



Реконструкция структуры генофонда казахов по данным об их родорасселении

М.К. Жабагин^{1, 2}✉, О.П. Балановский^{3, 4, 5}✉, Ж.М. Сабитов⁶, А.З. Темиргалиев, А.Т. Агджоян^{3, 4}, С.М. Кошель⁷,
Е.М. Раманкулов¹, Е.В. Балановская^{4, 5}

¹ Национальный центр биотехнологии, Астана, Казахстан

² National Laboratory Astana, Назарбаев Университет, Астана, Казахстан

³ Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия

⁴ Медико-генетический научный центр, Москва, Россия

⁵ Биобанк Северной Евразии, Москва, Россия

⁶ Евразийский Национальный университет им. Л.Н. Гумилева, Астана, Казахстан

⁷ Московский государственный университет им. М.В. Ломоносова, Москва, Россия

Использование для изучения генофонда квазигенетических маркеров – признаков небологической природы (фамилия, род), но четко наследующихся в поколениях, – одно из направлений в популяционной генетике человека. Если для популяций Западной Европы, европейской части России и Кавказа в роли квазигенетических маркеров выступают фамилии, то для популяций Центральной Азии эффективным квазигенетическим маркером является родовая принадлежность. В ряде исследований было показано, что для центральноазиатских популяций, в особенности казахов, во многих родах прослеживается происхождение большинства членов рода от общего биологического предка. В настоящей работе представлен анализ большого массива данных о 50 казахских родах у ~4.2 млн человек в начале XX века в сравнении с разнообразием Y-хромосомы в современных казахских популяциях. Анализ состоял из трех блоков: описание разнообразия квазигенетических маркеров стандартным методом изонимии; сравнение популяций по частотам квазигенетических маркеров; сравнение квазигенетических и генетических данных. Построено 50 карт частоты представленности каждого рода на территории Казахстана и ряда смежных территорий. Показано, что эти карты частот квазигенетических маркеров оказываются скоррелированы с картами генетических расстояний. О связи генетических и квазигенетических расстояний свидетельствует и тест Мантеля: обнаружена достоверная корреляция между матрицами географических и квазигенетических расстояний ($r = 0.60$; $p < 0.05$). Анализ межпопуляционной изменчивости выявляет наибольшее межгрупповое разнообразие между географическими территориями, соответствующими социально-территориальным объединениям казахского ханства – жузам, нежели другим историческим территориальным объединениям (феодальные государства, улусы, области), существовавшим на территории Казахстана в предшествующие и современную эпохи. С этими результатами согласуются графики главных компонент и многомерного шкалирования, на которых географические популяции объединяются в три кластера, соответствующие социально-территориальным объединениям – трем жузам. Это указывает на то, что окончательное структурирование казахского генофонда могло произойти в эпоху существования казахского ханства.

Ключевые слова: генофонд; геногеография; квазигенетические маркеры; казахи; родорасселение.

Reconstructing the genetic structure of the Kazakh from clan distribution data

M.K. Zhabagin^{1, 2}✉, O.P. Balanovsky^{3, 4, 5}✉,
Zh.M. Sabitov⁶, A.Z. Temirgaliyev, A.T. Agdzhoyan^{3, 4},
S.M. Koshel⁷, E.M. Ramankulov¹, E.V. Balanovska^{4, 5}

¹ National Center for Biotechnology, Astana, Kazakhstan

² National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan

³ Vavilov Institute of General Genetics, RAS, Moscow, Russia

⁴ Research Centre for Medical Genetics, Moscow, Russia

⁵ Biobank of North Eurasia, Moscow, Russia

⁶ L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

⁷ Lomonosov Moscow State University, Moscow, Russia

Applying quasigenetic markers – non-biological traits which are nevertheless inherited in generations – is one of the research fields within human population genetics. For the West European, East European, and Caucasus populations, surnames are typical quasigenetic markers. For Central Asian populations, particularly Kazakh, the clan affiliation serves as a good marker: a set of papers demonstrated that many clans include mainly persons which biologically descent from a recent common ancestor. In this study, we analyzed a large (~4.2 million persons) dataset on quasigenetic markers – the geographic distribution of 50 Kazakh clans at the beginning of the 20th century, and compared the dataset with the direct data of the Y-chromosomal diversity in modern Kazakh populations. The analysis included three steps: the isonymy method, which is standard for quasigenetic markers, comparing frequencies of quasigenetic markers, and comparing the quasigenetic and genetic datasets. We constructed 50 maps of frequency of the distribution of each clan and revealed that these maps correlate with the maps of genetic distances. The Mantel test also demonstrated a significant correlation between geographic and quasigenetic distances ($r = 0.60$; $p < 0.05$). The analysis of inter-population variability revealed the largest diversity between geographic territories corresponding to the social-territorial groups of the Kazakh Khanate (zhuzes) rather than to other historical groups that existed on the territory of Kazakhstan in preceding and modern epochs. The same is evidenced by the principal components and multidimensional scaling plots, which grouped geographic populations into three clusters corresponding to three zhuzes. This indicates that

the final structuring of the Kazakh gene pool might have occurred during the Kazakh Khanate period.

Key words: gene pool; gene geography; quasigenetic markers; Kazakh; clan.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Жабагин М.К., Балановский О.П., Сабитов Ж.М., Темиргалиев А.З., Агджоян А.Т., Кошель С.М., Раманкулов Е.М., Балановская Е.В. Реконструкция структуры генофонда казахов по данным об их родорасселении. Вавиловский журнал генетики и селекции. 2018;22(7):895-904. DOI 10.18699/VJ18.431

HOW TO CITE THIS ARTICLE:

Zhabagin M.K., Balanovsky O.P., Sabitov Zh.M., Temirgaliyev A.Z., Agdzhoyan A.T., Koshel S.M., Ramankulov E.M., Balanovska E.V. Reconstructing the genetic structure of the Kazakh from clan distribution data. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2018;22(7):895-904. DOI 10.18699/VJ18.431 (in Russian)

В популяционной генетике человека, в отличие от изучения генофондов других видов, широко используются не только данные по фенотипическому разнообразию и ДНК-полиморфизму, но и квазигенетические маркеры – признаки негенетической природы (Ревазов и др., 1986), которые, тем не менее, передаются непрерывно в поколениях благодаря культурным традициям. Наиболее эффективными среди квазигенетических маркеров являются фамилии, наследующиеся в основном по однородительскому типу – отцовской генеалогической линии. Изучение фамильного фонда проведено для популяций Западной Европы (King et al., 2006; McEvoy, Bradley, 2006; King, Jobling, 2009; Martínez-González et al., 2012; Solé-Morata et al., 2015; Martínez-Cadenas et al., 2016), европейской части России (Балановская и др., 2011) и Кавказа (Балановская и др., 2000). По сравнению с ДНК-маркерами квазигенетические маркеры имеют свои достоинства: они селективно нейтральны, и данные могут быть собраны не для небольших выборок, а для тотальной популяции (Балановская, Балановский, 2007). Установленное соответствие между квазигенетическими и генетическими маркерами позволяет использовать квазигенетические маркеры для оценки инбридинга (Cavalli-Sforza, Bodmer, 1971). Поэтому к настоящему моменту с помощью антропоники (данные о фамилиях, родах и т. п.) изучена генетическая структура многих популяций мира: Парагвая (Dipierri et al., 2011), Боливии (Rodriguez-Larralde et al., 2011), Чили (Barral et al., 2012), Албании (Mikerezi et al., 2013), Гондураса (Herrera et al., 2014), Аргентины (Dipierri et al., 2017), Восточной Европы (Scapoli et al., 2007), России (Балановская и др., 2005; Сорокина и др., 2007; Почешкова и др., 2008; Ельчинова и др., 2009; Tarskaia et al., 2009; Кучер и др., 2010; Лавряшина и др., 2011; Ульянова и др., 2014). Использование фамилий в сочетании с геногеографическим подходом позволяет наглядно увидеть изменчивость генофонда с помощью карт (Балановский и др., 2001).

Близким к фамилиям является понятие патрилинейного рода – группы людей, возводящих свое происхождение к общему предку. Однако это понятие прежде всего социальное: генеалогическая цепочка рода в одних случаях может соответствовать биологическому родству, а в других отличаться. В ситуациях, когда такое соответствие наблюдается, принадлежность к роду наследуется в поколениях так же, как и Y-хромосома – от отца к сыну. Последние исследования населения Центральной Азии по изучению связи Y-хромосомы и родовых групп преимущественно для казахов (Chaix et al., 2004; Bíró et al., 2009; Abilev et al., 2012; Жабагин и др., 2014; Balanovsky et al., 2015;

Balaresque et al., 2015; Жабагин и др., 2016; Zhabagin et al., 2017) выявили, что большинство родов действительно состоит в основном из людей, принадлежащих к общему биологическому предку по мужской линии. Поэтому род в популяциях Центральной Азии может рассматриваться как эффективный квазигенетический маркер – не менее, если не более, эффективный, чем фамилии в Европе.

Опираясь на эти твердо установленные факты тесной связи родовой принадлежности локальных казахских популяций и их генофондов по маркерам Y-хромосомы, мы предприняли следующий шаг и проанализировали обширные данные о расселении казахских родов и частоте их представленности в разных регионах Казахстана. Эти данные были, во-первых, проанализированы стандартными методами, используемыми для квазигенетических маркеров, и, во-вторых, сопоставлены с прямыми генетическими данными о распределении гаплогрупп Y-хромосомы в казахских популяциях. Кроме того, на основе исторической информации о принадлежности географических ареалов расселения казахских популяций к территориальным объединениям, существовавшим в историческом прошлом на территории Казахстана, была оценена межпопуляционная изменчивость для каждого из таких исторических объединений (феодалные государства, улусы, жузы, области) и определена их роль в формировании современной архитектуры генофонда казахов.

Материалы и методы

Материалом для исследования квазигенетических маркеров послужил свод статистических сведений по расселению казахских родов «Материалы по киргизскому (казахскому) землепользованию...» (~4.2 млн человек) (Тасилова, 2017), основанный на данных переписи населения Российской империи об особенностях хозяйства и землепользования в конце XIX и начала XX века, собранных под руководством Ф. Щербина, В. Кузнецова, П. Скрыплева, П. Хворостанского, В. Переплетчикова (Темиргалиев, 2010). Ни для более ранних (в связи с отсутствием систематических переписей), ни для более поздних периодов (в связи с отсутствием в последующих переписях вопроса о родовой принадлежности) не имеется столь же подробных и достоверных данных о численности представителей каждого рода в каждом из регионов. Настоящей работой мы вводим эти уникальные данные в более активный научный оборот, представив их в виде электронных таблиц и карт расселения родов (Приложения 1–3, размещены на сайте www.генофонд.рф). Всего изучено расселение 50 казахских родов. Проанализированы частоты родов в

33 популяциях (уездах), относящихся к восьми историческим областям Центральной Азии Российской империи а также частоты родов в шести губерниях, пограничных с центральноазиатскими областями Российской империи (рис. 1). Для сравнения использовалась база данных частот гаплогрупп Y-хромосомы в современных популяциях казахов (районы).

Анализ данных состоял из описания разнообразия квазигенетических маркеров стандартным методом изонимии; сравнения популяций по частотам квазигенетических маркеров; сравнения квазигенетических и генетических данных.

При описании разнообразия наименования всех 50 родов рассматривались в качестве аллелей нейтрального мультиаллельного локуса, наследующихся по одnorodительскому типу. На основе данных о частотах родов в популяциях с помощью программы Excel были рассчитаны такие параметры, как:

– случайная изонимия (I_r): $I_r = \sum q_i^2$, где q_i – частота i -го рода в популяции;

– случайный инбридинг (f_r): $f_r = 1/4 I_r$, соответствующий F_{ST} Райта (Crow, Mange, 1965), в том числе с вычетом вкладов более высоких уровней популяционной системы (Балановская и др., 2000) по формуле:

$$f_{r(\text{район-область})}^* = f_{r(\text{район})} - f_{r(\text{область})}$$

– показатель разнообразия (α): $\alpha = 1/I_r$ (Fisher, 1943; Barta et al., 1996);

– индекс миграции (v): $v = \alpha/(N + \alpha)$, где N – размер популяции (Karlin, McGregor, 1967; Zei et al., 1983; Piazza et al., 1985);

– энтропия распределения (H): $H = -\sum q_i \log_2 q_i$ (Barta et al., 1987, 1989) и показатель избыточности (R): $R = 100 \times (\log_2 q_i - H) \log_2 q_i$.

Расчет генетических расстояний между популяциями (по частотам квазигенетических маркеров) проводился согласно (Nei, 1987) в программе DJgenetic (Balanovsky et al., 2008). Генетические взаимоотношения между популяциями визуализировались в программе Statistica 8 (StatSoft, Tulsa, OK, США) методом многомерного шкалирования и в пространстве главных компонент. Корреляционный тест Мантеля рассчитан в программном пакете Arlequin 3.5 (Excoffier, Lischer, 2010), при этом статистическая значимость параметров была протестирована с помощью 1000 непараметрических перестановок. Расчет географических расстояний между уездами проводился по их географическим координатам с помощью программы DistGeo (Balanovsky et al., 2011). Геногеографические карты были созданы с помощью программы GeneGeo (Balanovsky et al., 2011; Koshel, 2012).

Результаты и обсуждение

Характеристика популяционного разнообразия по квазигенетическим маркерам

По обширному своду данных «Материалы по киргизскому (казахскому) землепользованию...» (Тасилова, 2017) были подсчитаны частоты родов (Приложение 1¹), рассматри-

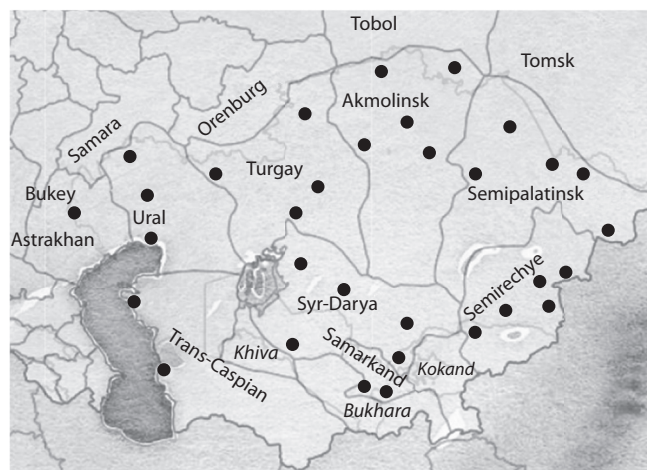


Fig. 1. Map of the Central Asian part of the Russian Empire at the end of the 19th – beginning of the 20th century (from Harvard World Map: <http://worldmap.harvard.edu/maps/russianempire> with modifications).

Oblasts and governorates are indicated by text blocks, and uyezd centers are shown as dots (Appendix 2).

ваемых далее в качестве квазигенетических маркеров, в популяциях казахов начала XX века (см. рис. 1).

Подробные характеристики разнообразия частот родов приведены в Приложении 2 (www.генофонд.рф). Здесь отметим только, что анализ инбридинга выполнен как для популяций (уездов, нижний административный уровень), так и для объединенных областных популяций (верхний административный уровень). При рассмотрении областных популяций самые высокие показатели изонимии и инбридинга (f_r^* до 0.156) отмечены на юго-западе (в Закаспийской области, которая также характеризуется наименьшим числом родов) и на северо-востоке (в Семипалатинской области). Напротив, самые низкие показатели инбридинга (f_r^* , от 0.003 до 0.018) отмечаются на северо-западе Казахстана (Уральская область) и на смежных территориях России, в Зауралье и нижнем Поволжье (регионы Оренбургская, Самарская, Астраханская, Букеевская Орда). Вероятно, эти наблюдения отображают следы исторического заселения казахами Заволжья в связи с образованием Букеевской Орды в XVIII веке и ростом роли Оренбурга как административного центра Центральной Азии.

Результаты расчета инбридинга (f_r) на нижнем административном уровне (уездные популяции) картографированы на рис. 2 в пределах современной территории Казахстана (исходные значения представлены в Приложении 2). Значения варьируют на целый порядок: от $f_r = 0.024$ для популяции KAZ_29 (Перовский уезд) до $f_r = 0.227$ для популяции KAZ_25 (Красноводский уезд). Паттерн случайного инбридинга как на уровне уездов, так и при анализе областей выявляет высокие значения на юго-западе, но также показывает, что вторая область максимумов на северо-востоке является географически более обширной и захватывает часть центральных районов Казахстана.

Для Центральной Азии, по данным других исследований, использовавших метод изонимии на основе ин-

¹ Приложения 1–3 см. по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2018-22/appx15.pdf>

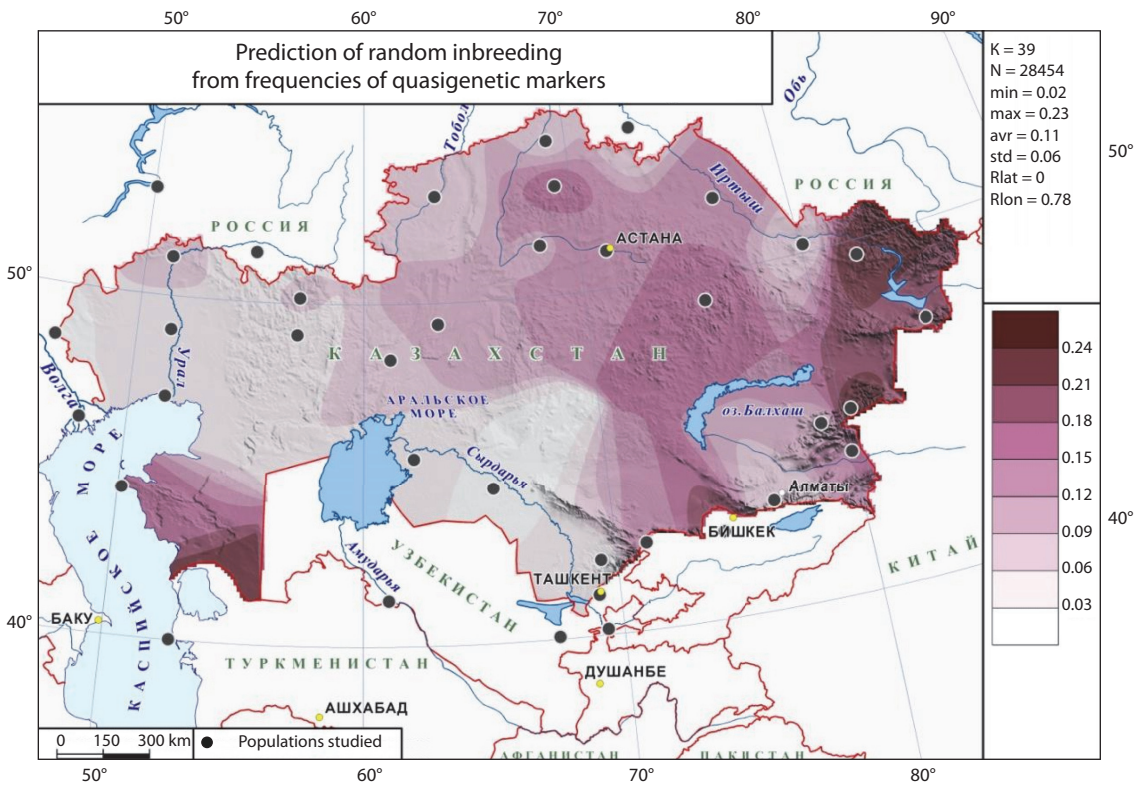


Fig. 2. Random inbreeding distribution predicted from data on quasigenetic markers.

The color scale illustrates the values of a character: purple for maximum values to white for minimum ones. The right upper corner of the legend shows statistical parameters of the numerical model of the character: K, the number of reference sites; N, the number of map grid nodes; min, minimum; max, maximum; avg, average value; std, standard deviation. Correlation coefficients: Rlat, with latitude; Rlon, with longitude.

формации о родовой принадлежности, в сельских популяциях установлен высокий уровень инбридинга и в Туркмении, и в Узбекистане (Бочков и др., 1984), причем в этих же популяциях отмечается и высокая отягощенность наследственной патологией. Поэтому обнаруженный нами максимальный инбридинг в данном исследовании в пределах юго-западных и северо-восточных казахских популяций позволяет прогнозировать высокий генетический груз в этих регионах Казахстана, и, соответственно, указывает на перспективность их прямого генетико-эпидемиологического обследования.

Сравнение карт распространения родов и генетических данных

По данным Приложения 1 была построена серия из 50 карт распространения отдельных родов (Приложение 3, www.генофонд.рф). На рис. 3 представлены примеры таких карт (для родов Кыпшак, Байулы, Найман, Дулат). Являясь лишь одним из результатов нашего исследования, представленные карты имеют первостепенную новизну для этнографов и историков, поскольку подобные карты родорасселения последний раз реконструировались в научных работах полувековой давности (Востров, Муканов, 1968), а новые карты, основанные на четких алгоритмах расчета и визуализации, созданы впервые.

Эти карты родорасселения крайне полезны в исследованиях структуры генофонда казахов. На рис. 4 показана тесная связь между распространением квазигенетиче-

ского маркера (расселением рода Аргын), частоты отдельного генетического маркера (гаплогруппы *G1-M285* Y-хромосомы) и характеристикой генофонда Каркаралинского района Казахстана (где доля представителей рода Аргын особенно высока) по совокупности генетических маркеров. Практически полное совпадение этих карт иллюстрирует и подтверждает то, что данные о расселении родов могут быть эффективны для прогноза генофонда в популяциях, где прямые генетические данные пока не получены. Такое совпадение карт позволяет сделать еще один важный вывод. Поскольку карты родорасселения получены на материалах начала XX века, а генетические данные представляют современное население, это служит косвенным аргументом, что население большинства регионов сохраняло генетическую преемственность с конца XIX до начала XXI века.

Роль исторических популяционных объединений в формировании структуры казахского генофонда

Межпопуляционная изменчивость по всей совокупности (39 популяций, включающих 33 уездные популяции и 6 сопредельных губерний) достигает высоких значений $F_{ST} = 0.35$. Для количественной оценки связи межпопуляционной изменчивости и географических расстояний проведен тест Мантеля. Между 33 уездами были определены географические расстояния. Между этими же уездами на основе частот родов были рассчитаны квазигенетические расстояния аналогично генетическим расстояниям Нея.

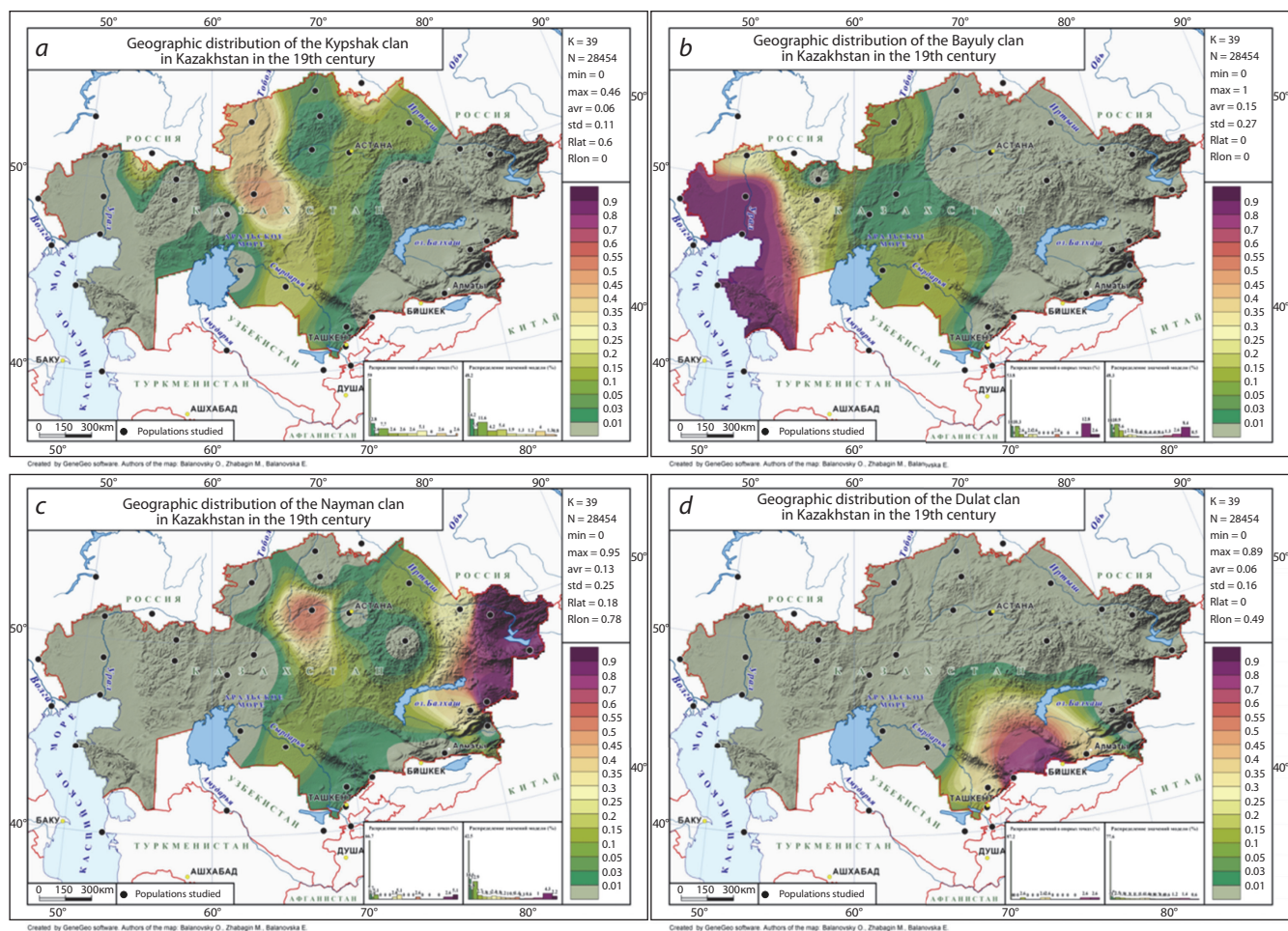


Fig. 3. Geographic distribution of some Kazakh clans: (a) Kypshak; (b) Bayuly; (c) Nayman; (d) Dulat.

Dots indicate Kazakh populations studied. The color scale illustrates the proportions of a clan in the population from purple (100%) to light-green (1%). Other panels of the legend show statistical parameters of mapping.

Сопоставление полученных двух матриц – географических и квазигенетических расстояний – демонстрирует достоверную высокую корреляцию между ними: $r = 0.60$ ($p < 0.05$).

Этот результат указывает на низкую миграционную активность казахского населения и позволяет перейти к анализу роли различных исторических социально-территориальных объединений. Для реализации этой цели проведен анализ AMOVA (таблица), популяции были сгруппированы согласно данным об исторических образованиях на территории Казахстана.

Первая группировка популяций («государственная») отражает три государственных образования в эпоху раннего феодализма (750–1050 гг.): Кимакское государство, Карлукское государство, Огузское государство. Предполагается, что они могли выступить основополагающим фактором формирования структуры генофонда современных казахов.

Вторая группировка популяций («улусы») – территориально-административное наследие империи Чингисхана – улусы Джучи, Угэдея и Чагатай, существовавшие на территории расселения исследуемых популяций с XIII по XV век. Длительный период существования та-

кого территориального деления Монгольской империи и смена сложившихся раннефеодальных взаимосвязей могли стереть сформировавшуюся изначальную структуру («государственную») и заложить другую основу («улусы») для структурирования генофонда.

Третья группировка популяций («жузы») – социально-территориальное подразделение казахского ханства (1465–1847) – Старший, Средний, Младший жузы, существующие поныне. Вероятно, именно социально-территориальные отношения, сложившиеся в казахском ханстве, выступили в качестве нового основного фактора в структурировании генофонда.

Четвертая группировка популяций («губернии») отражает территориально-административное деление Российской империи в Центральной Азии: Степное генерал-губернаторство, Туркестанское генерал-губернаторство и Зауральская киргизская степь (известна также как Оренбургское ведомство или Казахи Оренбургского ведомства). Несомненно, одной из целей колониального территориально-административного деления был перевод кочевников к оседлости. Это могло также выступить значимым фактором в дальнейшем переструктурировании генофонда.

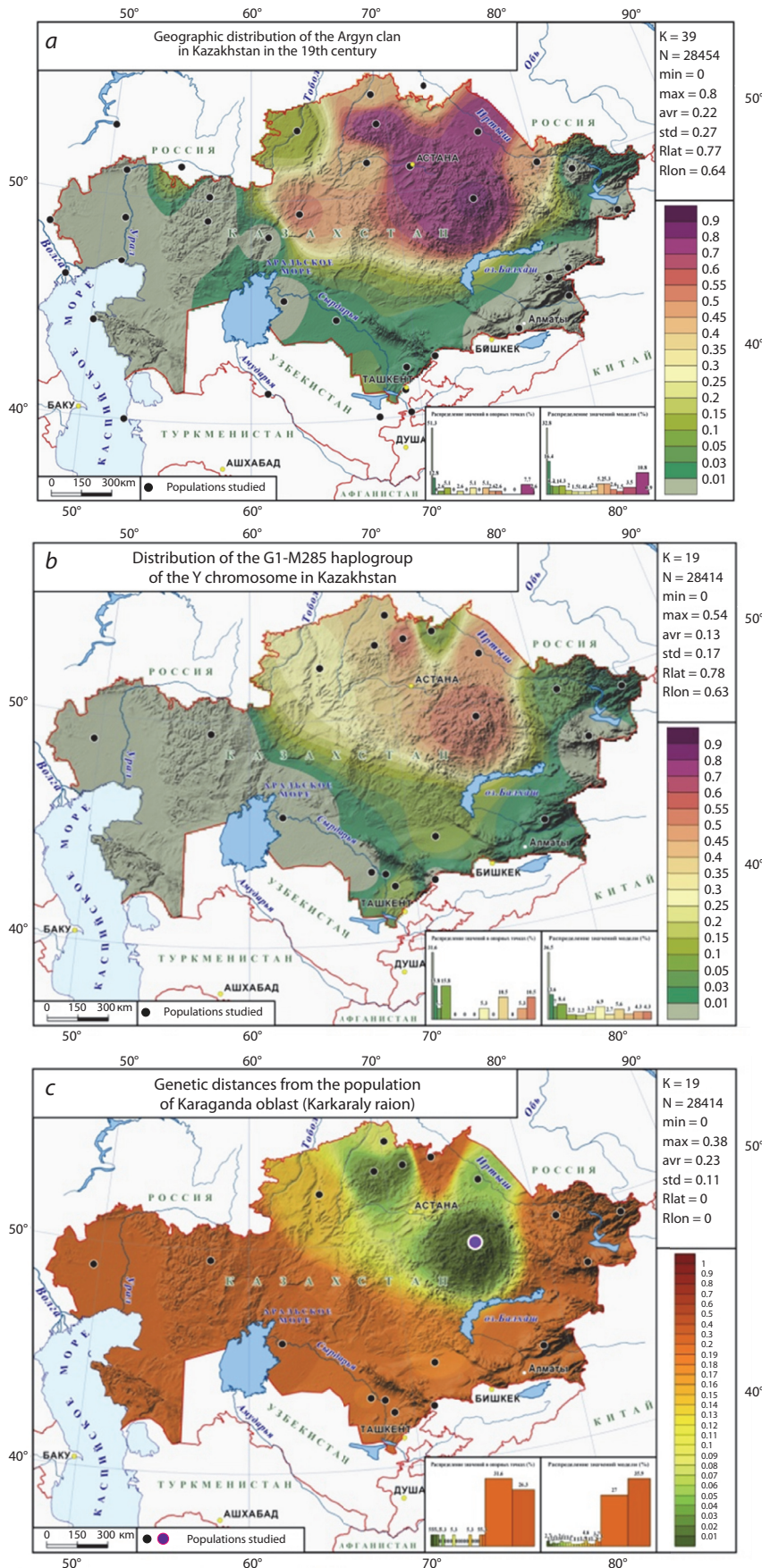


Fig. 4. Gene geography of the Argyn clan: (A) geographic distribution of the clan, the shares of Argyn people in the total populations of regions being mapped; (B) frequency distribution of haplogroup G1; (C) genetic distances from the Karkaraly raion of Kazakhstan.

Пятая группировка («современная») отражает современное распределение популяций между границами государств (Казахстана, России и стран Средней Азии).

Наименьшая межгрупповая изменчивость (3.6 %) была выявлена именно при пятой («современной») группировке (см. таблицу). Это указывает на то, что между современными казахскими популяциями различных государств нет существенных различий при допущении, что большая часть современных популяций является потомками населения тех же уездов.

Максимальное значение межгрупповой изменчивости (12.3 %) получено при объединении популяций, согласно социально-территориальному подразделению времен казахского ханства (жузы), существовавшему не менее 550 лет тому назад. Такой результат свидетельствует о том, что окончательное структурирование казахского генофонда произошло во время казахского ханства.

Результаты группировки популяций по более ранним географическим подразделениям указывают на значимое наследие Монгольской империи («улусы», 10.7 %) и позволяют предположить, что уже в эпоху Монгольской империи (Золотая Орда) начинает формироваться пространственная структура современного генофонда казахов.

Полученные результаты склоняют чашу весов в пользу гипотезы Т.И. Султанова, указывающей, что жузы возникли после распада Золотой Орды (Кляшторный, Султанов, 1992; Сабитов, 2014), нежели в сторону альтернативной гипотезы (Аманжолов, 1959), постулирующей, что казахи подразделились на жузы еще в X–XII веках, до объединения Чингисханом тюрков и монголов в одну империю.

Результаты анализа AMOVA подтверждаются и другими методами. Так, при анализе главных компонент на верхнем (областном) уровне административного деления географические популяции объединяются в три кластера, соответствующие социально-территориальным объединениям – трем жузам казахов: α -кластер отражает старший жуз; β -кластер – средний; γ -кластер – младший жуз (рис. 5).

Анализ на нижнем уровне административного деления (уезды) методом

The proportion of interpopulation variation (F_{ST}) among Kazakh populations grouped according to five different historical models of the territorial structure

Basis of division	Groups	Number of populations	Percentage of variation (%)	
			among populations within groups	among groups
Modern states	Kazakhstan	27	33.07	3.58
	Russia	5		
	Central Asian States	7		
Administrative subdivision of the Russian Empire	Steppe Governorate General	17	29.40	7.07
	Turkestan Governorate General	10		
	Transural Kirghiz steppe	12		
Socioterritorial division of the Kazakh Khanate	Senior zhuz	10	25.21	12.26
	Middle zhuz	16		
	Junior zhuz	13		
Administrative division of the Mongolian Empire heritage	Ulus of Jochi	25	27.16	10.72
	Ulus of Chagatai	7		
	Ulus of Ögedei	7		
Division in the epoch of early feudalism	Kimak Khanate	23	29.99	6.78
	Oguz Yabghu State	6		
	Karluk Yabghu	10		



Fig. 5. The principal component analysis plot of regional Kazakh populations from the dataset on quasigenetic markers.

многомерного шкалирования демонстрирует также отчетливые различия между взаимоотношениями уездов: определяются те же самые кластеры (рис. 6). В пределах β-кластера обнаруживается сближение юго-восточных популяций KAZ_8, KAZ6, KAZ_3, KAZ_4 (Семипалатинской и Семиреченской областей) и связь с KAZ_32

(Самаркандской области), вероятно, связанное с миграцией казахов в Самарканд (История народов Узбекистана, 1947). Примечательно, что наблюдается также четкое разделение между популяциями вдоль Сырдарьи – низовые (KAZ_28, KAZ_29, KAZ_31) и верховые (KAZ_26, KAZ_27, KAZ_30) уезды Сырдарьинской области.

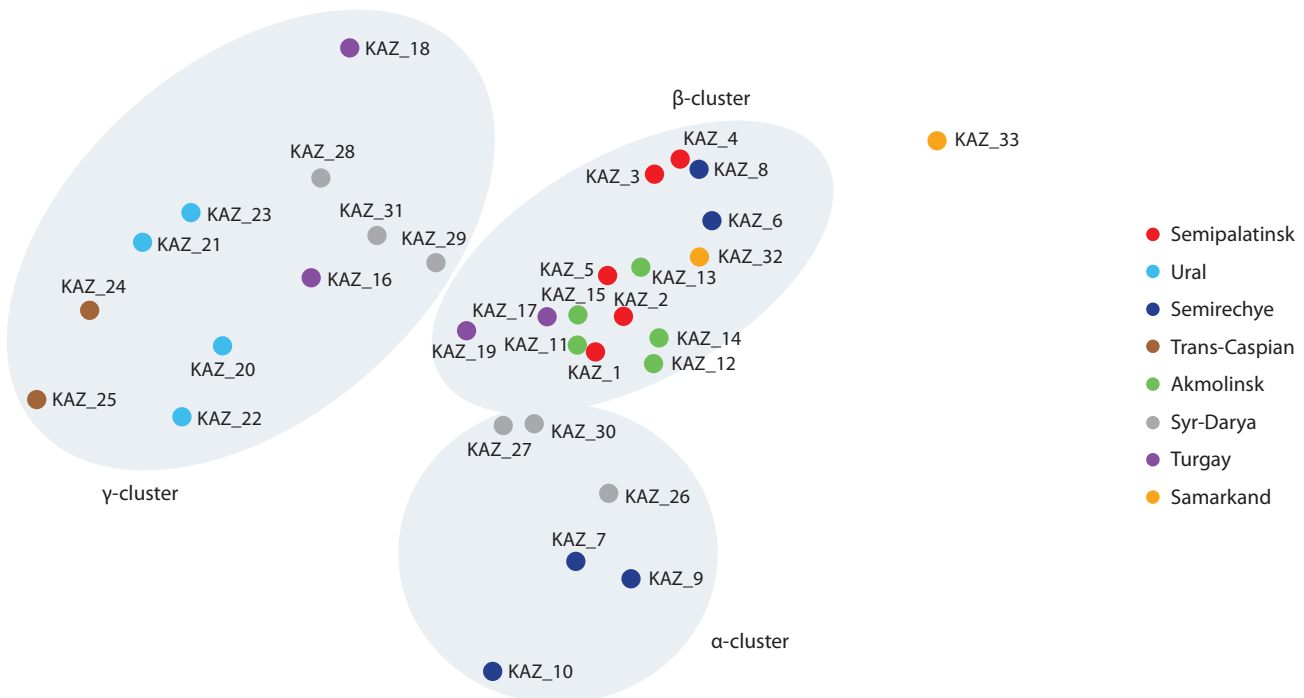


Fig. 6. Location of Kazakh subpopulations of lower administrative units (uyezds) in the multidimensional scaling plot according to quasigenetic markers (stress value 0.18, alienation coefficient 0.19).

Colored circles indicate to what oblasts uyezds belong. The designations of uyezd populations are shown in Appendix 1 (www.генофонд.рф).

Таким образом, особенность архитектуры казахской популяции – ее родоплеменная структура – позволила нам использовать рода в качестве квазигенетических маркеров, впервые ввести в научный оборот большой массив данных о родах у казахов и выявить сходство паттернов изменчивости квазигенетических маркеров и прямых генетических данных о полиморфизме Y-хромосомы. Изучение структуры генофонда казахов на основе геногеографического подхода и метода изонимии обнаружило более высокий уровень инбридинга в популяциях юго-запада и северо-востока. Высокие межпопуляционные различия проявляются больше всего между регионами, относившимися к разным социально-государственным образованиям в позднемонгольский период и период Казахского ханства, и поэтому могли сформироваться в эти эпохи.

Acknowledgement

Исследование выполнено при финансовой поддержке Министерства образования и науки Республики Казахстан (№ AP05134955), в рамках темы Государственного задания ФАНО России для Медико-генетического научного центра и Государственного задания ФАНО России для ИОГен РАН (тема АААА-А16-116111610171-1) и при финансовой поддержке РФФИ по научному проекту № 17-304-50005 «мол_нр».

Conflict of interest

The authors declare no conflict of interest.

Список литературы / References

Аманжолов С.А. Вопросы диалектологии и истории казахского языка. Алматы, 1959. [Amanzholov S.A. The Questions of Dia-

lectology and the History of the Kazakh Language. Almaty, 1959. (in Russian)]

Балановская Е.В., Балановский О.П. Русский генофонд на Русской равнине. М.: Луч, 2007. [Balanovska E.V., Balanovsky O.P. The Russian Gene Pool on the East European Plain. Moscow: Luch Publ., 2007. (in Russian)]

Балановская Е.В., Почешхова Э.А., Балановский О.П., Гинтер Е.К. Геногеографический анализ подразделенности популяции: II. География случайного инбридинга (по частотам фамилий у адыгов). Генетика. 2000;36(8):1126-1139. [Balanovskaya E.V., Balanovskii O.P., Ginter E.K., Pocheshkhova E.A. Gene-geographic analysis of a subdivided population. II. Geography of random inbreeding based on surname frequencies in Adygs. Rus. J. Genetics. 2000;36(8):936-948.]

Балановская Е.В., Романов А.Г., Балановский О.П. Однофамильцы или родственники? Подходы к изучению связи между гаплогруппами Y-хромосомы и фамилиями. Молекуляр. биология. 2011;45(3):473-485. [Balanovska E.V., Romanov A.G., Balanovsky O.P. Namesakes or relatives? Approaches to investigating the relationship between Y-chromosomal haplogroups and surnames. Molekulyarnaya Biologiya = Molecular Biology (Moscow). 2011;45(3):430-441. DOI 10.1134/S0026893311030022.]

Балановская Е.В., Соловьева Д.С., Балановский О.П., Чурнов М.И., Сорокина И.Н., Евсеева И.В., Аболмасов Н.Н., Почешхова Э.А., Серегин Ю.А., Пшеничников А.С. «Фамильные портреты» пяти русских регионов. Мед. генетика. 2005;1:2-10. [Balanovska E.V., Solovyeva D.S., Balanovsky O.P., Churnov M.I., Sorokina I.N., Evseeva I.V., Abolmasov N.N., Pocheshkova E.A., Sereyogin Y.A., Pshenichnov A.S. "Family portraits" of five regions of Russia. Meditsinskaya Genetika = Medical Genetics. 2005;1:2-10. (in Russian)]

Балановский О.П., Бужилова А.П., Балановская Е.В. Русский генофонд. Геногеография фамилий. Генетика. 2001;37(7):974-990. [Balanovsky O.P., Buzhilova A.P., Balanovskaya E.V. The Russian gene pool: gene geography of surnames. Rus. J. Genet. 2001;37(7):807-822. DOI 10.1023/A:1016755111586.]

- Бочков Н.П., Захаров А.Ф., Иванов А.И. Медицинская генетика. М.: Медицина, 1984. [Bochkov N.P., Zakharov A.F., Ivanov A.I. Medical Genetics. Moscow: Meditsina Publ., 1984. (in Russian)]
- Востров В.Б., Муканов М.С. Родоплеменной состав и расселение казахов (конец XIX – начало XX в.). Алма-Ата, 1968. [Vostrov V.B., Mukanov M.S. The Tribal Structure and Settlement of the Kazakhs (late XIX – early XX). Alma-Ata, 1968. (in Russian)]
- Ельчинова Г.И., Тереховская И.Г., Осетрова А.А., Порядина О.А., Зинченко Р.А. Распределение фамилий и случайный инбридинг в Кировской области. Генетика. 2009;45(10):1411-1419. [El'chinova G.I., Terekhovskaya I.G., Osetrova A.A., Poryadina O.A., Zinchenko R.A. Surname distribution and random inbreeding in Kirov oblast. Rus. J. Genet. 2009;45(10):1247-1255. DOI 10.1134/S1022795409100135.]
- Жабагин М.К., Дибирова Х.Д., Фролова С.А., Сабитов Ж.М., Юсупов Ю.М., Утевская О.М., Тарлыков П.В., Тажигулова И.М., Балаганская О.А., Нимадава П., Захаров И.А., Балановский О.П. Связь изменчивости Y-хромосомы и родовой структуры: генофонд степной аристократии и духовенства казахов. Вестн. Моск. ун-та. Сер. XXIII. Антропология. 2014;1:96-101. [Zhabagin M.K., Dibirova Kh.D., Frolova S.A., Sabitov Z.M., Yusupov Y.M., Utevskaia O.M., Tarlykov P.V., Tazhigulova I.M., Balaganskaya O.A., Nimadava P., Zakharov I.A., Balanovsky O.P. The relation between Y chromosome variation and the clan structure: the gene pool of the steppe aristocracy and the steppe clergy of the Kazakhs. Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Vestnik of Moscow University. Ser. XXIII. Anthropology. 2014;1:96-101. (in Russian)]
- Жабагин М.К., Сабитов Ж.М., Агджоян А.А., Юсупов Ю.М., Богуну Ю.В., Лаврышина М.Б., Схаляхо Р.А., Балаганская О.А., Тажигулова И.М., Акильжанова А.Р., Жумадилов Ж.Ш., Балановский О.П., Балановская Е.В. Генезис крупнейшей родоплеменной группы казахов – аргынов – в контексте популяционной генетики. Вестн. Моск. ун-та. Сер. XXIII. Антропология. 2016; 4:59-68. [Zhabagin M.K., Sabitov Zh.M., Agdzhoyan A.A., Yusupov Y.M., Bogunov Y.V., Lavryashina M.B., Tazhigulova I.M., Akil'zhanova A.R., Zhumadilov Zh.Sh., Balanovsky O.P., Balanovska E.V. Genesis of Argyns, the largest tribal-clan group of Kazakhs, in the context of population genetics. Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Bulletin of Moscow University. Ser. XXIII. Anthropology. 2016;4:59-68. (in Russian)]
- История народов Узбекистана. Ташкент, 1947. [History of the Peoples of Uzbekistan. Tashkent, 1947. (in Russian)]
- Кляшторный С.Г., Султанов Т.И. Казахстан: летопись трех тысячелетий. Алматы, 1992. [Klyashstornyy S.G., Sultanov T.I. Kazakhstan: Record of Three Millennia. Almaty, 1992. (in Russian)]
- Кучер А.Н., Данилова А.Л., Конева Л.А., Ноговицина А.Н. Структура браков в якутских популяциях: миграционные процессы. Генетика. 2010;46(5):692-699. [Kucher A.N., Danilova A.L., Koneva L.A., Nogovitsina A.N. Marriage structure of Yakut populations: ethnic composition and isonymy inbreeding. Russ. J. Genet. 2010;46(3):408-416. DOI 10.1134/S1022795410030142.]
- Лаврышина М.Б., Ульянова М.В., Толочко Т.А., Балаганская О.А., Романов А.Г., Балановская Е.В. Шорцы: сходство и различие территориальных групп по данным фонда фамилий и аутомных ДНК маркеров. Вестн. Моск. ун-та. Сер. XXIII. Антропология. 2011;2:66-77. [Lavryashina M.B., Ul'yanova M.V., Tolochko T.A., Balaganskaya O.A., Romanov A.G., Balanovska E.V. The Shors: similarities and differences among territorial groups according to the surname range and autosomal DNA markers. Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Bulletin of Moscow University. Ser. XXIII. Anthropology. 2011;2:66-77. (in Russian)]
- Почешхова Э.А., Балановская Е.В., Серегин Ю.А., Голубцов В.И., Балановский О.П. Динамика генофонда во времени по данным о фамилиях и родословных. Мед. генетика. 2008;7(8):25-29. [Pocheshkhova E.A., Balanovska E.V., Seregin Yu.A., Golub-tsov V.I., Balanovsky O.P. Temporal dynamics of gene pool reconstructed from genealogical and surname data. Meditsinskaya Genetika = Medical Genetics. 2008;7(8):25-29. (in Russian)]
- Резавов А.А., Парадеева Г.М., Русакова Г.И. Пригодность русских фамилий в качестве «квазигенетического» маркера. Генетика. 1986;22(4):699-703. [Revazov A.A., Paradeeva G.M., Rusakova G.I. Possibility of using Russian family names as a quasi-genetic marker. Genetika. 1986;22(4):699-703. (in Russian)]
- Сабитов Ж.М. Казахские жузы и клановая система Золотой Орды. Вестн. Евраз. нац. ун-та им. Л.Н. Гумилева. 2014;3(100):201-207. [Sabitov Zh.M. Kazakh zhuzes and the Golden Horde clan system. Vestnik Evraziyskogo Natsionalnogo Universiteta im. L.N. Gumileva = Bulletin of L.N. Gumilyov Eurasian National University. 2014;3(100):201-207. (in Russian)]
- Сорокина И.Н., Балановская Е.В., Чурносоев М.И. Генофонд населения Белгородской области. I. Дифференциация всех районных популяций по данным антропонимии. Генетика. 2007;43(6):841-849. [Sorokina I.N., Balanovska E.V., Churnosov M.I. The gene pool of the Belgorod oblast population. I. Differentiation of all district populations based on anthroponymic data. Russ. J. Genet. 2007; 43(6):697-704. DOI 10.1134/S1022795407060142.]
- Тасилова Н. «Материалы по киргизскому (казахскому) землепользованию...» – как источник по истории Казахстана (конец XIX в. – начало XX в.). Алматы, 2017. [Tasilova N. "Materials on Kyrgyz (Kazakh) Land Use ..." – as a Source on the History of Kazakhstan (late XIX century – early XX century). Almaty, 2017. (in Russian)]
- Темиргалиев А. Волости, уезды... Казахи: Со схематической картой низовых административно-территориальных делений проживания казахов в 1897–1915 гг. Алматы, 2010. [Temirgaliev A. Volosts, uyezds... Kazakhs: with a Schematic Map of Lower Administrative and Territorial Divisions of the Kazakhs' Residence in 1897–1915. Almaty, 2010. (in Russian)]
- Ульянова М.В., Лаврышина М.Б., Николаев В.В., Октябрьская И.В., Дружинин В.Г. Коренное население северных районов Алтая: отражение демографических процессов XIX – начала XXI века в динамике фамильного состава. Археология, этнография и антропология Евразии. 2014;3(59):128-140. [Ul'yanova M.V., Lavryashina M.B., Nikolaev V.V., Oktyabr'skaya I.V., Druzhinin V.G. Native populations of the northern Altai: demographic processes of the late 19th – early 21st century as reflected in surname dynamics. Arkheologiya, etnografiya i antropologiya Evrazii = Archaeology, Ethnology, and Anthropology of Eurasia. 2014;42(3): 128-140. DOI 10.1016/j.aae.2015.04.015. (in Russian)]
- Abilev S., Malyarchuk B., Derenko M., Wozniak M., Grzybowski T., Zakharov I. The Y-chromosome C3* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan. Hum. Biol. 2012;84(1):79-89. DOI 10.3378/027.084.0106.
- Balanovsky O., Dibirova K., Dybo A., Mudrak O., Frolova S., Pocheshkhova E., Haber M., Platt D., Schurr T., Haak W., Kuznetsova M., Radzhabov M., Balaganskaya O., Romanov A., Zakharova T., Soria Hernanz D.F., Zalloua P., Koshel S., Ruhlen M., Renfrew C., Wells R.S., Tyler-Smith C., Balanovska E., Genographic C. Parallel evolution of genes and languages in the Caucasus region. Mol. Biol. Evol. 2011;28(10):2905-2920. DOI 10.1093/molbev/msr126.
- Balanovsky O., Rootsi S., Pshenichnov A., Kivisild T., Churnosov M., Evseeva I., Pocheshkhova E., Boldyreva M., Yankovsky N., Balanovska E., Villems R. Two sources of the Russian patrilineal heritage in their Eurasian context. Am. J. Hum. Genet. 2008;82(1):236-250. DOI 10.1016/j.ajhg.2007.09.019.
- Balanovsky O., Zhabagin M., Agdzhoyan A., Chukhryaeva M., Zaporozhchenko V., Utevska O., Highnam G., Sabitov Z., Greenspan E., Dibirova K., Skhalyakho R., Kuznetsova M., Koshel S., Yusupov Y., Nymadawa P., Zhumadilov Z., Pocheshkhova E., Haber M., Zalloua P.A., Yepiskoposyan L., Dybo A., Tyler-Smith C., Balanovska E. Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y-chromosome

- and reveals migrations of Iranic speakers. *PLoS One*. 2015;10(4): e0122968. DOI 10.1371/journal.pone.0122968.
- Balaresque P., Poulet N., Cussat-Blanc S., Gerard P., Quintana-Murci L., Heyer E., Jobling M.A. Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations. *Eur. J. Hum. Genet.* 2015;23(10):1413-1422. DOI 10.1038/ejhg.2014.285.
- Barrai I., Barbujani G., Beretta M., Maestri I., Russo A., Formica G., Pinto-Cisternas J. Surnames in Ferrara: distribution, isonymy and levels of inbreeding. *Ann. Hum. Biol.* 1987;14(5):415-423.
- Barrai I., Formica G., Barale R., Beretta M. Isonymy and migration distance. *Ann. Hum. Genet.* 1989;53(3):249-262.
- Barrai I., Rodriguez-Larralde A., Dipierri J., Alfaro E., Acevedo N., Mamolini E., Sandri M., Carrieri A., Scapoli C. Surnames in Chile: a study of the population of Chile through isonymy. *Am. J. Phys. Anthropol.* 2012;147(3):380-388. DOI 10.1002/ajpa.22000.
- Barrai I., Scapoli C., Beretta M., Nesti C., Mamolini E., Rodriguez-Larralde A. Isonymy and the genetic structure of Switzerland. I. The distributions of surnames. *Ann. Hum. Biol.* 1996;23(6):431-455.
- Bíró A.Z., Zalán A., Völgyi A., Pamjav H. A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *Am. J. Phys. Anthropol.* 2009;139(3):305-310. DOI 10.1002/ajpa.20984.
- Cavalli-Sforza L.L., Bodmer W.F. *The genetics of human populations*. San Francisco: W.H. Freeman and Co; 1971. XVI; 965.
- Chaix R., Austerlitz F., Khégay T., Jacquesson S., Hammer M.F., Heyer E., Quintana-Murci L. The genetic or mythical ancestry of descent groups: Lessons from the Y chromosome. *Am. J. Hum. Genet.* 2004;75(6):1113-1116. DOI 10.1086/425938.
- Crow J.F., Mange A.P. Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen. Quart.* 1965;12(4):199-203. DOI 10.1080/19485565.1965.9987630.
- Dipierri J., Rodríguez-Larralde A., Alfaro E., Scapoli C., Mamolini E., Salvatorelli G., Caramori G., De Lorenzi S., Sandri M., Carrieri A., Barrai I. A study of the population of Paraguay through isonymy. *Ann. Hum. Genet.* 2011;75(6):678-687. DOI 10.1111/j.1469-1809.2011.00676.x.
- Dipierri J.E., Rodríguez-Larralde A., Barrai I., Redomero E.G., Alonso-Rodríguez C., Alfaro E.L. Consanguinity by random isonymy and socioeconomic development in Argentina: a population study. *J. Biosoc. Sci.* 2017;49(3):322-333. DOI 10.1017/S0021932016000444.
- Excoffier L., Lischer H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Res.* 2010;10(3):564-567. DOI 10.1111/j.1755-0998.2010.02847.x. PubMed PMID: 21565059.
- Fisher R.A. The relation between the number of species and the number of individuals in a random sample of animal population. *J. Anim. Ecol.* 1943;12(1):42-58. DOI 10.2307/1411.
- Herrera Paz E.F., Scapoli C., Mamolini E., Sandri M., Carrieri A., Rodriguez-Larralde A., Barrai I. Surnames in Honduras: A study of the population of Honduras through isonymy. *Ann. Hum. Genet.* 2014;78(3):165-177. DOI 10.1111/ahg.12057.
- Karlin S., McGregor J. The number of mutant forms maintained in a population. *Proc. 5th Berkeley Symp. Math., Stat. Prob.* 1967;4: 415-438.
- King T.E., Ballereau S.J., Schürer K.E., Jobling M.A. Genetic signatures of coancestry within surnames. *Curr. Biol.* 2006;16(4):384-388. DOI 10.1016/j.cub.2005.12.048.
- King T.E., Jobling M.A. Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. *Mol. Biol. Evol.* 2009;26(5):1093-1102. Epub 2009/02/09. DOI 10.1093/molbev/msp022.
- Koshel S.M. *Geoinformation technologies in genogeography*. Eds. I.K. Lure, V.I. Kravtsova. Modern Geographic Cartography. Moscow, 2012;158-166.
- Martínez-Cadenas C., Blanco-Verea A., Hernando B., Busby G.B., Brion M., Carracedo A., Salas A., Capelli C. The relationship between surname frequency and Y chromosome variation in Spain. *Eur. J. Hum. Genet.* 2016;24(1):120-128. Epub 2015/04/22. DOI 10.1038/ejhg.2015.75.
- Martínez-González L.J., Martínez-Espín E., Álvarez J.C., Albar-daner F., Rickards O., Martínez-Labarga C., Calafell F., Lorente J.A. Surname and Y chromosome in Southern Europe: a case study with Colom/Colombo. *Eur. J. Hum. Genet.* 2012;20(2):211-216. DOI 10.1038/ejhg.2011.162.
- McEvoy B., Bradley D.G. Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum. Genet.* 2006;119(1-2):212-219. DOI 10.1007/s00439-005-0131-8.
- Mikerezi I., Xhina E., Scapoli C., Barbujani G., Mamolini E., Sandri M., Carrieri A., Rodriguez-Larralde A., Barrai I. Surnames in Albania: a study of the population of Albania through isonymy. *Ann. Hum. Genet.* 2013;77(3):232-243. DOI 10.1111/ahg.12015.
- Nei M. *Molecular Evolutionary Genetics*. New York: Columbia University Press, 1987.
- Piazza A., Mayr W.R., Contu L., Amoroso A., Borelli I., Curtioni E.S., Marcello C., Moroni A., Olivetti E., Richiardi P. Genetic and population structure of four Sardinian villages. *Ann. Hum. Genet.* 1985; 49(1):47-63.
- Rodriguez-Larralde A., Dipierri J., Gomez E.A., Scapoli C., Mamolini E., Salvatorelli G., De Lorenzi S., Carrieri A., Barrai I. Surnames in Bolivia: a study of the population of Bolivia through isonymy. *Am. J. Phys. Anthropol.* 2011;144(2):177-184. Epub 2010/08/25. DOI 10.1002/ajpa.21379. PubMed PMID: 20740661.
- Scapoli C., Mamolini E., Carrieri A., Rodriguez-Larralde A., Barrai I. Surnames in Western Europe: a comparison of the subcontinental populations through isonymy. *Theor. Popul. Biol.* 2007;71(1):37-48. DOI 10.1016/j.tpb.2006.06.010.
- Solé-Morata N., Bertranpetit J., Comas D., Calafell F. Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. *Eur. J. Hum. Genet.* 2015;23(11):1549-1557. Epub 2015/02/18. DOI 10.1038/ejhg.2015.14.
- Tarskaia L., El'chinova G.I., Scapoli C., Mamolini E., Carrieri A., Rodriguez-Larralde A., Barrai I. Surnames in Siberia: a study of the population of Yakutia through isonymy. *Am. J. Phys. Anthropol.* 2009;138(2):190-198. DOI 10.1002/ajpa.20918.
- Zei G., Guglielmino C.R., Siri E., Moroni A., Cavalli-Sforza L.L. Surnames as neutral alleles: observations in Sardinia. *Hum. Biol.* 1983; 55(2):357-365.
- Zhabagin M., Balanovska E., Sabitov Z., Kuznetsova M., Agdzhoyan A., Balaganskaya O., Chukhryaeva M., Markina N., Romanov A., Skhalyakho R., Zaporozhchenko V., Saroyants L., Dalimova D., Davletchurin D., Turdikulova S., Yusupov Y., Tachigulova I., Akil-zhanova A., Tyler-Smith C., Balanovsky O. The connection of the genetic, cultural and geographic landscapes of Transoxiana. *Sci. Rep.* 2017;7. DOI 10.1038/s41598-017-03176-z.

ORCID ID

M.K. Zhabagin orcid.org/0000-0003-3414-0610
O.P. Balanovsky orcid.org/0000-0003-4218-6889
Zh.M. Sabitov orcid.org/0000-0001-7186-156X
A.Z. Temirgaliyev orcid.org/0000-0002-6082-9668
A.T. Agdzhoyan orcid.org/0000-0002-8776-2934
S.M. Koshel orcid.org/0000-0002-4540-2922
E.M. Ramankulov orcid.org/0000-0001-6791-4497
E.V. Balanovska orcid.org/0000-0002-3882-8300