

# Полное секвенирование отдельной гаплогруппы измеряет мутации и выявляет миграции

[Надежда Маркина](#)

Полное секвенирование Y-хромосомы гаплогруппы G1 позволило новым способом определить скорость мутирования и указало на направления миграций

**Российские и казахские генетики провели подробное исследование Y-хромосомной гаплогруппы G1. На основе полного секвенирования Y-хромосомы они реконструировали филогенетическое дерево этой гаплогруппы. На дереве выделились пять ветвей, географически специфичных для разных регионов, и показывающих направления миграций. Структура одной из ветвей – свойственной казахскому роду аргынов – точно совпала с генеалогией этого рода. Используя достоверные исторические даты, удалось получить объективную оценку скорости мутирования Y-хромосомы.**

## Гаплогруппа с необычным распространением

Y-хромосомная гаплогруппа G1 в некоторых популяциях Юго-Западной и Центральной Азии достигает частоты 80%, но до сих пор она не привлекала большого внимания генетиков, так как в среднем по миру встречается редко. В новой работе распространение этой гаплогруппы изучили в 27 популяциях Евразии, проанализировали ее у 367 носителей и секвенировали Y-хромосому у 20 человек. На основе этих данных исследователи описали географическое разнообразие гаплогруппы G1 и впервые построили ее подробное филогенетическое дерево. Возникло предположение, что в степную Евразию она была принесена с миграцией ираноязычных народов за несколько тысячелетий до нашей эры.

Эта работа, [опубликованная 7 апреля в журнале PLoS ONE](#), проведена международной группой специалистов под руководством Олега и Елены Балановских (Институт общей генетики РАН и Медико-генетический научный центр), в ней участвовали также специалисты из Казахстана, Украины, Монголии, Армении, США. Секвенирование Y-хромосомы проводилось в американской компании FamilyTreeDNA, но для обработки полногеномных данных российские специалисты разработали и свой собственный алгоритм, специально для целей филогенетического анализа.

Гаплогруппа G1-M285 (M285 обозначает мутацию, которая служит маркером этой гаплогруппы) привлекла внимание исследователей тем, что ее географическое распространение оказалось сложнее, чем считали ранее. Новые данные показали, что она достигает высокой частоты не только среди мажарцев в Казахстане, но и в некоторых популяциях Иранского и Армянского нагорий. Возникло предположение, что это может отражать процессы формирования этих популяций много веков назад и таким образом, указывать на генетические связи между двумя частями древнего ареала ираноязычных народов – нагорий Юго-Западной Азии и степей Центральной Азии.

## Новый подход – новые возможности

До сих пор не было установлено место происхождения гаплогруппы G1, не описано, насколько отличаются ее разные ветви и не определен их возраст. Теперь появилась возможность получить ответы на некоторые из этих вопросов. Для глубокого филогеографического анализа гаплогруппы G1 ученые использовали как традиционные подходы (генотипирование по однонуклеотидному полиморфизму – SNP и коротким tandemным повторам – STR), так и мощный потенциал нового подхода – полного секвенирования Y-хромосомы. Сначала они генотипировали 5346 человек в 27 евразийских популяциях на маркер M285 (это одна из мутаций, маркирующая искомую гаплогруппу) и выявили 367 носителей гаплогруппы G1. Разнообразие Y-хромосом этих 367 человек изучили по 17 STR маркерам.

Из всех образцов отобрали 19 для полного секвенирования Y-хромосомы, при отборе исходили из того, чтобы в них было максимально представлено гаплогруппическое разнообразие G1 и охвачен весь ее географический ареал. Для сравнения секвенировали также один образец Y-хромосомы, принадлежащий гаплогруппе G2, братской по отношению к G1. Для прочтения последовательности нуклеотидов воспользовались методом BigY, разработанным компанией FamilyTreeDNA. Он позволяет секвенировать участки, составляющие в сумме 11 миллионов нуклеотидов (11 Mb), на сегодня это «золотой стандарт» для анализа нерекombинирующей Y-хромосомы. Эти участки секвенировали с высоким покрытием (в среднем каждый нуклеотид был прочитан 67 раз).

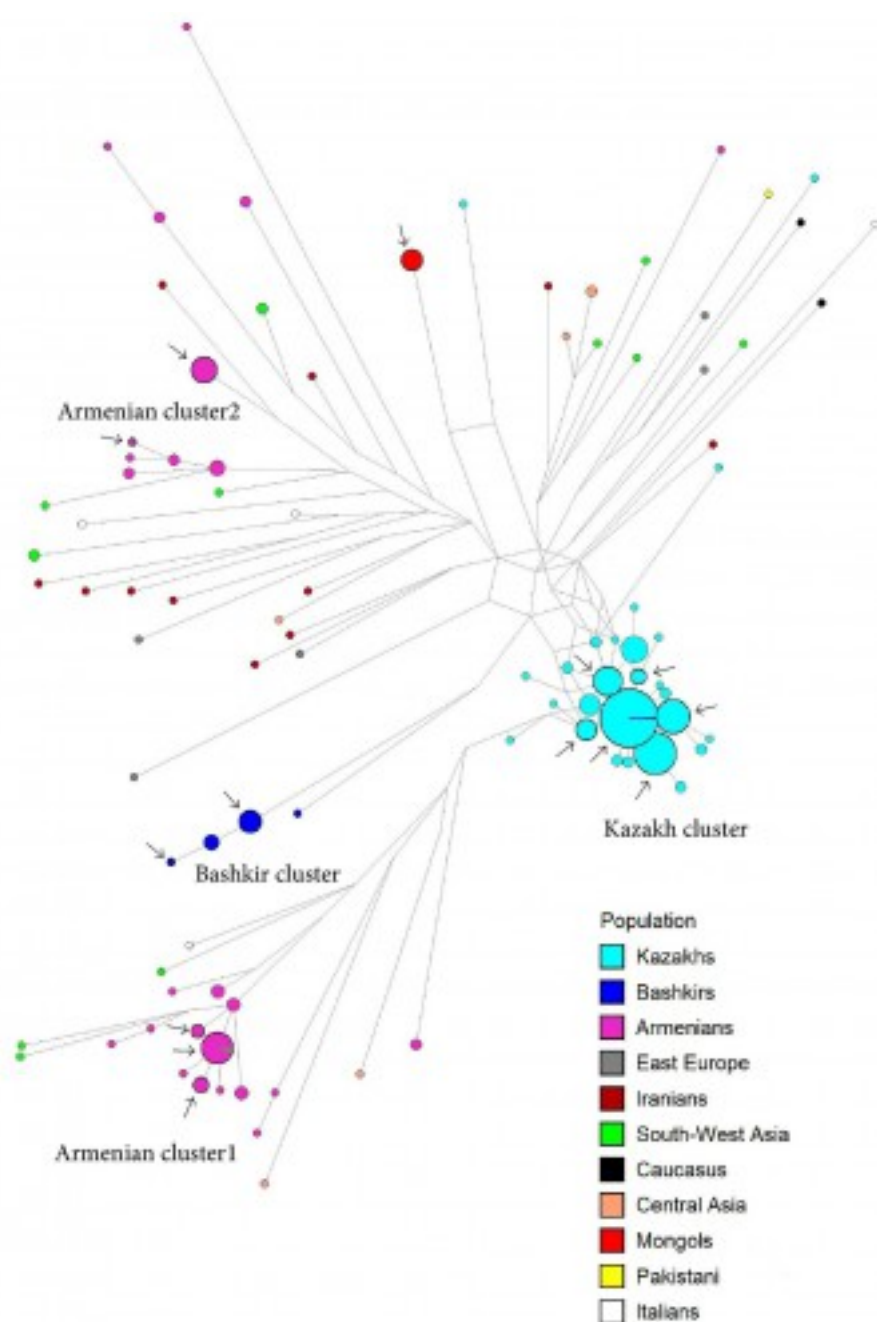
## Карта G1

Исследователи впервые составили карту распространения гаплогруппы G1-M285, используя свои данные по 27 популяциям, данные из литературы по 33 популяциям и опубликованные данные по 266 евразийским популяциям, где ее частота равняется 0. Карта приведена сверху. Черными точками обозначены популяции, для которых известна частота этой гаплогруппы, на остальных территориях показаны интерполированные значения. Справа приведена цветовая шкала частот.

Как показывает карта, гаплогруппа G1 распространена на обширной территории, от Италии на западе до Монголии на востоке, но высокой частоты достигает в ограниченном ареале, включающем центральноазиатские степи, Иранское нагорье и Армянское нагорье. Два пика частоты наблюдаются на противоположных частях этого ареала – в степях Северного Казахстана (до 80%), и в горах Армении (до 42%).

Авторы статьи отмечают, что ареал гаплогруппы G1 хорошо совпадает с ареалом ираноязычных популяций в первом тысячелетии до н.э. Миграция ираноязычных популяций между центральноазиатскими степями и возвышенностями Юго-Западной Азии — это один из важных вопросов в поиске прародины индоевропейцев. В соответствии с курганной гипотезой, носители иранских языков распространились с территории евразийских степей на территорию современного Ирана. Альтернативная анатолийская гипотеза помещает прародину индоевропейцев в восточную Анатолию, предполагая, что ираноязычные популяции именно оттуда мигрировали в степи. Таким образом, две гипотезы предполагают противоположные направления миграции.

## Кластеры в сети



Филогенетическая сеть гаплогруппы G1 по STR-гаплотипам.

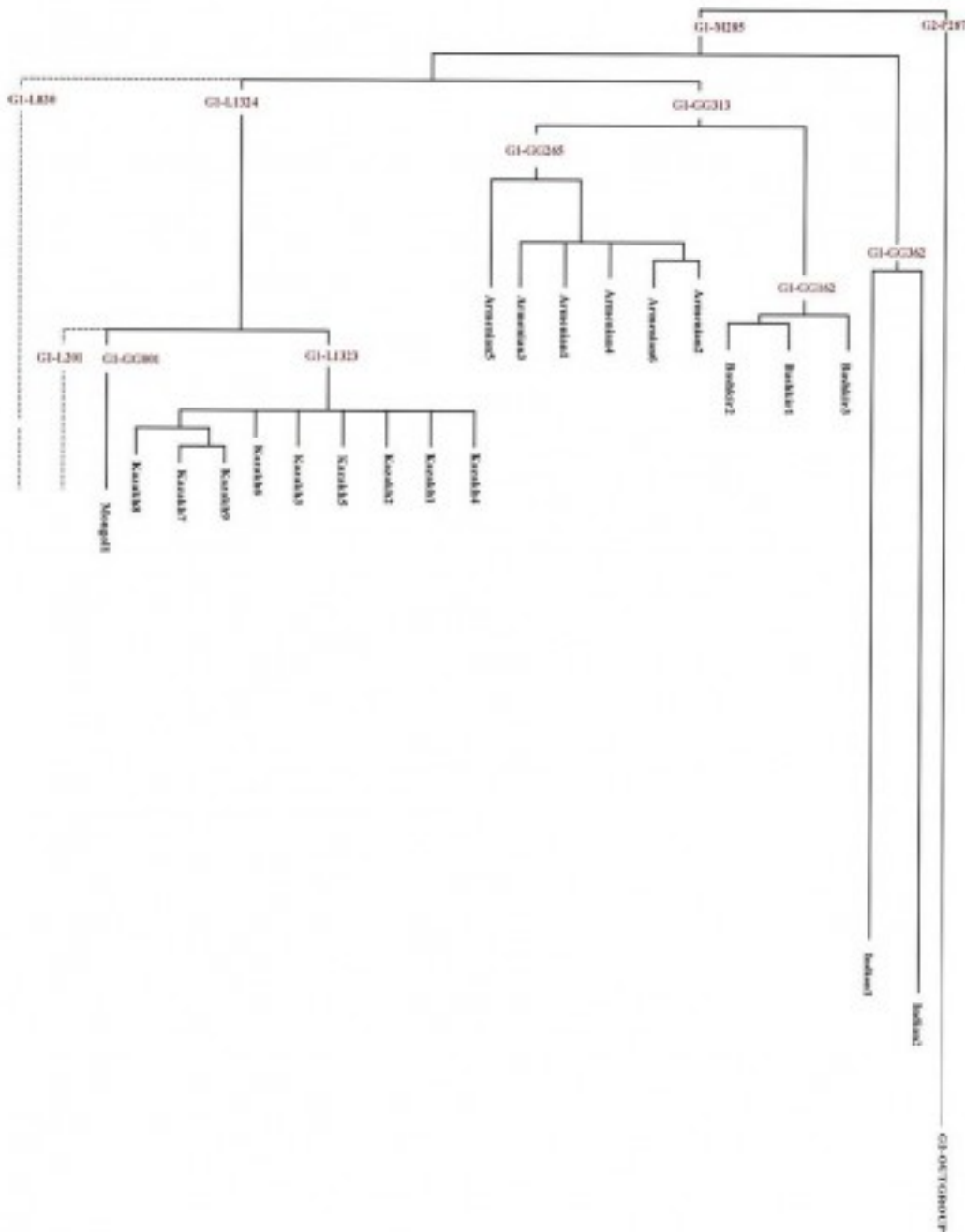
Изучив разнообразие STR-гаплотипов в пределах гаплогруппы G1, исследователи построили ее сеть, в которой на основании

сходства гаплотипов выделили несколько кластеров. Y-хромосомные STR-гаплотипы сгруппировались в четыре кластера: два из них были представлены образцами из Армении, один – самый большой – образцами из Казахстана, и один – из Башкирии. Гаплотипы из других популяций представлены в сети разрозненно и не группировались в кластеры.

Оказалось, что у казахов гаплогруппа G1 типична для рода аргынов (90% ее носителей принадлежали к нему), у армян она часта в роду амшен, у башкир все носители гаплогруппы G1 принадлежат к роду канглы.

### Пять ветвей дерева

Данные по секвенированию Y-хромосомы позволили реконструировать филогенетические события и построить дерево G1 гаплогруппы. Для этого исследователи использовали два метода — парсиномию («максимальной экономии») и байесовский метод. И поскольку деревья, построенные двумя методами, хорошо совпали, они делают вывод о надежности филогенетических реконструкций по данным полного секвенирования. Для построения итогового дерева гаплогруппы G1 авторы дополнили собственные данные информацией из проекта «1000 геномов» (1000 Genomes Project) в котором были секвенированные с низким покрытием два образца гаплогруппы G1.



Филогенетическое дерево гаплогруппы G1.

На дереве выделяются три крупные ветви, соответствующие кластерам, полученным по STR-гаплотипам, — казахская, армянская и башкирская. Хотя при этом два армянских кластера объединились в единую ветвь. Видно, что армянская и башкирская ветви происходят от общего предка, в то время как казахская представляет собой независимую линию. Кроме того, на дереве представлена независимая ветвь, представленная только в Западной Индии, а также монгольский образец, родственник казахской ветви.

Казахская ветвь соответствует ранее описанной G-L1323, в то время как армянская (G1-GG265), башкирская (G1-GG162), монгольская (G1-GG1) и индийская (G1-GG362) в этой работе представлены впервые. Аббревиатура GG в названиях маркера означает геногеографию (**G**ene **G**eography) – принято, что имена маркеров начинаются с букв, показывающих, каким научным коллективом они были открыты.

Исследователи отмечают, что все ветви на дереве сходны по длине, и это наводит на мысль о постоянстве скорости мутаций. Единственное исключение – очень длинные ветви из образцов проекта «1000 геномов», вероятно, это связано с тем, что используемые фильтры не совсем подходят к образцам, секвенированным с малым покрытием.

### **Генетика совпала с генеалогией**

В данной работе исследователи получили уникальную возможность сравнить данные по Y-хромосомному секвенированию с достоверными историческими источниками по казахским кланам и откалибровать их по времени.

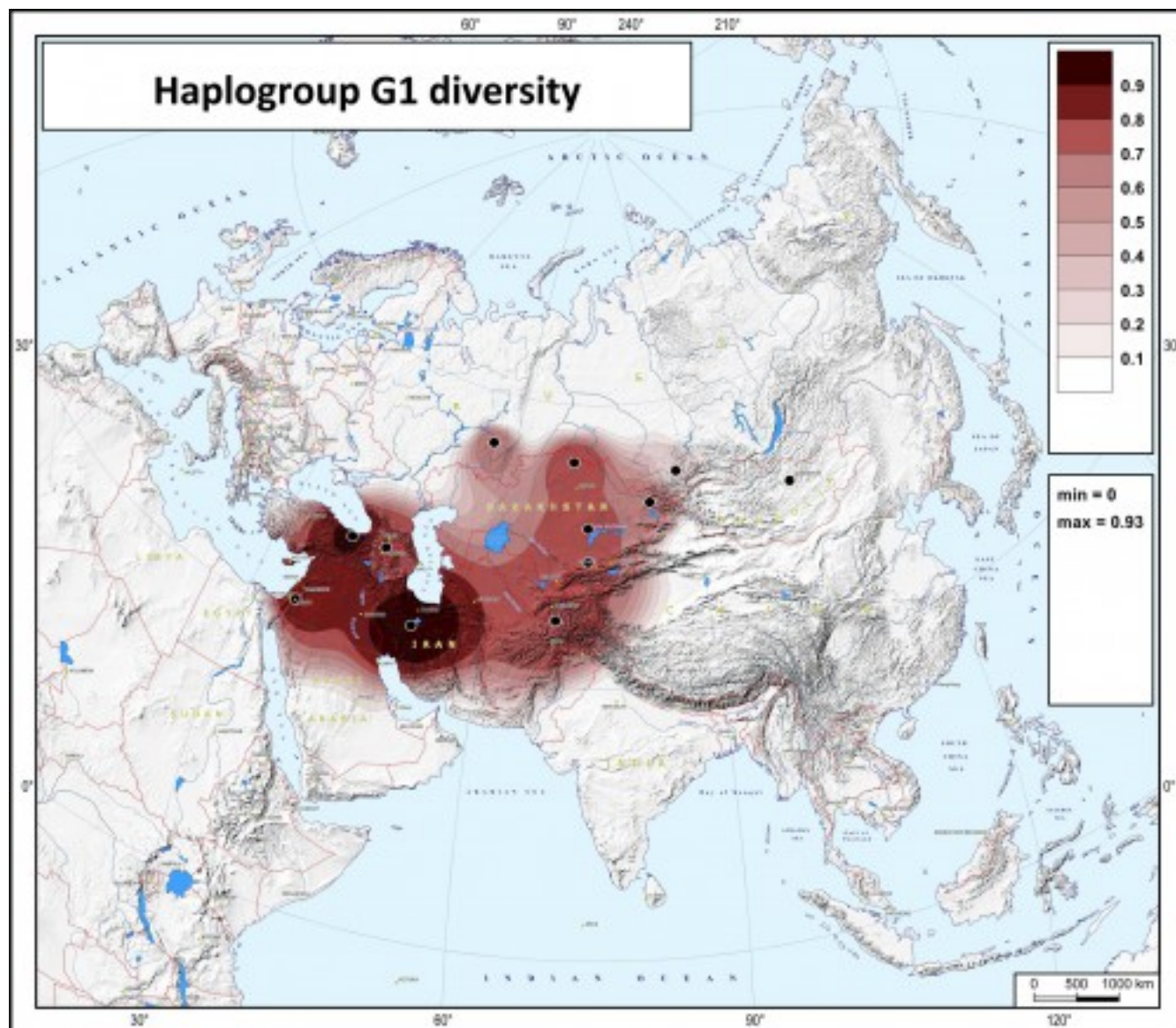
Род аргынов, в котором доминирует гаплогруппа G1, по преданиям состоит из потомков единственного мужчины (Аргын). Исторических свидетельств реального существования Аргына нет, но Караходжа, (по легенде, правнук Аргына) – это историческая персона и в источниках упоминается как посланник хана Тохтамыша – предводителя Золотой орды — к Тамерлану в 1405 году (предположительно, в возрасте 50 лет). Большинство из рода аргынов считают себя потомками Караходжи, а другие считают, что происходят от его брата Сомдыка (Токал-аргын).



1982 год. Таким образом, наблюдаемые на дереве мутации накапливались с 1355 по 1982 годы, то есть в течение 627 лет. Авторы учли также общую длину Y-хромосомных сегментов, секвенированных в каждом образце (9972660 нуклеотидов), и среднее число мутаций у каждого ныне живущего представителя рода по сравнению с родоначальников (4,89 мутаций). По этим исходным данным они вычислили скорость мутаций в изученном участке Y-хромосомы. Скорость составила  $0,78 \times 10^{-9}$  на нуклеотид в год (то есть каждый год с вероятностью 0,78 мутирует каждый миллиардный нуклеотид).

Тот же подход применили и к Y-хромосомным STR. По данным от 15 Y-STR аргынов просчитали 21 поколение, за которое произошло 0,68 мутаций. Скорость мутаций по Y-STR оценили в 0,0022 на локус на поколение. Это значение оказалось очень близким к «генеалогической» скорости мутаций, показанной во многих исследованиях при сравнении пар «отец-сын».

### Гаплогруппа указала на историю миграций



Карта гаплотипического разнообразия гаплогруппы G1.

Исследователи подчеркивают, что паттерн географического распределения гаплогруппы G1 своеобразен, поэтому ее нельзя назвать однозначно восточноевразийской или западноевразийской. Область распространения G1 хорошо соотносится с ареалом древних ираноязычных групп, которые обитали как на Иранском плато, так и в евразийских степях. В то же время из построенного дерева становится совершенно ясно, что все пять ветвей G1 географически специфичны: Западная Индия, Казахстан, Монголия, Башкирия и Армения.

Возникает вопрос, где прародина гаплогруппы G1 – в горах или степях. Намного большее разнообразие по STR в западной части Иранского и Армянского нагорий делает горную прародину более вероятным кандидатом. Это соотносится с анатolianской гипотезой возникновения индоевропейцев. Что важно, миграции из Ирана в Центральную Азию также не противоречат палеоантропологическим данным по древним популяциям Таджикистана и Туркмении.

Но сама по себе гаплогруппа G1 не может служить маркером распространения индоевропейцев, поскольку при любой



миграции перемещается часть популяции, состоящая из носителей разных гаплогрупп. Авторы подчеркивают, что ее можно рассматривать только как генетический компонент, принесенный волной ираноязычной миграции,двигающейся на север к степям Евразии. Генетические данные свидетельствуют, что все основные ветви уже существовали к тому времени, как миграция началась. Действительно, последняя развилка на башкирский и армянский кластеры произошла 8000 лет назад, а ираноязычные популяции появились значительно позже (армянская лингвистическая ветвь отделилась около 4600 лет назад, а индоиранские языки отделились около 4200 лет назад).

Гаплогруппа G1, вероятно, оставалась редким генетическим компонентом среди многих евразийских и особенно ираноязычных популяций. Несколько выше ее частота на предполагаемой прародине – в Иране и Армении. Когда иранские языки в степях были замещены тюркскими (вероятно, в середине 1-го тысячелетия н.э.), носители гаплогруппы G1, скорее всего, оказались ассимилированы в тюркские и потом – в монгольские популяции.

Экспансия этой гаплогруппы (быстрый рост численности ее носителей) у казахов генетически датируется временем 470-750 лет назад, в зависимости от того, какую скорость мутации использовать. Генеалогический предок аргынов жил 600 лет назад, как раз посередине этого интервала. Распространение гаплогруппы от одного человека к полумиллиону людей за 600 лет (20 поколений) не слишком удивительно, пишут авторы, при том что два выживших сына в каждом поколении дают полмиллиона в 19-м поколении, а у казахов в семьях было в среднем 3,5 ребенка.

Экспансия этой гаплогруппы у армян-амшененов генетически датируется временем 1150 лет назад. Это хорошо соотносится с историческими источниками, говорящими, что амшены произошли от родственников и слуг князя Шапух Аматауни, который приехал в 791 году из Аббасидского халифата.

Экспансия гаплогруппы G1 среди башкир (у рода канглы) генетически датируется XV веком н.э. Считается, что этот род берет начало у печенегов около VIII века н.э., а затем участвует в этногенезе башкир. Рост численности гаплогруппы может быть связан с демографическими изменениями в XIV веке, когда Башкирия стала частью Золотой орды, и в XVI веке, когда она стала частью России. Авторы отмечают, что несмотря на географическое соседство Башкирии и Казахстана, носители G1 в этих двух регионах не имеют недавних общих предков. Эти ветви (как и третья монгольская ветвь) присутствовали в евразийских степях, очевидно, начиная с эпохи скифов.

По словам первого автора статьи Олега Балановского, данная работа — всего лишь второй случай, когда полное секвенирование Y-хромосомы позволило выявить историю отдельной гаплогруппы. Первый относится к [изучению популяции евреев ашкенази](#).

Уникально также полученное в работе совпадение генетических и генеалогических данных, что позволило использовать исторические источники для объективной калибровки данных во времени. Результат – независимая оценка скорости мутаций Y-хромосомы, авторы назвали ее «клановой». Они подчеркивают, что «клановая» скорость мутаций была протестирована только внутри определенного временного периода (несколько веков), но считают, что она может быть надежным инструментом для уточнения датировки событий по археологическим данным.

К счастью, пишут исследователи, несмотря на разницу в подходах, все скорости мутаций, полученные при полном секвенировании Y-хромосомы, лежат в интервале  $0,6-1,0 \times 10^{-9}$  на нуклеотид в год, и в ходе дальнейших исследований этот интервал будет сужаться.

#### **Источник:**

#### **Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y-chromosome and reveals migrations of Iranian speakers**

Oleg Balanovsky, Maxat Zhabagin, Anastasiya Agdzhoyan, Marina Chukhryaeva, Valery Zaporozhchenko, Olga Utevska, Gareth Highnam, Zhaxylyk Sabitov, Elliott Greenspan, Khadzizhat Dibirova, Roza Skhalyakho, Marina Kuznetsova, Sergey Koshel, Yuldash Yusupov, Pagbajabyn Nymadawa, Zhaxybay Zhumadilov, Elvira Pocheshkhova, Marc Haber, Pierre A. Zalloua, Levon Yepiskoposyan, Anna Dybo, Chris Tyler-Smith, Elena Balanovska

PLoS ONE, April 7, 2015.

Аффилиация авторов – по ссылке

<http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0122968&representation=PDF>