

ГЕНОМИКА.
ТРАНСКРИПТОМИКА

УДК 575.174:599.9

ОДНОФАМИЛЬЦЫ ИЛИ РОДСТВЕННИКИ? ПОДХОДЫ К ИЗУЧЕНИЮ СВЯЗИ МЕЖДУ ГАПЛОГРУППАМИ Y-ХРОМОСОМЫ И ФАМИЛИЯМИ

© 2011 г. Е. В. Балановская*, А. Г. Романов, О. П. Балановский

Медико-генетический научный центр Российской академии медицинских наук, Москва, 115478

Поступила в редакцию 17.07.2010 г.

Принята к печати 02.11.2010 г.

Популяционная генетика с успехом применяет фамилии в качестве квазигенетических маркеров для оценки сходства популяций и случайного инбридинга. Используемый при этом коэффициент изонимии предполагает, что каждая фамилия монофилетична: она восходит к единому основателю и поэтому все однофамильцы – родственники. С другой стороны, большинство исследователей убеждены, что типичная русская фамилия полифилетична: она восходит к множеству линий происхождения и практически все однофамильцы – не родственники. Возможность разрешения этого спора появилась лишь недавно, благодаря параллельным исследованиям Y-хромосомы и фамилий. В работе обсуждаются подходы к статистически корректному анализу гаплогрупп Y-хромосомы у носителей одинаковых фамилий. Предложен показатель накопленной частоты гаплогрупп Y-хромосомы у однофамильцев, исключающий случайное “популяционное” повышение частоты гаплогруппы у однофамильцев благодаря вычитанию ее популяционной частоты (характерной именно для тех популяций, в которых встречены однофамильцы). Из суммарной выборки (1244 человека, представляющих 13 популяций исторического ареала русского народа) были отобраны носители 14 фамилий с наибольшим числом однофамильцев (123 человека). Сравнение частот гаплогрупп в 14 когортах однофамильцев и в сформированных из суммарной выборки 14 контрольных когортах показало, что даже у носителей распространенных фамилий наблюдается неслучайное накопление определенных гаплогрупп Y-хромосомы. Степень родства однофамильцев можно оценить при переходе от анализа гаплогрупп (SNP-маркеры) к анализу гаплотипов Y-хромосомы (STR-маркеры), чему будет посвящена следующая работа.

Ключевые слова: полиморфизм Y-хромосомы, гаплогруппы, SNP, однофамильцы, русский генофонд.

NAMESAKES OR RELATIVES? APPROACHES TO INVESTIGATING THE RELATIONSHIP BETWEEN Y CHROMOSOMAL HAPLOGROUPS AND SURNAMEN, by *E. V. Balanovska**, *A. G. Romanov*, *O. P. Balanovsky* (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, 115478 Russia; *e-mail: balanovska@mail.ru). Population genetics successfully applies surnames as quasi-genetic markers when estimating similarity between populations and calculating a measure of random inbreeding. These calculations are based on an isonymy coefficient which assumes that every surname is monophyletic: that it originated from single common ancestor and all namesakes are therefore relatives. On the other hand, there is a general opinion that a typical Russian surname is polyphyletic: it originated multiple times and most namesakes are therefore not related to each other. Combined studies of Y chromosomes and surnames now allow us to address this issue. In this study, we discuss approaches for statistical evaluation of Y chromosomal haplogroup frequencies in groups of people bearing the same surname (namesakes). We propose an ‘Index of Accumulated Haplogroup Frequency’, which allows for errors due to random (artifactual) effects increasing a haplogroup frequency in a group of namesakes by subtracting the population frequency of this haplogroup. This population frequency is calculated as the weighted average of the frequencies of this haplogroup in the populations that the carriers of this surname come from. From the total sample (comprising 1244 persons from 13 populations of the historical Russian area) we chose 123 persons carrying 14 surnames which were the most frequent in the total sample. Haplogroup frequencies in these 14 “surname” groups were compared with the respective 14 “population” control groups compiled from the total sample as described above. We found that even these widespread surnames exhibit non-random accumulation of specific Y chromosomal haplogroups. More detailed analyses of the relationships between namesakes could be carried out using Y-STR haplotypes rather than Y-SNP haplogroups, and will be the subject of a future study.

Keywords: Y chromosomal variation, haplogroups, SNP, namesakes, persons bearing the same surname, Russian gene pool.

* Эл. почта: balanovska@mail.ru

ВВЕДЕНИЕ

Анализ исторической связи между фамилией и генами становится реальностью благодаря проведению молекулярных исследований на популяционном уровне. Совместное изучение фамилий и ДНК-маркеров основано на анализе нерекombинирующей части (NRY) Y-хромосомы [1, 2]. NRY передается целостным гаплотипом и поэтому сохраняет память о мутациях, которые приводят не к возникновению нового равноправного аллеля (как у аутосомных маркеров), а к появлению новой ветви на дереве гаплогрупп Y-хромосомы, причем корень древа находится на этапе появления человека на планете. Вторая черта маркеров Y-хромосомы – высокая дифференциация: различия популяций по NRY оказываются в среднем на порядок выше, чем различия между теми же популяциями по митохондриальной ДНК и аутосомным маркерам [3]. Благодаря этим двум чертам – сохранению генетической памяти и высокой дифференциации популяций – гаплогруппы Y-хромосомы стали одними из наиболее широко изучаемых ДНК-маркеров.

То, что Y-хромосома наследуется так же, как и фамилии – по отцовской линии – вызывает интерес к их совместному изучению на стыке лингвистики и генетики. Лингвистике такие комплексные исследования позволяют находить новые подходы к ранее неразрешимым проблемам, например, к определению времени возникновения фамилий у каждого народа. Генетические исследования обогащаются использованием уникальных свойств фамилий, например, разнообразием их семантики и морфологии, что позволяет более детально анализировать генофонд [4]. Параллельные исследования Y-хромосомы и фамилий вызывают живой интерес ученых [5–16].

Ранее было показано, что структура генофонда многих народов России, выявляемая по фамилиям, хорошо согласуется с закономерностями, обнаруженными при использовании классических и ДНК-маркеров [17–21]. В частности, изучая русский генофонд [4], мы обнаружили высокий коэффициент корреляции ($r = 0.6$) между показателями попарного сходства популяций, рассчитанных по частотам гаплогрупп Y-хромосомы и частотам фамилий в тех же популяциях. Столь тесная связь между фамилиями и ДНК-маркерами, во-первых, дает основания для генетического изучения фамилий в масштабе всего русского генофонда, а не только одного–двух соседних районов, как предполагалось ранее [22], а во-вторых, позволяет использовать паттерны изменчивости Y-хромосомы и фамилий для взаимного прогноза. С одной стороны, анализ полных списков фамилий дает прогноз распространения Y-хромосомы во всех популяциях всего народа (дорогостоящий анализ Y-хромосомы позволяет изучить лишь несколько избранных популяций). С другой стороны, анализ Y-хромосомы дает возмож-

ность оценить наиболее общие черты фонда фамилий и структуры древних миграций, сформировавших генофонд, но нечетко улавливаемых фамилиями из-за их многообразия.

Важно, что изучать связь между фамилиями и Y-хромосомой можно не только на популяционном уровне, как в приведенном примере русских популяций, но и на индивидуальном уровне, чему посвящена данная работа. Связь этих двух систем на популяционном уровне означает, что *популяции*, сходные по фонду фамилий, будут сходны и по фонду Y-хромосомы. Эта связь в случае русских фамилий, как говорилось выше, оказалась высокой (коэффициент корреляции 0.6). Существование связи на индивидуальном уровне означало бы, что *индивиды* со сходными фамилиями (т.е. однофамильцы) будут обладать и сходными вариантами Y-хромосомы. Изучение этой связи для русских фамилий и составило задачу данной публикации.

Данная работа посвящена постановке проблемы, поиску подходов к ее решению, разработке алгоритмов изучения связи между гаплогруппами Y-хромосомы и фамилиями, а также их иллюстрации на ряде примеров. Ее цель – не столько описание конкретных результатов, сколько разработка подходов к изучению связи между гаплогруппами Y-хромосомы и фамилиями на индивидуальном уровне. Поэтому фактический материал и отдельные результаты рассматриваются как иллюстрация подходов, а не как итоги, подробное описание которых потребует отдельной публикации.

Необходимость изучения связи между фамилиями и вариантами Y-хромосомы

На популяционном уровне маркеры Y-хромосомы обычно изучают на небольших выборках, редко превышающих 100 человек (см., например, табл. 1, где средняя выборка – 85 человек), поскольку популяционная выборка требует забора биологических образцов и дорогостоящего лабораторного анализа. Поэтому из-за небольшого объема выборок всегда остро стоит вопрос об их репрезентативности: правильно ли представляет данная выборка ту популяцию, в которой она собрана?

Фамилии же, используемые в популяционной генетике как аналог генов (“квазигенетические” маркеры [23]), могут быть изучены не по малой выборке, а во всей популяции (см. табл. 2, где средняя выборка – 190 тысяч человек). Столь массовый характер изучения фамилий обеспечивает заведомую репрезентативность результатов.

То, что у народов с патронимической традицией фамилии наследуются по отцовской линии, так же как и Y-хромосома, позволяет рассчитывать на тесную связь между именно этой системой биологических маркеров (Y-хромосома) и маркерами “социальными” (фамилиями). Эта связь обеспечивает

Таблица 1. Условное название популяций, их административное положение и частоты восьми наиболее частых гаплогрупп NRY

Условное название	Область	Район	Выборка, N	R1a	N1c1	I1b	R1b1c	I1	J2	N1b	E1b1b1
Мезень	Архангельская	Лешуконский (бассейн р. Мезень)	54	0.353	0.368	0	0	0	0	0.059	0
Пинега	Архангельская	Карпогорский (бассейн р. Пинега)	114	0.352	0.211	0.039	0.125	0.008	0	0.141	0
Красноборск	Архангельская	Красноборский и Ленский	91	0.152	0.273	0.045	0.045	0.076	0.076	0.045	0
Унжа	Костромская	Мантуровский и Межевской (бассейн р. Унжа)	52	0.258	0.106	0.091	0.076	0.091	0.030	0	0.045
Кашин	Тверская	Кашинский	73	0.482	0.084	0.060	0.060	0.024	0.036	0	0.036
Порхов	Псковская	Порховский и Дедовичевский	57	0.423	0.127	0.085	0.042	0.028	0	0.014	0.014
Остров	Псковская	Островский	75	0.382	0.236	0.079	0.022	0.056	0.011	0	0.034
Рославль	Смоленская	Рославльский и Ершицкий	107	0.405	0.116	0.091	0.099	0.017	0.025	0.008	0.066
Ливны	Орловская	Ливненский	110	0.556	0.040	0.121	0.032	0.073	0.008	0.008	0.008
Пристен	Курская	Пристенский	45	0.424	0.102	0.136	0.017	0.034	0.017	0	0.017
Репьевка	Воронежская	Репьевский	96	0.514	0.055	0.147	0.046	0.028	0.009	0	0.009
Белгород	Белгородская	Яковлевский, Прохоровский, Красненский	143	0.541	0.108	0.115	0.025	0.032	0.038	0.006	0.006
Кубанские казаки	Адыгея	Майкопский	90	0.408	0.058	0.146	0.068	0.039	0.039	0.010	0.029
В СРЕДНЕМ			85	0.404	0.145	0.089	0.051	0.039	0.023	0.022	0.020

Таблица 2. Пять регионов исторического ареала русских, в которых собраны однофамильцы и дана популяционная характеристика изменчивости и Y-хромосомы, и фамилий

Регион	Численность изученных по фамилиям	Область
ЗАПАД	101 000	Смоленская
ВОСТОК	25 000	Костромская
ЦЕНТР	108 000	Тверская
СЕВЕР	54 000	Архангельская
ЮГ	655 000	Курская, Белгородская, Воронежская, Орловская
В среднем	188 600	

прогноз изменчивости маркеров Y-хромосомы на основании массово изученных фамилий [10], что особенно важно в случае популяций, прямое генетическое изучение которых затруднено. Например, подробный анализ изменчивости популяций в пределах отдельного региона, часто представляющий большой научный и практический интерес, обычно невозможен в связи с высокой стоимостью сбора и генотипирования многотысячных выборок, необходимых для решения этой задачи.

Принципиальным моментом, таким образом, становится наличие или отсутствие реальной связи

между фамилиями и Y-хромосомой. Ряд аргументов указывает на проблематичность такой связи. Например, фамилии, в отличие от гаплотипов Y-хромосомы, исторически молоды: считается, что самые древние фамилии возникли в Китае около 5000 лет назад [7]. Основная масса современных фамилий возникла в последнюю тысячу лет, а у некоторых народов – в пределах последних веков. Одинаковые гаплотипы Y-хромосомы встречаются у людей с разными фамилиями. И, наоборот, однофамильцы часто имеют разные гаплотипы. Связь между фамилией и Y-гаплотипом нарушается по

разным причинам: из-за полифилетичности — независимого многократного возникновения фамилии; из-за “мутации” фамилии — изменения в ее написании (Есенин—Ясенин—Весенин) или же принятия потомками новой фамилии [6, 7]; в случае “не-отцовства” — в англоязычной литературе термином “*non-paternity*” обозначают усыновленных и внебрачных детей, а также детей, получивших фамилию от матери (оценки “не-отцовства” варьируют от 1.3 до 5% за поколение [5]).

Однако недавнее происхождение фамилий можно рассматривать не только как недостаток, но и как важное достоинство: оно позволяет нам отделить древние изменения генофонда от недавних [11]. Поскольку одна из сложных задач популяционно-генетических исследований состоит в датировке возникновения наблюдаемого разнообразия, фамилии могут стать инструментом, датирующим популяционные события недавнего прошлого.

Анализ связи между генотипом и фамилией важен и для применения фамилий в прогнозе груза наследственной патологии. Предпосылка этого научного направления [20, 24–27] состоит в постулировании связи между конкретной фамилией и конкретной Y-хромосомой *на индивидуальном уровне*: расчет коэффициента изонимии, на основе которого прогнозируется груз наследственной патологии, опирается на предположение, что все однофамильцы — родственники. Но доказательства этой связи получены только *на популяционном уровне* — при сравнении популяционных данных о фамилиях с популяционными же частотами генов. Поэтому связь фамилий и генов на индивидуальном уровне (в какой мере носители одной фамилии являются биологическими родственниками) — актуальная проблема не только для лингвистики, истории и популяционной генетики, но и для медицинской генетики.

Кратко изложим существующие подходы к изучению связи фамилий и маркеров Y-хромосомы на популяционном и на индивидуальном уровне.

Связь фамилии и Y-генотипа на популяционном уровне

Работы этого направления нацелены на анализ различий между популяциями. Здесь фамилия выступает не как маркер конкретной Y-хромосомы, а как маркер популяции. Фамилия, как признак культуры, ставится в один ряд с другими характеристиками популяции — лингвистическими, этнографическими, генетическими и т.д. Как сходные узоры на традиционной одежде, сходные особенности говора или же сходные мотивы в мифах, так и сходство набора фамилий указывает на возможное родство популяций. При популяционном подходе случаи “не-отцовства” не влияют на популяционные частоты фамилий, поскольку компенсируют друг

друга. Во всех работах подчеркивается, что преимуществом фамилий по сравнению с другими маркерами является возможность изучения всего населения, а не выборка из него.

Так, фамилии Сардинии распределили по трем регионам их происхождения [8], а затем проанализировали генофонды популяций в пределах этих регионов. Обнаружили генетические различия между популяциями, фамилии которых принадлежат разным регионам.

В генофонде Восточной Германии [11] были выделены группы лиц с фамилиями немецкого, славянского и смешанного происхождения. Обнаружена генетическая близость носителей славянских фамилий к генофонду поляков и их генетические отличия от носителей остальных групп фамилий.

К изучению горного изолята в Колумбии [10] привлекли данные о частотах фамилий на разных временных срезах. Сходство распределения фамилий в XVIII и XX веках при росте численности населения в 107 раз показало, что основной фактор популяционной динамики — собственный рост популяций, а не иммиграция населения из Европы.

Шапсугов Северного Кавказа изучили по родословным на протяжении шести поколений [28]. Обобщенные карты изменчивости всех фамилий обнаружили преемственность генофонда: коэффициент корреляции между картами первого и шестого поколений достиг 0.74.

Связь фамилии и Y-генотипа на индивидуальном уровне

Это направление рассматривает связь между конкретной фамилией и конкретной Y-хромосомой.

Сравнение гаплотипов Y-хромосомы [12] в парах английских однофамильцев выявило существование корреляции, хотя и невысокой. При этом лица с редкими фамилиями чаще оказывались родственниками, чем носители частых фамилий.

В работе [6] анализировали Y-хромосомы носителей фамилии Sykes. Ожидалось, что эта фамилия полифилетична и возникала многократно, поскольку происходит от диалектного Йоркширского “ручей”. Однако 44% однофамильцев имеют один и тот же гаплотип Y-хромосомы, причем следующий по частоте гаплотип встречается в 4 раза реже. Это позволило авторам датировать появление своей фамилии.

В Ирландии одной из причин полифилетичности фамилий была их “англизация”. Например, фамилии *Mac Fhíodhbuidhe* и *Mac an Bheatha*, которые не перепутает ни один ирландец, англичане записали одинаково — McEvoy. По данным о NRY у носителей этой фамилии четко выделяются два основателя, которых разделяет более 500 лет [13].

Иногда мотивом для изучения однофамильцев служит “проверка мифа”. Так легенду о происхожде-

дении всего корейского народа от одного предка по имени Taangoon проверяли в работе [7] и, как можно было ожидать, получили отрицательный результат.

Два наиболее крупных исследования посвящены ирландским и британским фамилиям. С использованием стандартной панели из 17 STR-маркеров изучены 1125 однофамильцев, представляющих 43 исторические фамилии Ирландии [13], а также 1678 однофамильцев, носителей 40 британских фамилий [15, 16], причем анализ проведен как на уровне гаплотипов (STR-маркеры), так и на уровне гаплогрупп (SNP-маркеры). Авторы подчеркивают сходство данных, полученных с помощью анализа как гаплогрупп (SNP), так и гаплотипов (STR), что важно для понимания результатов нашей работы. Контрольная популяционная выборка составила 110 человек, однако ее не использовали для коррекции частот гаплогрупп и гаплотипов, в отличие от предлагаемого нами подхода. Рассчитанные средние оценки генетического возраста британских (650 лет) и ирландских (990 лет) фамилий оказались близкими к историческим данным о возникновении как британских (700 лет), так и ирландских (900 лет) фамилий.

Связь фамилии и Y-генотипа с учетом двух уровней — индивидуального и популяционного

Особенность предлагаемого нами подхода заключается в учете обоих уровней изменчивости — индивидуального и популяционного.

Для проведения такого анализа необходимо: 1) собрать образцы ДНК носителей одной и той же фамилии (когорты однофамильцев); 2) проанализировать маркеры Y-хромосомы для этих образцов; 3) определить частоту гаплогрупп Y-хромосомы в каждой когорте (носителей одной фамилии); 4) определить частоту этих же гаплогрупп Y-хромосомы в тех популяциях, из которых родом однофамильцы (общий фон частоты гаплогрупп Y-хромосомы). Если окажется, что частота гаплотипов у однофамильцев превышает популяционный (фоновый) уровень, то мы можем утверждать — связь фамилии с конкретным вариантом Y-хромосомы неслучайна и фамилия не полностью полифилетичная. На основе этого сравнения можно дать количественную оценку **степени монофилетичности** фамилии, т.е. оценить, какая часть носителей фамилии восходит к одному общему предку, передавшего им и свою фамилию, и свою Y-хромосому. В случае полного совпадения гаплотипов NRY в группе однофамильцев мы получаем свидетельство о прямом родстве однофамильцев, хотя надежность такого свидетельства зависит от числа маркеров Y-хромосомы и их частоты в популяциях.

Если таким образом проанализировать одну—две фамилии, то результат будет иметь значение для частной истории только этих фамилий. Но если рассмат-

ривать выборки, репрезентативные для многих русских фамилий, то результат будет иметь общее значение для изучения феномена русских фамилий и генофонда русского народа, а, учитывая немногочисленность подобных зарубежных работ, — и для понимания феномена европейских фамилий в целом.

Предлагаемый подход позволяет ответить на вопросы: существует ли реальная связь между русскими фамилиями и генотипом? И если существует, то сколь велика полифилетичность русских фамилий? Каковы генотипы основателей, передавших и фамилию, и Y-хромосому своим потомкам? Эти простые по форме вопросы до сих пор не получили разрешения для совокупности фамилий, а от ответа на них зависит, в каком направлении будет дальше развиваться комплексное изучение фамилий и генофонда.

Реализация этого подхода наталкивается на целый ряд затруднений, и дальнейшее описание посвящено конкретным способам их устранения.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Охват всего ареала генофонда. Уникальный по своей полноте банк данных о географии русских фамилий включает 70 тысяч фамилий, охватывающих 1.2 млн человек из более чем 1000 популяций разных регионов Центральной России, Русского Севера, Северного Кавказа, Южной Сибири. Из этого массива данных выделены пять регионов (табл. 2) в пределах исторического ареала русского народа. Для минимизации эффекта недавних миграций анализировали только сельское население.

Показано [4, 29], что каждый регион характеризуется собственным “фамильным портретом”. Например, в Западном регионе преобладают календарные фамилии (Ивановы, Васильевы, Федоровы и т.д.), а Южный регион отличается обилием профессиональных фамилий (Кузнецовы, Поповы, Бондаревы и т.д.). Эти различия позволяют предполагать, что в каждом регионе собственная традиция образования фамилий может привести к собственному уровню полифилетичности и, следовательно, к разной доле генетически сходных однофамильцев. Поэтому только в том случае, когда ареал изучаемых однофамильцев охватывает все основные регионы, мы можем выявить закономерности, общие для русских фамилий в целом.

Включение в выборку только неродственных представителей коренного населения. В этих же пяти регионах (дополненных кубанскими казаками) изучен полиморфизм Y-хромосомы: образцы ДНК от 1631 мужчины [30], представляющих 19 русских популяций из всех частей исторического ареала русских (табл. 1). Обследуемых отбирали таким образом [30, 31], чтобы они представляли только коренное население и не включали родственников, что

принципиально важно в данном случае. Это достиглось следующими методами.

Во-первых, составляли родословную каждого обследуемого и включали его в выборку только в том случае, если все его предки в трех поколениях происходили из данной популяции и считали себя русскими. Опыт показывает, что именно у обследуемых, которые по всем линиям (а не только по отцовской, маркируемой Y-хромосомой) принадлежат к данной популяции, с большей вероятностью нет предков из других регионов и в более дальних поколениях.

Во-вторых, родство исключали одновременно тремя методами: по родословным; с помощью информаторов; прямым опросом (каждому кандидату предъявляли список прошедших обследованных, и он должен был указать, входят ли в список его близкие или дальние родственники). При наличии в списке однофамильцев обычно задавали вопрос — не являются ли они родственниками? Такой метод прямо указывает, что однофамильцы в наших выборках были уверены в отсутствии родственных отношений между ними.

Анализируемые группы однофамильцев. В изученном по Y-хромосоме массиве популяционных данных 228 фамилий встречались у двух и более человек. Число носителей этих неуникальных фамилий составило 954 человека. Однофамильцев из них отобрали по двум критериям.

1) Все однофамильцы должны быть только из регионов, детально изученных в отношении гаплогрупп Y-хромосомы [30], и фамилий [4, 29]. Этот критерий обеспечивает знание закономерностей распространения фамилий и Y-хромосомы, необходимое для понимания и правильной интерпретации получаемых результатов. Этому критерию удовлетворяют 13 популяций (табл. 1).

2) Число однофамильцев должно быть больше пяти. Обоим критериям удовлетворяют 14 фамилий (число носителей указано в скобках): Булыгин (16), Иванов (15), Васильев (10), Ильин (10), Кузьмин (10), Павлов (8), Тарасов (8), Попов (7), Семенов (7), Соколов (7), Черноусов (7), Павлов (6), Кувалдин (6), Михайлов (6).

Анализируемый набор гаплогрупп Y-хромосомы. В отобранных 13 популяциях (1244 человека) обнаружены 22 гаплогруппы Y-хромосомы [30].

Из 22 обнаруженных гаплогрупп только восемь (табл. 1) удовлетворяют критерию полиморфизма (т.е. средняя частота выше 1%). Это гаплогруппы **R1a** (40.4%), **N1c1** (14.5%), **I1b** (8.9%), **R1b1c** (5.1%), **I1** (3.9%), **J2** (2.3%), **N1b** (2.2%), **E1b1b1** (2.0%). Поэтому на диаграммах анализируется распределение этих восьми гаплогрупп в каждой группе однофамильцев.

Из табл. 1 мы видим, что популяции значительно различаются по частотам гаплогрупп, что подтверждено картографированием русских популяций [4,

30]. Это означает, что в качестве ожидаемой случайной (популяционной) частоты той или иной гаплогруппы среди однофамильцев мы не можем использовать среднюю частоту этой гаплогруппы, характерную для русского народа в целом, а должны учесть ее частоту именно в тех популяциях, где встречены носители данной фамилии. С этой целью разработан специальный алгоритм.

Алгоритм оценки частот гаплогрупп Y-хромосомы у однофамильцев. Введем три понятия частоты гаплогруппы у однофамильцев: наблюдаемая (f_o), ожидаемая (f_e), накопленная (f_n).

Наблюдаемая (f_o) частота гаплогрупп у однофамильцев рассчитывается как отношение числа носителей гаплогруппы среди носителей фамилии к общему числу носителей фамилии. Например, зная, что из 10 носителей данной фамилии четверо обладают гаплогруппой **R1a**, трое — гаплогруппой **N1c1** и трое — гаплогруппой **N1b**, $f_o(\mathbf{R1a}) = 0.40$, $f_o(\mathbf{N1c1}) = 0.30$, $f_o(\mathbf{N1b}) = 0.30$. Казалось бы, данный блок однофамильцев отличается повышенной частотой **R1a** и в равной мере содержит гаплогруппы **N1c1** и **N1b**. Однако такая гипотеза (обозначим ее как гипотезу № 1) будет неверна, если популяционные частоты этих гаплогрупп различны. Чем выше частота гаплогруппы в популяции, тем больше вероятность, что гаплогруппа у однофамильцев совпала случайно, а не потому, что они имеют общего предка по мужской линии — основателя данной фамилии. Например, в популяции народа фон (Бенин) представлена только одна гаплогруппа **E3a** [9], поэтому какую бы фамилию мы не взяли, все ее носители будут иметь эту гаплогруппу, независимо от того, происходят ли они от общего основателя или нет. Поэтому необходимо ввести понятие ожидаемой популяционной частоты гаплогруппы для блока однофамильцев.

Ожидаемая (f_e) частота гаплогрупп у однофамильцев целиком зависит от частоты встречаемости гаплогруппы во всей популяции и не имеет отношения к родству однофамильцев. Смысл ожидаемой частоты f_e — именно такой должна быть частота гаплогруппы в выборке, если ее члены выбраны из популяции случайно по отношению к их генотипу. Если же мы увидим, что реальная наблюдаемая (f_o) частота гаплогруппы у однофамильцев выше ожидаемой f_e (“случайной”), то это означает, что они объединены неслучайно, т.е. выбор по признаку “совпадение фамилии” повлек за собой увеличение частоты определенных гаплогрупп. Иными словами — фамилия связана с генотипом. Но если среди однофамильцев частоты гаплогрупп окажутся такими же, как и в популяции (наблюдаемая f_o равна ожидаемой f_e), то это значит, что они действительно лишь однофамильцы, но не родственники.

Возьмем для приведенного выше примера в качестве ожидаемых средние русские частоты гаплогрупп

(табл. 1): $f_e(\mathbf{R1a}) = 0.40$, $f_e(\mathbf{N1c1}) = 0.15$, $f_e(\mathbf{N1b}) = 0.02$. Мы сразу увидим, что выдвинутая выше гипотеза № 1, основанная только на наблюдаемой частоте f_o , ошибочна. У однофамильцев нет накопления гаплогруппы **R1a** (поскольку $f_o = f_e = 0.40$), есть двукратное увеличение гаплогруппы **N1c1** ($f_o = 2f_e$), а гаплогруппа **N1b** встречается в 15 раз чаще, чем в популяции ($f_o = 15f_e$). Такое неслучайное накопление **N1b** позволяет выдвинуть гипотезу № 2, согласно которой обладатели гаплогруппы **N1b** состоят в родстве между собой, а обладатели гаплогруппы **R1a** — просто однофамильцы.

Однако и эта гипотеза может оказаться неверной, поскольку она не учитывает, что конкретные популяции, из которых происходят однофамильцы, могут значительно отличаться по частотам гаплогрупп от средних русских частот. Например, на Русском Севере преобладает гаплогруппа **N1c1**, а в Центральной и Южной России — гаплогруппа **R1a** (табл. 1). И наши выводы о родстве однофамильцев будут зависеть от того, в какой популяции — на севере или на юге — встречены однофамильцы с этими гаплогруппами.

Поэтому ожидаемая частота каждой гаплогруппы в данном блоке однофамильцев должна рассчитываться как средневзвешенная частот этой гаплогруппы именно в тех популяциях, из которых происходят однофамильцы, а весом будет выступать общее число однофамильцев (безотносительно к генотипу), встреченных в каждой из популяций. Обозначим популяционные частоты данной гаплогруппы: в популяции **1** как f_{pop1} ; в популяции **2** как f_{pop2} ; в популяции **k** как f_{popk} . А число всех носителей данной фамилии в популяции **1** обозначим как N_{pop1} ; в популяции **2** как N_{pop2} ; в популяции **k** как N_{popk} . Тогда искомая ожидаемая частота f_e данной гаплогруппы рассчитывается по формуле:

$$f_e = (N_{pop1}f_{pop1} + N_{pop2}f_{pop2} + \dots + N_{popk}f_{popk}) / (N_{pop1} + N_{pop2} + \dots + N_{popk}).$$

Накопленная частота гаплогрупп NRY (f_n). Выше мы уже рассчитывали частоту накопления как разность между наблюдаемой и ожидаемой частотами. Причем для обнаружения родства однофамильцев важны лишь превышения наблюдаемой частоты над ожидаемой. Поэтому, если f_n имеет отрицательное значение, то ее можно принять за 0:

$$f_n = f_o - f_e, \text{ если } f_o < f_e, \text{ то } f_n = 0.$$

Средние географические расстояния между популяциями однофамильцев (\bar{d}_{geo}). Эта вспомогательная величина помогает оценить “географическое накопление” однофамильцев и рассчитывается как среднее географическое расстояние (в км) между всеми популяциями, в которых встречены носители данной фамилии.

$$\bar{d}_{geo} = (d_{geo1-2} + d_{geo1-3} + \dots + d_{geok-n}) / k,$$

(где d_{geok-n} — расстояния между k -й и n -й популяциями, в которых встречены однофамильцы).

Если носители фамилии обнаружены только в одной популяции, то $\bar{d}_{geo} = 0$, и это локальная фамилия. Если носители фамилии встретились во всех изученных нами популяциях, то $\bar{d}_{geo} = \bar{d}_{MEAN}$, и данная фамилия может считаться широко распространенной. Если носители фамилии обнаружены только на противоположных концах ареала, то $\bar{d}_{geo} > \bar{d}_{MEAN}$, и ареал данной фамилии является разорванным. Этот показатель важен потому, что для широко распространенных фамилий можно ожидать большую степень полифилетичности, чем для локальных фамилий. Географические расстояния удобно использовать в виде индекса, показывающего отношение наблюдаемого \bar{d}_{geo} к среднему: $I_{geo} = \bar{d}_{geo} / \bar{d}_{MEAN}$.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Рассмотрим применение изложенных подходов на примере нескольких фамилий. Представим частоты гаплогрупп в выборке однофамильцев в виде полигона (рис. 1, рис. 2), т.е. диаграммы в виде звезды, каждый луч которой соответствует одной из восьми гаплогрупп Y-хромосомы. На каждом луче отложена частота соответствующей гаплогруппы в выборке носителей данной фамилии: в центре — нулевые значения (у однофамильцев нет данной гаплогруппы); на конце луча — максимальная встреченная популяционная частота (0.6). На рисунках представлены полигоны распределения гаплогрупп Y-хромосомы в выборках однофамильцев — **Булыгиных, Кувалдиных, Ильиных, Семеновых** (рис. 1), **Ивановых, Поповых, Соколовых, Павловых** (рис. 2), а общие характеристики этих блоков однофамильцев приведены в табл. 3.

Прежде всего, обратим внимание на различия между наблюдаемыми частотами f_o (рис. 1а и рис. 2а) и накопленными частотами f_n (рис. 1б и рис. 2б) гаплогрупп у однофамильцев. Эти различия ярко показывают значение предложенного нами алгоритма, избавляющего от влияния популяционных частот, фон которых имитирует родство однофамильцев.

Фамилии с высоким накоплением гаплогрупп Y-хромосомы

На рис. 1 приведены две локальные редкие фамилии — **Булыгины** и **Кувалдины** — у которых можно ожидать меньшую полифилетичность, чем у частых русских фамилий, представленных на рис. 1 — **Ильиных** и **Семеновых** (соответственно 26 и 17 место в общем списке русских фамилий, табл. 3), большую полифилетичность которых можно ожидать. Например, обнаружено увеличение полифилетичности британских фамилий с ростом их распространенности [15]. У ирландцев такая связь не

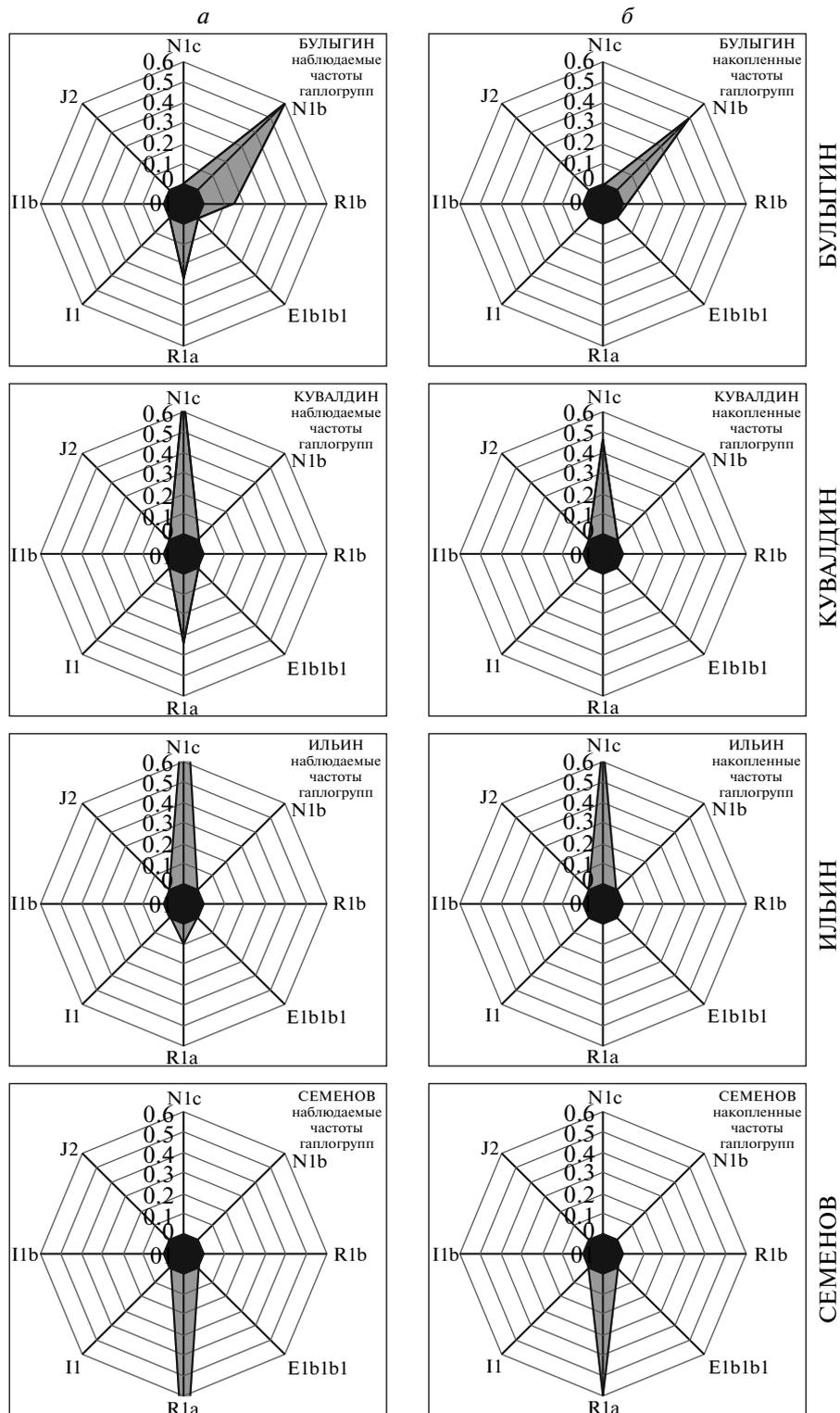


Рис. 1. Распределение гаплогрупп Y-хромосомы (R1a, I1, I1b, J2, N1c1, N1b, R1b1c, E1b1b1) в четырех блоках однофамильцев: Булыгины, Кувалдины, Ильины, Семеновы (перечислены сверху вниз). Для каждого блока однофамильцев приведены наблюдаемые f_o (a) и накопленные f_n (б) частоты гаплогрупп.

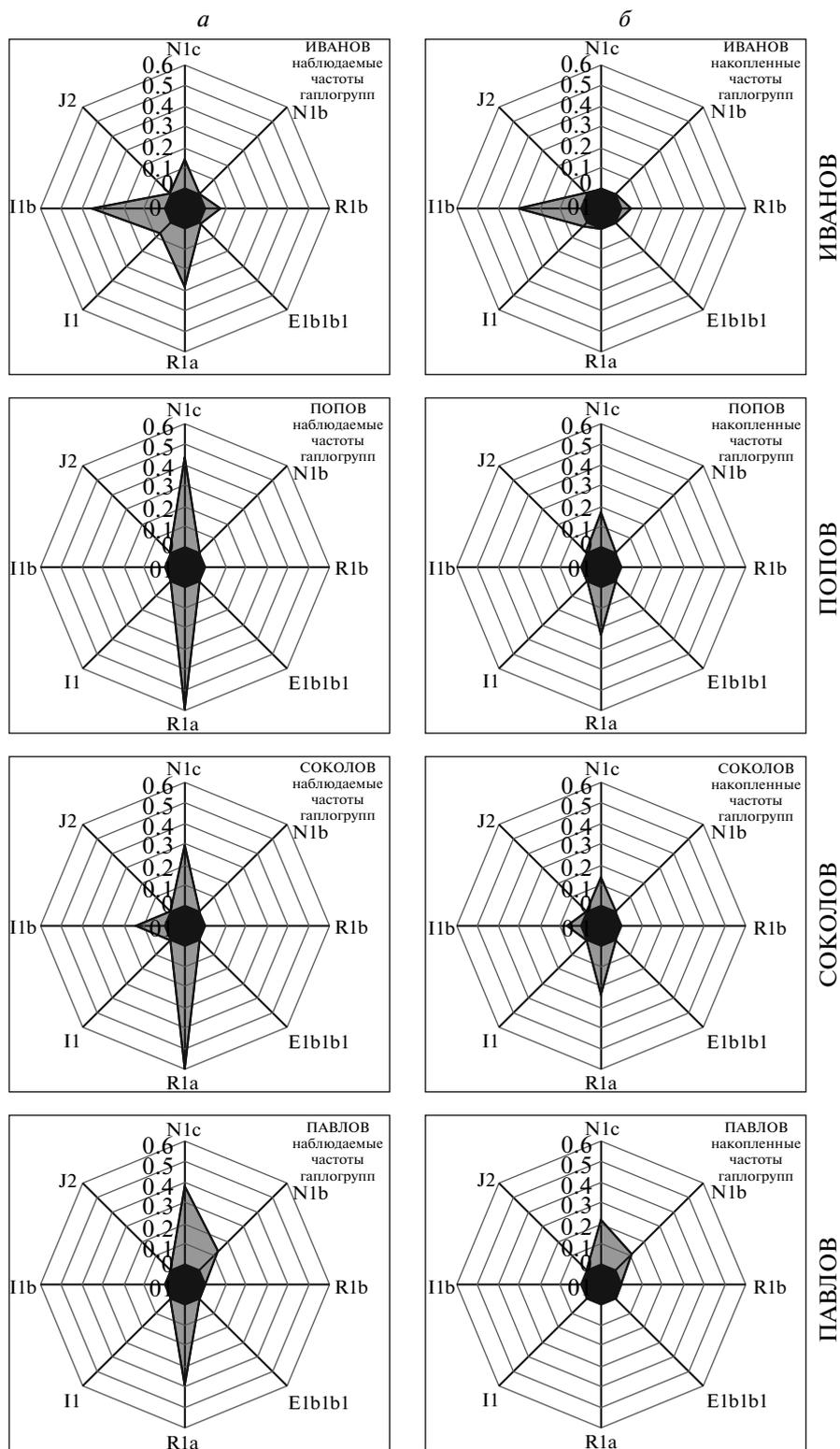


Рис. 2. Распределение гаплогрупп Y-хромосомы (R1a, I1, I1b, J2, N1c1, N1b, R1b1c, E1b1b1) в четырех блоках однофамильцев: Ивановы, Поповы, Соколовы, Павловы (перечислены сверху вниз). Для каждого блока однофамильцев приведены наблюдаемые f_o (a) и накопленные f_n (б) частоты гаплогрупп.

Таблица 3. Характеристики блоков однофамильцев

Фамилии	Порядковый номер фамилий*	Однофамильцы, N	Популяции, N	Галлогруппы, N	Встреченные галлогруппы**	Характерная галлогруппа			Среднее расстояние между популяциями, км d_{geo}	Индекс расстояний, I_{geo}^{***}	Район****
						галлогруппа	накопленная частота, f_n	наблюдаемая частота, f_o			
Булыгины	35	15	1	3	N1b, R1a, R1b1c	N1b	0.46	0.60	0	0	Пинега (15)
Кувалдины	245	6	1	2	N1c1, R1a	N1c1	0.46	0.67	0	0	Пинега (6)
Ильины	26	10	2	2	N1c1, R1a	N1c1	0.69	0.90	1202	1.3	Пинега (9), Остров (1)
Семеновы	17	7	2	1	R1a	R1a	0.60	1.00	87	0.1	Остров (4), Порхов (3)
Ивановы	2	14	3	6	I1b, R1a, N1c1, R1b1c, I1, E3b3	I1b	0.28	0.36	344	0.4	Остров (8), Порхов (5), Рославль (1)
Поповы	5	7	3	2	R1a, N1c1	R1a	0.22	0.57	1121	1.2	Мезень (3), Красноборск (2), Белгород (2)
Соколовы	4	7	6	3	R1a, N1c1, I1b	R1a	0.24	0.57	1021	1.1	Унжа (2), Красноборск (1), Пинега (1), Кашин (1), Порхов (1), Кубанские казаки (1)
Павловы	15	8	5	4	N1c1, R1a, N1b, G2	N1c1	0.22	0.38	1127	1.2	Остров (3), Кубанские казаки (2), Рославль (1), Пинега (1), Кашин (1)
<i>В среднем</i>		9	3	3			0.40	0.63	613	0.7	

* Номер фамилии указывает на ее относительную частоту: он означает место данной фамилии в списке всех русских фамилий, расположенных по убыванию средней частоты.

** Галлогруппы приведены в порядке убывания их частот в данном блоке однофамильцев.

*** $I_{geo} = \bar{d}_{geo} / \bar{d}_{MEDIAN}$

**** В скобках указано число однофамильцев, встреченных в данном районе.

Примечание. Фамилии приведены в том же порядке, что и на рис. 1 и рис. 2; обозначения галлогрупп согласно классификации 2008 г. [Karafet et al., 2008].

обнаружена [13], хотя это может быть связано с тем, что рассматривали только “исторические” ирландские фамилии. В случае четырех русских фамилий мы видим, что независимо от распространенности фамилий накопленные частоты всех четырех — и редких, и частых — фамилий (рис. 1 и табл. 3) указывают на монофилетичность, выявляя лишь одну гаплогруппу с очень высокой неслучайной частотой f_n . Именно эта гаплогруппа может считаться характерной для данной фамилии: у **Булыгиных** — это гаплогруппа **N1b** (с накопленной частотой $f_n = 0.46$, при наблюдаемой частоте $f_o = 0.60$); у **Кувалдиных** — **N1c1** ($f_n = 0.46, f_o = 0.67$); у **Ильиных** — **N1c1** ($f_n = 0.69, f_o = 0.90$); у **Семеновых** — **R1a** ($f_n = 0.60, f_o = 1.00$).

При этом как все 15 **Булыгиных**, так и все шесть **Кувалдиных** оказались родом из одной изолированной популяции — Пинежского района Архангельской области ($\bar{d}_{geo} = 0$). На Пинеге однофамильцев настолько много, что кроме фамилии, имени и отчества широко используются и “царские” обозначения: Булыгин Иван Ильич I, Булыгин Иван Ильич IV, причем эти обозначения фигурируют даже в местных официальных документах. Поэтому на Пинеге мы особенно тщательно отслеживали, чтобы однофамильцы, включаемые в выборку, не считали себя родственниками. Однако вопреки локальности этих фамилий около трети Булыгиных и Кувалдиных обладают иными гаплогруппами, кроме основной (табл. 3): **Булыгины** несут еще **R1a** ($f_o = 0.27$) и **R1b1c** ($f_o = 0.13$), а **Кувалдины** — еще **R1a** ($f_o = 0.33$).

Распространенная фамилия **Ильин** встречена в двух удаленных популяциях ($\bar{d}_{geo} = 1202$ км, $I_{geo} = 1.3$). Однако 9 из 10 **Ильиных** обнаружены все на той же Пинеге и все они содержат гаплогруппу **N1c1**. Таким образом, мы видим, что независимо от общей “распространенности” фамилии, высокая концентрация ее носителей в одной популяции может привести к тому, что большая часть однофамильцев восходит к одному основателю: для архангельских Ильиных может быть выдвинута гипотеза общего основателя с гаплогруппой **N1c1**.

Сходную ситуацию мы видим и у семи **Семеновых**, хотя четверо из них встречены в Островском районе, а трое — в Порховском, однако оба района расположены недалеко друг от друга в пределах Псковской области ($\bar{d}_{geo} = 87$ км, $I_{geo} = 0.1$). Поэтому для псковских Семеновых можно выдвинуть гипотезу общего основателя с гаплогруппой **R1a**.

Эти гипотезы необходимо проверить с помощью STR-гаплотипов Y-хромосомы (конкретных вариантов этих гаплогрупп), время возникновения которых сопоставимо с возрастом фамилий, в отличие от возраста гаплогрупп.

Фамилии с низким накоплением гаплогрупп Y-хромосомы

На рис. 2 специально отобраны блоки однофамильцев (**Ивановы, Поповы, Соколовы, Павловы**), которые отличаются от предыдущих (рис. 1) по всем основным параметрам (табл. 3). Во-первых, однофамильцы встречены в трех и более районах, представляющих разные регионы. Во-вторых, расстояния между популяциями велики ($\bar{d}_{geo} \approx 350$ км у **Ивановых**; $\bar{d}_{geo} \approx 1100$ км у **Поповых, Соколовых и Павловых**). В-третьих, все фамилии относятся к самым распространенным русским фамилиям — они занимают в общерусском списке фамилий соответственно 2, 5, 4 и 15 места.

Таким образом, на рис. 2 представлены выборки носителей самых частых русских фамилий, которые встречены в удаленных друг от друга популяциях. Все эти факторы считаются причиной полифилетичности — многократного возникновения одной фамилии от разных основателей. В наших данных это должно было бы проявиться в чисто случайном распределении гаплогрупп Y-хромосомы: накопленные частоты гаплогрупп при полифилетичности должны быть равны 0.

Действительно, мы видим некоторый сдвиг в сторону полифилетичности в сравнении с фамилиями на рис. 1: число встреченных гаплогрупп в каждом блоке однофамильцев (рис. 2, табл. 3) выросло почти в 2 раза, а средняя частота самой частой гаплогруппы упала примерно в 2 раза (с 0.55 до 0.24). Тем не менее, в каждом блоке фамилий наблюдается явное накопление гаплогрупп Y-хромосомы, указывая на неслучайность их распределения. У **Ивановых** — это гаплогруппа **I1b**, у **Поповых** и **Соколовых** — **R1a** и **N1c1**, у **Павловых** — **N1c1** и **N1b**.

Это указывает на ограниченность явления полифилетичности даже среди распространенных русских фамилий и позволяет сделать вывод: **русские фамилии имеют ограниченное число основателей, и каждая фамилия обладает характерным профилем гаплогрупп Y-хромосомы.**

Анализ гаплотипов Y-хромосомы с помощью STR-маркеров

Рассматривая накопление гаплогрупп Y-хромосомы, мы можем ответить только на вопрос о степени полифилетичности фамилии, но не можем достоверно отличить родственников от однофамильцев. Наличие одной и той же гаплогруппы у однофамильцев увеличивает вероятность того, что они родственники, но гаплогруппы намного древнее, чем фамилии, поэтому для подтверждения родства необходим анализ более поздних мутаций Y-хромосомы. С этой целью, как и в наиболее детальных исследованиях народов мира [13, 15], мы проанализировали 17 STR-маркеров Y-хромосомы, об-

ладающих высокой скоростью мутаций [32]. При различии в трех и более локусах из 17 мы уже далеко выходим за историческое время существования русских фамилий, что позволяет выявить родственников среди однофамильцев.

В качестве примера приведем анализ STR-маркеров у носителей фамилии **Тюрюхановы**, собранных в различных населенных пунктах Забайкалья. Все эти Тюрюхановы не считают себя родственниками, однако анализ STR-маркеров показал существование родственной связи между всеми проанализированными однофамильцами. Из семи Тюрюхановых у четверых оказались полностью идентичные 17 STR-маркеров, а остальные трое отличались от “базового” STR-гаплотипа по одному из 17 STR-маркеров каждый. Это позволяет выдвинуть гипотезу о близкой степени родства четырех Тюрюхановых и более отдаленного родства остальных троих Тюрюханов — как между собой, так и по отношению к основной группе. Используя данные о скорости мутирования данных STR-маркеров [32], время расхождения с предковым гаплотипом оценивается в 300–400 лет. Полученный результат соответствует преданию об основателе — Тюрюханове, переселившемся в Забайкалье около 400 лет назад. Основной вывод STR-анализа — все обследованные однофамильцы являются родственниками с двумя различными степенями родства.

Изложенные подходы, алгоритмы анализа и примеры их использования позволяют сделать три вывода.

1) При анализе однофамильцев необходимо учитывать **популяционные частоты** гаплогрупп Y-хромосомы; использовать популяционные частоты для **локальных популяций**, откуда происходят однофамильцы (согласно предложенному алгоритму); и использовать при интерпретации результатов связи между фамилией и Y-хромосомой только **“накопленные” частоты** гаплогрупп. Игнорирование популяционного фона (популяционного профиля гаплогрупп Y-хромосомы) создает ложные связи, имитирует генетическое сходство и завышает степень родства однофамильцев. Нами предложен новый алгоритм — коррекцию частот гаплотипов и гаплогрупп с использованием популяционного фона ранее не проводили даже в наиболее детальных исследованиях [13, 15].

2) Анализ накопленных частот гаплогрупп Y-хромосомы показал, что каждый блок однофамильцев восходит к ограниченному числу основателей: **впервые выявлена ограниченная степень полифилетичности русских фамилий**. Этот вывод относится не только к локальным, но и к самым распространенным русским фамилиям.

3) При определении степени родства однофамильцев анализ гаплогрупп Y-хромосомы необходимо дополнять изучением набора STR-маркеров Y-хромосомы.

Авторы благодарят Рихарда Виллемса за возможность анализа гаплогрупп Y-хромосомы на базе Эстонского биоцентра, С.А. Фролову за анализ STR-маркеров, глав администраций и медицинский персонал районов за помощь в организации обследования. Эта работа была бы невозможна без 1300 представителей коренного русского населения, давших согласие на изучение образцов их ДНК.

Исследование выполнено при поддержке Российского гуманитарного научного фонда (07-06-00448а) и Российского фонда фундаментальных исследований (10-07-00515а, 08-06-00102а, 07-07-000890а).

СПИСОК ЛИТЕРАТУРЫ

1. Karafet T.M., Mendez F.L., Meilerman M.B., Underhill P.A., Zegura S.L., Hammer M.F. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* **18(5)**, 830–838.
2. Харьков В.Н., Степанов В.А., Медведева О.Ф., Спиридонова М.Г., Максимова Н.Р., Ноговицына А.Н., Пузырев В.П. 2008. Происхождение якутов: анализ гаплотипов Y-хромосомы. *Молекуляр. биология.* **42**, 226–237.
3. Pshenichnov A., Balanovska E., Balanovsky O., Atramentova L., Churnosov M., Soloviova D., Zaporozhchenko V. 2007. Variation of paternal, maternal and autosomal genetic markers on intra-ethnic (Ukrainians) and inter-ethnic (Europe) level supports Y chromosomal marker bias. *Intern. Soc. Appl. Biol. Sci. Split.* 128.
4. Балановская Е.В., Балановский О.П. 2007. *Русский генофонд на Русской равнине*. М.: Луч.
5. Macintyre S., Sooman A. 1991. Non-paternity and prenatal genetic screening. *Lancet.* **5, 338(8771)**, 869–871.
6. Sykes B., Irven I. 2000. Surnames and the Y chromosome. *Am. J. Hum. Genet.* **66(4)**, 1417–1419.
7. Jobling M.A. 2001. In the name of the father: surnames and genetics. *Trends Genet.* **17(6)**, 353–357.
8. Zei G., Lisa A., Fiorani O., Magri C., Quintana-Murci L., Semino O., Santachiara-Benerecetti A.S. 2003. From surnames to the history of Y chromosomes: the Sardinian population as a paradigm. *Eur. J. Hum. Genet.* **11(10)**, 802–807.
9. Luis J.R., Rowold D.J., Regueiro M., Caeiro B., Cinnioglu C., Roseman C., Underhill P.A., Cavalli-Sforza L.L., Herrera R.J. 2004. The Levant versus the Horn of Africa: Evidence for bidirectional corridors of human migrations. *Am. J. Hum. Genet.* **74**, 532–544.
10. Bedoya G., Montoya P., Garcia J. 2006. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc. Nat. Acad. Sci. USA.* **103(19)**, 7234–7239.
11. Immel U.D., Krawczak M., Udolph J., Richter A., Rodig H., Kleiber M., Klintschar M. 2006. Y-chromosomal STR haplotype analysis reveals surname associated strata in the East-German population. *Eur. J. Hum. Genet.* **14(5)**, 577–582.

12. King T.E., Ballereau S.J., Schürer K.E., Jobling M.A. 2006. Genetic signatures of coancestry within surnames. *Curr. Biol.* **21**, 384–388.
13. McEvoy B., Bradley D.G. 2006. Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum. Genet.* **119**(1–2), 212–219.
14. Moore L.T., McEvoy B., Cape E., Simms K., Bradley D.G. 2006. A Y-chromosome signature of hegemony in Gaelic Ireland. *Am. J. Hum. Genet.* **78**(2), 334–338.
15. King T.E., Jobling M.A. 2009. Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. *Mol. Biol. Evol.* **26**(5), 1093–1102.
16. King T.E., Jobling M.A. 2009. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet.* **25**(8), 351–360.
17. Рычков Ю.Г., Балановская Е.В. 1990. Этническая генетика: Соотношение адаптивной и нейтральной генетической дифференциации этносов. *Генетика.* **26**, 541–549.
18. Ельчинова Г.И., Кадошников М.Ю., Мамедова Р.А. 1991. Выявление особенностей генетической структуры популяций с помощью метода описания “генетического ландшафта”. *Генетика.* **27**, 1994–2001.
19. Почешхова Э.А. 2008. Топография фамилий и генетических расстояний (адыги Западного Кавказа). *Мед. генетика.* **7**, 21–29.
20. *Наследственные болезни в популяциях человека.* 2002. Под ред. Е.К. Гинтера. М.: Медицина.
21. Чурносов М.И., Сорокина И.Н., Костоглодова И.Н., Песик В.Ю., Рудых Н.А., Аристов И.К., Жерлицына М.С., Балановская Е.В. 2004. Структура генофонда населения Белгородской области: квазигенетические, биохимические, молекулярно-генетические и популяционно-демографические параметры. В сб.: *Экология и демография человека в прошлом и настоящем.* М.: 289–290.
22. Ельчинова Г.И., Кадошников М.Ю., Мамедова Р.А., Букина А.М., Петрова Н.В., Старцева Е.А. 1991. О частотном критерии выбора фамилий для изучения генетической структуры популяций. *Генетика.* **27**, 358–360.
23. Ревазов А.А., Парадеева Г.М., Русакова Г.И. 1986. Пригодность русских фамилий в качестве “квазигенетического” маркера. *Генетика.* **22**, 699–703.
24. Cavalli-Sforza L.L., Bodmer W.F. 1971. *The genetics of human populations.* San Francisco: Freeman.
25. Гинтер Е.К. 1993. Влияние генетической структуры на груз наследственных болезней в русских популяциях. *Вестник РАМН.* **9**, 23–26.
26. Зинченко Р.А., Ельчинова Г.И., Балановская Е.В., Нурбаев С.Д., Гинтер Е.К. 2000. Влияние генетической структуры популяций на размеры груза моногенных наследственных болезней в российских популяциях. *Вестник РАМН.* **5**, 5–10.
27. Зинченко Р.А., Ельчинова Г.И., Гинтер Е.К. Факторы, определяющие распространение наследственных болезней в российских популяциях. 2009. *Мед. генетика.* **8**, 7–23.
28. Почешхова Э.А., Балановская Е.В., Серегин Ю.А., Голубцов В.И., Балановский О.П. 2008. Динамика генофонда во времени по данным о фамилиях и родословных. *Мед. генетика.* **7**, 3–10.
29. Балановская Е.В., Балановский О.П., Соловьева Д.С., Сорокина И.Н., Лавряшина М.Б., Почешхова Э.А. 2007. “Фамильный портрет” русского генофонда. В сб.: *Человек в культурной и природной среде.* М.: Наука, 247–267.
30. Balanovsky O., Rootsi S., Pshenichnov A., Kivisild T., Churnosov M., Evseeva I., Pocheshkhova E., Boldyreva M., Yankovsky N., Balanovska E., Villems R. 2008. Two sources of the Russian patrilineal heritage in their Eurasian context. *Am. J. Hum. Genet.* **82**, 236–250.
31. Соловьева Д.С., Балановская Е.В., Кузнецова М.А., Васинская О.А., Фролова С.А., Почешхова Э.А., Евсева И.В., Болдырева М.Н., Балановский О.П. 2010. Русский генофонд: геногеография ALU-инсерций (ACE, APOA1, B65, PV92, TRA25). *Молекуляр. биология.* **44**, 393–400.
32. Ge J., Budowle B., Aranda X.G., Planz J.V., Eisenberg A.J., Chakraborty R. 2009. Mutation rates at Y chromosome short tandem repeats in Texas populations. *Forensic Sci. Int. Genet.* **3**, 179–184.