Дискуссии о датировках

Олег Балановский

Из книги О.П.Балановского "Генофонд Европы"

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 нерекомбинирующих («однородительских») генетических систем перед аутосомными (рекомбинирующими) генетическими системами. В тоже время получение надежных датировок является большой проблемой, ставшей предметом многих научных, околонаучных 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 (отличающие 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. 14. 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.

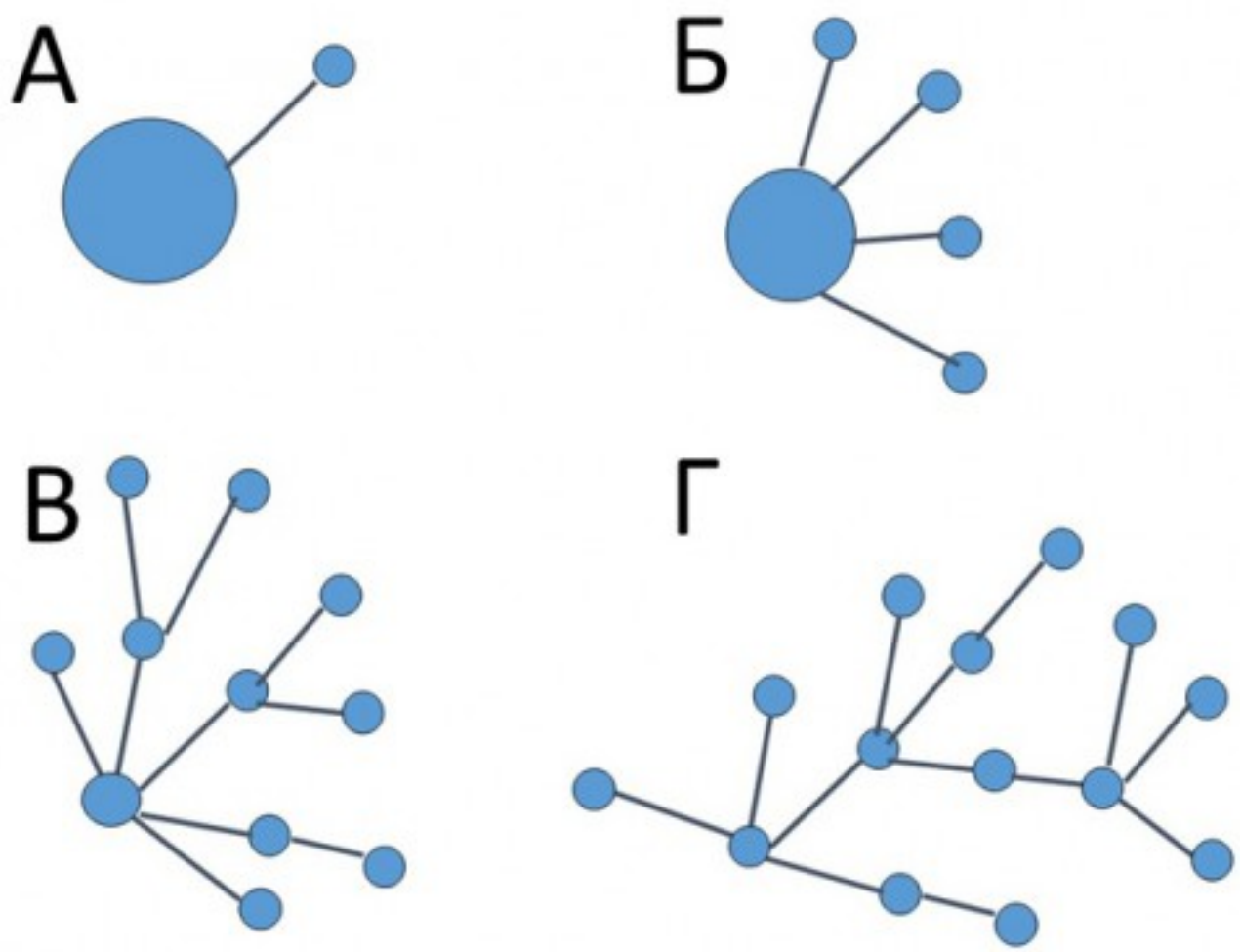

Рис. 14. Молекулярные часы: нарастание дерева гаплотипов. Показан постепенный рост дерева гаплотипов: возникновение новых гаплотипов в результате мутаций. А. Один исходный гаплотип (встречен у многих образцов, поэтому большой кружок) и один производный (только у одного образца, маленький кружок). Б. 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.

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 90-е годы прошлого века. Филогенетические сети были построены в программе 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 кластеров гаплотипов, выявление и датировка времени их происхождения стали 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.

- 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).
- Этот узловой гаплотип рассматривался в качестве предкового гаплотипа-основателя (founder) для своего кластера. Выбор гаплотипа-основателя важен при расчете возраста с использованием показателя ρ.
- Рассматривались только кластеры, содержащие 10 или более образцов, чтобы избежать ошибок в расчете возраста из-за малых объемов выборок.
- 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, раздел 4.3) и BEAST (как один из методов датирования ветвей филогенетического дерева, полученного по полным сиквенсам Y-хромосомы, раздел 5.4). BATWING кратко описан ниже в этом разделе, параметры применения BEAST – в разделе 5.4. 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, и, те которые используются 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 ρ 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 ρ.

Этот классический метод [Forster et al., 1996] основан на подсчете среднего числа мутаций, накопившихся в пределах кластера (это число и обозначается ρ). 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 делятся 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. 15. 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 ρ, а возможные погрешности

обсуждаются в [Cox, 2008].

Рис. 15. Расчет возраста кластера гаплотипов с помощью показателя ρ. 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 ρ – 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 ρ в данном примере составляет 0,7.

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 принимается не предполагаемый гаплотип-основатель, а 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, чем 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 заместившихся 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. 11 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 ρ.

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 ρ. Но и он не свободен от недостатков. Ведь для его использования требуется — не всегда известное — древо родства гаплотипов.

Резюмируем: когда дерево родства гаплотипов реконструировано надежно, оптимален метод ρ, 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-маркерный набор Y-filer) составляет $2.1*10^{-3}$ на локус за поколение. То есть для среднестатистического STR-маркера вероятность мутировать 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 дотоле необитаемую Новую Зеландию не позднее 800 лет назад, и в популяции цыган Болгарии, 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 мутировать 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-маркеров, 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-гаплотипа 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. 16). 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 мутировать 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 многочисленным 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).

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.

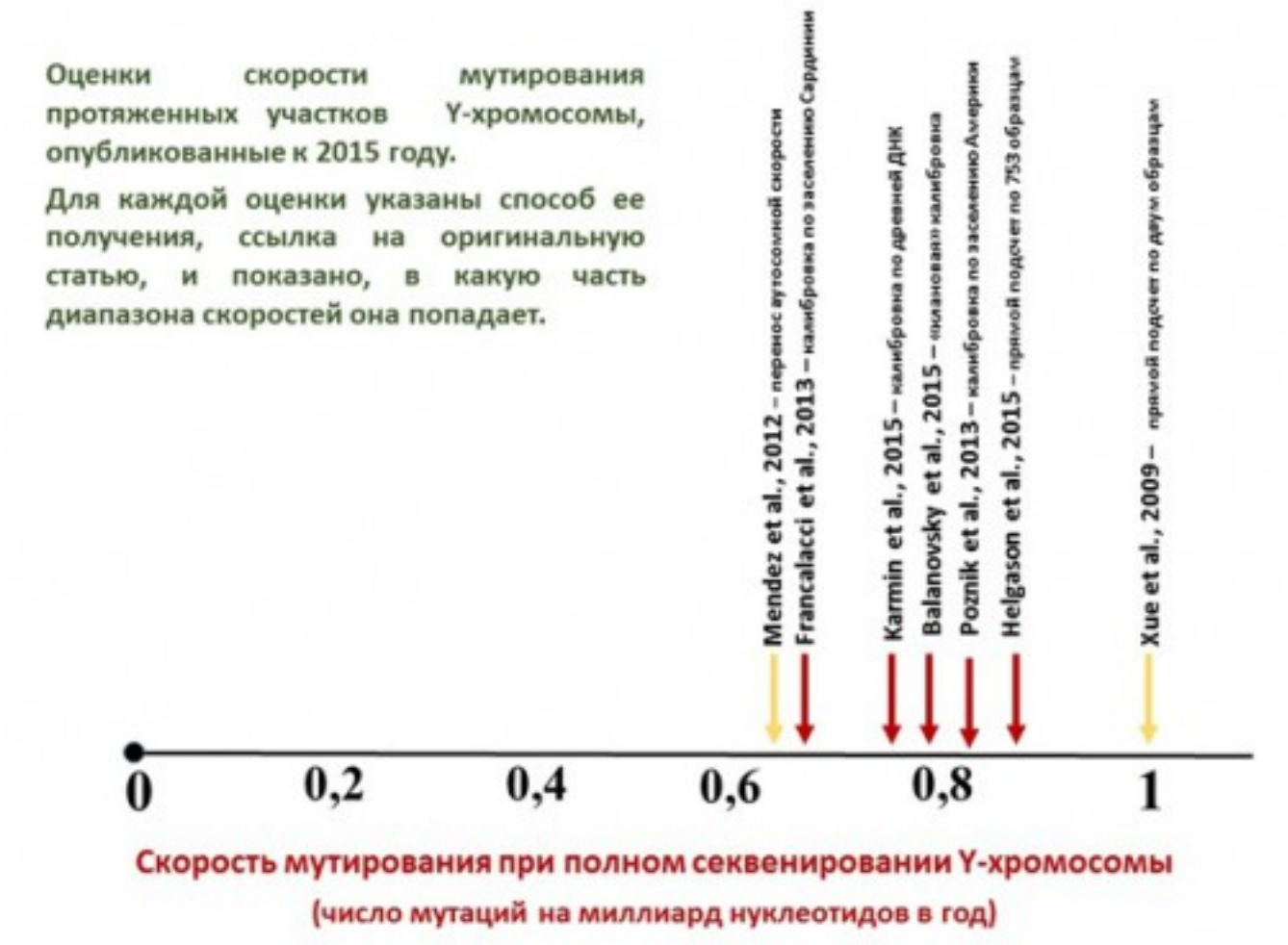

Рис. 16. Оценки «полногеномной» скорости мутирования на Y-хромосоме. [Xue et al., 2009] – скорость 1*10-9 мутаций на нуклеотид в год. Проведен прямой подсчет мутаций по 1 родословной между 2 образцами, разделенными 13 поколениями. [Mendez et al., 2012] – скорость 0,62*10-9 мутаций на нуклеотид в год. На Y-хромосому просто перенесен 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-аргынов (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.

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, возникавшее 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 – демографической 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 в куда более хорошем согласии с лингвистическими и историческими датировками, чем полученные с помощью «эволюционной» скорости. 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 примерно с 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 быстромутирующих 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 околонаука – 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 компенсировать 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 – путь 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 последовательности 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 Запорожченко; 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 индивидуальных, 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, №1 (170), 13 января 2015 г.), «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-генеалогию 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 подпадают под очарование запредельного упрощения А.А. Клесовым генетических данных, методов и результатов. Нам придется подробнее обсудить это в главе 3, при рассмотрении генофонда славян – ведь именно историю славян «научный патриотизм» (термин А.А. Клесова) ДНК-генеалогии перекраивает наиболее активно. Здесь же 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 не научные 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-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 занизят наблюдаемую долю производных гаплотипов по сравнению с долей реально возникавших 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 генетике 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 последовательность, 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 неловко 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 5.4.,, а по данным о STR-гаплотипах наиболее последовательно и систематически применялись при исследовании генофонда Кавказа (раздел 4.3.). 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 SD — вместе с научным координатором проекта «Генография» Давидом Сория (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 гаплотипов, как описано в работах Cruciani с соавторами [Cruciani 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 анализируются, 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 монофилетическими 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 потомков (рис. 14 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 4.3.

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 условий, стимулирующих или замедляющих 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, дамокловым 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 4.3), 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 этот 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 мутирований конкретного локуса обычно не важна, поскольку всегда анализируются их наборы. Поэтому нужно знать скорость мутирования именно набора локусов (то есть вероятность того, что смутирует 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 лишь 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 дуплицированные маркеры DYS385a и DYS385b типируются 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 ρ, 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.