

ДОПУСТИМОСТЬ ИСПОЛЬЗОВАНИЯ ТАТАРСКИХ ФАМИЛИЙ В КАЧЕСТВЕ КВАЗИГЕНЕТИЧЕСКОГО МАРКЕРА В ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Г.И. Ельчинова, Р.А. Зинченко

Медико-генетический научный центр РАМН, Москва

В работе проведена проверка допустимости использования татарских фамилий в качестве квазигенетического маркера. В Арском районе Татарстана на основании 2333 фамилий и 3482 брачных записей за 1990–2000 гг. подсчитаны стандартные характеристики популяционно-генетической структуры: параметры Барраи (случайная изонимия, индекс миграций, показатель разнообразия фамилий, энтропия, избыточность распределения), параметры изоляции расстоянием Малеко (локальный инбридинг, степень изоляции расстоянием, среднеквадратичная миграция), матрицы фамильных, инбредных и миграционных расстояний. На основании сравнения полученных данных сделан вывод о возможности применения татарских фамилий в качестве биологического маркера при популяционно-генетических исследованиях.

Ключевые слова: татарская фамилия, корреляция, параметры Барраи, изоляция расстоянием Малеко, схема фамильного ландшафта

УДК 572

Фамилии тюркоязычных народов априорно считаются непригодными для их использования в качестве биологического маркера вследствие их, как правило, недавнего возникновения и, соответственно, недолгого существования. Приступая к комплексному медико-популяционно-генетическому изучению населения Республики Татарстан, проводимому в соответствии с разработанным нами протоколом [Зинченко и др., 2007], и изучая разнообразные литературные источники, мы выяснили, что фамилии татарских мурз, служивших татар и отдельных мишарских родов появились еще в XVI веке [<http://ru.wikipedia.org>]. Большинство татарских фамилий представляют собой производные от имен какого-либо предка мужского пола с добавлением русских окончаний (-ов, -ин и пр.). Окончание -ев свидетельствует о древности и уважительности рода. Кроме того, фамилии у татар могут быть образованы от профессий (Урманчиев, Аракчеев), и от русских корней (например, Клейменовы получили такую фамилию за участие в Пугачевском восстании) [<http://ru.wikipedia.org>]. Таким образом, вопрос о непродолжительности использования татарских фамилий оказался снят, а соблазн использования в работе такого удобного маркера, как фамилии, был велик. Для выяснения допустимости использования татарских

фамилий в популяционно-генетических исследованиях мы провели работу, аналогичную проделанной А.А. Ревазовым с соавторами [Ревазов и др., 1986] по проверке пригодности использования русских фамилий в качестве квазигенетического маркера в русских северных популяциях.

Материалы и методы

Для проверки нами был выбран Арский район, расположенный на севере Республики Татарстан, с численностью населения 51.6 тыс. чел, 92.7 % которых – татары, 6.3 % – русские [<http://ru.wikipedia.org>]. Все расчеты выполнены стандартными методами [Ельчинова, Кривенцова, 2009; Ельчинова 2004] для популяций ранга сельсовет в предположении, что татарская фамилия пригодна для использования в качестве биологического маркера при популяционно-генетических исследованиях. Если наше предположение ошибочно, то, следуя законом логики, рано или поздно мы должны прийти к противоречию, т.е. получить отличный от ожидаемого результат, объяснение которому мы найти не сможем. Фамилии выкопированы из списков избирателей тотально (2333 фами-

лии). Брачные записи проанализированы за 1990–2000 годы, из тотального списка удалены браки пострепродуктивного возраста (невеста старше 45 лет) и браки жителей других регионов, приезжающих играть свадьбу на историческую родину. Анализируемая выборка составила 3482 брака. В сравнение вошли параметры Барраи [Barrai et al., 1992], параметры изоляции расстоянием Малекко [Morton, 1977], матрицы фамильных и инбредных расстояний, матрица миграционных дистанций и матрица корней квадратных из географических расстояний.

Параметры Барраи: случайная изонимия рассчитана как $I_r = \sum q_i^2$, где q_i – частота i -ой фамилии в популяции, индекс миграций рассчитан как $v = (1 - I_r) / [I_r(N - 1)]$, где N – число индивидов в популяции, показатель разнообразия фамилий в популяции определяется из формулы $\alpha = Nv / (1 - v)$, энтропия (мера априорной неопределенности) в распределении фамилий рассчитана по формуле: $H = -\sum q_i \log_2 q_i$, избыточность распределения R рассчитывается по формуле: $R = 100(1 - H / H_0)$, $H_0 = \log_2 N$.

Параметры изоляции расстоянием Малекко вычисляются следующим образом. Зависимость коэффициента родства Φ от расстояния d задается формулой

$\Phi(d) = ae^{-bd}$, где $\Phi(d)$ – средний коэффициент родства для пары индивидов, родившихся на расстоянии d друг от друга, a – средний коэффициент родства в локальной популяции (локальный инбридинг), т.е. $\Phi(0) = a$, который в элементарной популяции зависит от ее эффективного размера и уровня иммиграции в нее и задается уравнением:

$$a = 1 / (4N_e m_e + 1),$$

где N_e – эффективный размер популяции (т.е. это та часть популяции, в которой заключаются браки и которая создает следующее поколение), обычно принимаемый за треть от ее цензового, т.е. общего, размера.

Эффективное давление миграций

$$m_e = \sqrt{m(m + 2k)},$$

где m – половина доли дальних миграций, т.е.

$$m = 1/2P(d > 4\sigma / \sqrt{2}),$$

а k – половина доли промежуточных миграций, т.е. $k = 1/2P(d > \sigma' / 10\sqrt{2})$. Степень изоляции

расстоянием b вычисляется как

$$b = \sqrt{2m_e} / \sigma'$$

В модели изоляции расстоянием Малекко принято делить миграции на длинные, короткие и промежуточные. За длинные принимаются такие, где расстояние между местами рождения супругов

$$d > 4\sigma / \sqrt{2}, \sigma = \sqrt{\sum \frac{d_{ii}^2}{n}},$$

за короткие – где $d < \sigma' / 10\sqrt{2}$,

$$\text{а } \sigma' \text{ это } \sqrt{\sum \frac{d_{ii}^2}{n}},$$

где d_i – множество миграций, не включающих дальние, т.е. любое d_{ii} таково, что $d_{ii} < 4\sigma / \sqrt{2}$; за промежуточные миграции принимаются такие, в которых расстояние между местами рождения супругов d удовлетворяет неравенству:

$$\sigma' / 10\sqrt{2} < d < 4\sigma / \sqrt{2}.$$

Матрица фамильных расстояний рассчитана на основании частот частых фамилий (частота превышает 0.1%) по формуле:

$$D_{12} = \frac{2}{\pi} \arccos\left(\sum_k \sqrt{p_{1k} p_{2k}}\right)$$

где p_{1k} и p_{2k} – частоты k -ой фамилии в изучаемых 1-ой и 2-ой популяциях [Cavalli-Sforza, Bodmer, 1971].

Матрица инбредных дистанций рассчитывалась по оригинальной метрике [Ельчинова, 2000]. За расстояние между двумя популяциями принимается сумма двух элементов, первым из которых является разность между локальным инбридингом a , подсчитанным в первой популяции, и инбридингом, который рассчитывается при помощи формулы Малекко $\Phi(d) = ae^{-bd}$ на расстоянии d до второй популяции, т.е. падение инбридинга. Вторым слагаемым в этой сумме будет аналогичная разность между локальным инбридингом, подсчитанным во второй популяции, и инбридингом, подсчитанным по формуле Малекко на, естественно, том же расстоянии d от второй до первой популяции, но уже с параметрами изоляции расстоянием второй популяции. Формальная формула выглядит следующим образом:

$$D_{12} = [a_1 - a_1 e^{-b_1 d}] + [a_2 - a_2 e^{-b_2 d}]$$

где a_i , b_i – параметры Малекко в i -ой популяции, $i = 1, 2$.

Элементами матрицы миграционных дистанций являются значения $1 - m_{ij}$, где m_{ij} – суммарная доля мигрантов из i -ого в j -ый и из j -ого в i -ый сельсовет.

Матрицы в работе не приводятся вследствие их громоздкости (30 × 30). Коэффициенты корреляции, подсчитанные и используемые нами в данном анализе подчеркнуты, в отличие от приводимых для сравнения.

Результаты и их обсуждение

Коэффициент линейной корреляции между случайной изонимией I_r (табл. 1), отличающейся от случайного инбридинга Райта лишь множителем, и локальным инбридингом a (табл. 2) составил 0.64 ± 0.15 , что по знаку соответствует ожидаемому. Для Афанасьевского района Кировской области аналогичный коэффициент корреляции составил 0.48 ± 0.22 [Ельчинова и др., 1992], в Тарасовском районе Ростовской области 0.55 ± 0.30 . Корреляция энтропии H и локального инбридинга ожидаемо отрицательна, $r = -0.67 \pm 0.14$. Корреляция степени изоляции расстоянием b , определяемой г.о. случайными факторами, и индекса миграций n ожидаемо не отличается от нуля, $r = 0.12 \pm 0.19$.

Коэффициент корреляции между матрицей фамильных дистанций и матрицей корней квадратных из географических расстояний $r = 0.37 \pm 0.04$. Корреляция невысока, однако, положительна, что и ожидалось, и значима. При анализе схемы фамильного ландшафта (рис. 1) прослеживается ориентация кластеров вдоль путей сообщения, что и объясняет столь невысокие значения r . Аналогичный коэффициент корреляции в Костромской области составил от 0.28 ± 0.10 (Галичский район) до 0.87 ± 0.14 (Пыщугский район) [Ельчинова, 1989], в Атинском районе Татарстана 0.58 ± 0.10 .

Коэффициент корреляции матриц инбредных и фамильных расстояний составил 0.66 ± 0.04 . В республике Марий Эл аналогичные коэффициенты корреляции составили от 0.34 ± 0.07 (Моркинский район) до 0.75 ± 0.07 (Медведевский район), а в Афанасьевском районе Кировской области 0.27 ± 0.07 [Ельчинова, 2000].

Коэффициент корреляции матриц миграционных и фамильных дистанций $r = 0.58 \pm 0.04$, хотя при изучении республики Марий Эл он не отличался от нуля (0.01 ± 0.50) [Ельчинова и др., 1996], а при изучении русских северных популяций был значимо отрицательным [Резавов и др., 1986]. Возможно, в нашем случае это связано с тем, что в Арском районе 30 сельсоветов, рассмотренное нами число браков велико, а из 435 значений матрицы миграционных дистанций 159 (более трети) оказались единицей, что свидетельствовало об отсутствии соответствующих миграций, минимальное из рассмотренных значений составило 0.972 при среднем 0.9977, кроме этого, различия в значениях наблюдались на 5–6 знаке мантисы, что было за пределами рассматриваемой нами значимости. Возможно, при столь низком уровне гаметного обмена между сельсоветами миграци-

онная матрица, составленная для популяций ранга сельсовет, непригодна для сравнительного анализа.

Таким образом, получив 6 раз положительный и 1 раз спорный ответ на вопрос о допустимости использования татарских фамилий в качестве квазигенетического маркера при популяционно-генетических исследованиях в соответствии с критерием знаков Диксона и Муда [Закс, 1976] мы можем принять нашу гипотезу и использовать татарские фамилии для анализа популяционно-генетической структуры населения Республики Татарстан. При этом данное утверждение относится лишь к татарским фамилиям, а не к фамилиям всех тюркоязычных народов.

Завершая работу, отметим, что наиболее распространенными фамилиями в районе являются Валиев (частота 1.01%), Гарипов (1.05%), Закиров (1.13%), Зарипов (1.05%), Сабиров (1.33%), Хайруллин (1.06%), Шакиров (1.57%), индекс эндогамии (без учета этнической принадлежности) 0.61, а параметры Барраи и Малек, приведенные в таблицах для популяций ранга сельсовет, для популяции ранга район составили, соответственно $I_r = 0.003536$, $v = 0.0074$, $\alpha = 283.9$, $H = 9.21$, $R = 39.46$; $\sigma = 59.13$, $a = 0.000124$, $b = 0.012965$.

Благодарности

Работа выполнена при частичном финансировании РФФИ (08-04-00534) и федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2012 годы», государственный контракт № 02.512.11.0001.

Библиография

- Ельчинова Г.И. Реконструкция популяционной структуры из негенетических параметров. Дисс. ... канд. биол. наук. М.: ИМГ АМН СССР, 1989.
- Ельчинова Г.И., Кадошников М.Ю., Мамедова Р.А., Брусинцева О.В. Подсчет инбридинга через повторяющиеся пары браков в популяциях русского Нечерноземья // Генетика. 1992. Т. 28. № 2. С. 157–159.
- Ельчинова Г.И. Метрика, построенная через параметры изоляции расстоянием Малек, как характеристика генетического сходства популяций // Генетика. 2000. Т. 36. № 6. С. 856–858.
- Ельчинова Г.И. Методы обработки популяционно-генетических данных: структура брачных миграций // Медицинская генетика. 2004. Т. 3. № 4. С. 185–192.

Таблица 1. Параметры Барраи для сельсоветов Арского района Татарстана

Поселение	N	I_r	v	α	H	R
1. Апазовское с.п.	1445	0.009995	0.069	106.42	7.23	31.12
2. Венетинское с.п.	185	0.035793	0.146	31.73	5.24	30.42
3. Казанбашское с.п.	904	0.009488	0.116	118.17	7.06	28.11
4. Качелинское с.п.	898	0.009521	0.116	117.81	7.20	26.61
5. Кошлаучское с.п.	179	0.031241	0.174	37.76	5.29	29.31
6. Купербашское с.п.	997	0.010623	0.094	102.85	7.02	29.53
7. Наласинское с.п.	667	0.012922	0.115	86.41	6.62	29.44
8. Новокинерское с.п.	1975	0.008728	0.058	120.57	7.47	31.77
9. Новокишитское с.п.	943	0.012441	0.084	86.78	6.81	31.08
10. Новокрылайское с.п.	683	0.013464	0.107	82.21	6.66	29.27
11. Нусинское с.п.	521	0.013929	0.136	82.11	6.56	27.31
12. Сизинское с.п.	684	0.007686	0.189	159.43	7.30	22.49
13. Сикертанское с.п.	444	0.011139	0.200	111.27	6.82	22.45
14. Смак-Корсинское с.п.	772	0.008668	0.148	134.46	7.29	24.00
15. Среднеатынское с.п.	830	0.011102	0.107	99.92	6.85	29.36
16. Среднекорсинское с.п.	759	0.010778	0.121	104.57	6.98	27.05
17. Среднешалымское с.п.	690	0.010452	0.137	109.92	6.98	25.98
18. Староашитское с.п.	630	0.015208	0.103	72.30	6.56	29.46
19. Старокырайское с.п.	588	0.013096	0.128	86.61	6.60	28.26
20. Старочурилинское с.п.	890	0.007247	0.154	162.12	7.48	23.66
21. Сюрдинское с.п.	727	0.012051	0.113	92.54	6.79	28.57
22. Ташкичинское с.п.	1135	0.008238	0.106	134.81	7.35	27.58
23. Тюбяк-Чекурчинское с.п.	497	0.007656	0.261	175.83	7.37	17.72
24. Урнякское с.п.	934	0.007048	0.151	166.13	7.57	23.28
25. Утар-Атынское с.п.	792	0.013519	0.092	80.49	6.59	31.56
26. Училиное с.п.	717	0.012113	0.114	92.17	6.82	28.10
27. Шурабашское с.п.	791	0.011982	0.104	92.19	7.08	26.46
28. Шушмабашское с.п.	1345	0.007855	0.094	139.51	7.44	28.42
29. Янга-Салское с.п.	343	0.016771	0.171	70.96	6.17	26.74
30. Арск, город	15027	0.003261	0.020	312.02	9.23	33.48

Таблица 2. Параметры изоляции расстоянием Малеко для сельсоветов Арского района Татарстана

Поселение	N_e	σ	a	b
1. Апазовское с.п.	688	48.5	0.00365	0.01563
2. Венетинское с.п.	88	28.4	0.01586	0.03428
3. Казанбашское с.п.	430	46.7	0.00354	0.02377
4. Качелинское с.п.	428	58.4	0.00613	0.01014
5. Кошлаучское с.п.	85	40.2	0.01501	0.02921
6. Купербашское с.п.	475	49.8	0.00444	0.01580
7. Наласинское с.п.	318	73.5	0.00734	0.01494
8. Новокинерское с.п.	940	38.9	0.00252	0.01650
9. Новокишитское с.п.	449	47.4	0.00424	0.01645
10. Новокырлайское с.п.	325	42.3	0.00577	0.01784
11. Нусинское с.п.	248	52.0	0.00702	0.02078
12. Сизинское с.п.	326	45.3	0.00549	0.01760
13. Сикертанское с.п.	211	32.9	0.01339	0.02043
14. Смак-Корсинское с.п.	368	29.3	0.00710	0.02247
15. Среднеатынское с.п.	410	82.0	0.00624	0.01173
16. Среднекорсинское с.п.	361	50.0	0.00655	0.02061
17. Среднепшалымское с.п.	329	49.5	0.00641	0.01452
18. Староашитское с.п.	300	41.6	0.00666	0.02635
19. Старокырлайское с.п.	280	51.7	0.00702	0.01471
20. Старочурилинское с.п.	424	53.2	0.00546	0.01188
21. Сюрдинское с.п.	346	43.4	0.00633	0.01728
22. Ташкичинское с.п.	540	58.6	0.00670	0.00817
23. Тюбяк-Чекурчинское с.п.	237	49.2	0.01483	0.00971
24. Урнякское с.п.	445	69.2	0.00719	0.00832
25. Утар-Атынское с.п.	377	57.5	0.00510	0.01879
26. Училинское с.п.	341	72.9	0.00656	0.01067
27. Шурабашское с.п.	377	48.6	0.00549	0.02065
28. Шушмабашское с.п.	640	55.6	0.00294	0.01708
29. Янга-Салское с.п.	163	47.6	0.01794	0.01605
30. Арск, город	7156	77,1	0,00034	0,00875

Примечание. Учтены миграции до 500 км.

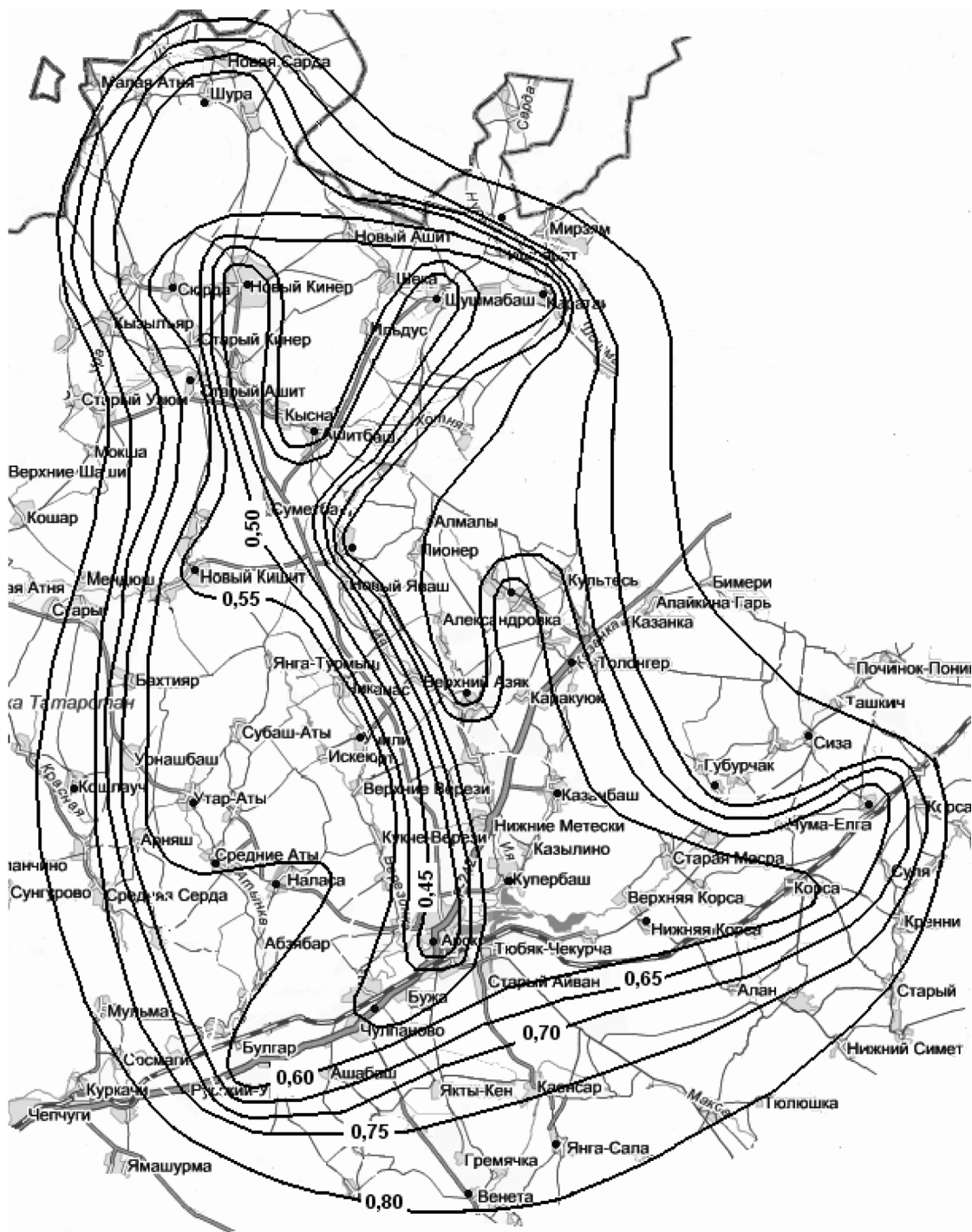


Рис. 1. Схема фамильного ландшафта Арского района

- Ельчинова Г.И., Кривенцова Н.В. Распределение фамилий в Ростовской области // Вестник Московского университета. Серия XXIII. Антропология. 2009. № 2 С. 51–64.
- Закс Л. Статистическое оценивание. М.: Статистика, 1976.
- Зинченко Р.А., Ельчинова Г.И., Петрова Н.В и др. Генетическая структура удмуртской популяции // Генетика. 2007. Т. 43. № 8. С. 1107–1119.
- Ревазов А.А., Парадеева Г.М., Ельчинова (Русакова) Г.И. Пригодность русских фамилий в качестве «квазигенетического» маркера // Генетика. 1986. Т. 22. № 4. С. 699–704.
- Barrai I., Formica G., Scapoli C., Beretta M., Mamolini E., Volinia S., Barale R., Ambrosino P., Fontana F. Microevolution in Ferrara: Isonymy 1890–1990 // Ann. Human Biol. 1992. Vol. 19. N 4. P. 371–385.
- Cavalli-Sforza L.L., Bodmer W.F. The Genetics of Human populations. San Francisco: Ed. W.H.Freeman and Company. 1971.
URL: <http://ru.wikipedia.org/wiki/...> (дата обращения 08.09.2009).
- Morton N.E. Isolation by distance in human populations // Ann. Hum.Genet. 1977. Vol. 40. P. 361–365.

Контактная информация:

Ельчинова Г.И. 115478, Москва, ул. Москворечье, д. 1, Медико-генетический научный центр РАМН. Раб. тел. (495) 324-12-24.
E-mail: elchinova@med-gen.ru;
Зинченко Р.А. 115478, Москва, ул. Москворечье, д. 1, Медико-генетический научный центр РАМН. Раб. тел. (495) 324-12-24.
E-mail: renazinchenko@mail.ru.

POSSIBILITY TO USE TATAR SURNAMES AS A QUASIGENETIC MARKER IN POPULATION- GENETIC RESEARCHES

G.I. El'chinova, R.A. Zinchenko

Research Center for Medical Genetics of Russian Academy of Medical Science, Moscow

The admissibility of use of Tatar surnames as a quasigenetic marker has been checked in this work. In the Arsky area of Tatarstan on the basis of 2,333 surnames and 3,482 marriage records for 1990-2000 the standard characteristics of population-genetic structure are estimated: parameters of Barraï (casual isonomy, an index of migrations, an indicator of a variety of surnames, entropy, redundancy of distribution), parameters of isolation in distance of Malekot (local inbreeding, degree of isolation in distance, root-mean-square migration), matrixes family, inbreed and migratory distances. The results show the possibility of application of Tatar surnames as a biological marker in population-genetic studies.

Key words: Tatar surname, correlation, parameters of Barraï, isolation in distance of Malekot, the scheme of surname landscape