

СРАВНИТЕЛЬНЫЙ АНАЛИЗ КРИТЕРИЕВ ОТБОРА ЧАСТЫХ ФАМИЛИЙ ПРИ ПРОВЕДЕНИИ ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЙ НАСЕЛЕНИЯ

И. Н. Сорокина, В. И. Евдокимов, М. И. Чурносов

ФГАОУ ВПО Белгородский государственный национальный исследовательский университет
Поступила в редакцию 10.04. 2012 г.

Аннотация. Проведен сравнительный анализ эффективности использования “демографического” и “частотного” критериев отбора частых фамилий для описания популяционно-генетической структуры населения Центральной России. Установлено, что частые фамилии, отобранные на основе применения как “демографического” критерия (включение в анализ фамилий, число носителей которых, достигших репродуктивного возраста, превышает в районе пять человек), так и “частотного” критерия 0,001 позволяют корректно оценивать уровень подразделенности (f_r) населения крупной областной популяции юга Центральной России.

Ключевые слова: частые фамилии, демографический критерий, частотный критерий, генофонд

Abstract. The summary: the comparative analysis of efficiency of use of “demographic” and “frequency” criteria of selection of frequent surnames for the description of populjatsionno-genetic structure of the population of the Central Russia is carried out. It is established that the frequent surnames which have been selected on the basis of application as “demographic” criterion (inclusion in the analysis of the surnames, which number of the carriers which have reached of reproductive age, exceeds in area four persons), and “frequency” criterion 0,001 allow to estimate correctly level podrazdelennost (f_r) the population of large regional population of the south of the Central Russia.

Keywords: frequent surnames, demographic criterion, frequency criterion, a genofund

ВВЕДЕНИЕ

При изучении генетических характеристик населения традиционно в качестве аналогов генетических маркеров используются фамилии (“квазигенетические” маркеры). Фамилии позволяют определять уровень инбридинга среди населения, оценивать генетические расстояния между популяциями, изучать генетические соотношения между различными группами народонаселения, выявлять действие таких эволюционных факторов, как дрейф генов и миграции, описывать “генетический ландшафт” населения на основе геногеографических технологий [1]. Однако одной из особенностей проведения популяционно-генетических исследований населения с использованием данных антропоники является большая численность фамилий, что затрудняет их обработку [1-3]. Кроме этого редкие фамилии исторически случайны и поэтому мало информативны при изучении структуры генофонда [1]. С учетом этого исследователи ограничиваются ана-

лизом какой-то части фамилий, определив при этом достаточный уровень их частоты с помощью определенного критерия. К настоящему времени в отечественной популяционной генетике применяются три критерия, разделяющие фамилии на редкие и распространенные: “демографический” критерий [1]; “частотный” критерий [3]; “территориальный” критерий [4].

В настоящем сообщении представлены результаты сравнительного анализа двух критериев отбора распространенных фамилий (“демографический”, “частотный” критерии) и возможности их использования для изучения отдельных популяционно-генетических характеристик населения.

МЕТОДИКА ЭКСПЕРИМЕНТА

Сравнительный анализ эффективности использования “демографического” и “частотного” критериев отбора частых фамилий при популяционно-генетическом исследовании населения проведен на модели Белгородской области. Вся база данных по фамилиям включает 48902 фамилии среди 822316 человек, проживающих в 20 райо-

нах Белгородской области. Анализ популяционно-генетических характеристик осуществлялся на уровне района (элементарная популяция [5]).

Проведен анализ распределения среди населения частых фамилий (ЧФ), отобранных в соответствии двумя критериями: “демографический” [1] и «частотный» ([2,3]) критерии.

В соответствии с “демографическим” критерием, в исследуемую выборку включались лишь те фамилии, которые встречались в рассматриваемой элементарной популяции (район) 5 и более раз. Распространенность частых фамилий (ЧФ) определялась как отношение числа носителей определенной частой фамилии в изучаемой популяции к общему числу жителей старше 18 лет в данной популяции [1]. “Частотный” критерий предполагает включение в анализ фамилии, которые встречались в рассматриваемой элементарной популяции (район) с частотой 0.1% [3].

Случайный инбридинг по частотам фамилий оценивался с помощью коэффициента изонимии f_r , предложенного Grow и Mange [6]. Значения f_r^* рассчитывались как $f_r^*_{(район-область)} = \bar{f}_r_{(район)} - f_r_{(область)}$, где $\bar{f}_r_{(район)}$ – это средний уровень инбридинга на уровне элементарной популяции (район), полученный в результате усреднения оценки f_r по районам, а $f_r_{(область)}$ – случайный инбридинг в области в целом [7].

Результаты, полученные на основе данных о частых фамилиях, отобранных по “демографическому” и “частотному” критериям, сравнивались как между собой, так и с аналогичными показателями, полученными нами ранее по всем фамилиям [8,9]. Эффективным считался тот критерий отбора распространенных фамилий, применяя который получали ЧФ 100% репрезентативно (то есть полученные на основе ЧФ популяционно-генетические данные полностью соответствовали результатам по всем фамилиям [8]) описывавшие популяционно-генетические характеристики (уровень подразделенности) населения.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

При использовании “демографического” критерия (включение в анализ фамилий, число носителей которых, достигших репродуктивного возраста, превышает в районе 5 человек) к базе всех фамилий Белгородской области был исключен достаточно большой массив редких фамилий. Из 48902 фамилий после использования “демографического” критерия для исследования осталось лишь 12499 фамилий, т.е. более 74% фамилий

(36403 фамилии) были исключены из рассмотрения (табл. 1). Другой критерий – “частотный”, оказался более жестким. Применение “частотного” критерия (включение в анализ фамилий, встречающихся с частотой 0.1%) ко всей базе фамилий привело к исключению еще большего количества редких фамилий по сравнению с “демографическим” критерием. Было исключено более 95% фамилий (46690 фамилии), т.е. в дальнейшем рассматривалось лишь 2212 распространенных фамилий из 48902 фамилий (табл.1).

Исследование распределения ЧФ, оставшихся после применения “демографического” критерия, выявило достаточно широкое колебание числа ЧФ по районам (16.3 раз), при среднем значении по области 1493 фамилии. Минимальное количество ЧФ наблюдалось в Красненском районе (358 ЧФ), а максимальное – в Старооскольском районе (5825 ЧФ) (табл. 1). Аналогичная вариабельность числа фамилий (в 17.5 раз) наблюдалась и при анализе всех фамилий: наименьшее число фамилий регистрируется в Красненском районе (1272 фамилии), а самое большое в Старооскольском районе (22247 фамилий). Удельный вес частых фамилий по районам области варьировал незначительно – от 24.1% в Прохоровском районе до 34.1% в Красногвардейском районе (1.6 раз). При использовании “демографического” критерия в анализ включается в среднем 27.7% фамилий. Несколько иная ситуация наблюдалась после применения “частотного” критерия 0.001. Так, вариабельность числа ЧФ по районам, оставшихся после применения “частотного” критерия 0.001, была незначительной (1.9 раз), что в 8.6 и 9 раз ниже вариабельности данного показателя по “демографическому” критерию и по всем фамилиям, соответственно. Минимальное количество ЧФ наблюдалось в Старооскольском районе (130 ЧФ), а максимальное – в Красногвардейском районе (243 ЧФ) при среднем значении по области 201 фамилия (табл. 1).

После использования “демографического” критерия не происходит существенных потерь в объеме изучаемого населения. Количество населения с ЧФ в целом по области составило более 673 тыс. человек при общей численности населения старше 18 лет 822316 человек. Удельный вес населения с ЧФ от общей базы фамилий (т.е. населения старше 18 лет) в среднем по области составил 81.8 % при минимальном значении данного показателя в Борисовском районе (74.7%) и максимальном в Красногвардейском районе

Таблица 1

Характеристика распределения фамилий и уровня подразделенности (f_r^*) в районных популяциях Белгородской области

популяции (районы)	В целом			По данным списков избирателей с использованием "демографического" критерия отбора ЧФ (5 и более фамилий в популяции)				По данным списков избирателей с использованием "частотного" критерия 0,001 и выше					f_r^*		
	Численность населения (тыс. человек)	Кол-во населения старше 18 лет	Кол-во фамилий	Кол-во ЧФ	% ЧФ	Кол-во населения с ЧФ	Удел. вес населения с ЧФ от населения старше 18 лет, %	Кол-во ЧФ	% ЧФ	Кол-во населения с ЧФ	Удел. вес населения с ЧФ от населения старше 18 лет, %	Мин кол-во населения, включаемого в анализ	По всем фамилиям	По ЧФ, отобранным по "демографическому" критерию "5 и более"	По ЧФ, отобранным по "частотному" критерию 0.001
Алексеевский	66.0	50858	5401	1775	32.9	43933	86.4	203	3.8	21506	42.3	51	0.00028	0.00029	0.00029
Борисовский	26.1	19366	3500	902	25.8	14471	74.7	213	6.1	8446	43.6	20	0.00022	0.00022	0.00021
Валуийский	74.0	56461	7628	2075	27.2	46268	81.9	172	2.3	18511	32.8	57	0.00015	0.00016	0.00014
Вейделевский	25.8	19324	2825	859	30.4	15566	80.6	217	7.7	9614	49.8	20	0.00033	0.00034	0.00033
Волоконовский	36.6	28230	3704	1206	32.6	23408	82.9	216	5.8	12486	44.2	29	0.00024	0.00024	0.00023
Грайворонский	27.7	20606	3388	913	26.9	15997	77.6	207	6.1	9659	46.9	21	0.00029	0.0003	0.00029
Губкинский	120.4	87926	9999	2865	28.7	73688	83.8	178	1.8	32032	36.4	88	0.00019	0.0002	0.00019
Ивнянский	24.0	18802	2749	709	25.8	14913	79.3	228	8.3	10637	56.6	19	0.00048	0.00049	0.00049
Корочанский	40.1	30125	4388	1150	26.2	24041	79.8	212	4.8	13369	44.4	31	0.00022	0.00023	0.00022
Красненский	15.5	12737	1272	358	28.1	11120	87.3	207	16.3	9999	78.5	13	0.00125	0.00126	0.00128
Красногвардейский	40.7	35027	3562	1215	34.1	30704	87.7	243	6.8	17540	50.1	36	0.00025	0.00026	0.00025
Краснояржский	15.2	11589	2064	569	27.6	8734	75.4	201	9.7	6112	52.7	12	0.00052	0.00052	0.00052
Новооскольский	47.3	38108	5507	1528	27.7	30808	80.8	180	3.3	14006	36.8	39	0.00019	0.00019	0.00018
Прохоровский	27.8	22878	3531	850	24.1	17973	78.6	215	6.1	11767	51.4	23	0.00039	0.00040	0.0004
Ракитянский	35.2	26744	3687	1004	27.2	21682	81.1	229	6.2	13601	50.9	27	0.00029	0.0003	0.0003
Ровенской	25.5	18507	2311	695	30.1	15401	83.2	229	9.9	11029	59.6	19	0.00062	0.00062	0.00063
Старооскольский	250.6	187229	22247	5825	26.2	154389	82.5	130	0.6	45843	24.5	188	0.00007	0.00008	0.00007
Чернянский	33.9	25871	3987	1126	28.2	20460	79.1	212	5.3	11283	43.6	26	0.00025	0.00025	0.00025
Шебекинский	93.2	73120	9424	2579	27.4	60186	82.3	162	1.7	23653	32.3	74	0.00016	0.00017	0.00017
Яковлевский	50.2	38862	6684	1664	24.8	29284	75.4	169	2.5	11988	30.8	39	0.00009	0.00010	0.00009
В среднем по области	53.8	41116	5393	1493	27.7	33651	81.0	201	3.7	15654	45.4	42	0.00032	0.00033	0.00033
В целом по области	1075.8	822316	48902	12499	25.6	673026	81.8	2212	4.5	313081	38.1	-	0.00013	0.00012	0.00008

(87.7%). Применение же “частотного” критерия, напротив, приводит к значительным потерям в объеме исследуемого населения. Количество населения с ЧФ в целом по области составило лишь 313081 человек. Доля населения, описываемого с помощью данного критерия в 2 раза меньше, чем по “демографическому” критерию, и в среднем по области составила лишь 38.1% от населения старше 18 лет. Минимальный удельный вес населения с ЧФ, полученными после применения “частотного” критерия, наблюдается в Старооскольском районе (24.5%), а максимальный в Красненском районе (78.5%). Значительно варьирует по районам области и минимальное количество населения, включаемого в анализ после применения критерия (14.5 раз). Так, в Краснояружском районе фамилия будет считаться частой, если отмечена у 12 человек, тогда как в Старооскольском районе она должна встречаться у 188 человек. В среднем данный показатель по области составил 42 человека (табл.1).

Далее на основе частот “коренных” фамилий, полученных после применения “демографического” и “частотного” критериев, нами был рассчитан уровень подразделенности f_r^* (табл. 1) в районных популяциях, который сравнивался с аналогичным показателем по всем фамилиям, полученным ранее [8]. Уровень f_r^* , рассчитанный по ЧФ, полученным по “демографическому” и “частотному” критериям, во всех рассмотренных районных популяциях и варибельность данного показателя по районам не отличались от аналогичных значений по всем фамилиям (табл. 1). Минимальный уровень подразделенности как по “демографическому”, так и по “частотному” критериям отмечен в Красненском районе (0.00126 и 0.00128, соответственно), а максимальный в Старооскольском районе (0.0008 и 0.00007, соответственно) при среднем значении по области 0.00033. Коэффициент корреляции Спирмена между f_r^* по всем фамилиям и f_r^* по ЧФ, отобранным по “демографическому” и “частотному” критериям, составил 0.99 ($p < 0.001$). При использовании “демографического” критерия в целом по области f_r составило 0.00012, что соответствует областному значению f_r по всем фамилиям (0.00013). При использовании “частотного” критерия 0.001 значение f_r по области было равно 0.00008.

ЗАКЛЮЧЕНИЕ

Итак, использование “демографического” критерия приводит к значительному отсеву ред-

ких фамилий – в анализ включается примерно одна четверть фамилий (в целом по области 25.6% ЧФ). При этом объем исследуемого населения в целом по области остается достаточно большим – 81.8% от населения старше 18 лет или 62.6 % от всего населения. Использование более “жесткого” “частотного” критерия 0.001 приводит к значительному отсеву, как редких фамилий, так и населения, включаемого в анализ. После применения критерия в целом по области остается лишь 4.5% ЧФ, которые встречаются у 38.1% населения старше 18 лет. Априорно количество фамилий, включаемых в анализ, в первую очередь зависит от общей численности жителей в изучаемом районе. В соответствии с этим минимальное число населения, включаемого в анализ, по районам существенно различается – более чем на порядок (от 12 до 188 человек). Однако, несмотря на различия в “жесткости” отбора данных критериев, абсолютные значения уровня подразделенности f_r^* , полученным после применения данных критериев совпадают и практически не отличаются от аналогичных показателей по всем фамилиям.

Таким образом, и более “мягкий” “демографический” (фамилия встречается в элементарной популяции у 5 и более человек) критерий отбора распространенных фамилий, и более “жесткий” “частотный” критерий 0.001, несмотря на различия в отборе фамилий и объеме исследуемого населения, могут быть эффективно использованы для адекватной оценки уровня подразделенности населения юга Центральной России.

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.

СПИСОК ЛИТЕРАТУРЫ

1. Балановская Е.В. Русский генофонд. Взгляд в прошлое / Е.В. Балановская, О.П. Балановский. — Луч, 2007. — 352 с.
2. Ельчинова, Г.И. Опыт применения методов популяционно-генетического анализа при изучении популяций России с различной генетико-демографической структурой: автореф. дисс. ... д. биол. наук (03.00.15) / Г.И. Ельчинова. — М.: МГНЦ РАМН, 2001. — 48 с.
3. Ельчинова, Г.И. О частотном критерии выбора фамилий для изучения генетической структуры популяций Г.И. Ельчинова [и др.] // Генетика. — 1991. — Т. 27, №2. — С. 358–360.

4. Бужилова А.П. География русских фамилий // Восточные славяне. Антропология и этническая история / А.П. Бужилова. — М.: Научный мир, 1999. — С. 135-152.

5. Чурносов, М.И. Генофонд населения Белгородской области. Динамика индекса эндогамии в районных популяциях / М.И. Чурносов, И.Н. Сорокина, Е.В. Балановская // Генетика. — 2008. — Т. 44, № 8. — С. 1117–1125.

6. Crow J.F. Measurement of inbreeding from the frequency of marriages between person of the same surname / J.F. Crow, A.P. Mange // Eugen. Quart. — 1965. — V.12. — P. 199-203.

7. Балановская Е.В. Геногеографический анализ подразделенной популяции П. География

случайного инбридинга (по частотам фамилий у адыгов) / Е.В. Балановская [и др.] // Генетика. — 2000. — Т.36, №8. — С. 1126-1139.

8. Сорокина, И.Н. Генофонд населения Белгородской области. I. Дифференциация всех районных популяций по данным антропоники / И.Н. Сорокина, Е.В. Балановская, М.И. Чурносов // Генетика. — 2007. — Т. 43, № 6. — С. 841-849.

9. Сорокина, И.Н. Генофонд населения Белгородской области. Описание “генетического” ландшафта 22 районных популяций / И.Н. Сорокина, М.И. Чурносов, Е.В. Балановская // Генетика. — 2009. — Т.45, №5. — С. 700-710.

Сорокина Инна Николаевна — д.б.н., профессор кафедры медико-биологических дисциплин факультета лечебного дела и педиатрии медицинского института ФГАОУ ВПО Белгородский государственный национальный исследовательский университет; e-mail: sorokina@bsu.edu.ru

Sorokina Inna N. — Ph.D., professor department of medico-biological studies of the faculty of medical business and pediatrics medical institute FGAOU VPO Belgorod State National Research University; e-mail: sorokina@bsu.edu.ru Department of Medico-Biological Studies; e-mail: sorokina@bsu.edu.ru

Евдокимов Валерий Иванович — д.м.н., профессор, заведующий кафедрой медико-профилактических дисциплин факультета лечебного дела и педиатрии медицинского института ФГАОУ ВПО Белгородский государственный национальный исследовательский университет; e-mail: evdokimov@bsu.edu.ru

Evdokimov Valery I. — M.D., professor, head of the department of medico- preventive disciplines of the faculty of medical business and pediatrics medical institute FGAOU VPO Belgorod State National Research University; e-mail: evdokimov@bsu.edu.ru

Чурносов Михаил Иванович — д.м.н., профессор, заведующий кафедрой медико-биологических дисциплин факультета лечебного дела и педиатрии медицинского института ФГАОУ ВПО Белгородский государственный национальный исследовательский университет; e-mail: churnosov@bsu.edu.ru

Churnosov Mikhail I. — M.D., professor, head of the department of medico-biological studies of the faculty of medical business and pediatrics medical institute FGAOU VPO Belgorod State National Research University; e-mail: churnosov@bsu.edu.ru