



ИСПОЛЬЗОВАНИЕ ЧАСТЫХ ФАМИЛИЙ В ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЯХ

**И.Н. СОРОКИНА, А.В. ВЕРЗИЛИНА
А.В. ЕЛЫКОВА, Н.А. ДЕМАКОВА
М.И. ЧУРНОСОВ**

*Белгородский
государственный
национальный
исследовательский
университет*

e-mail: sorokina@bsu.edu.ru

В статье даны характеристики трех основных критериев (“частотный”, “территориальный”, “демографический” критерии) отбора распространенных фамилий и возможности их использования при проведении популяционно-генетических исследований.

Ключевые слова: частые фамилии, критерии отбора частых фамилий.

В предыдущих работах нами была дана характеристика фамилий как “квазигенетических” маркеров [1], показаны возможности использования фамилий как для описания отдельных популяционно-генетических характеристик населения [2, 3], так и генетических взаимоотношений между элементарными популяциями [4-6]. В настоящей работе представлены результаты анализа возможностей использования частых фамилий, отобранных в соответствии с тремя основными критериями отбора распространенных фамилий (“частотный”, “территориальный”, “демографический” критерии) при проведении популяционно-генетических исследований.

При работе с фамилиями одной из трудностей является их большая численность. В многочисленных популяциях фамилий бывает достаточно много (часто до нескольких тысяч или десятков тысяч), что затрудняет их обработку, а кроме того, вносит определенный шум, поскольку единичные носители какой-либо фамилии (а их бывает до 10%) являются, как правило, мигрантами [7, 8]. Для того, чтобы использовать фамилии для изучения истории генофонда, необходимо исключить «информационные шумы» – то есть редкие для данной популяции фамилии, которые в ней исторически случайны, эфемерны и потому мало информативны для изучения истории ее генофонда [9]. Поэтому ряд исследователей ограничивается изучением какой-то части фамилий, определив при этом достаточный уровень их частоты с помощью определенного критерия. В целой серии работ [8-11] посвященных описанию генетической структуры различных популяций (Кировская, Костромская, Смоленская, Тверская, Архангельская, Курская области и др.) убедительно показано, что исключение редких и использование только распространенных фамилий дает адекватную оценку структуры генофонда.

Работая с базой фамилий, единственной информацией, которой располагает исследователь о каждом человеке, является его фамилия и его населенный пункт, а также количество однофамильцев в других населенных пунктах. В связи с этим возникает вопрос, как отсеять “пришлого” население с редкими фамилиями? Есть несколько подходов – критериев для отбора “коренных” фамилий, и все они основываются на том, что пришельцев не может быть слишком много – пришлые фамилии в среднем более редкие, чем коренные.

В отечественной популяционной генетике используются три критерия, разделяющие фамилии на редкие и распространенные: “частотный” критерий [8]; “территориальный” критерий [12]; “демографический” критерий [9].

В соответствии с “частотным” критерием фамилия включается в исследование, если её распространенность превышает определенный порог. Различные варианты “частотного” критерия отбора фамилий и территориальные границы его использования при изучении популяционной структуры населения детально разработаны в работах Ельчиновой Г.И. [7, 8]. При изучении генетической структуры популяций Кировской, Костромской областей и республики Адыгея была оценена допустимость использования только частых фамилий (частота более 0,001) и очень частых фамилий (частота более 0,01) для изучения генетической структуры популяций. Авторами предпола-



гается, что за счет отбрасывания единичных и редких фамилий, носителями которых являются, как правило, мигранты последних лет, получается выборка, в которой частоты генов (фамилий) относятся к коренному населению [7, 8]. В результате работы Ельчиновой Г.И. [7] показана хорошая корреляция между матрицами генетических расстояний, подсчитанных по всем фамилиям, частым фамилиям (ЧФ) (0,93-0,99) и очень частым фамилиям (ОЧФ) (0,63-0,90). Однако отмечается, что при использовании ОЧФ более значительно меняется среднее значение матрицы генетических расстояний и дисперсия, чем при использовании ЧФ. На основе полученных данных автор указывает, что число частых фамилий, если их рассматривать как аллели одного локуса, является достаточным для характеристики популяции, в то же время число ОЧФ представляется недостаточным, хотя обычно число биологических систем, которыми располагают исследователи, ближе к числу ОЧФ.

Возможность использования частых фамилий (с частотой более 0,001) для описания популяционно-генетической структуры Курской области была показана и Ивановым В.П. и др. [11]. Авторами на основе проведенного подсчета коэффициента инбридинга и его составляющих с использованием данных о распределении как всех фамилий, так и только часто встречающихся фамилий установлено, что в большинстве рассматриваемых районов показатели случайного инбридинга F_{st} , определенные на основе всех фамилий и на основе только частых фамилий, были примерно одинаковы. В то же время в двух районах (Кореневском и Льговском) случайный инбридинг F_{st} , вычисленный на основе частых фамилий был в 2-3 раза меньше, чем F_{st} , рассчитанный по всем фамилиям. Исходя из полученных данных, авторы отмечают, что в районах, с долей населения с частыми фамилиями 50 и более процентов, популяционная структура адекватно описывается с использованием частых фамилий, а в районах с минимальным удельным весом населения с частыми фамилиями для корректной оценки популяционной структуры следует использовать данные о распределении всех фамилий.

Другим подходом к отбору «коренных» фамилий является исключение фамилий, встречающихся только в одной – двух локальных популяциях («территориальный» критерий). Первое исследование в общерусском масштабе с использованием «территориального» критерия проведено А.П. Бужиловой для 75 распространенных фамилий [12]. Было показано, что географическая изменчивость распространенных русских фамилий отражает не только события районного масштаба, но и этническую историю русского народа в целом: обнаружена связь между зонами накопления определенных фамилий и основными антропологическими типами русского этноса. В этом исследовании отбор фамилий для анализа определялся по «качественному» принципу: наличие или отсутствие фамилии в данной популяции. Фамилии считаются редкими, если встречаются только в одной популяции и распространенными, если встречаются одновременно в нескольких популяциях (двух и более). В списке распространенных фамилий отдельной группой выделены фамилии, встречающиеся более чем в двух популяциях одновременно (в трех и более). Причем фамилия включалась в анализ при её регистрации в двух локальных популяциях, распределенных по всему русскому ареалу [12].

Согласно «демографического» критерия отбора ЧФ фамилия включается в анализ, если число носителей данной фамилии, достигших репродуктивного возраста, превышает четыре человека в районе, а не в локальной популяции (селе, сельсовете) [9]. Это означает, что даже в том случае, если пять носителей фамилии разбросаны по разным весям одного района и в каждой из пяти отдаленных деревень есть лишь по одному носителю фамилии, то фамилия все равно будет включена в анализ. Авторы назвали его «демографическим», так как он соответствует генетико-демографическим представлениям о расширенном воспроизводстве семьи: двое родителей и более 2 детей, достигших репродуктивного возраста. Именно эти его свойства позволяют предположить, что этот критерий обеспечит отбор тех исторически «неслучайных» фамилий, у которых есть шанс закрепиться в данном генофонде [9].

«Демографический» критерий в отличие от «частотного» критерия, во-первых, не зависит от общей численности изучаемого района, которая чрезвычайно варьирует и часто определяется исторически случайными причинами. Во-вторых, и это главное, он не зависит от количества фамилий в районе: как правило, при большом числе фамилий их часто-

ты малы (вступает в силу зависимость частоты от $1/N$ фамилий). В этом случае множество “коренных” фамилий автоматически отбрасываются “частотным” критерием. А ведь соотношение “число фамилий/численность населения” часто является особой исторически сложившейся чертой региона. Использование частотного критерия в районах с разным исторически сложившимся соотношением числа фамилий и общей численности может привести к резкому искажению параметров генофонда [9].

“Демографический” критерий не является чрезмерно строгим. Достаточно всего пяти человек с данной фамилией из десятков тысяч жителей района для того, чтобы фамилия была включена в разряд “распространенных”, т.е. предположительно “коренных” фамилий. Но при этом “демографический” критерий все же позволяет отсеять те редкие фамилии, которые занесены миграционными потоками, но еще не успели основательно укорениться в данной популяции [13]. Данный критерий в среднем в пять раз мягче частотного. При численности района 25 000 человек 0,1% частотный критерий требует наличия двадцати пяти носителей фамилии, а демографический – лишь пяти. В отличие от “территориального” критерия, “демографический” не зависит от численности и случайных событий в локальных популяциях.

При применении данного критерия отбора ЧФ при анализе генофонда населения пяти регионов России (Восточный, Центральный, Западный, Северный, Южный) авторами показано, что при исключении огромного массива редких фамилий (от 71% до 80%, при среднем значении по пяти регионам 75%), не происходит значительного сокращения численности выборки – число анализируемых индивидуумов сокращается лишь на одну пятую [9]. Авторы, полагают, что дальнейшее использование “коренных” фамилий (то есть фамилий, уже “пустивших корни” в данной популяции, уже оставившие в ней генетический след) позволяет быть уверенным, что изучается история формирования народа, а не мимолетные потоки последних миграций.

Таким образом, проанализированные данные убедительно показывают, что исключение редких и использование частых фамилий позволяет проводить анализ популяционно-генетической структуры населения. Однако, следует отметить, что каждый авторский коллектив, используя определенный критерий отбора частых фамилий для исследования, во-первых, определяет возможности и ограничения только данного критерия для описания отдельных популяционно-генетических характеристик населения. Во-вторых, авторы, демонстрируя схожесть популяционных характеристик (их средние значения, изменчивость), полученных на основе всех фамилий и изучаемых ими частых фамилий не всегда рассматривают вопрос, насколько корректно частые фамилии будут описывать генетические соотношения между изучаемыми популяциями. В-третьих, в настоящее время остается неясным вопрос насколько результаты популяционно-генетического анализа (описание отдельных популяционно-генетических характеристик, оценка генетических соотношений между изучаемыми популяциями), полученные по фамилиям, отобраным на основе различных критериев, соотносятся между собой.

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы (государственный контракт № 14.740.11.0627 “Разработка новых подходов использования фамилий для популяционно-генетического анализа населения Центральной России”).

Литературы

1. Сорокина, И.Н. Фамилии как квазигенетические маркеры при популяционно-генетических исследованиях / И.Н. Сорокина, И.Н. Лепендина, Н.А. Рудых, А.В. Верзилина, М.И. Чурносков // Научные ведомости БелГУ. Серия: Медицина. Фармация. – 2010. – №22(93), выпуск 12. – С. 72-79.
2. Сорокина, И.Н. Генофонд населения Белгородской области. I. Дифференциация всех районных популяций по данным антропоники / И.Н. Сорокина, Е.В. Балановская, М.И. Чурносков // Генетика. – 2007. – Т.43, №6. – С. 841-849.
3. Сорокина, И.Н. Генофонд населения Белгородской области. II. «Фамильные портреты» в группах районов с разным уровнем подразделенности и роль миграций в их формировании / И.Н. Сорокина, М.И. Чурносков, Е.В. Балановская // Генетика. – 2007. – Т.43, №8. – С. 1120-1128.



4. Сорокина, И.Н. Генофонд населения Белгородской области. Описание “генетического” ландшафта 22 районных популяций / И.Н. Сорокина, М.И. Чурносос, Е.В. Балановская // Генетика. – 2009. – Т.45, №5. – С. 700-710.
5. Чурносос, М.И. Моделирование генетического ландшафта населения юга Центральной России / М.И. Чурносос, И.Н. Сорокина // Научные ведомости БелГУ. Серия: Медицина. Фармация. – 2010. – №10 (81), вып. 10. – С. 36-41.
6. Сорокина, И.Н. Использование фамилий для описания популяционно-генетического “ландшафта” населения / И.Н. Сорокина, И.С. Полякова, Н.А. Горяинова, И.В. Батлуцкая, В.И. Евдокимов, М.И. Чурносос // Научные ведомости БелГУ. Серия медицина. Фармация. – 2011. – №16(111), выпуск 15. – С. 218-222.
7. Ельчинова, Г.И. Опыт применения методов популяционно-генетического анализа при изучении популяций России с различной генетико-демографической структурой / Г.И. Ельчинова / Автореф. дисс....доктор. биол. наук. – М., МГНЦ РАМН, 2001. – 48 с.
8. Ельчинова, Г.И. О частотном критерии выбора фамилий для изучения генетической структуры популяций / Г.И. Ельчинова, М.Ю. Кадошников, Р.А. Мамедова и др. // Генетика. – 1991б. – Т.27, №2. – С. 358-360.
9. Балановская, Е.В. Русский генофонд на Русской равнине / Е.В. Балановская, О.П. Балановский – М.: Луч, 2007. – 416 с.
10. Балановская, Е.В. “Фамильные портреты” пяти русских регионов / Е.В. Балановская, Д.С. Соловьева, О.П. Балановский и др. // Мед. генетика. – 2005. – №1. – С.2-10
11. Иванов, В.П. Распределение фамилий в сельских районах Курской области / В.П. Иванов, М.И. Чурносос, А.И. Кириленко // Генетика. – 1998а. – Т.32, №8. – С. 1114-1119.
12. Бужилова, А.П. География русских фамилий / А.П. Бужилова // Восточные славяне. Антропология и этническая история. – М.: Научный мир, 1999. – С. 135-152.
13. Алтухов, Ю.П. Генетические процессы в популяциях / Ю.П. Алтухов. – М.: Наука, 2003. – 370 с.

USE OF FREQUENT SURNAMES IN POPULATION GENETIC RESEARCH

**I.N. SOROKINA, A.V. VERZILINA
A.V. ELYKOVA, N.A. DEMAKOVA
M.I. CHURNOSOV**

*Belgorod National
Research University*

e-mail: sorokinat@bsu.edu.ru

In article characteristics of three basic criteria ("frequency", "territorial", "demographic" criteria) selection of widespread surnames and possibility of their use are given at population genetic research.

Key words: Frequent surnames, criteria of selection of frequent surnames