

На правах рукописи

Жабагин Максат Кизатович

**АНАЛИЗ СВЯЗИ ПОЛИМОРФИЗМА Y-ХРОМОСОМЫ
И РОДОПЛЕМЕННОЙ СТРУКТУРЫ
В КАЗАХСКОЙ ПОПУЛЯЦИИ**

03.02.07 - генетика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2017

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Медико-генетический научный центр» (Москва, Россия) и Частном учреждении «National Laboratory Astana» автономной организации образования «Назарбаев Университет» (Астана, Казахстан).

Научный руководитель:

доктор биологических наук,
профессор РАН **Балановский Олег Павлович**

Официальные оппоненты:

Харьков Владимир Николаевич, доктор биологических наук
НИИ медицинской генетики, ФГБНУ «Томский НИМЦ РАН»,
ведущий научный сотрудник лаборатории эволюционной генетики

Дубова Надежда Анатольевна, доктор исторических наук, кандидат
биологических наук
ФГБНУ «ИЭА РАН», заведующая сектором этнической экологии

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение
высшего образования «Московский государственный университет имени
М.В.Ломоносова», биологический факультет, кафедра генетики

Защита состоится «__» _____ 2017 году в _____ часов на
заседании Диссертационного совета Д 002.214.01 при Федеральном
государственном бюджетном учреждении науки «Институт общей
генетики им. Н.И. Вавилова» Российской академии наук, 119991, ГСП-1,
Москва, ул. Губкина, д. 3, тел. (499) 135-62-13, факс (499) 132-89-62,
www.vigg.ru, e-mail: aspirantura@vigg.ru

С диссертацией можно ознакомиться в библиотеке Федерального
государственного бюджетного учреждения науки «Институт общей
генетики им. Н.И. Вавилова» Российской академии наук и на сайте
www.vigg.ru.

Автореферат разослан «__» _____ 2017 г.

Ученый секретарь диссертационного совета Д 002.214.01
по защите докторских и кандидатских диссертаций

кандидат биологических наук **Синельщикова Татьяна Аркадьевна**

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы.

Анализ Y-хромосомы стал одним из магистральных направлений в популяционной генетике человека, в последние годы усиленным возможностью полного секвенирования этого крупнейшего нерекombинирующего блока в геноме человека. Межпопуляционное разнообразие Y-хромосомы много выше, чем других генетических систем. Этот феномен объясняется не столько «однородительским» наследованием (присущим также и митохондриальной ДНК, разнообразие которой, однако, не столь велико), сколько свойственной большинству популяций человека патриликальностью. Поэтому многие тюркоязычные популяции Евразийской степи (включая казахов), подразделенные на патриликальные группы и традиционно отслеживающие генеалогию на семь мужских поколений, являются прекрасным модельным объектом для изучения связи биологического явления – межпопуляционных различий по Y-хромосоме – с его вероятными демографическими, социальными и историческими причинами.

Казахи представляют собой одну из крупнейших родоплеменных систем евразийской степи, записанную в традиционной казахской генеалогии «шежире». Ее структура и расселение является ценным историческим источником (Левшин, 1832; Вельямин-Зернов, 1864; Аристов, 1894, 1896; Валиханов, 1904; Тынышпаев, 1925; Востров, Муканов, 1968). Родоплеменная структура включала в себя различные по генезису племена, подразделявшиеся на кланы, в свою очередь состоящие из родов. Общими для всех элементов родоплеменной структуры являются передаваемые из поколения в поколение системы генеалогических сведений, возводящих всех потомков к единому родоначальнику – реальному или мифическому. Поэтому «род» – понятие социальное, но генеалогическая цепочка рода может иметь (или не иметь) связь с генетикой. Важно, что традиция передачи «рода» наследуется как Y-хромосома – от отца к сыну.

Данные о полиморфизме Y-хромосомы в казахской популяции в той или иной мере публиковались во многих работах: в исследованиях, проведенных в *общем мировом масштабе* (Underhill et al., 1997, 2000; Hammer et al., 2001; Wells et al., 2001); в *региональном масштабе* – Центральной Азии (Perez-Lezaun et al., 1999; Zerjal et al., 2002; Chaix et al., 2007; Segurel et al., 2008; Heyer

et al., 2009; Biro et al., 2015), Сибири (Karafet et al., 2002; Харьков, 2005, 2012; Балаганская и др., 2011a,b; Балановская и др., 2014), Волго-Урала (Roewer et al., 2007), Кавказа (Karafet et al., 2015); в *масштабе локальных территорий* расселения казахов – Казахстана (Tarlykov et al., 2013; Балмуханов и др., 2013), Алтая (Dulik et al., 2011), Китая (Shou et al., 2010; Zhong et al., 2010, 2011; Shan et al., 2014; Mei et al., 2016; Nothnagel et al., 2017); в работах посвященных *генетико-демографическим событиям* – экспансии «гена Чингисхана» (Zerjal et al., 2003, Деренко и др., 2007, Захаров-Гезехус 2013), заселению Америки (Karafet et al., 1999; Seielstad et al., 2003; Dulik et al., 2012), сокращению численности населения в неолите (Karmin et al., 2015), экспансии кочевых скотоводов и земледельцев Азии (Balaesque et al., 2015); в *детальных исследованиях филогеографии* отдельных гаплогрупп Y-хромосомы – гаплогруппы **R1a** (Underhill et al., 2010; 2015), гаплогруппы **N** (Rootsi et al., 2007; Ilumäe et al., 2016), гаплогруппы **C2** (Balaesque et al., 2009; Malyarchuk et al., 2010, 2012).

Но только в четырех работах (Chaix et al., 2004; Biro et al., 2009; Abilev et al., 2012; Аширбеков и др., 2014) изучалась *родоплеменная структура*, и показано, что «род» и «клан» могут быть информативными квазигенетическими маркерами. К такому же мнению приходят авторы работ по *генетической генеалогии* (Turuspekov et al., 2011; Сабитов, 2015). Однако этот феномен еще не подтвержден для большинства родоплеменных групп казахов.

Суммируя данные о генофонде казахов из всех вышеперечисленных работ, мы видим недостаточный географический охват обширной территории Казахстана, недостаточно сопоставимый набор маркеров и, как правило, небольшие выборки. В плане анализа родовой структуры подавляющее большинство казахских родов остаются не изученными, и роль родоплеменной структуры в структурировании генофонда остается не измеренной. Работы по полному секвенированию Y-хромосомы затронули лишь редкие для казахов линии **R1a** и **N**, а мажорные гаплогруппы **C2** и **G1** остаются слабо изученными. Наконец, отсутствует обобщающая характеристика генофонда казахов по данным изменчивости Y-хромосомы. Заполнению этих лакун и целостному изучению генофонда казахов по маркерам Y-хромосомы посвящена диссертационная работа.

Цель исследования. Изучить изменчивость Y-хромосомы в популяциях казахов и связь структуры генофонда с родоплеменной структурой населения.

Задачи исследования.

1. Создать «генетический портрет» популяций казахов на основе большого массива данных (~**2000 образцов**), единой обширной панели маркеров Y-хромосомы (**44 SNP и 17 STR**) и на трех уровнях иерархической популяционной системы: а) казахи в целом; б) социально-территориальная структура казахов (**3 жуза**); с) родоплеменная структура (**14 родоплеменных групп**).

2. Определить роль фактора родоплеменной структуры в дифференциации генофонда казахов как в абсолютном выражении, так и относительно фактора географической дифференциации.

3. Выявить положение генофонда казахов в генетическом контексте окружающих популяций Трансоксианы (исторический регион Центральной Азии) и определить закономерности структурирования генофонда населения этого региона.

4. Выявить пути вероятного происхождения генофондов ряда родоплеменных групп и сравнить эти генетические реконструкции с историческими гипотезами происхождения этих групп.

5. На основе секвенирования ~10 млн. п.н. Y-хромосомы провести детальный филогеографический и филогенетический анализ гаплогруппы G1, одной из наиболее частых у казахов, и применить эти данные для определения скорости мутирования Y-хромосомы.

Научная новизна.

Впервые Y-хромосомный генофонд казахов охарактеризован по большой выборке (~ **2000 образцов**), по широкой палитре маркеров (44 SNP и 17 STR), и с охватом большинства родоплеменных групп (14 родов).

Впервые показано, что в формировании межпопуляционной генетической изменчивости у кочевых обществ роль родоплеменной структуры может превышать роль географических расстояний.

Впервые проведен детальный филогеографический анализ гаплогруппы **G1-M285** на основе полногеномного анализа Y-хромосомы и обнаружена связь популяций Иранского нагорья и генофонда Центральной Азии, которая получила подтверждение в последующих публикациях по древней ДНК (Lazaridis et al., 2016).

Скорость мутирования SNP в Y-хромосоме до сих пор оценивалась четырьмя методами - прямым подсчетом в родословных (Xue et al., 2009; Mendez et al., 2013; Helgason et al., 2015), по сравнению с шимпанзе (Thomson et al., 2000; Kuroki et al., 2006), по археологическим датировкам (Francalacci et al., 2013, Poznik et al., 2013) и по древним образцам (Fu et al., 2014; Trombetta et al., 2015; Karmin et al., 2015; Адамов и др., 2015; Piumae et al., 2016). В данном исследовании разработан пятый метод – «клановый», основанный на использовании исторической даты жизни общего предка клана.

Научно-практическая значимость.

Полученные результаты значительно увеличивают объем данных о генофонде народонаселения Центральной Азии.

Для судебно-медицинской экспертизы обширный массив данных по изменчивости Y-хромосомы казахов обеспечивает создание надежной референсной базы для ДНК-идентификации. Ее важная особенность – определение не только географической, но и вероятной *родоплеменной принадлежности* неизвестного лица по образцу его ДНК, что значительно расширяет возможности экспертизы.

Для медико-генетических исследований сформированные коллекции образцов послужат для формирования строгих контрольных выборок. Реализованы принципы и практика организации Биобанка народонаселения.

Полученные результаты имеют междисциплинарный характер и представляют интерес для специалистов смежных наук (антропологов, археологов, этнографов, лингвистов, демографов, историков).

Вывод о ключевой роли родовой структуры в формировании генофонда важен для исследователей не только Центральной Азии, но и других регионов (Урала, Сибири, Дальнего Востока), где также сохраняется родоплеменная структура.

Результаты работы находят широкое применение в учебно-педагогическом процессе при подготовке курсов лекций и семинаров для студентов биологических, медицинских, исторических специальностей, и в популяризации науки.

Основные положения, выносимые на защиту.

1. Родоплеменная структура является ключевым фактором в формировании структуры генофонда казахов. Впервые выявлено: различия между родовыми популяциями в 1.5 раза больше, чем между районными популяциями (AMOVA); наличие достоверной корреляции между генетическими и квазигенетическими (частота родов в популяциях) расстояниями и ее отсутствие между генетическими и географическими расстояниями.

2. Географический ландшафт Трансоксианы, несмотря на его контрастность (пустыни и плодородные бассейны рек, предгорья и низменности) не оказывает прямого влияния на генетический ландшафт (AMOVA: $F_{ST}=0.01$ и $F_{ST}=0.00$). Основную роль в структурировании генофонда казахских и соседних популяций бассейнов Амударьи и Сырдарьи играет хозяйственно-культурный тип: земледелие и кочевое скотоводство (AMOVA: $F_{ST}=0.03$).

3. Верификация генеалогических легенд родоплеменных групп (на примере аргын, торе, кожа-сунак) с помощью полиморфизма Y-хромосомы демонстрирует широкие возможности использования генетических данных для других популяций, где сохраняется родоплеменная структура.

4. Полученное соответствие филогенетического древа гаплогруппы G1 (на основе анализа протяженных участков MSY региона ~10 млн. п.н.) и традиционной генеалогии (на основе шежира аргын), восходящей к историческому лицу с известным временем жизни, дает независимую оценку скорости SNP мутаций Y-хромосомы: $0.78 \cdot 10^{-9}$ на нуклеотид в год.

Апробация работы. Основные результаты исследования были представлены в форме **устных докладов** на Международной конференции «Проблемы генетики населения и этнической антропологии» памяти Ю.Г. Рычкова (Москва, 2013), на VI Съезде Вавиловского общества генетиков и селекционеров и на ассоциированных генетических симпозиумах (Ростов-на-Дону, 2014), Республиканской научно-практической конференции «Модернизация отечественной исторической науки в контексте национальной идеи «Мәңгілік ел»» (Астана, 2014), Международной конференции UNESCO «Great Migrations in Asia Minor: Circulation, exchange and social transformation» (Paris, 2016) и в форме **постерных докладов** на X Конгрессе этнографов и антропологов России (Москва, 2013), American Society of Human Genetics 63rd

Annual Meeting (Boston, 2013), 5-ой Международной конференции "Алексеевские чтения" памяти академиков Т.И. Алексеевой и В.П. Алексеева (Москва, 2013), Nazarbayev University Research Week (Astana, 2013), DNA in Forensics – 9th International Y-chromosome workshop & 6th International EMPOP meeting (Brussels, 2014), The 19th Congress of the European Anthropological Association "Anthropology: Unity in Diversity" (Moscow, 2014), Human Genome Meeting (Kuala Lumpur, 2015), VII Съезде Российского общества медицинских генетиков (Санкт-Петербург, 2015), Second International Scientific Conference Personalized medicine & Global health (Astana, 2015), The 13th International Congress of Human Genetics (Kyoto, 2016), Human Evolution: Fossils, Ancient and Modern Genomes (Wellcome Genome Campus, Hinxton, Cambridge, 2017).

Личный вклад автора. Автор принимал непосредственное участие во всех экспериментальных этапах исследования: экспедиционное обследование ряда популяций Казахстана; выделение ДНК, определение ее концентрации и формирование ДНК-коллекций; генотипирование SNP маркеров, пробоподготовка для фрагментного анализа STR маркеров. Автор самостоятельно провел анализ данных: формирование базы данных по родоплеменной структуре; статистический анализ (расчет генетического разнообразия, генетических расстояний, анализ главных компонент и многомерного шкалирования, AMOVA, корреляционный тест Мантеля); филогенетический анализ по STR-маркерам (построение филогенетических сетей, их датировка методами rho и ASD); обработка филогенетических деревьев, полученных по данным полного секвенирования Y-хромосомы, а также подготовка данных для картографического анализа.

Публикации. По теме диссертации опубликовано 12 публикаций, в том числе 6 статей в рецензируемых научных изданиях, рекомендованных ВАК при Минобрнауки России для защиты диссертаций, и 6 тезисов докладов, представленных на международных конференциях.

Структура и объем работы. Диссертационная работа изложена на 148 страницах и состоит из трех глав, включая введение, материалы и методы, результаты, заключение, выводы и список литературы. Данные проиллюстрированы в 26 таблицах и на 39 рисунках. Список литературы содержит 324 источника, из которых 215 зарубежные.

МАТЕРИАЛЫ И МЕТОДЫ

Материалы. Образцы венозной крови мужчин были собраны в ходе экспедиционных работ **2009-2012** гг. под рук. проф. Е.В. Балановской; в **2011-2012** гг. под рук. к.б.н. Е.В. Жолдыбаевой; в **2012-2016** гг. под рук. автора. Всего обследовано 19 географических популяций казахов (N=1982), три географических популяции узбеков (N=217), по одной популяции каракалпаков (N=100), туркмен рода йомуд (N=83), дунган (N=31) и монгольского рода борджигин (N=88). Суммарная выборка N=2501.

В обследование включались лишь те пробанды, которые согласно родословным на протяжении трех поколений принадлежали к исследуемой популяции в пределах одной этнической группы и административных границ изучаемого и соседних районов, исключая родственников индивидов в пределах третьей степени родства (Балановская и др., 2016).

Родоплеменная структура казахской популяции (табл.1) представлена тремя социально-территориальными объединениями - «жузами», в составе которых изучены 12 родоплеменных групп: *уйсун, жалайыр, канлы* – старший жуз; *аргын, керей, конырат, кыпшак, найман, уак* – средний жуз; *алимулы, байулы, жетиру* – младший жуз; а также 2 родоплеменные группы, не входящие в жузы: *кожа-сунак* - степное духовенство; *торе* - степная аристократия. Всего изучено 49 кланов (табл. 1).

Методы молекулярно-генетического анализа. Выделение ДНК проведено методом экстракции с помощью органических растворителей (фенол-хлороформ). Концентрация ДНК определена методами спектрофотометрии (NanoDrop 2000), флуориметрии (Qubit 2.0) и ПЦР в реальном времени (Quantifiler Human DNA Kit).

Генотипирование Y-хромосомы проведено методом ПЦР в реальном времени с использованием TaqMan зондов на 44 SNP: M174, P37, M35, M78, M123, M217, M48, M407, M285, P15, P303, M377, M69, M170, M253, M267, P58, M172, M47, M67, M92, M12, M231, M178, M128, P43, M119, M122, P201, M134, M242, M120, M143, M378, M198, M458, M343, P297, M269, M124, M27, M357, M317, M70.

Анализ 17 STR-маркеров Y-хромосомы проведен методом фрагментного анализа на капиллярном секвенаторе ABI 3130xl с использованием набора Y-filer PCR Amplification Kit.

Секвенирование Y-хромосомы проведено с использованием коммерческого продукта «BigY», предлагаемого компанией Gene By Gene, Ltd. по селективному обогащению целевых участков Y-хромосомы, так называемого Extended Gold Standard региона (общая длина 10.5Mb) на платформе Illumina HiSeq2000/2500.

Таблица 1. Исследованные родоплеменные группы.

Социально-территориальные объединения	Родо-племенные группы	Кланы в составе родоплеменных групп*	Объем выборки
старший жуз	жалайыр	каракалпак, шуманак, сырманак	103
	канлы	караканлы, кызылканлы	27
	уйсун	албан, дулат, ошакты, ысты, шапырашты, сиргели, суан	248
средний жуз	аргын	бес мейрам, жети момын, токал аргын, таракты	384
	керей	абак, ашамайлы	102
	конырат	коктунулы, котенши	90
	кыпшак	бултын, кара, кулан	133
	найман	сарыжомарт, толегетай	336
	уак	шога	45
младший жуз	алимулы	каракесек, карсакал, кете, торткара, шекти, шомекей	145
	байулы	адай, алаша, байбакты, бериш, кызылкурт, тана, ыссык	130
	жетиру	табын, тама, телеу, жагабайлы	55
кожа-сунак		кожа, сунак	107
торе-туленгиты		торе	28
другие группы		кыргыз, кулеген, толенгиты	32
нет информации			17
Всего изучено казахов			1982

Примечание: *указаны названия кланов, объемы выборок из которых составляют не менее 5 образцов.

Методы статистического анализа. С помощью программы Arlequin 3.5.1.3 (Excoffier, Lischer, 2010) проведен анализ молекулярной изменчивости (GD), анализ AMOVA, тест Мантеля. Географические и генетические расстояния Нея (Nei, 1975, 1987) рассчитаны в программах DistGeo и DJgenetic (Balanovsky et al., 2008, 2011, 2015); они визуализированы на графике многомерного шкалирования в программе Statistica 6.0 (StatSoft).

Методы филогенетического и картографического анализа.
Филогенетический анализ. Поиск родственных гаплотипов Y-хромосомы проводили в программе Haplomatch (Чухряева и др., 2016). Медианные сети гаплотипов строили по методу Бандельта, используя алгоритмы reduced-median (Bandelt et al., 1995) и median-joining (Bandelt et al., 1999) в программе Network 4.1.1.2 (Forster et al., 2001; Polzin, Daneschmand, 2003). Полученные графики редактировались в Network Publisher (Fluxus Engineering, Clare, U.K.). Возраст медианных кластеров определялся методом ASD (Goldstein et al., 1995a; Goldstein et al., 1995b) в программе Y TMRCA Calculator, производной от Ytime (Behar et al., 2003), а также показателем rho (Forster et al., 1996; Saillard et al., 2000) с использованием «генеалогической» скорости мутирования $2,1 \times 10^{-3}$ (Ge et al., 2009), принимая длину поколения равной 30 лет (Fenner, 2005). *Картографический анализ* проведен с помощью программы GeneGeo (Balanovsky et al., 2011; Koshel, 2012; Балановский, 2015) методом средневзвешенной интерполяции.

Методы полногеномного анализа деревьев Y-хромосомы.
Компиляция данных проводилась с помощью программного обеспечения NGSConv, разрабатываемого В.В. Запороженко под руководством О.П. Балановского. Филогенетические деревья по извлеченным генотипам строились в программе Phylomurka, разработанной В.В. Запороженко, и в программном пакете BEAST (Drummond et al., 2012; Bouckaert et al., 2014).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Генетические портреты казахов. Генотипировано 1982 образца из 19 популяций казахов по 44 SNP и 17 STR маркерам Y-хромосомы. Созданы «генетические портреты» казахов на трех уровнях популяционной системы: 1) тотальная популяция; 2) социально-территориальная структура (жузы); 3) родоплеменные группы.

«Тотальная популяция» ($N=1982$) характеризуется высоким разнообразием гаплогрупп ($GD=0.89$). Однако большая часть (87%) изменчивости Y-хромосомы распределена между 7 гаплогруппами с частотой $> 5\%$: **C2-M217** – 41%; **G1-M285** – 15%; **O2a2b1-M134** – 8%; **R1a1a-M198** – 7%; **R1b1a1a-P297** – 6%; **J2-M172** – 5%; **N1a1a-M178** – 5%. В генетическом пространстве Евразии (график многомерного шкалирования приведен в диссертации) популяция казахов располагается в Центрально-азиатском кластере. Генетически наиболее близко население регионов – Восточная Сибирь ($d=0.29$), степи Европы ($d=0.77$), Южная Сибирь ($d=0.82$).

«Жузы» ($N=1798$) и вне-жузовые группы ($N=167$). Социально-территориальная структура казахов определяется делением на три «жуза» (исторически сложившиеся географические объединения казахских кланов) – старший, средний и младший (рис. 1). Генетический портрет жузов характеризуется низким уровнем межгрупповых различий между ними ($F_{ST}=0.06$) и меньшим уровнем разнообразия гаплогрупп (средняя по всем жузам $GD=0.74$), чем для всей популяции.

«Родоплеменная структура» ($N=1791$, табл. 2). Генетический портрет родоплеменных групп характеризуется низким уровнем разнообразия Y-хромосомных линий (средняя $GD=0.62$), высоким уровнем генетических расстояний между группами (средняя $d=1.83$) и межгруппового разнообразия ($F_{ST}=0.31$). Это значение F_{ST} является одной из самых высоких величин различий между популяциями внутри одного народа в Евразии (сравнительная таблица приведена в диссертации).

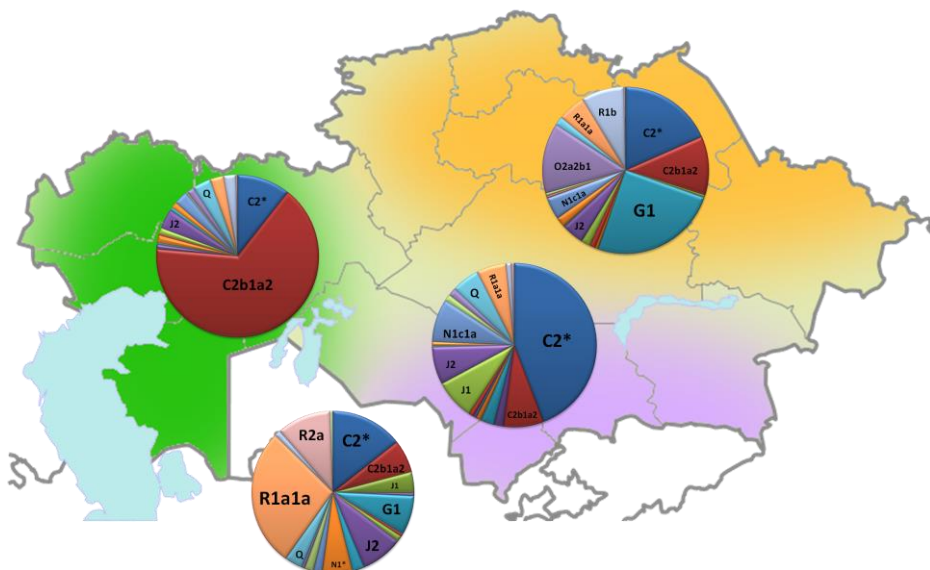


Рисунок 1. Генетические портреты «жузов»

Обозначения: Общий цветовой паттерн отражает ареалы жузов: фиолетовый – старший жуз; желтый – средний жуз; зеленый – младший жуз. Портрет вне-жузовых кланов представлен в виде диаграммы за пределами территории Казахстана. На диаграммах подписаны гаплогруппы с частотой встречаемостью не менее 4% в жузе. Маркеры, определяющие гаплогруппы, приведены в табл. 2.

Частоты гаплогрупп представлены в табл. 2. Самые частые гаплогруппы, указанные при описании тотальной популяции, в составе отдельных родоплеменных групп имеют еще более повышенное накопление: **C2*** - у керей (66%), уйсун (50%), жалайыр (38%), торе (36%), жетиру (25%); **C2b1a2** – у алимұлы (77%), байұлы (69%), жетиру (27%), найман (27%); **C2c1a1a1** у коньрат (86%). **G1** – у аргын (67%); **O2a2b1** – у найман (42%); **R1a1a*** - кожа и сунак (32%); **R1b1a1a*** - у кыпшақ (47%) (а также **J2*** - 22%); **N1a1a** - у групп уак (64%) и жалайыр (22%); **Q*** - у канлы (48%).

Ганглогруппа/ Маркер		Казахи N=1982	Старший жуз			Средний жуз					Маллший жуз			Торе N=28	Кожа- Сунак N=107	
			Уйсун N=248	Жалайыр N=103	Канлы N=27	Аргын N=384	Керей N=102	Коньрат N=90	Кыпшак N=133	Найман N=336	Уак N=45	Алимулы N=145	Байулы N=130			Жетиру N=55
D*	M174	0.01	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0.07
D1b	P37.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E1b1b1*	M35	0	0	0.04	0	0	0	0	0	0	0	0	0	0.02	0	0.01
E1b1b1a1	M78	0	0.01	0.01	0	0	0.01	0	0	0	0	0	0.02	0	0	0
E1b1b1b2a1	M123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C2*	M217	0.17	0.50	0.38	0.07	0.03	0.66	0.02	0.02	0.10	0.02	0.02	0.13	0.25	0.36	0.07
C2b1a2	M48	0.19	0.11	0.02	0	0.05	0.09	0.01	0.02	0.27	0.07	0.77	0.69	0.27	0.04	0.03
C2c1a1a1	M407	0.04	0.01	0.01	0	0.01	0	0.86	0	0.01	0.04	0	0	0.02	0	0
G1	M285	0.15	0.03	0	0.07	0.67	0.04	0	0.05	0.01	0.04	0	0	0.02	0.04	0.07
G2a*	P15	0	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0
G2a2b2a	P303	0.01	0	0	0	0.01	0	0	0.01	0	0.04	0.02	0	0.02	0	0
G2b1	M377	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0
H1	M69	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0
I*	M170	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0.01
I1	M253	0.01	0.01	0.01	0	0	0	0	0.01	0.02	0.02	0	0	0	0	0
I2a1	P37.2	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0
J1*	M267	0.03	0.12	0	0.04	0.02	0.04	0.01	0	0	0.04	0.01	0.02	0	0.07	0
J1a2a1a2	P58	0	0	0	0	0.01	0.01	0	0	0	0	0	0	0	0	0
J2*	M172	0.05	0.08	0.03	0.07	0.03	0.02	0	0.22	0.01	0	0.01	0.02	0.13	0	0.10
J2a1a	M47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J2a1b*	M67	0	0	0.01	0	0	0	0	0	0	0	0	0.02	0	0	0.01
J2a1b1	M92	0	0	0	0.07	0	0	0	0	0	0	0	0	0	0	0
J2b	M12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01
N*	M231	0.01	0	0	0	0.01	0.01	0	0	0.03	0	0	0	0.02	0	0.01
N1a1a	M178	0.05	0.04	0.22	0	0.02	0	0.03	0.01	0.02	0.64	0.01	0.03	0.07	0	0.01
N1a2a	M128	0	0.01	0	0	0	0	0	0	0	0	0	0.01	0	0	0.02
N1a2b	P43	0.01	0	0.01	0	0	0	0	0	0.02	0	0.01	0	0.02	0	0.02
O1b	P31	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0
O2*	M122	0	0.01	0.01	0	0	0	0	0	0	0	0.01	0	0	0	0.03
O2a2*	P201	0	0	0.01	0	0	0	0	0	0.02	0	0	0	0	0	0
O2a2b1	M134															

На рис. 2 представлены генетические взаимоотношения между родоплеменными группами (по данным 44 SNP Y-хромосомы). В генетическом пространстве выделяются два кластера: первый (I) объединяет родоплеменные группы *жалайыр*, *жетиру*, *керей*, *уйсун*, *торе*; второй (II) кластер – *алимулы*, *байулы*, *найман*. Родоплеменные группы *аргын*, *канлы*, *кожа-сунак*, *конырат*, *кыпшак* и *уак* характеризуются специфическими особенностями генофонда, не объединяясь с другими изученными родоплеменными группами.

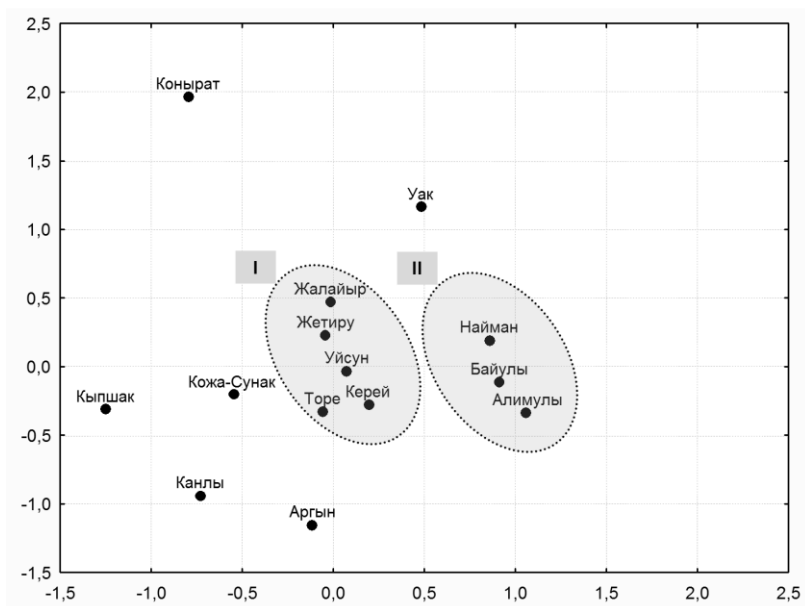


Рисунок 2. Родоплеменные группы казахов в генетическом пространстве (многомерное шкалирование, величина стресса = 0.14, алиенации = 0.18).

Связь генетической и родоплеменной структуры генофонда казахов. Уже анализ генетических портретов указывает на значительную связь между генетической и родоплеменной структурой генофонда казахов. Однако необходимо подтверждение этой связи количественными методами. С этой целью мы изучили генетические различия между различными группами методом AMOVA (табл. 3), где одни и те же образцы были сгруппированы

двумя способами: родоплеменные группы и географические популяции.

Роль родоплеменной структуры в формировании генофонда казахской популяции оказалась более важной, чем даже географический фактор (играющий главную роль в генофондах большинства регионов мира): генетические различия между родоплеменными группами (31%) в 1,5 раза больше, чем между географическими популяциями (15-21%).

Таблица 3. Анализ эффективности родоплеменной и географической классификаций методом AMOVA.

Классификация	Группы	Изменчивость в %	
		Внутри группы	Между группами
Родоплеменная структура	14 кланов	68.8*	31.2*
География	3 жуза	84.2*	15.8*
	12 областей	81.3*	18.7*
	19 районов	78.6*	21.4*

Примечание: *Р-значения<0.001

Еще более ярко ключевая роль родоплеменной структуры выявляется независимым методом – корреляционным тестом Мантеля. Частная корреляция между матрицами генетических и квазигенетических расстояний (отражающих сходство районных популяций по родовому составу) велика и достоверна ($r=0.55$, p -значения<0.001), в то время как частная корреляция между матрицами генетических и географических расстояний нулевая ($r=-0.07$, $p=0.69$).

Положение казахов в генетическом пространстве Центральной Азии. Одним из ключевых регионов Центральной Азии является **Трансоксиана** – исторический ареал тесного контакта казахов с популяциями Средней Азии. В контексте детального исследования их генетических взаимоотношений этот регион заслуживает отдельного изучения его генофонда. Наше исследование изменчивости Y-хромосомы Трансоксианы включило собственные данные о генофондах 4 популяций казахов, 3 популяций узбеков и популяций каракалпаков, туркмен и дунган. На графике многомерного шкалирования, где эти 10 популяций

рассмотрены в контексте 69 окружающих популяций, выделились кластеры, соответствующие региональному делению Азии (рис. 3). Голубыми линиями объединены популяции вдоль рек Амударьи и Сырдарьи, однако положения этих генетических «рек» лишь отдаленно схожи с их географическими прототипами.

Земледельческие популяции узбеков и таджиков, а также киргизы, генетически отдалены от большинства кочевых популяций – казахов, монголов, туркмен, хазарейцев. Популяции туркмен, хотя и происходят из трех разных стран (Узбекистана, Ирана и Афганистана) формируют свой собственный, четко обособленный кластер.

Анализ роли культурного ландшафта (земледельческое или скотоводческое население) и двух аспектов географического ландшафта (рельеф и бассейны рек) в формировании генофонда проведен двумя независимыми методами. Согласно методу AMOVA (табл. 4) роль культурного ландшафта (3%) в организации генофонда в три раза выше, чем роль географического ландшафта (0% и 1%). Тест Мантеля подтверждает, что генетические расстояния между популяциями Трансоксианы, в отличие от большинства других регионов мира, не скоррелированы с географическими расстояниями ($r=-0.006$, $P=0.44$).

Таблица 4. Анализ генетического разнообразия (AMOVA): роль географического и культурного ландшафта в генофонде Трансоксианы.

Классификация	Группы	Изменчивость (%)		
		Внутри популяции	Внутри группы	Между группами
Бассейны рек	Амударья / Сырдарья	84.18*	16.04*	-0.22 (p-value=0.39)
Рельеф	Предгорья / равнины	82.43*	16.52*	1.05 (p-value=0.27)
Тип хозяйства	Земледелие/ кочевое скотоводство	82.82*	14.33*	2.85 (p-value=0.096)

Примечание: *P-значения < 0.001

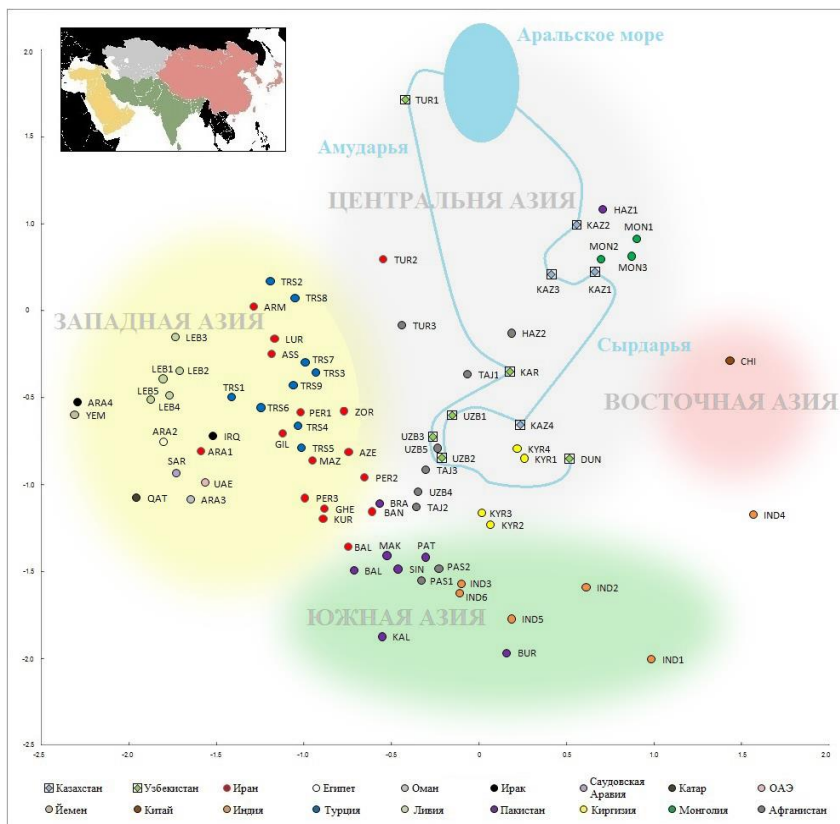


Рисунок 3. Положение популяций Трансоксианы в генетическом пространстве Азии (многомерное шкалирование, величина стресса = 0.16, алиенации = 0.17)

Обозначения: Популяции на графике представлены кружками, популяции, изученные в диссертации, выделены ромбиками в квадратах. Цвета кружков определяют принадлежность популяций к одной из 18 стран. Цветовыми облаками выделены географические кластеры. Врезка: выделение регионов Азии по классификации ООН (в той же цветовой шкале). Расшифровка кода популяций представлена в диссертации.

Филогеография одной из мажорных гаплогрупп у казахов.

Ареал гаплогруппы **G1**, второй по частоте в генофонде казахов, простирается от Италии до Западной Монголии. Однако пики частоты **G1** (рис. 4А) приходятся на степи Центральной Азии (в Северном Казахстане 80%) и на Иранское и Армянское нагорья (в Армении 42%). Была построена карта гаплотипического

разнообразия **G1** (рис. 4Б), которая выявляет его четкое снижение от западного Ирана к восточным окраинам Юго-Западной Азии и далее на север к Евразийским степям. Высокое разнообразие гаплотипов **G1** в западной части Ирана в сочетании с высокой частотой **G1** может указывать на иранский ареал как на прародину гаплогруппы **G1**. Таким образом, **G1** могла бы обозначить древнюю генетическую связь между ираноязычным населением Юго-Западной Азии и популяциями центральноазиатских степей, в которых во II-I тыс. до н.э. тоже преобладали языки иранской группы.

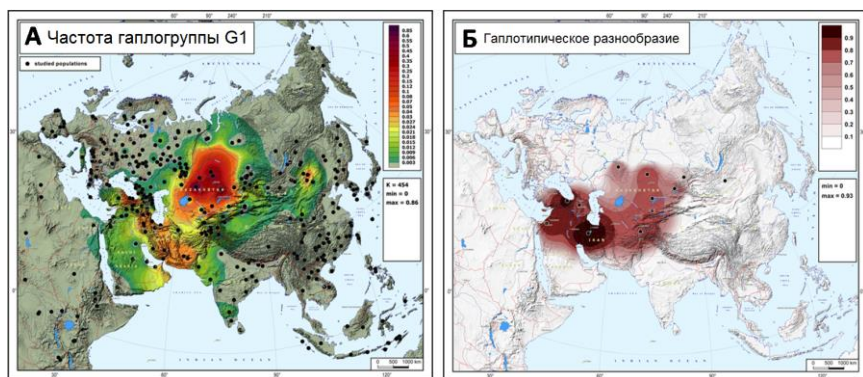


Рисунок 4. Филогеография гаплогруппы G1-M285.

А) Карта распространения **G1**; Б) Карта разнообразия гаплотипов **G1**.

Использование родоплеменной структуры для определения скорости мутирования на Y-хромосоме. Совпадение филогенетического дерева гаплогруппы **G1** у аргынов (фактического биологического родства представителей этой родоплеменной группы) и их традиционного генеалогического дерева (рис. 5) позволило создать «клановый» метод оценки скорости мутирования Y-хромосомы. Этот метод использует генеалогическое дерево в качестве модели дивергенции носителей **G1** от их общего предка (золотоордынский эмир Караходжа), дата жизни которого известна из исторических источников. При расчете использованы параметры: промежуток времени 627 лет между Караходжой (средняя дата рождения его сыновей - 1385 г.) и исследованными образцами казахов (средняя дата рождения – 1982 г.); общая длина Y-хромосомных сегментов, секвенированных в

каждом образце (9972660 нуклеотидов); среднее число мутаций у исследованных казахов аргынов по сравнению с родоначальником (4.89 мутаций). На основе этих величин получена оценка скорости мутирования Y-хромосомы: 0.78×10^{-9} на нуклеотид на год.

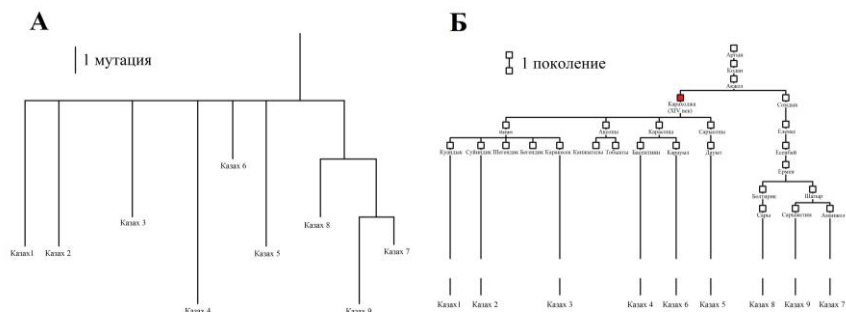


Рисунок 5. Генетическая и генеалогическая реконструкции изученных представителей казахского рода аргын:

А) Филогенетическое древо на основе данных секвенирования Y-хромосомы (~10 млн. п.н.); Б) Генеалогическое древо аргын.

Генетическая реконструкция происхождения родоплеменных групп. Изучение связи генетической и родоплеменной структуры имеет значение как для решения фундаментальных проблем этногенеза, так и для решения частных задач, например, генезиса крупнейшей родоплеменной группы казахов – аргын и двух внежузовых групп: степной аристократии – торе и духовенства – кожа-сунак.

Историческая наука выдвигает две гипотезы происхождения аргын: традиционная «шежире» возводит всех аргын к общему предку, а альтернативная версия рассматривает их как союз племен различного происхождения. Генетические портреты разных родов аргын (рис. 6), выявляющие высокую частоту гаплогруппы **G1** (красный цвет) практически во всех генеалогических линиях аргын, указывают на реальность существования их единого биологического предка. Начало экспансии **G1** в генофонде казахов, оцененное по разнообразию SNP и STR маркеров, датируется 750-470 лет назад и совпадает со временем жизни генеалогического предка аргын. Таким образом, данные генетики свидетельствуют в пользу гипотезы общего предка, а не гипотезы «союза племен».

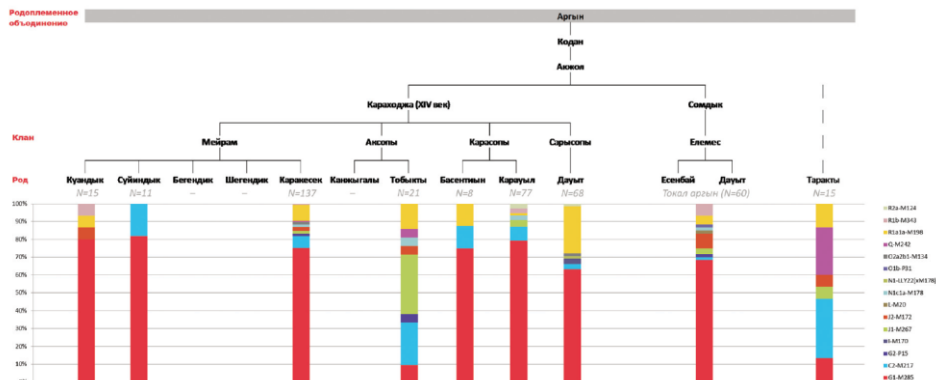


Рисунок 6. Структура родоплеменного объединения аргын и генетические портреты изученных родов.

Распространение ислама является одной из самых мощных культурных экспансий в Западной, Южной и Центральной Азии. Степное духовенство по своим генеалогическим сказаниям происходит от близких родственников пророка Мухаммеда по мужской линии. Однако исследование разнообразия Y-хромосомы кожа-сунак выявило высокий уровень разнообразия гаплогрупп ($GD=0.86$) и разнообразия гаплотипов ($HD=0.98$), что визуальное продемонстрировано на их медианной сети (представлена в диссертации). Можно предположить, что традиционная генеалогия духовенства первоначально отражала не биологическое родство, а цепочку «учитель-ученик». Но у кочевников, благодаря родовой организации и традиции патронимии, учениками часто становились сыновья, что проявилось в возникновении ряда мини-кластеров гаплотипов среди генеалогических линий степного духовенства.

Исторически важную роль в формировании Казахского ханства сыграл род торе – степная аристократия, по своим генеалогическим преданиям восходящая к Чингисхану. Генетический портрет рода составляют три основные гаплогруппы: $C2^*$ – 36%, $R2a$ – 25%, $R1a1a^*$ – 18%. Сравнительный анализ с монгольским родом борджигин, из которого происходит Чингисхан (генеалогическая линия Кият) выявил сходство по гаплогруппе $C2^*$ - 39%, но практически полное отсутствие у борджигинов гаплогрупп $R2a$ и $R1a1a$. Причем модальный гаплотип - «стар-кластер» гаплогруппы

C2*, который, как считается, имеет отношение к Чингисхану и его близким родственникам по отцовской линии (Zerjal et al., 2003), встречен не далеко во всех династических линиях (как у торе, так и у борджигинов), что может быть связано с включением в состав правящей средневековой элиты групп, биологически не являющихся прямыми потомками чингизидов.

ВЫВОДЫ

1. Генетические различия между разными казахскими субпопуляциями (от 14 до 19 выборок, 1982 образца) определяются родоплеменной структурой казахской популяции больше, чем ее географической подразделенностью. Это подтверждается и AMOVA (различия между родами $F_{ST}=0.31$, различия между районными субпопуляциями $F_{ST}=0.21$), и тестом Мантеля (частная корреляция генетических расстояний с родоплеменной структурой высока, 0.55, с географическими расстояниями недостоверна, -0.07).
2. Генофонд казахских и других популяций бассейнов Амударьи и Сырдарьи, в отличие от большинства регионов мира, структурирован не столько по географическим факторам, сколько по фактору хозяйственно-культурной деятельности (AMOVA: земледелие или кочевое скотоводство $F_{ST}=0.03$; предгорья или равнины $F_{ST}=0.01$; бассейн Амударьи или Сырдарьи $F_{ST}=0.00$).
3. Особенности генофондов родоплеменных групп позволили верифицировать исторические гипотезы их происхождения: для крупнейшей группы аргын подтвердилась «гипотеза одного предка», а не «союза племен»; для группы «степного духовенства» кожа-сунак не подтвердилась гипотеза единого предка (пророка Мухаммеда); для группы «степной аристократии» торе не для всех династических линий прослеживается связь с общим предком (Чингисханом).
4. Новая оценка скорости SNP мутаций на Y-хромосоме (0.78 мутаций на миллиард нуклеотидов в год) получена на основе совпадения филогенетического дерева гаплогруппы G1 у аргынов (построенного по результатам секвенирования ~10 млн. п.н.) и их традиционной генеалогии, восходящей к историческому лицу с известной датой жизни.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи, опубликованные в рецензируемых научных изданиях, рекомендованных ВАК МОН РФ

1. **Zhabagin M.**, Balanovska E., Sabitov Zh., Kuznetsova M., Agdzhoyan A., Balaganskaya O., Chukhryaeva M., Markina N., Romanov A., Skhalyakho R., Zaporozhchenko V., Saroyants L., Dalimova D., Davletchurin D., Turdikulova Sh., Yusupov U., Tazhigulova I., Akilzhanova A., Tyler-Smith C., Balanovsky O. *The Connection of the Genetic, Cultural and Geographic Landscapes of Transoxiana* // Scientific Reports. 2017. 7(1):3085.
2. **Жабегин М.К.**, Сабитов Ж.М. Агджоян А.А, Юсупов Ю.М., Богунов Ю.В., Лавряшина М.Б., Схалыхо 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. С.59-68
3. Балановская Е.В., **Жабегин М.К.**, Агджоян А.Т., Чухряева М.И., Маркина Н.В., Балаганская О.А., Схалыхо Р.А., Юсупов Ю.М., Утевская О.М., Богунов Ю.В., Асылгужин Р.Р., Долинина Д.О., Кагазежева Ж.А., Дамба Л.Д., Запороженченко В.В., Романов А.Г., Дибирова Х.Д., Кузнецова М.А., Лавряшина М.Б., Почешхова Э.А., Балановский О.П. *Популяционные биобанки: принципы организации и перспективы применения в геногеографии и персонализированной медицине* // Генетика. 2016. Т. 52. № 12. С. 1371-1387.
4. Схалыхо Р.А., **Жабегин М.К.**, Юсупов Ю.М., Агджоян А.Т., Сабитов Ж., Гурьянов В.М., Балаганская О.А., Далимова Д., Давлетчурин Д., Турдикулова Ш., Асылгужин Р.Р., Акильжанова А.Р., Балановский О.П., Балановская Е.В. *Генофонд туркмен Каракалпакстана в контексте популяций Центральной Азии (полиморфизм Y-хромосомы)* // Вестн. Моск. ун-та. Сер. XXIII. Антропология, 2016. №3. С.86-96.
5. 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., A Zalloua P., 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.
6. **Жабегин 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. С.96-101.

Тезисы докладов

7. **Zhabagin M.**, Balanovsky O., Zaporozhchenko V., Sabitov Zh., Zhumadilov Zh., Balanovska E. *The genetic connection between Western and Central Asian populations* // Proceedings of The International Conference Great Migrations in Asia Minor: Circulation, Exchange and Social Transformation. UNESCO, Paris, 29 -30 November 2016. P.176-177.
8. **Zhabagin M.**, Yusupov Yu., Sabitov Zh., Shalyakho R., Agdzhoyan A., Zhumadilov Zh., Balanovska E., Balanovsky O. *Geographical and Cultural Influences on Genetic Diversity: Patterns of the Y-Chromosomal Variation in Populations with Patronymic Tradition* // Abstracts of The 13th International Congress of Human Genetics. Kyoto, 3-7 April 2016. Tue(3)-P-172.
9. **Zhabagin M.**, Balanovsky O., Sabitov Z., Tazhigulova I., Askapuli A., Dibirova K., Chukhryaeva M., Agdzhoyan A., Markina N., Yusupov Y., Tarlykov P., Zholdybaeva E., Akilzhanova A., Zhumadilov Z., Balanovska E. *Gene geography of Kazakh populations from the Y-chromosomal data* // Abstract book of The Human Genome Meeting. Kuala Lumpur, 14-17 March 2015. P.133.
10. **Жабегин М.К.**, Балановская Е.В., Сабитов Ж.М., Богунов Ю., Фролова С.А., Дибирова Х.Д., Захаров-Гезехус И.А., Балановский О.П. *«Род» как квазигенетический маркер Y-хромосомы в популяционно-генетических и исторических исследованиях* // Тезисы докладов VI Съезда Вавиловского Общества Генетиков и Селекционеров и ассоциированные генетические симпозиумы. Ростов-на-Дону, 15-20 июня 2014 г. С.93.
11. **Zhabagin M.**, Sabitov Zh., Balanovsky O. *Genetic polymorphism of 17 Y-chromosomal STR loci in the Kozha and Tore tribes of Kazakh population.* // Abstract book of DNA in Forensics. 9th Y-User Workshop and 6th EMPOP meeting. Brussels, 14-16 May 2014. P.107.
12. **Жабегин М.К.**, Сабитов Ж.М., Тажигулова И.М., Тарлыков П.В., Агджоян А.Т., Дибирова Х.Д., Шанько А.В., Балаганская О.А., Богунов Ю.В., Захаров-Гезехус И.А., Балановский О.П. *Генофонд казахского народа в контексте этнотерриториальных объединений казахов (по данным о полиморфизме Y-хромосомы)* // Тезисы Международной конференции памяти Ю.Г. Рычкова Проблемы генетики населения и этнической антропологии. Москва, 19-21 ноября 2013. С.23