sci.* 2001 Jun 7;268(1472):1197-206.
324. Nasidze I, Sarkisian T, Kerimov A, Stoneking M. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Hum Genet.* 2003 Mar;112(3):255-61. Epub 2002 Dec 14.
325. Nasidze I, Ling EY, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O, Sarraf-Zadegan N, Naderi GA, Asgary S, Sardas S, Farhud DD, Sarkisian T, Asadov C, Kerimov A, Stoneking M. Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann Hum Genet.* 2004 May;68(Pt 3):205-21.
326. Nasidze I, Quinque D, Dupanloup I, Rychkov S, Naumova O, Zhukova O, Stoneking M. Genetic evidence concerning the origins of South and North Ossetians. *Ann Hum Genet.* 2004 Nov;68(Pt 6):588-99.
327. Nasidze I, Quinque D, Dupanloup I, Cordaux R, Kokshunova L, Stoneking M. Genetic evidence for the Mongolian ancestry of Kalmyks. *Am J Phys Anthropol.* 2005 Dec;128(4):846-54.
328. Nei M. Molecular population genetics and evolution. *Front Biol.* 1975;40:1-288.
329. Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S, Piskácková T, Balasćák I, Peltonen L, Jakkula E, Rehnström K, Lathrop M, Heath S, Galan P, Schreiber S, Meitinger T, Pfeufer A, Wichmann HE, Melegh B, Polgár N, Toniolo D, Gasparini P, D'Adamo P, Klovins J, Nikitina-Zake L, Kucinskas V, Kasnauskienė J, Lubinski J, Debniak T, Limborska S, Khrunin A, Estivill X, Rabionet R, Marsal S, Julif A, Antonarakis SE, Deutsch S, Borel C, Attar H, Gagnebin M, Macek M, Krawczak M, Remm M, Metspalu A. Genetic structure of Europeans: a view from the North-East. *PLoS One.* 2009;4(5):e5472.
330. Nikolayev, S. L. & Starostin, S. A. *A North Caucasian Etymological Dictionary*, Asterisk, Moscow 1994.
331. Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Pääbo S, Pritchard JK, Rubin EM. Sequencing and analysis of Neanderthal genomic DNA. *Science.* 2006 Nov 17;314(5802):1113-8.
332. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. Genes mirror geography within Europe. *Nature.* 2008 Nov 6;456(7218):98-101. Epub 2008 Aug 31.
333. Novotná P, Blažek V. Glottochronology and its application to the Balto-Slavic languages. *Baltistica* 42/2: 185–210; *Baltistica* 42/3: 323–346. *Baltistica.* 2007;42: 323–346. Bouckaert et al., 2012]
334. Onofri V., Tagliabracci A., Boschi I., Brisighelli F., Scarnicci F., Pascali V. L., Ferri G., Pelotti S. & Capelli C. 2008. Y chromosome J2 subtyping in an Italian sample: Population and forensic implications. *Forensic Sci. Int. Genet. Suppl. Ser.*,1:233–234.
335. Opdal SH, Rognum TO, Vege A, Stave AK, Dupuy BM, Egeland T. Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. *Acta Paediatr.* 1998 Oct;87(10):1039-44.
336. Orekhov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov P, Yankovsky N. Mitochondrial DNA sequence diversity in Russians. *FEBS Lett.* 1999 Feb 19;445(1):197-201.
337. Pääbo S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A.* 1989 Mar;86(6):1939-43.
338. Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. Mitochondrial DNA evidence for admixed origins of central Siberian populations. *Am J Phys Anthropol.* 2003 Mar;120(3):211-24.
339. Pakendorf B, Stoneking M. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet.* 2005;6:165-83.
340. Perić M, Lauc LB, Klarić IM, Rootsi S, Janićijević B, Rudan I, Terzić R, Colak I, Kvesić A, Popović D, Sijacki A, Behluli I, Dorđević D, Efremska L, Bajec DD, Stefanović BD, VILLEMS R, Rudan P. High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. *Mol Biol Evol.* 2005 Oct;22(10):1964-75. Epub 2005 Jun 8.
341. Pereira L, Cunha C, Amorim A. Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. *Int J Legal Med.* 2004 Jun;118(3):132-6. Epub 2004 Feb 11.

342. Pereira V, Gomes V, Amorim A, Gusmão L, João Prata M. Genetic characterization of uniparental lineages in populations from Southwest Iberia with past malaria endemicity. *Am J Hum Biol.* 2010 Sep-Oct;22(5):588-95.
343. Picornell A, Giménez P, Castro JA, Ramon MM. Mitochondrial DNA sequence variation in Jewish populations. *Int J Legal Med.* 2006 Sep;120(5):271-81. Epub 2006 May 18.
344. Pilipenko AS, Trapezov RO, Zhuravlev AA, Molodin VI, Romaschenko AG (2015) MtDNA Haplogroup A10 Lineages in Bronze Age Samples Suggest That Ancient Autochthonous Human Groups Contributed to the Specificity of the Indigenous West Siberian Population. *PLoS ONE* 10(5): e0127182. doi:10.1371/journal.pone.0127182
345. Pfeiffer H, Brinkmann B, Hühne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P. Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med.* 1999;112(5):291-8.
346. Poetsch M, Wittig H, Krause D, Lignitz E. Mitochondrial diversity of a northeast German population sample. *Forensic Sci Int.* 2003 Nov 26;137(2-3):125-32.
347. Popova SN, Slominsky PA, Pocheshnova EA, Balanovskaya EV, Tarskaya LA, Bebyakova NA, Bets LV, Ivanov VP, Livshits LA, Khusnutdinova EK, Spitsyn VA, Limborska SA. Polymorphism of trinucleotide repeats in loci DM, DRPLA and SCA1 in East European populations. *Eur J Hum Genet.* 2001 Nov;9(11):829-35.
348. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PL, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Pääbo S. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.* 2014 Jan 2;505(7481):43-9. doi: 10.1038/nature12886. Epub 2013 Dec 18. PubMed PMID: 24352235; PubMed Central PMCID: PMC4031459.
349. Pshenichnov A, Balanovsky O, Utevska O, Metspalu E, Zaporozhchenko V, Agdzhoyan A, Churnosov M, Atramentova L, Balanovska E. Genetic affinities of Ukrainians from the maternal perspective. *Am J Phys Anthropol.* 2013 Dec;152(4):543-50. doi: 10.1002/ajpa.22371. Epub 2013 Sep 30.
350. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, Sica L, Mouguiama-Daouda P, Comas D, Tzur S, Balanovsky O, Kidd KK, Kidd JR, van der Veen L, Hombert JM, Gessain A, Verdu P, Froment A, Bahuchet S, Heyer E, Dausset J, Salas A, Behar DM. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A.* 2008 Feb 5;105(5):1596-601. Epub 2008 Jan 23.
351. Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG, Behar DM. Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture *Am J Hum Genet.* 2010 Apr 9;86(4):611-20. Epub 2010 Mar 25.
352. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Vilems R, Nielsen R, Jakobsson M, Willerslev E. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature.* 2014. 2. 505(7481):87-91.
353. Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk HV, Parik J, Loogväli EL, Derenko M, Malyarchuk B, Bermisheva M, Zhadanov S, Pennarun E, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechanina E, Mikerezi I, Moisan JP, Chaventré A, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Rickards O, De Stefano GF, Papiha S, Beckman L, Janicijevic B, Rudan P, Anagnou N, Michalodimitrakis E, Koziel S, Usanga E, Geberhiwot T, Herrnstadt C, Howell N, Torroni A, Vilems R. Origin and diffusion of mtDNA haplogroup X. *Am J Hum Genet.* 2003 Nov;73(5):1178-90. Epub 2003 Oct 20.
354. Rexová K, Frynta D, Zrzavý J. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics.* 2003;19: 120–127. doi:10.1111/j.1096-0031.2003.tb00299.x
355. Richard C, Pennarun E, Kivisild T, Tambets K, Tolk HV, Metspalu E, Reidla M, Chevalier S, Giraudet S, Lauc LB, Perićić M, Rudan P, Claustres M, Journel H, Dorval I, Müller C, Vilems R, Chaventré A, Moisan JP. An mtDNA perspective of French genetic variation. *Ann Hum Biol.* 2007 Jan-Feb;34(1):68-79.
356. Richards M, Côté-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet.* 1996 Jul;59(1):185-203.
357. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Vilems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt HJ. Tracing European founder lineages in the Near Eastern mtDNA pool *Am J Hum Genet.* 2000. V.67. N.5 P. 1251-1276.
358. Richards M, Macaulay V, Torroni A, Bandelt HJ. In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet.* 2002 Nov;71(5):1168-74. Epub 2002 Sep 25.
359. Ringe D, Warnow T, Taylor A. Indo-European and Computational Cladistics. *Trans Philol Soc.* 2002;100: 59-129. doi:10.1111/1467-968X.00091

360. Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet.* 2005 Mar;116(4):279-91. Epub 2005 Jan 20.
361. Rogaev EI, Grigorenko AP, Faskhutdinova G, Kittler EL, Moliaka YK. Genotype analysis identifies the cause of the “royal disease”. *Science.* 2009 Nov 6;326(5954):817. Epub 2009 Oct 8.
362. Rogaev EI, Grigorenko AP, Moliaka YK, Faskhutdinova G, Goltsov A, Lahti A, Hildebrandt C, Kittler EL, Morozova I. Genomic identification in the historical case of the Nicholas II royal family. *Proc Natl Acad Sci U S A.* 2009 Mar 31;106(13):5258-63. Epub 2009 Feb 27.
363. Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinniodlu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Füredi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, VILLEMS R, Semino O. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet.* 2004 Jul;75(1):128-37. Epub 2004 May 25.
364. Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumäe AM, Khusnutdinova EK, Voevoda MI, Osipova LP, Stoneking M, Lin AA, Ferak V, Parik J, Kivisild T, Underhill PA, VILLEMS R. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet.* 2007 Feb;15(2):204-11. Epub 2006 Dec 6.
365. Rootsi S, Myres NM, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Varendi K, Sahakyan H, Behar DM, Khusainova R, Balanovsky O, Balanovska E, Rudan P, Yepiskoposyan L, Bahmanimehr A, Farjadian S, Kushniarevich A, Herrera RJ, Grugni V, Battaglia V, Nici C, Crobu F, Karachanak S, Hooshar Kashani B, Houshmand M, Sanati MH, Toncheva D, Lisa A, Semino O, Chiaroni J, Di Cristofaro J, VILLEMS R, Kivisild T, Underhill PA. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur J Hum Genet.* 2012 20(12):1275-82.
366. Rootsi S, Behar DM, Järve M, Lin AA, Myres NM, Passarelli B, Poznik GD, Tzur S, Sahakyan H, Pathak AK, Rosset S, Metspalu M, Grugni V, Semino O, Metspalu E, Bustamante CD, Skorecki K, VILLEMS R, Kivisild T, Underhill PA. Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. *Nat Commun.* 2013;4:2928. doi: 10.1038/ncomms3928.
367. Rosser Z. H., Zerjal T., Hurler M. E., Adojaan M., Alavantic D., Amorim A., Amos W., Armenteros M., Arroyo E., Barbujani G., Beckman G., Beckman L., Bertranpetit J., Bosch E., Bradley D. G., Brede G., Cooper G., Côrte-Real H. B., de Knijff P., Decorte R., Dubrova Y. E., Evgrafov O., Gilissen A., Glisic S., Gölge M., Hill E. W., Jeziorowska A., Kalaydjieva L., Kayser M., Kivisild T., Kravchenko S. A., Krumina A., Kucinskas V., Lavinha J., Livshits L. A., Malaspina P., Maria S., McElreavey K., Meitinger T. A., Mikelsaar A. V., Mitchell R. J., Nafa K., Nicholson J., Nørby S., Pandya A., Parik J., Patsalis P. C., Pereira L., Peterlin B., Pielberg G., Prata M. J., Previdere C., Roewer L., Rootsi S., Rubinsztein D. C., Saillard J., Santos F. R., Stefanescu G., Sykes B. C., Tolun A., VILLEMS R., Tyler-Smith C., Jobling M. A. Y chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language // *Am. J. Hum. Genet.* 2000. V. 67. P. 1526-1543.
368. Rubicz R, Zlojutro M, Sun G, Spitsyn V, Deka R, Young KL, Crawford MH. Genetic architecture of a small, recently aggregated Aleut population: Bering Island, Russia. *Hum Biol.* 2010 Dec;82(5-6):719-36.
369. Rubinstein S, Dulik MC, Gokcumen O, Zhadanov S, Osipova L, Cocca M, Mehta N, Gubina M, Posukh O, Schurr TG. Russian Old Believers: genetic consequences of their persecution and exile, as shown by mitochondrial DNA evidence. *Hum Biol.* 2008 Jun;80(3):203-37.
370. Ruhlen MA. *Guide to the World's Languages. Classification.* V. 1. Stanford: Stanford University Press. 1987.
371. Saillard J, Forster P, Lynnerup N, Bandelt HJ, Nørby S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet.* 2000 Sep;67(3):718-26. Epub 2000 Aug 2.
372. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T, Issel-Tarver L, Di Rienzo A, Pääbo S. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* 1995 Aug;5(1):42-52.
373. Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, Martí M, Bertranpetit J, Lalueza-Fox C. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci.* 2007 Sep 7;274(1622):2161-7.
374. Sánchez-Diz P, Alves C, Carvalho E, Carvalho M, Espinheira R, García O, Pinheiro MF, Pontes L, Porto MJ, Santapa O, Silva C, Sumita D, Valente S, Whittle M, Yurrebaso I, Carracedo A, Amorim A, Gusmão L; GEP-ISFG (The Spanish and Portuguese Working Group of the International Society for Forensic Genetics). Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study. *Int J Legal Med.* 2008 Nov;122(6):529-33. Epub 2008 Jul 24.
375. Santos C, Lima M, Montiel R, Angles N, Pires L, Abade A, Aluja MP. Genetic structure and origin of peopling in the Azores islands (Portugal): the view from mtDNA. *Ann Hum Genet.* 2003 Sep;67(Pt 5):433-56.
376. Schneider S, Roessli D, Excoffier L. *Arlequin vers. 2.000: a software for population genetics data analysis.* Geneva, Switzerland: Genetics and Biometry Laboratory, Department of Anthropology and Ecology. Univ. of Geneva. 2000.
377. Seguin-Orlando A, Korneliusen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, Albrechtsen A, Ko A, Margaryan

- A, Moiseyev V, Goebel T, Westaway M, Lambert D, Khartanovich V, Wall JD, Nigst PR, Foley RA, Lahr MM, Nielsen R, Orlando L, Willerslev E. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science*. 2014 Nov 28;346(6213):1113-8. doi: 10.1126/science.aaa0114. Epub 2014 Nov 6.
378. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*. 2000 Nov 10;290(5494):1155-9.
379. Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet*. 2004 May;74(5):1023-34. Epub 2004 Apr 6.
380. Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet*. 2006 Feb;78(2):202-21. Epub 2005 Dec 16.
381. Shabrova EV, Khusnutdinova EK, Tarskaia LA, Mikulich AI, Abolmasov NN, Limborska SA. DNA diversity of human populations from Eastern Europe and Siberia studied by multilocus DNA fingerprinting. *Mol Genet Genomics*. 2004 Apr;271(3):291-7.
382. Schleicher A. Compendium der vergleichenden Grammatik der indogermanischen Sprachen. Weimar: H. Böhlau; 1861.
383. Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC. Mitochondrial DNA variation in Koryaks and Ite'emen: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *Am J Phys Anthropol*. 1999 Jan;108(1):1-39.
384. Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet*. 2000 Jan;66(1):262-78.
385. Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC. mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. *Am J Hum Genet*. 1998 Nov;63(5):1473-91.
386. Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, Wallace DC. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet*. 2005 Jan;69(Pt 1):67-89.
387. StatSoft, Inc. 2001. STATISTICA (data analysis software system), version 6. www.statsoft.com.
388. Starostin G. S. A lexicostatistical approach towards reconstructing Proto-Khoisan // *Mother Tongue*, Vol. 8, 2003, pp. 81-126.
389. Stepanov VA. Genomes, populations and diseases: ethnic genomics and personalized medicine. *Acta Naturae*. 2010 Oct;2(4):15-30.
390. Stepanov VA, Melnikov AV, Lash-Zavada AY, Kharkov VN, Borinskaya SA, Tyazhelova TV, Zhukova OV, Schneider YV, Shil'nikova IN, Puzyrev VP, Rybakova AA, Yankovsky NK. Genetic variability of 15 autosomal STR loci in Russian populations. *Leg Med (Tokyo)*. 2010 Sep;12(5):256-8. Epub 2010 Jul 13.
391. Stepanov VA, Balanovsky OP, Melnikov AV, Lash-Zavada AY, Khar'kov VN, Tyazhelova TV, Akhmetova VL, Zhukova OV, Shneider YV, Shil'nikova IN, Borinskaya SA, Marusin AV, Spiridonova MG, Simonova KV, Khitrinskaya IY, Radzhabov MO, Romanov AG, Shtygasheva OV, Koshe' SM, Balanovskaya EV, Rybakova AV, Khusnutdinova EK, Puzyrev VP, Yankovsky NK. Characteristics of Populations of the Russian Federation over the Panel of Fifteen Loci Used for DNA Identification and in Forensic Medical Examination. *Acta Naturae*. 2011 Apr;3(2):56-67.
392. Sussex R, Cubberley P. *The Slavic Languages (Cambridge Language Surveys)*. Cambridge University Press; 2006
393. Swadesh M. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*. 1955, 21: 121-137.
394. Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli EL, Tolk HV, Reidla M, Metspalu E, Pliss L, Balanovsky O, Pshenichnov A, Balanovska E, Gubina M, Zhadanov S, Osipova L, Damba L, Voevoda M, Kutuev I, Bermisheva M, Khusnutdinova E, Gusar V, Grechanina E, Parik J, Pennarun E, Richard C, Chaventre A, Moisan JP, Barac L, Pericic M, Rudan P, Terzic R, Mikerezi I, Krumina A, Baumanis V, Koziel S, Rickards O, De Stefano GF, Anagnou N, Pappa KI, Michalodimitrakis E, Ferak V, Füredi S, Komel R, Beckman L, Villems R. The western and eastern roots of the Saami – the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet*. 2004 Apr;74(4):661-82. Epub 2004 Mar 11.
395. Tetzlaff S, Brandstätter A, Wegener R, Parson W, Weirich V. Mitochondrial DNA population data of HVS-I and HVS-II sequences from a northeast German sample. *Forensic Sci Int*. 2007 Oct 25;172(2-3):218-24. Epub 2007 Feb 28.
396. Tillmar AO, Coble MD, Wallerström T, Holmlund G. Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations. *Int J Legal Med*. 2010 Mar;124(2):91-8. Epub 2009 Jul 10.
397. Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet*. 2004;49(4):187-93. Epub 2004 Mar 2.

398. Tofanelli S, Ferri G, Bulayeva K, Caciagli L, Onofri V, Taglioli L, Bulayev O, Boschi I, Alù M, Berti A, Rapone C, Beduschi G, Luiselli D, Cadenas AM, Awadelkarim KD, Mariani-Costantini R, Elwali NE, Verginelli F, Pilli E, Herrera RJ, Gusmão L, Paoli G, Capelli C. J1-M267 Y lineage marks climate-driven pre-historical human displacements. *Eur J Hum Genet.* 2009 Nov;17(11):1520-4. Epub 2009 Apr 15.
399. Tolk HV, Barac L, Pericic M, Klaric IM, Janicijevic B, Campbell H, Rudan I, Kivisild T, VILLEMS R, Rudan P. The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications. *Eur J Hum Genet.* 2001 Sep;9(9):717-23.
400. Tonks S., Winney B.J., Evseeva I. Comparison of sex-linked and autosomal markers in Orkney and other North European populations // Data deposited in Genbank. 2006. / URL: www.ncbi.nlm.nih.gov.
401. Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus ML, Bonn -Tamir B, Scozzari R. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet.* 1998 May;62(5):1137-52.
402. Turchi C, Buscemi L, Previder  C, Grignani P, Brandst tter A, Achilli A, Parson W, Tagliabracci A; Ge.F.I. Group. Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing. *Int J Legal Med.* 2008 May;122(3):199-204. Epub 2007 Oct 20.
403. Underhill P.A., Myres N.M., Rootsi S., Chow C.T., Lin A.A., Ollilar R.P., King R., Zhivotovsky L.A., Balanovsky O., Pshenichnov A., Ritchie K.H., Cavalli-Sforza L.L., Kivisild T., Villems R., Woodward S.R., 2007. New Phylogenetic Relationships for Y-chromosome Haplogroup I: Reappraising its Phylogeography and Prehistory. In: Mellars P, Boyle K, Bar-Yosef O, Stringer C, Eds, *Rethinking the Human Revolution*, McDonald Institute for Archaeological Research, Cambridge, UK, 2007, pp. 33-42.
404. Underhill PA, Myres NM, Rootsi S, Metspalu M, Zhivotovsky LA, King RJ, Lin AA, Chow CE, Semino O, Battaglia V, Kutuev I, J rve M, Chaubey G, Ayub Q, Mohyuddin A, Mehdi SQ, Sengupta S, Rogaev EI, Khusnutdinova EK, Pshenichnov A, Balanovsky O, Balanovska E, Jeran N, Augustin DH, Baldovic M, Herrera RJ, Thangaraj K, Singh V, Singh L, Majumder P, Rudan P, Primorac D, Villems R, Kivisild T. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur J Hum Genet.* 2010 Apr;18(4):479-84. Epub 2009 Nov 4.
405. Underhill PA, Poznik GD, Rootsi S, J rve M, Lin AA, Wang J, Passarelli B, Kanbar J, Myres NM, King RJ, Di Cristofaro J, Sahakyan H, Behar DM, Kushniarevich A, Sarac J, Saric T, Rudan P, Pathak AK, Chaubey G, Grugni V, Semino O, Yepiskoposyan L, Bahmanimehr A, Farjadian S, Balanovsky O, Khusnutdinova EK, Herrera RJ, Chiaroni J, Bustamante CD, Quake SR, Kivisild T, Villems R. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet.* 2015; 23(1):124-31.
406. Verbenko DA, Kekeeva TV, Pogoda TV, Khusnutdinova EK, Mikulich AI, Kravchenko SA, Livshits LA, Bebyakova NA, Limborska SA. Allele frequencies for D1S80 (pMCT118) locus in some East European populations. *J Forensic Sci.* 2003 Jan;48(1):207-8. .
407. Verbenko DA, Pogoda TV, Spitsyn VA, Mikulich AI, Bets LV, Bebyakova NA, Ivanov VP, Abolmasov NN, Pocheshkhova EA, Balanovskaya EV, Tarskaya LA, Sorensen MV, Limborska SA. Apolipoprotein B 3'-VNTR polymorphism in Eastern European populations. *Eur J Hum Genet.* 2003 Jun;11(6):444-51.
408. Verbenko DA, Slominsky PA, Spitsyn VA, Bebyakova NA, Khusnutdinova EK, Mikulich AI, Tarskaia LA, Sorensen MV, Ivanov VP, Bets LV, Limborska SA. Polymorphisms at locus D1S80 and other hypervariable regions in the analysis of Eastern European ethnic group relationships. *Ann Hum Biol.* 2006 Sep-Dec;33(5-6):570-84.
409. Varzari A., Kharkov V., Stephan W., Dergachev V., Puzyrev V., Weiss E., Stepanov V. Searching for the origin of Gagauzes: Inferences from Y-chromosome analysis. *Am J. Hum. Biol.* 2009. V. 21. № 3. P. 326-336.
410. Vernot B, Akey JM Resurrecting Surviving Neandertal Lineages from Modern Human Genomes *Science*, 29 January 2014 / Page 8 / 10.1126/science.1245938
411. Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV, Wallace DC, Sukernik RI. Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocenic peopling of the Americas. *Am J Hum Genet.* 2008 May;82(5):1084-100. Epub 2008 May 1.
412. Wiik K. Eurooppalaisten juuret. Atenakustannus Oy, Jyv skyl  2002.
413. Wilson IJ, Weale ME, Balding DJ. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J. R. Statist. Soc.* Vol. 166, № 2. 2003, pp. 155-201.
414. Womble WH. 1951. Differential systematic. *Science.* 114: 315-322.
415. Xing J, Watkins WS, Witherspoon DJ, Zhang Y, Guthery SL, Thara R, Mowry BJ, Bulayeva K, Weiss RB, Jorde LB. Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Res.* 2009 May;19(5):815-25.
416. Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurler ME, Yang H, Tyler-Smith C: Modelling male prehistory in East Asia using BATWING. In *Simulations, Genetics and Prehistory*. Edited by: Matsumura S, Forster P, Renfrew C. Cambridge: McDonald Institute; 2008:81-90.
417. Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet.* 2002 Mar;70(3):635-51. Epub 2002 Feb 8.
418. Yunusbayev B, Kutuev I, Khusainova R, Guseinov G, Khusnutdinova E. Genetic structure of Dagestan populations: a study of 11 Alu insertion polymorphisms. *Hum Biol.* 2006 Aug;78(4):465-76.

419. Yunusbayev B, Metspalu M, Järve M, Kutuev I, Rootsi S, Metspalu E, Behar DM, Varendi K, Sahakyan H, Khusainova R, Yepiskoposyan L, Khusnutdinova EK, Underhill PA, Kivisild T, Villems R. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol.* 2012 Jan;29(1):359-65. Epub 2011 Sep 13.
420. Yunusbayev B, Metspalu M, Metspalu E, Valeev A, Litvinov S, Valiev R, Akhmetova V, Balanovska E, Balanovsky O, Turdikulova S, Dalimova D, Nymadawa P, Bahmanimehr A, Sahakyan H, Tambets K, Fedorova S, Barashkov N, Khidiyatova I, Mihailov E, Khusainova R, Damba L, Derenko M, Malyarchuk B, Osipova L, Voevoda M, Yepiskoposyan L, Kivisild T, Khusnutdinova E, Villems R. The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet.* 2015 Apr 21;11(4):e1005068. doi: 10.1371/journal.pgen.1005068. eCollection 2015.
421. Zakharov IA, Derenko MV, Maliarchuk BA, Dambueva IK, Dorzhu CM, Rychkov SY. Mitochondrial DNA variation in the aboriginal populations of the Altai-Baikal region: implications for the genetic history of North Asia and America. *Ann N Y Acad Sci.* 2004 Apr;1011:21-35.
422. Zhivotovsky LA, Underhill PA, Cinniodlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet.* 2004 Jan;74(1):50-61. Epub 2003 Dec 19.
423. Zhivotovsky LA, Veremeichyk VM, Mikulich AI, Udina IG, Atramentova LA, Kotova SA, Kartel NA, Tsybovsky IS. A comprehensive population survey on the distribution of STR frequencies in Belarus. *Forensic Sci Int.* 2007 Oct 25;172(2-3):156-60. Epub 2007 Feb 15.
424. Zhivotovsky LA, Akhmetova VL, Fedorova SA, Zhirkova VV, Khusnutdinova EK. An STR database on the Volga-Ural population. *Forensic Sci Int Genet.* 2009 Sep;3(4):e133-6. Epub 2008 Dec 23.
425. Zhivotovsky LA, Malyarchuk BA, Derenko MV, Wozniak M, Grzybowski T. Developing STR databases on structured populations: the native South Siberian population versus the Russian population. *Forensic Sci Int Genet.* 2009 Sep;3(4):e111-6. Epub 2008 Sep 12.
426. Zimmermann B, Brandstätter A, Duftner N, Niederwieser D, Spiroski M, Arsov T, Parson W. Mitochondrial DNA control region population data from Macedonia. *Forensic Sci Int Genet.* 2007 Dec;1(3-4):e4-9. Epub 2007 May 9.
427. Zlojutro M, Rubicz R, Devor EJ, Spitsyn VA, Makarov SV, Wilson K, Crawford MH. Genetic structure of the Aleuts and Circumpolar populations based on mitochondrial DNA sequences: a synthesis. *Am J Phys Anthropol.* 2006 Mar;129(3):446-64.

СОДЕРЖАНИЕ

Предисловие.....	3
Благодарности	4
Глава 1. СОЗДАНИЕ БАЗ ДАННЫХ О ГЕНОФОНДАХ МИРА	5
1.1. База данных по Y-хромосоме	5
Краткая характеристика БД Y-base	5
Программная реализация БД Y-base	8
Программа Harlomat для поиска сходных гаплотипов	15
Предиктор SNP-гаплогруппы по STR-гаплотипу	19
1.2. Базы данных по мтДНК и аутосомным маркерам	21
Краткая характеристика БД MURKA	21
Программирование и интерфейс	22
Базы данных по аутосомным маркерам	24
Применение созданных баз данных	26
1.3. Собственные данные	30
Стратегия сбора выборок	31
Банк образцов «Коренное население Северной Евразии»	32
Изученные популяции Восточной Европы, Кавказа, Центральной Азии и Сибири	32
Анализ древней ДНК	38
Особенности генотипирования Y-хромосомы и мтДНК	39
1.4. Картографический и статистический анализ	44
Программа GeneGeo	45
Этап 1: построение карт отдельных признаков	46
Этап 2: анализ отдельной карты	49
Этап 3: расчет обобщенных карт	51
Создание картографического образа	51
Методы статистического анализа	54
1.5. Дискуссии о датировках	56
Методы филогенетического анализа	56
Датировки по гаплотипам	57
Датировки по полным сиквенсам Y-хромосомы	60
Научные дискуссии	62
Околонаучные и лженаучные дискуссии	63
Датировки в этой книге	66
Глава 2. СТРУКТУРА ГЕНОФОНДА ЕВРОПЫ В ЗЕРКАЛЕ Y-ХРОМОСОМЫ	70
2.1. Предшествовавшие исследования	70
Обобщающие и региональные работы	70
Гомогенность и гетерогенность Европы	71
Предыдущие исследования основных гаплогрупп	72
2.2. Генеогеография гаплогрупп Y-хромосомы в Европе	75
Организация данных	75
Гаплогруппа R1a и ее субварианты	80
Гаплогруппа R1b и ее субварианты	84
Гаплогруппа I и ее субварианты	93
Гаплогруппа N и ее субварианты	95
Гаплогруппа E и ее субварианты	97
Гаплогруппа G и ее субварианты	100
Гаплогруппа J и ее субварианты	102
Редкие гаплогруппы L, Q, T	104
Y-генофонд Европы – принцип паззлов	106
2.3. Генетические взаимоотношения популяций: три уровня	107
Массивы данных для статистического анализа	107
Региональный уровень	110
Этнический уровень: 17 гаплогрупп, многомерное шкалирование	112
Этнический уровень: 17 гаплогрупп, главные компоненты	115

Этнический уровень: 40 гаплогрупп, многомерное шкалирование	118
Субэтнический уровень: этнические облака	119
2.4. Географическая и лингвистическая структурированность – закономерности	124
Глава 3. СТРУКТУРА ГЕНОФОНДА ЕВРОПЫ В ЗЕРКАЛЕ МИТОХОНДРИАЛЬНОЙ ДНК	129
3.1. Предшествовавшие исследования изменчивости мтДНК в Европе	130
3.2. Генетические взаимоотношения популяций: три масштаба	132
3.3. Географическая и лингвистическая структурированность – тенденции	136
3.4. Гаплотипическое разнообразие и мониторинг генофонда	139
Гаплотипическое разнообразие мтДНК	139
Эколого-генетический мониторинг	141
Глава 4. СТРУКТУРА ГЕНОФОНДА ЕВРОПЫ В ЗЕРКАЛЕ ПОЛНОГЕНОМНЫХ МАРКЕРОВ	147
4.1. Особенности широкогеномного анализа	147
4.2. Генетические взаимоотношения популяций: три вариации на тему	149
4.3. Анализ предковых компонентов (ADMIXTURE)	153
Глава 5. ПАНОРАМА НАРОДОВ НА ФОНЕ ЕВРОПЫ	160
5.1. Народы Северо-Восточной Европы (серия I)	161
5.2. Восточные и западные славяне (серия II)	169
5.3. Неславянские народы Восточной Европы (серия III)	177
5.4. На севере Балкан (серия IV)	184
5.5. Южные славяне (серия V)	188
5.6. Обрамление Европы (серия VI)	192
Глава 6. ГЕНОФОНД СЛАВЯН	195
6.1. Основные черты балто-славянского генофонда	195
Три зеркала балто-славянского генофонда	196
«Предковые» генетические компоненты	199
Общие фрагменты в геномах	200
Древо славянских языков и его проекция на генофонд	202
Синтез	205
6.2. Отдельные черты генофонда восточных славян	208
6.3. Митохондриальный генофонд украинцев	215
Как формировались популяционные выборки украинцев	215
Украинцы на фоне Европы: карты генетических расстояний	217
Анализ методом «географических» групп гаплотипов	219
6.4. Проблема своеобразия генофонда Русского Севера	223
Квазигенетические маркеры	224
Аутосомные маркеры	225
Митохондриальная ДНК	226
Y-хромосома	228
Глава 7. ГЕНОФОНДЫ И ЯЗЫКИ	231
7.1. Структура и история народонаселения Кавказа	231
Народонаселение Кавказа по данным археологии и антропологии	231
Народонаселение Кавказа по данным лингвистики	233
Методология лингвистических датировок	234
7.2. Сравнение лингвистической и генетической структуры Кавказа	236
Сравнение методом дендрограмм	236
Сравнение методом генетических границ	237
Сравнение матриц расстояний	238
Сравнение межпопуляционного разнообразия	240
7.3. Филогенетические датировки на Кавказе	241
Датировки популяционно-специфичных кластеров гаплотипов	241
Сравнение исторических, лингвистических и генетических данных	245
Параллелизм лингвистической и генетической микроэволюции	245
7.4. Возможности полного секвенирования Y-хромосомы	247

Обзор методов и первых результатов	247
Исследование гаплогруппы G1: скорость мутаций и миграции иранской группы	249
7.5. Три модели взаимосвязи генофондов и языков: Кавказ, славяне, тюрки	257
Глава 8. ДРЕВНЯЯ ДНК ЕВРОПЫ	261
8.1. Особенности анализа древней ДНК	261
8.2. Средний палеолит (неандертальцы)	266
8.3. Верхний палеолит и мезолит	271
Геном из Восточной Европы (Костенки)	271
Европеоидный компонент в Сибири (Мальта)	274
Мезолит Восточной Европы (Южный Олений остров)	277
8.4. Неолитизация Европы	284
Первые земледельцы (митохондриальный генофонд)	284
Три источника генофонда европейцев	289
Неолитизация в Пиренеях	293
8.5. Эпоха металла	297
Ранний металл на севере Европы	297
Бронзовый век в Европе	299
Бронзовый век в Евразии	309
Скифы Дона	314
Заключение	316
Данные и методы их анализа	317
Структура генофонда Европы	320
История генофонда Европы	325
О славянах и прочих частностях	327
Список литературы	332
Содержание	352

Балановский Олег Павлович

ГЕНОФОНД ЕВРОПЫ

М.: Товарищество научных изданий КМК. 2015. 354 с.

Отпечатано в ООО "Галлея-принт"

Подписано в печать 20.11.2015. Тираж 250 экз.

Объем 44 уч. изд. л.



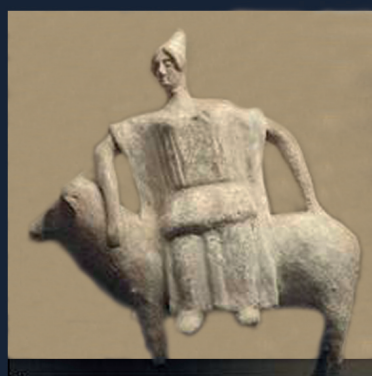
Автор «Генофонда Европы»

Балановский Олег Павлович,
доктор биологических наук,
заведующий лабораторией геномной
географии Института общей генетики
им. Н.И. Вавилова РАН.
Автор более 300 научных публикаций.



Структура генофонда Европы

На моих глазах карта Европы показывала не только таинственные перемены, но и удивительную неизменность. Неизменность была в картах ее современного генофонда – как ни менялась научная мода, как ни расширялись технологии анализа, европейский генофонд являл все те же черты – однородность большую, чем другие регионы мира, а его изменчивость была мала и следовала очертаниям «морей и полуостровов».



Терракотовая статуэтка
«Похищение Европы»,
Беотия, 470 лет до н.э.

История генофонда Европы

Перемены же случаются ежегодно – но касаются они генофондов былых эпох, реконструкций генетической карты древней Европы. В последние годы, когда анализ древней ДНК набрал силу и набрал объем выборок, калейдоскоп непохожих кадров – моментальных снимков генофондов разных эпох – начал складываться в захватывающий фильм истории генофондов Европы.

ISBN 978-5-9907157-0-7



9 785990 715707