Как складывался генофонд славян и балтов

Олег Балановский, Надежда Маркина

Реконструкция генетической и лингвистической истории балто-славянских популяций

Эта наиболее полная работа по генофонду славянских и балтских народов подводит итоги многолетних исследований. Для реконструкции давней истории народов, говорящих на родственных языках, применен междисциплинарный подход. Генетики и лингвисты проследили пути формирования генофонда всех групп славян и балтов одновременно по трем генетическим системам: по Y-хромосоме (отцовские линии наследования), по митохондриальной ДНК (материнские линии наследования) и по широгеномным данным об аутосомных маркерах (где отцовские и материнские линии представлены равноправно). Прослежено, какие местные популяции впитывал генофонд славян при их расселении по Европе: именно этот глубинный субстрат сформировал основные различия генофондов разных ветвей славян. Корреляция генетического разнообразия с лингвистическим оказалась велика, но еще больше — с географическим соседством популяций. Результатом исследования стало и уточнение древа балто-славянских языков.

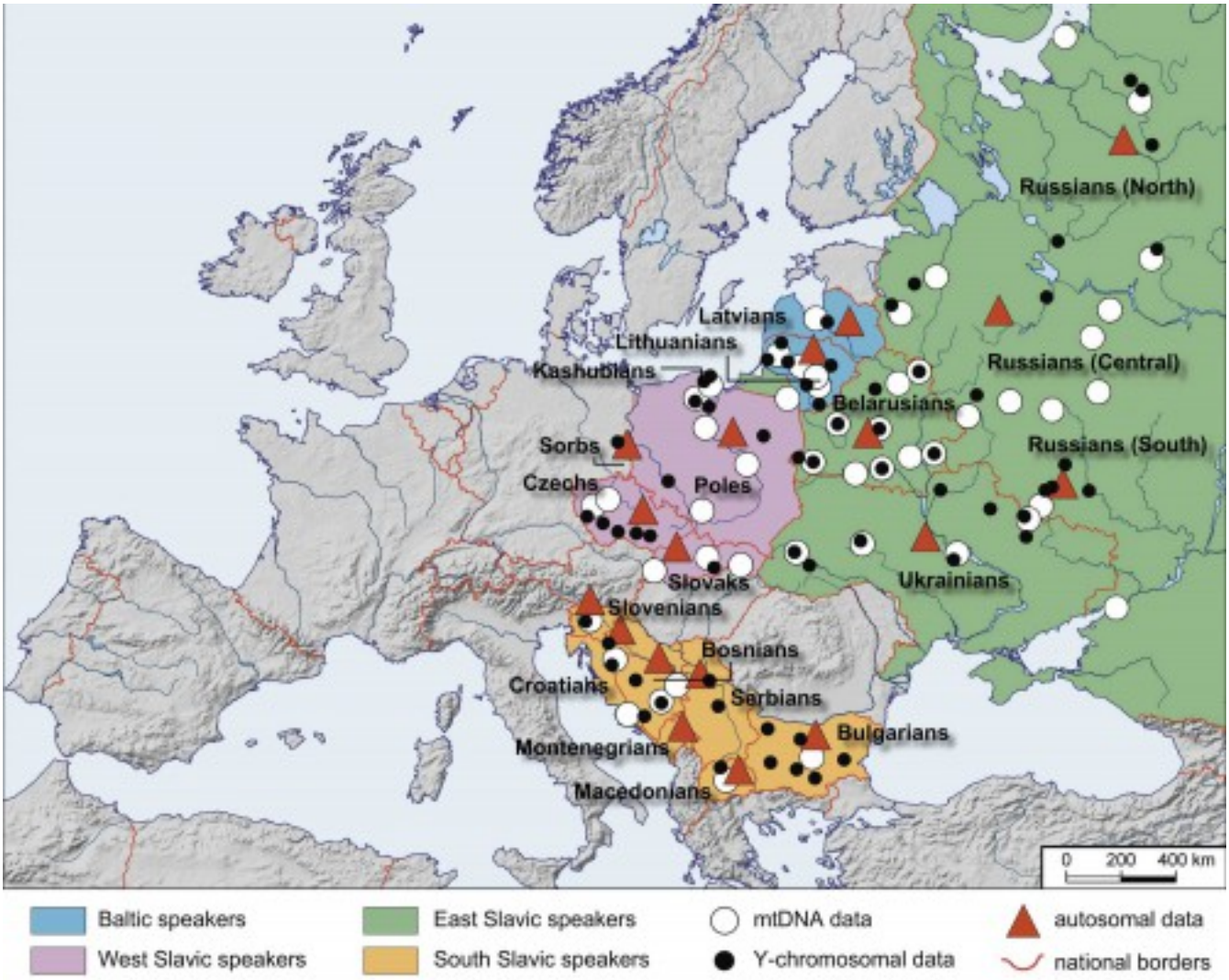
Отклики на исследование в СМИ и на научно-популярных сайтах — в конце текста

ответ О.П.Балановского А.А.Клесову читайте здесь

Формирование генофонда балто-славянских популяций исследовала большая международная группа генетиков и лингвистов. Статья с результатами их работы опубликована в журнале PLoS ONE. Исследование проведено под руководством доктора биол. наук О.П. Балановского (Институт общей генетики и Медико-генетический научный центр) и академика Рихарда Виллемса (Эстонский биоцентр и Тартусский университет). В нем участвовали исследователи из многих стран, в которых славянские и балтские народы составляют большинство населения — России, Украины, Белоруссии, Литвы, Хорватии, Боснии и Герцеговины, а также ученые Эстонии, Великобритании и консорциум международного проекта Genographic. Это наиболее полная работа по генофонду славянских и балтских народов подводит итог многолетним исследованиям многочисленных авторов статьи и учитывает данные других научных коллективов.

На балто-славянских языках говорит примерно треть современного населения Европы, а по площади балтские и славянские народы занимают около половины Европы. Лингвисты сходятся во мнении, что балтские и славянские языки не только родственны, но и имеют общий корень в семье индоевропейских языков. По их оценкам прото-балто-славянский язык отделился от других индоевропейских языков в интервале от 7000 до 4500 лет назад, и произошло это, вероятнее всего, в Центральной Европе. Расхождение балтской и славянской языковых ветвей датируется временем 3500-2500 назад. Дальнейшее разделение славянских языков происходило уже относительно недавно — 1700-1300 лет назад. 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 разноязыких популяций.

Но как эти изменения в культуре Европы, фиксируемые распространением славянских языков, повлияли на генофонд Европы? Именно это стало главным вопросом исследования. Ведь до сих пор оставалась недостаточно изученной генетическая история балто-славянских популяций и их взаимодействие с генофондами популяций, говоривших на других языках — финно-угорских, германских, тюркских.



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).

Что изучали

Для максимально полного изучения балто-славянских популяций ученые использовали все три генетические системы, которые на данный момент наиболее информативны для исследования генофонда.

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 точки генетического разнообразия (однонуклеотидного полиморфизма, SNP), которые разбросаны по всему геному, и расположены на аутосомах – неполовых хромосомах.

> Для анализа ADMIXTURE использовано 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 были 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.

Значительная часть этих внушительных массивов данных получена авторами 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;

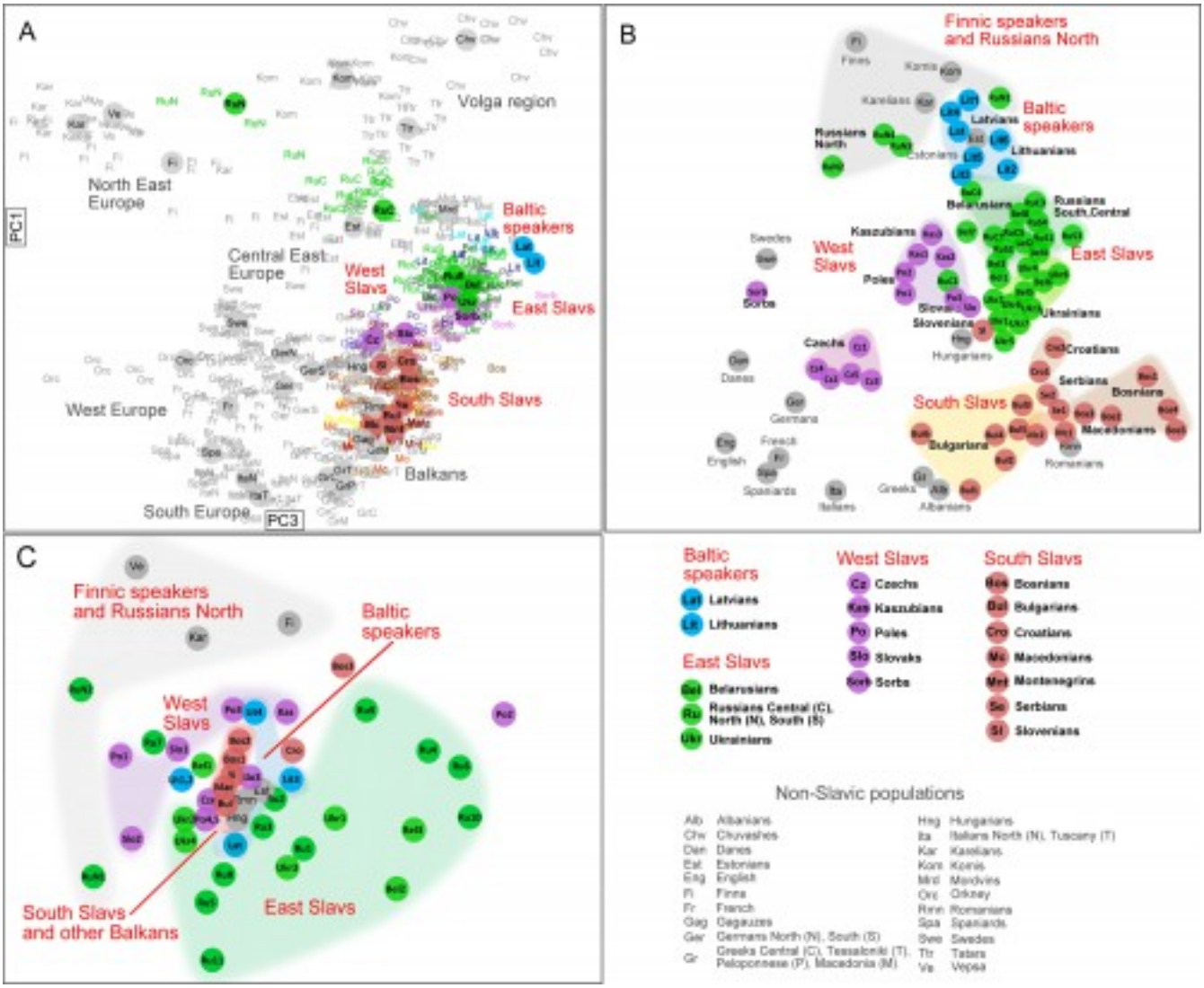
восточные славяне — белорусы, русские, украинцы;

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, а общая – 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 популяций.

Генетический ландшафт славян через три призмы



Генетическая структура балто-славянских популяций в сравнении с другими народами Европы по трем генетическим

системам: A) по аутосомным SNP-маркерам, B) по Y-хромосоме, C) по митохондриальной ДНК.

Генетические соотношения друг с другом всех изученных популяций, установленные в результате исследования, показаны на рисунках.

> Рис. A представляет результаты для широкогеномных (аутосомных) SNP-маркеров. Эти маркеры называются аутосомными потому, что находятся на неполовых хромосомах (аутосомах). А широкогеномными они называются потому, что равномерно разбросаны по всему геному.
>
> Рис. B представляет результаты по Y-хромосоме, полученные на основании частот ее гаплогрупп.
>
> Рис. C отражает результаты, полученные по частотам гаплогрупп митохондриальной ДНК (мтДНК).
>
> Чтобы показать относительную близость и удаленность разных популяций на двумерном графике, в популяционной генетике 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, и по Y-хромосоме (A и B) большинство балто-славянских популяций выстраиваются вдоль оси север-юг.

Восточные славяне – русские, белорусы и украинцы — отчетливо группируются, Они образуют свой кластер, хотя внутри него русские, белорусы и украинцы не перекрываются полностью друг с другом. Исключение составляют северные русские, 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 географически 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 некоторым восточным славянам (белорусам). Оказалось также, что балтские популяции близки к волжской группе финно-угорских народов (особенно к мордве). Авторы уточняют, что это может отражать исторические события – в древности ареал балто-язычных популяций простирался далеко на восток и почти доходил до нынешнего ареала мордвы.

Важно, что все перечисленные закономерности выявлены по независимым и, казалось бы, совершенно разным генетическим системам – Y-хромосоме и широкогеномным аутосомным маркерам.

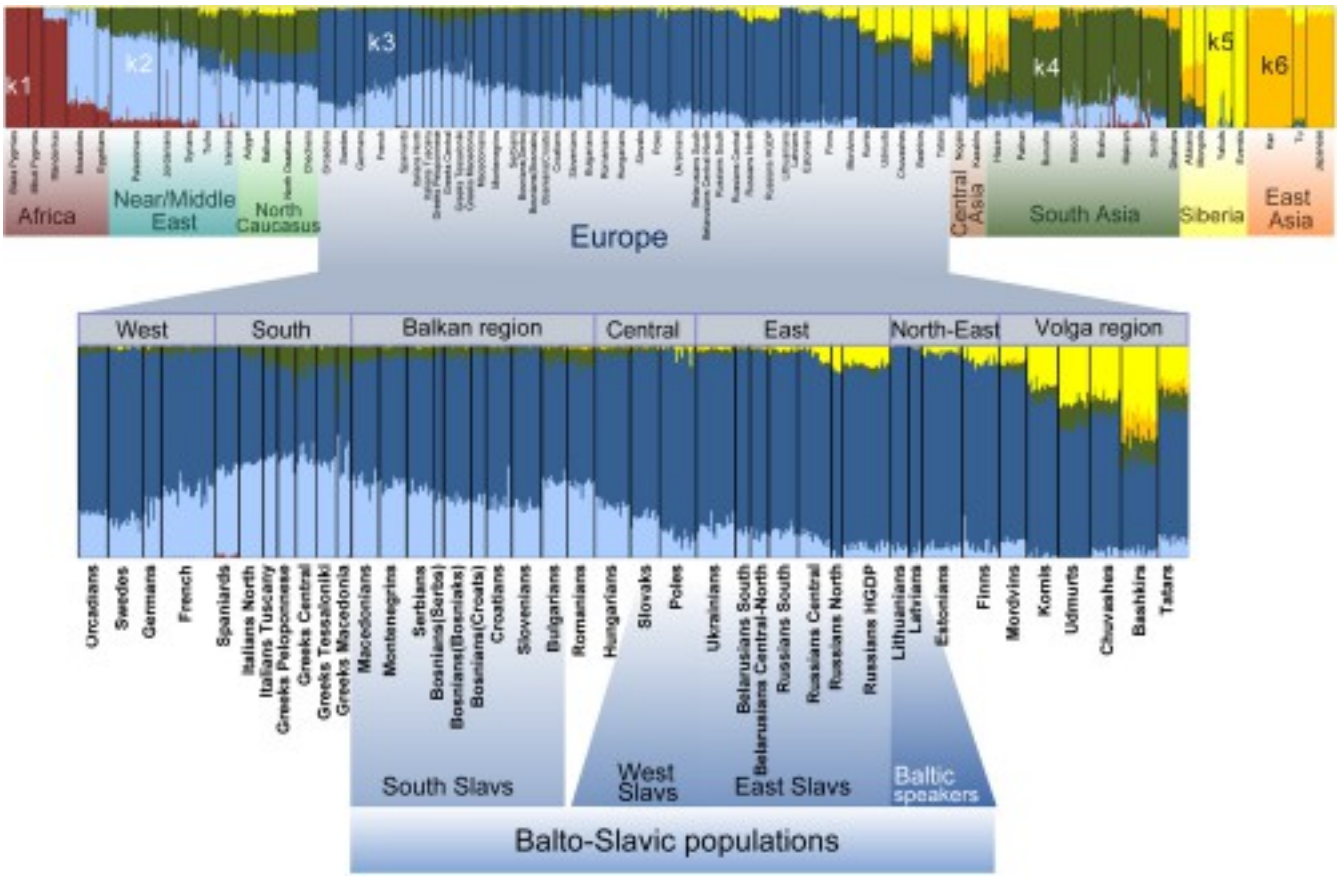
По мтДНК (рисунок C), как обычно, степень структурированности генофонда выражена значительно слабее, что связано с более низким филогенетическим разрешением в имеющихся данных по мтДНК. Но, хотя и не так четко выраженные, в результатах по мтДНК проглядывают те же самые закономерности. Например, и на графике по мтДНК большинство восточно-славянских популяций перекрываются друг с другом, северные русские отделены от них, а южные славяне генетически сходны со своими неславяноязычными соседями по Балканам.

Сравнивая степень выраженности одних и тех же закономерностей в результатах по разным генетическим системам, авторы подчеркивают, что Y-хромосома часто выявляет закономерности более детально, чем не только мтДНК, но и более модные широкогеномные маркеры.

Поиск глубинных предков

Чтобы сравнить популяции по составу их предковых компонентов, часто используется программа ADMIXTURE («смешение», или «состав»). В нее 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. Например, при анализе данных в масштабе мира всегда первым выделяется африканский компонент, который составляет почти 100% в африканских популяциях южнее Сахары, а его доля в других популяциях мира хорошо соответствует степени их прямого или опосредованного смешения с популяциями Африки.

В данной работе также был применен метод ADMIXTURE – авторы задавали разное число предковых популяций и опубликовали все соответствующие графики, но специальный тест показал, что статистически наиболее обоснованные результаты получены в случае, когда число предковых компонентов задавалось равным шести (K=6). В этом случае авторы получили вот такую картину.



Результаты оценки спектра предковых компонентов популяций мира по методу 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 – различиям между индивидами в пределах этноса. (Russians HGDP соответствует сборной выборке русских из проекта «Разнообразие генома человека – Human Genome Diversity Project).

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 регионов 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 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 (новое опровержение расхожего представления о большом влиянии на русский генофонд татаро-монгольского ига). Темно-зеленый компонент k4 характеризует популяции Южной Азии, 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 геномов славян и их соседей

Чтобы ответить на этот вопрос, авторы провели 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 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 и по аутосомным маркерам – зная скорость рекомбинаций, можно по длине сохранившихся общих гаплотипов оценить, сколько времени прошло от общих предков, то есть давно ли существовала общность генофондов.

Было подсчитано число общих гаплотипов между «западно-восточными» славянами (авторам пришлось пользоваться этим неуклюжим термином за неимением лучшего) и восемью другими группами народов Европы:

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 — их авторы условно называют «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 – языке романской группы);
6) греками;
7) популяциями волжского региона и Приуралья (башкиры, коми, мордва, татары, удмурты, чуваши);
8) северокавказскими популяциями (адыги, балкарцы, ногайцы).

Если принять за эталон число общих гаплотипов между западно-восточными и южными славянами, то часть окружающих неславянских популяций будет (по числу общих гаплотипов) выше этого эталона, часть ниже, а часть равняться ему. Ниже эталона (то есть имеют меньшее родство с западно-восточными славянами, чем южные славяне) оказались народы Поволжья, Западной Европы, Кавказа, а также греки.

Казалось бы, можно говорить о большем родстве славянских генофондов друг с другом, чем с окружающими неславянскими

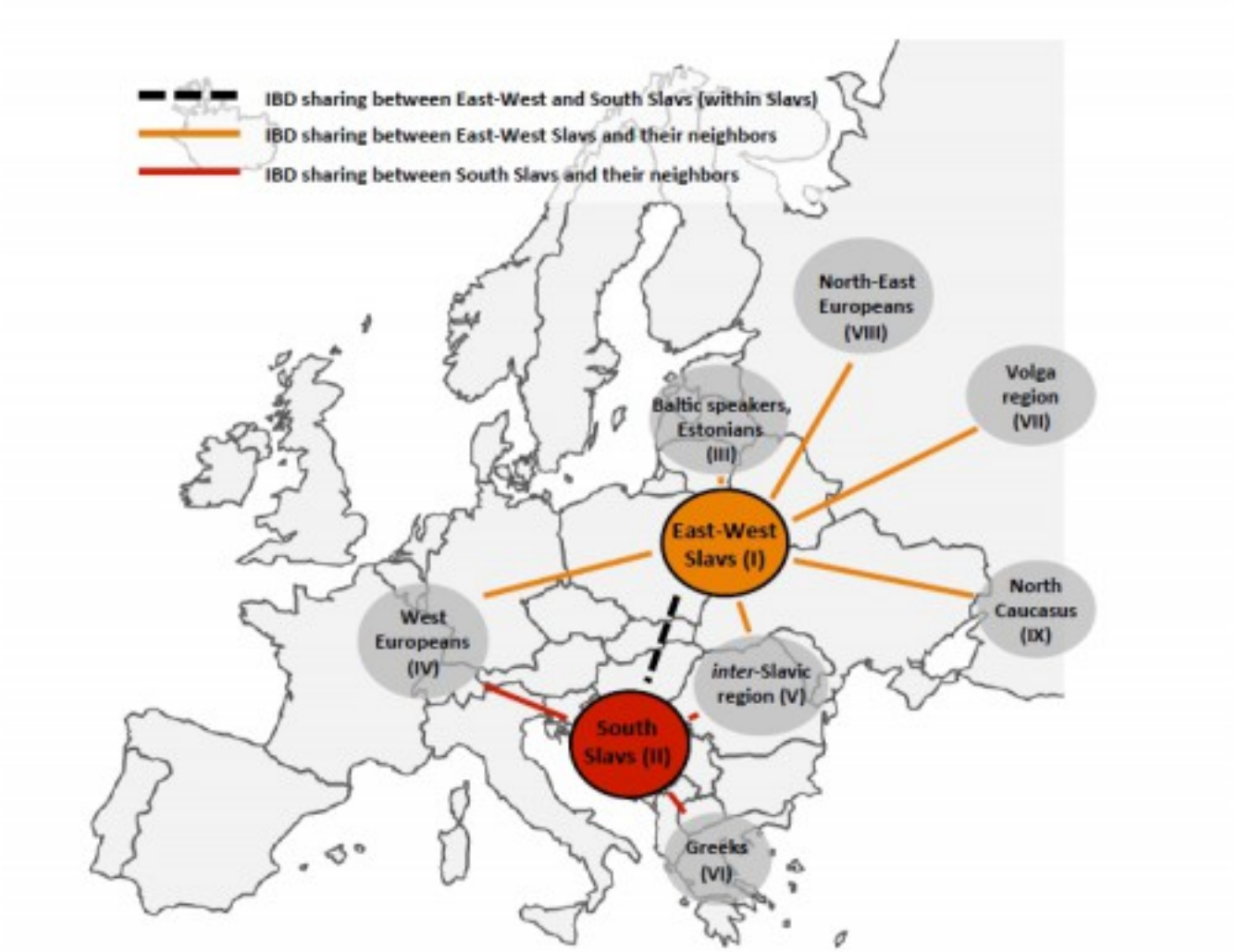
народами. Отчасти это так, но все не так просто – в два раза выше эталона оказалось родство генофондов балтов и популяций северо-восточной Европы (вепсы, карелы, финны, латыши, литовцы, северные русские, эстонцы). Можно впасть в противоположную крайность и считать, что «западно-восточные» славяне генетически 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).

Поэтому авторы провели еще один аналогичный анализ, но теперь поставив в центр рассмотрения южных славян. 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. Учтем, что западно-восточные славяне географически 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 языковое 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 разнятся по длине, 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.

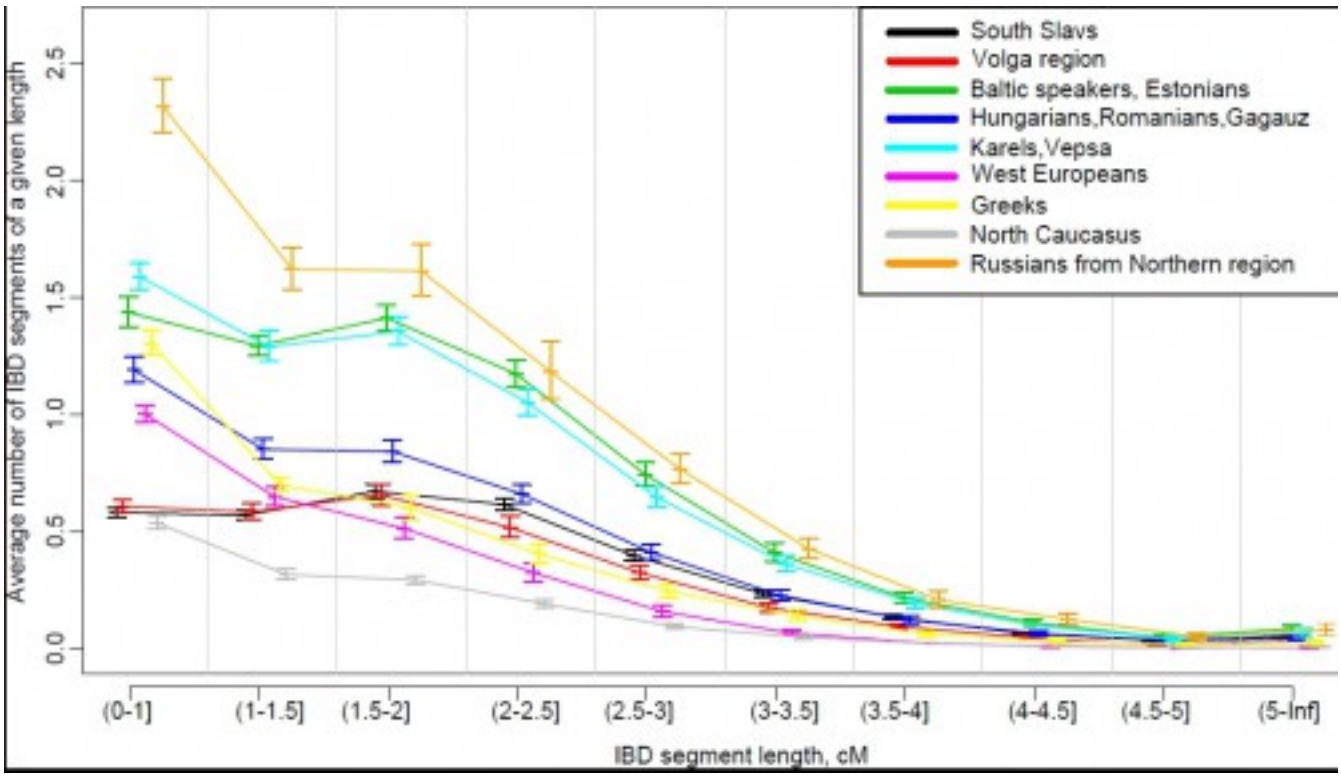
Эти результаты по славянам, из которых нельзя сделать однозначных выводов, следует сравнить с недавним похожим исследованием тюркоязычных популяций (Yunusbaev et al., 2015). Казалось бы, в обоих случаях идет быстрое распространение носителей языков (соответственно, тюркских или славянских) по обширным территориям, которое не может не сопровождаться ассимиляцией местного (дотюркского или 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 родственна, по крайней мере, балтским группам.

В целом, из этого анализа общих фрагментов 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.

Схема и результаты анализа общих фрагментов генома (IBD)



Схематическое представление групп популяций, использованных в анализе IBD (подсчет числа общих фрагментов генома).

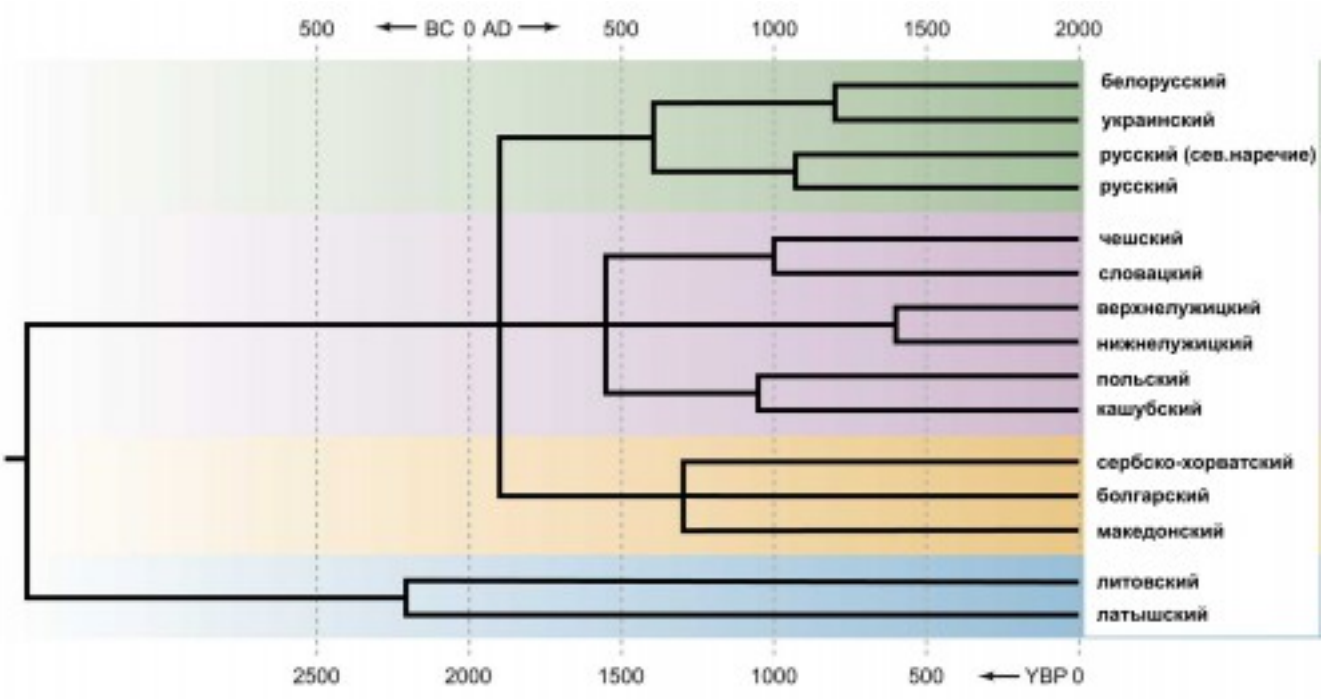


Графики среднего числа общих фрагментов генома (IBD сегментов) между группами западно-восточных славян и их географическими соседями (северные русские рассматриваются отдельно от группы популяций северо-востока Европы). По

оси X показано 10 классов IBD сегментов по их длине (в сантиморганидах, cM); по оси Y – среднее число общих фрагментов для пары индивидов в пределах каждого класса IBD.

Реконструкция дерева языков.

В коллектив авторов входили не только генетики, но и ведущие российские лингвисты. Это позволило использовать в этой работе уточненное древо родства балто-славянских языков благодаря обновлению и перепроверке массива лексикостатистических данных. Лексикостатистика занимается выявлением скорости языковых изменений и определением времени разделения родственных языков и степени родства между ними. Исходным материалом послужили лексические списки (списки Сводеша) 20 современных балто-славянских языков и диалектов.



Реконструированное филогенетическое дерево балто-славянских языков, созданное 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.

После разделения балтской и славянской ветвей самая первая развилка на славянской ветви оказалась тройной – разделение славян на западную, восточную и южную ветви – и датированной около 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 дерева

Поскольку лингвистическое дерево славянских языков так точно построено, появилась возможность проанализировать, как распределено по этому дереву генетическое разнообразие славянских популяций, оцененное по частотам гаплогрупп Y-хромосомы. Этот анализ проведен с помощью стандартной процедуры теста 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. 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 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 — около 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 как бы находим ее центр тяжести, точку происхождения, реконструируем генофонд этой предковой популяции.

Но все это, как сказано выше, работает лишь тогда, когда популяции предоставлены сами себе и мало взаимодействуют с соседями. Однако для славян величины генетического разнообразия на разных иерархических уровнях не одинаковы: при эквидистантности они должны быть 0.01, 0.01, 0.01, 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 языки 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 не унаследованных, 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 чрезвычайно 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, авторы рассмотрели частную корреляцию, чтобы различить прямое и непрямое влияние географии на две другие системы. При исключении географического фактора, частная корреляция с лингвистикой стала 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.

Два субстрата в славянских генофондах

Генетики считают, что, распространяясь по Европе, славяне ассимилировали местные популяции, которые жили на данных территориях в дославянские времена. Это тот генетический субстрат, которые они впитали в себя, и этот субстрат различается на разных территориях. Итоги работы позволили выделить два основных субстрата. «Центрально-восточноевропейский субстрат» приняли в себя западные и восточные славяне (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.

Синтез данных по разным генетическим и не генетическим системам.

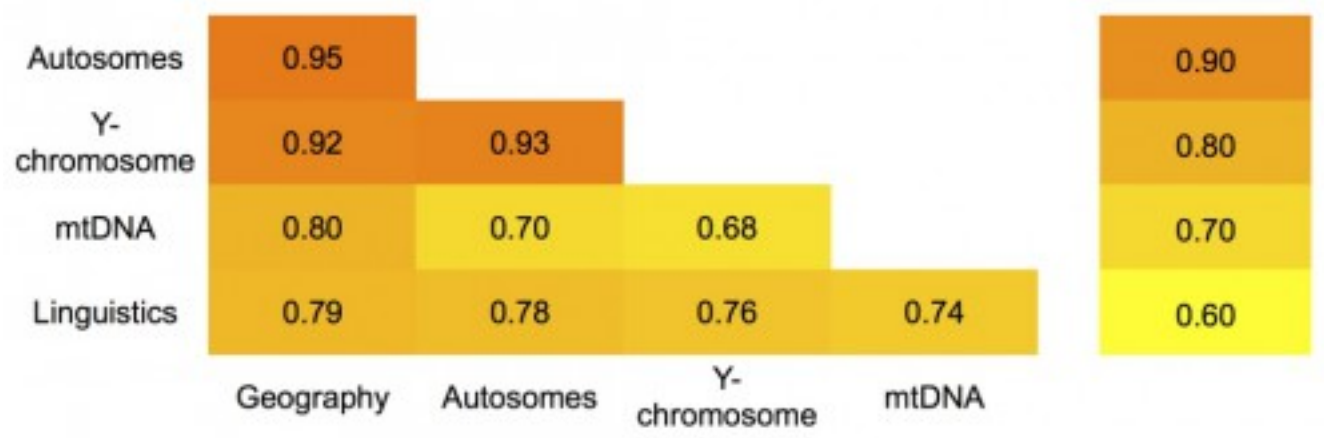
В генетических работах на каждом шагу встречается слово «анализ», и очень редко — «синтез». Здесь же «синтез» вынесен даже в название работы. Что это значит?

Уже говорилось о том, что это исследование уникально тем, что чуть ли не впервые для большой группы родственных народов изучен каждый народ, причем изучен по всем трем современным генетическим системам, да вдобавок количественно оценено лингвистическое родство между ними. И это позволяет на примере славян посмотреть, как связаны между собой три разные генетические системы, лингвистика и география – и синтезировать эти разнородные данные в общие выводы. Это тем более важно, что сравнение генетических и лингвистических реконструкций с географией имеет давнюю традицию в популяционной генетике.

Корреляции всех пяти систем (трех генетических, лингвистической и географической) друг с другом показаны на рисунке. 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 достигают максимально возможного потолка (корреляция 0,95). Особенное соответствие отмечаются 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 рисунке).

Соответствие результатов по разным системам признаков подтверждает надежность таких результатов. Оно также указывает на перспективность так называемого полисистемного подхода. Этот подход состоит, во-первых, в параллельном 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 данной статье авторы широко пользовались полисистемным подходом. Формулируя утверждения о генетическом сходстве или различии тех или иных народов, авторы 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 совпадения 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).



Корреляции между матрицами генетических, лингвистических и географических расстояний для балто-славянских популяций. Столбец справа – цветовые обозначения интервалов коэффициентов корреляции.

История славянских генофондов: итоги исследования

Итак, для генетического изучения славян авторы использовали возможность синтеза разных наук. Какие же выводы позволил сделать такой синтез и столь обширное исследование?

Прежде всего, это преобладание в славянских популяциях дославянского субстрата — двух ассимилированных ими генетических компонентов – восточноевропейского для западных и восточных славян и южноевропейского для южных славян. (Слишком длинные названия «центрально-восточноевропейский» и» южно-восточноевропейский» для краткости удобнее называть восточноевропейский и южноевропейский, помня при этом, что на Западную Европу они не простираются, а находятся в восточной ее половине при дихотомическом делении Европы).

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 западно-славянские 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.

В работе высказывается осторожное предположение, 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 в этой работе 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.

Источник:

Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data

Alena Kushniarevich, Olga Utevska, Marina Chuhryaeva, Anastasia Agdzhoyan, Khadizhat Dibirova, Ingrida Uktverite, Märt Möls, Lejla Kovačević, Andrey Pshenichnov, Svetlana Frolova, Andrey Shanko, Ene Metspalu, Maere Reidla, Kristiina Tambets, Erika Tamm, Sergey Koshel, Valery Zaporozhchenko, Lubov Atramentova, Vaidutis Kučinskas, Oleg Davydenko, Lidya Tegako, Irina Evseeva, Michail Churnosov, Elvira Pocheshchova, Bayazit Yunusbaev, Elza Khusnutdinova, Damir Marjanović, Pavao Rudan, Siiri Rootsi, Nick Yankovsky, Phillip Endicott, Alexei Kassian, Anna Dybo, The Genographic Consortium, Chris Tyler-Smith, Elena Balanovska, Mait Metspalu, Toomas Kivisild, Richard Villems and Oleg Balanovsky

PLOS ONE, Sept 2, 2015

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0135820

Отклики в СМИ и на научно-популярных сайтах:

http://antropogenez.ru/single-news/article/507/

http://www.gazeta.ru/science/2015/09/03_a_7734953.shtml

http://www.strf.ru/material.aspx?CatalogId=222&d_no=103794#.VelbuC6A__c

http://chrdk.ru/news/2015/09/03/uchenye_rekonstruirovali_geneticheskuu_i_lingvisticheskuu_istoriu_slavyan/

http://www.vz.ru/news/2015/9/3/764945.html

http://www.vladtime.ru/nauka/444371-uchenye-rekonstruirovali-lingvo-geneticheskuyu-istoriyu-slavyanskih-narodov.html

http://svopi.ru/nauka/60092-uchenye-dokazali-geneticheskoe-edinstvo-russkih-ukraincev-i-belorusov.html

http://polit.ru/article/2015/09/06/ps_balto_slavic/

http://lenta.ru/articles/2015/09/15/balto/

Передачи на радио:

Олег Балановский в передаче «Родина слонов» (радиостанция «Говорит Москва»)

http://xn--c1acc6aafa1c.xn--p1ai/wp-content/uploads/2015_09_15_Rodina_slonov.mp3

Олег Балановский в передаче «Наука в фокусе» (радиостанция «Эхо Москвы»)

http://1.cdn.echo.msk.ru/snd/2015-09-18-naukafokus-1605.mp3

Олег Балановский на радиостанции «Спутник» (быв. Голос России), агентство «Россия сегодня»

http://xn--c1acc6aafa1c.xn--p1ai/wp-content/uploads/151008_interview_balanovsky_genofond_researches.mp3

Передачи на телевидении:

Олег Балановский в передаче «Гамбургский счет», Общественное телевидение России (ОТР)

http://www.otr-online.ru/programmi/oleg-balanovskii-geneticheskii-45818.html

https://www.youtube.com/watch?v=mgsKpoOdeZc&list=PLWLmNuEnIXrk60R4z4b2nvFXCw4LRH7vO&index=1