

ЧУХРЯЕВА Марина Игоревна

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ ГЕНОФОНДОВ ПОПУЛЯЦИЙ
ИНДОЕВРОПЕЙСКОЙ И ДРУГИХ ЛИНГВИСТИЧЕСКИХ СЕМЕЙ
В ЗОНАХ ИХ КОНТАКТОВ**

03.02.07 – генетика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2018

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте общей генетики им. Н.И. Вавилова Российской академии наук и Федеральном государственном бюджетном научном учреждении «Медико-генетический научный центр»

Научный руководитель:

Доктор биологических наук, профессор РАН **Балановский Олег Павлович;**

Официальные оппоненты:

Прохорчук Егор Борисович, доктор биологических наук, профессор,
Федеральное государственное учреждение «Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук», заведующий лабораторией геномики и эпигеномики позвоночных.

Пежемский Денис Валерьевич, кандидат биологических наук,
Музей антропологии Московского государственного университета им. М.В. Ломоносова, старший научный сотрудник.

Ведущая организация: Научно-исследовательский институт медицинской генетики (НИИ медицинской генетики) Федерального государственного бюджетного научного учреждения "Томский национальный исследовательский медицинский центр Российской академии наук" (Томский НИМЦ)

Защита состоится «20» декабря 2018 г. в ____ часов на заседании Диссертационного ученого совета Д 002.214.01 в Федеральном государственном бюджетном учреждении науки «Институт общей генетики им. Н.И. Вавилова» Российской академии наук (119991, Москва, ул. Губкина, дом 3).

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки «Институт общей генетики им. Н.И. Вавилова» Российской академии наук (119991, Москва, ул. Губкина, дом 3).

Автореферат разослан « _____ » _____ 2018 г.

Учёный секретарь диссертационного совета Д 002.214.01

по защите докторских и кандидатских диссертаций, доктор биологических наук Горячева И.И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Изучение структуры генофондов популяций человека является одним из ключевых направлений современной генетики. Наиболее часто для подобных исследований используются маркеры с одnorodительским типом наследования: митохондриальная ДНК и особенно Y-хромосома, которая обладает самым низким уровнем полиморфизма среди других хромосом [Hammer et al., 2003], но при этом ее межпопуляционное разнообразие максимально. Кроме одnorodительских маркеров, все большую популярность в популяционно-генетических исследованиях приобретают полногеномные аутосомные панели и секвенирование полных геномов. Наибольшее число исследований велось с использованием серии панелей Illumina, в том числе и для российских популяций [Fedorova et al., 2013; Yunusbaev et al., 2015; Triska et al., 2017], разработанных для медицинских целей. Однако существуют и две полногеномные панели, предназначенные специально для популяционно-генетических исследований: GenoChip [Elhaik et al., 2013] и Axiom Human Origins [Lazaridis et al., 2014]. Для обеих панелей показана высокая дифференцирующая способность [ArunKumar et al., 2015; Naak et al., 2015].

Большинство этнических популяций мира к настоящему времени уже изучено, хотя и в разной степени, по маркерам Y-хромосомы или полногеномным панелям, и эти генетические данные – хотя и с разной степенью убедительности – применяются к решению многих междисциплинарных научных проблем, касающихся истории популяций человека. Одной из наиболее значимых, старых и сложных является «индоевропейская проблема», то есть вопрос о месте прародины индоевропейской языковой семьи (крупнейшей в мире по ареалу распространения и числу носителей в мире) и о миграциях индоевропейцев. В разработку этой проблемы основной вклад внесли лингвисты и археологи, выдвинув ряд конкурирующих гипотез о локализации индоевропейской прародины и путях миграций индоевропейцев, а в последние годы к ним присоединились и генетики, в том числе и с данными по древней ДНК [Damgaard et al., 2018; Nielsen et al., 2017].

Так, ряд исследователей высказывал предположения, что миграции индоевропейцев маркируются отдельными гаплогруппами Y-хромосомы. Чаще всего на эту роль выдвигают гаплогруппу R1a, в этом случае более вероятной оказывается переднеазиатская прародина [Semino et al., 2000; Underhill et al., 2015]; хотя в работе [Naak et al., 2015] были представлены генетические аргументы в пользу причерноморской прародины (носители ямной археологической культуры), в генофонде которых доминирует гаплогруппа R1b.

Для поиска генетических следов миграций лингвистических групп в мировой науке несколько раз продемонстрировал свою эффективность подход пар популяций – когда

сравниваются популяции, в генофонд которых был направлен поток миграций исследуемой группы и их географические соседи, не испытывавшие такого влияния [Zalloua et al., 2008; Yunusbayev et al., 2015]. Но для индоевропейской языковой семьи такой подход до сих пор не был задействован.

Зато неоднократно исследовалась степень согласованности между структурой генофонда популяций и их лингвистической принадлежностью. Чаще всего оказывалось, что роль лингвистического родства меньше, чем роль географического соседства, но для населения некоторых регионов мира закономерность оказывалась обратной. Для индоевропейцев в целом такой анализ не проводился, но неоднократно исследовались отдельные индоевропейские популяции или их группы. Причем мнения исследователей разделились: одни приходили к заключению, что генетическое сходство между разными индоевропейскими популяциями незначительно, а, следовательно, индоевропейские языки распространялись путём языкового заимствования, без массовых миграций и поэтому без значимых отпечатков в генофонде [Belledi et al., 2000; Rosser et al., 2000; Yunusbayev et al., 2012; Pagani et al., 2017], тогда как другие авторы обнаруживали четкую согласованность генетических и лингвистических границ [Kayzer et al., 2005].

Экспансия индоевропейцев – это не только события новокаменного или бронзового века. Она продолжалась тысячелетиями и продолжается сегодня: завоевания Александра Македонского в античности, экспансия славян в Средние века, колонизация Дикого Запада и походы казаков в Сибирь в Новое время, строительство БАМа в новейшее – все эти процессы распространяли индоевропейские языки и влияли на формирование генофондов того множества популяций, которые сейчас говорят на этих языках. И чтобы охватить разные звенья этой цепи, важно изучить генофонд индоевропейцев не только в целом (что отразит наиболее ранние или наиболее общие черты истории их генофондов), но и отдельных групп индоевропейцев и их отдельных популяций. В нашем исследовании мы сосредоточились на балто-славянской и армянской группах индоевропейской семьи.

Если западноевропейские и индийские индоевропейские популяции изучены достаточно хорошо, то обобщающего исследования генофонда народов балто-славянской языковой группы до нашего исследования не было проведено, несмотря на большой массив работ об отдельных её представителях [Pericic et al., 2005; Luca et al., 2006; Balanovsky et al., 2008; Battaglia et al., 2008; Lappalainen et al., 2008; Noveski et al., 2009; Underhill et al., 2010; Myres et al., 2010; Rebala et al., 2012; Karachanak et al., 2013; Kushniarevich et al., 2013; Rootsi et al., 2012; Mittnik et al., 2018]. Особенный интерес представляет исторически хорошо изученное распространение восточных славян по Восточно-европейской равнине, сопровождавшееся ассимиляцией дославянского населения (главным образом финно-угорской языковой группы) в том числе на территории

Верхнего Поволжья. Важно также понимать, какие популяционно-генетические процессы шли в ходе дальнейших миграций русского народа, в том числе при формировании популяций казаков.

Армяне, представители армянской языковой группы, одной из наиболее рано отделившихся ветвей на лингвистическом дереве индоевропейских языков, также лишь отчасти изучены генетиками. Их генофонд по маркерам Y-хромосомы был описан в работах [Rosser et al., 2000; Weale et al., 2001, Nasidze, 2003, Wells et al., 2004; Yunusbayev et al., 2012], но с недостаточным уровнем филогенетического разрешения. Отдельные популяции армян изучены также в работах [Grugni et al., 2012; Margaryan et al., 2012; Herrera et al., 2012; Lowery et al. 2012; Теучеж и др., 2013; Hovhannisyan A. et al., 2014; Margaryan et al., 2017].

Цель исследования:

Изучить генофонд ряда популяций, говорящих на языках индоевропейской языковой семьи, в сравнении с генофондами их географических соседей, принадлежащих к иным языковым семьям; выявить степень согласованности лингвистического сходства индоевропейских популяций и их сходства по маркерам Y-хромосомы и широкогеномным аутосомным панелям.

Задачи исследования

1. Определить степень генетического сходства популяций индоевропейцев по полногеномным данным путем анализа комплексов сравнения (индоевропейских и географически соседних народов других языковых семей) и провести поиск SNP-маркеров, характерных для популяций индоевропейской лингвистической семьи.
2. Количественно оценить степень сходства лингвистических и генетических реконструкций родства индоевропейских популяций (по полногеномным маркерам).
3. Изучить популяции балто-славянской лингвистической группы индоевропейской семьи по маркерам Y-хромосомы, включая популяции в зонах исторического контакта славянской и финно-угорских групп.
4. Изучить популяции армянской лингвистической группы индоевропейской семьи по маркерам Y-хромосомы.
5. Сравнить генофонды армянских популяций на исходной территории расселения и в диаспоре.

Научная новизна

Впервые проведено целенаправленное генетическое изучение популяций индоевропейской семьи на основе полногеномных данных: по единой обширной панели маркеров изучены популяции разных групп индоевропейской семьи и соседние с каждой группой не-индоевропейские популяции.

Впервые на основе данных полногеномного генотипирования дана количественная оценка степени согласованности изменчивости индоевропейских популяций по их генофонду, языковой близости и географическим расстояниям.

Впервые проведен анализ степени согласованности лингвистических и генетических (по маркерам Y-хромосомы) расстояний в популяциях балто-славянской группы индоевропейской семьи на основе всего массива опубликованных и собственных новых данных.

Впервые изучен генофонд донских казаков.

Впервые выявлено структурирование генофонда Юго-Западной Азии на горные и равнинные популяции.

Впервые показано сохранение генофонда армян в диаспоре (по маркерам Y-хромосомы).

Научно-практическая значимость

Совокупность полученной информации о структуре генофонда изученных народов обеспечит проведение генетико-демографического мониторинга, включая прогнозирование изменений генофонда в результате продолжающихся массовых миграций и межэтнических браков.

База данных по распределению панели 17 STR маркеров Y-хромосомы (набор Y-filer) в популяциях армян, грузин, донских казаков и русских Ярославской области может применяться в судебно-медицинской экспертизе в качестве референсной базы при идентификации личности и определении районов возможного происхождения искомым гаплотипов Y-хромосомы.

Полученные результаты важны для понимания путей распространения индоевропейских языков, структуры генофонда народов армянской и балто-славянской языковых групп. Полученные результаты важны, в том числе специалистам смежных отраслей: лингвистам, антропологам, археологам, историкам, этнографам при реконструкции истории народов индоевропейской языковой семьи. Результаты работы используются в научном и учебно-педагогическом процессе в российских и зарубежных организациях: ФГБНУ «Медико-генетический научный центр», Институт общей генетики им. Н.И. Вавилова РАН, ФГБУН Институт языкознания РАН, Российский государственный гуманитарный университет, Казанский и Харьковский государственные университеты, Кубанский государственный медицинский университет, Следственный комитет РФ, Эстонский биоцентр, Институт молекулярной биологии Национальной академии наук Армении.

Основные положения, выносимые на защиту

1. Популяции, говорящие на языках индоевропейской языковой семьи, и соседние с ними иноязычные популяции генетически похожи. Их генетические взаимоотношения объясняются в основном географическим фактором (корреляция между генетической и географической близостью популяций 0,72). Влияние лингвистического фактора

обнаруживается при рассмотрении взаимоотношений отдельных групп индоевропейцев (коэффициент корреляции до 0,56).

2. Общий компонент генофонда и общие SNP маркеры для всех народов индоевропейской языковой семьи по полногеномным данным (панель GenoChip) не обнаружены. Анализ комплексов сравнения показал, что распространение индоевропейских языков лучше объясняется моделью культурного заимствования, а не демической экспансии.

3. Частная корреляция генетических (Y-хромосомных) и лингвистических данных для народов балто-славянской языковой семьи равна нулю, для генетических и географических данных – 0,81. Следовательно, генофонд современных славян был сформирован преимущественно за счет автохтонного субстрата, а влияние древних славян было преимущественно культурным и языковым.

4. Генофонд армян диаспоры воспроизводит генофонд на исторической территории армян, представлен в основном следующими гаплогруппами Y-хромосомы: R-L23, J-M67, G-M285.

5. Генофонд армян сходен с современными переднеазиатскими популяциями, и отличается от генофонда народов Северного Кавказа.

Апробация работы

Работа представлялась на Конференции молодых ученых ФГБНУ «МГНЦ». (Москва, 2014, 2015); на конференции «Антропология и этнология Кавказа» (Тбилиси, 2016); European Human Genetics Conference (Глазго, Шотландия, 2015); на VII Съезде Российского общества медицинских генетиков (Санкт-Петербург, 2015); на конференции «Языковая политика и языковые конфликты в современном мире», (Москва, 2014); на конгрессе «The 19th Congress of the European Anthropological Association “Anthropology: Unity in Diversity”» (Москва, 2014); на конференции «Проблемы генетики населения и этнической антропологии», памяти Ю.Г. Рычкова. (Москва, 2013); на VI Съезде Вавиловского Общества Генетиков и Селекционеров. (Ростов-на-Дону, 2014); Международной конференции «Алексеевские чтения» памяти академиков Алексеевой Т.И. и Алексеева В.П. «Человек в окружающей среде: этапы взаимодействия» (Москва, 2014).

Личный вклад автора

Автором выполнена техническая часть работы, которая включает выделение ДНК из 250 образцов армян и 350 образцов русских фенол-хлороформным методом, измерение концентрации (NanoDrop, RealTime 7900HT), формирование рабочих, архивных и нормализованных (до 2 нг/мкл) ДНК-коллекций для указанных выше популяций, генотипирование SNP маркеров и фрагментный анализ STR маркеров Y-хромосомы (кроме технической работы на 16-ти капиллярном генетическом анализаторе 3130x1, Applied Biosystems) 250 образцов армян, 152 образцов грузин, 314 образцов донских казаков, 132

образцов русских Верхнего Поволжья; а также ведение баз данных анкетной информации и результатов генотипирования. Автор провел отбор и подготовку образцов для генотипирования по полногеномной аутосомной панели GenoChip.

Автор самостоятельно провел статистический анализ: расчет генетических расстояний, кластерный анализ, многомерное шкалирование, построение филогенетических сетей, картографический анализ, расчет теста Мантеля и расчет межгрупповых вариаций (AMOVA), поиск маркеров, характерных для индоевропейских народов.

Публикации

Основные результаты исследования представлены в 18 научных публикациях, в том числе в 4 статьях в ведущих рецензируемых научных журналах, рекомендованных ВАК РФ для защиты диссертаций.

Структура и объем работы

Работа изложена на 174 страницах и состоит из введения и 5 глав, включающих обзор литературы, материалы и методы, результаты исследования и их обсуждение, список литературы, содержащий 183 источника, из которых 143 иностранных. Работа иллюстрирована 16 таблицами и 29 рисунками, включая карты.

МАТЕРИАЛЫ И МЕТОДЫ

Материалы

Образцы венозной крови или слюны от представителей коренного населения изучаемых в работе популяций собраны в ходе серии экспедиционных работ 1998-2014 гг., проведенных под руководством Балановской Е.В., Балановского О.П., Почешховой Э.А. и Епископосяна Л.М. У всех обследуемых получено письменное информированное согласие под контролем Этической комиссии ФГБНУ «МГНЦ».

Для всех перечисленных выше видов анализа, помимо собственных данных, в работу были включены литературные данные. Для балто-славянской лингвистической группы в анализ включены данные по 6079 образцам, для армянской лингвистической группы – по 600 образцам, взятые из базы данных Y-base, разработанной под руководством Балановского О.П.

Методы экспериментального ДНК анализа

ДНК из образцов венозной крови выделена классическим методом, основанным на использовании протеиназы К с последующей фенол-хлороформной экстракцией. Установление концентрации ДНК проводилось сначала на спектрофотометре NanoDrop (для создания рабочих коллекций), а затем на основе рабочих коллекций точно определялась концентрация двухцепочечной ДНК для создания коллекций с нормализованной ДНК.

Для всех образцов на генетическом анализаторе 3130xl проанализированы 17 STR локусов Y-хромосомы, составляющие панель Y-filer™ PCR Amplification Kit. Анализ SNP маркеров проводился методом ПЦР в реальном времени на приборах StepOne и 7900HT с помощью Taqman зондов (Applied Biosystems).

Генотипирование по маркерам Y-хромосомы проведено для 250 образцов армян, 152 образцов грузин, 314 образцов донских казаков, 132 образцов русских Верхнего Поволжья.

Методы статистического и филогенетического анализа

На основе данных об STR гаплотипах Y-хромосомы были построены филогенетические сети с помощью программ Network 4.1.1.2 [Fluxus Technology Ltd.] (www.fluxus-engineering.com) и Network Publisher [Fluxus Engineering, Clare, U.K.]. Возраст кластеров гаплотипов определялся согласно публикациям [Forster et al., 1996; Saillard et al., 2000; Ge et al., 2009].

Генетические расстояния между анализируемыми популяциями рассчитаны в программе DJ [Balanovsky et al., 2008] и визуализированы в программе Statistica 6.0 [StatSoft. Inc., 2001]. Карты генетических расстояний и распространения отдельных гаплогрупп построены в ПО GeneGeo [Balanovsky et al., 2011].

Тест Мантеля и анализ межпопуляционной изменчивости (AMOVA) проведены в программе Arlequin 3.5.

Методы анализа полногеномных данных

Проведено генотипирование 245 образцов (таблица 1) по полногеномной панели GenoChip на платформе Illumina.

Все собственные данные организованы в так называемые комплексы сравнения (6 комплексов) – географически соседние популяции, одна или несколько из которых принадлежат к индоевропейской языковой семье, а остальные принадлежат к иным лингвистическим семьям. Вместе с собственными, в анализ включены данные по 943 образцам [Elhaik et al., 2013], образцам предоставленным Pasteur Institute (Франция) и Эстонским Биоцентром (Эстония). С ними количество комплексов сравнения удалось увеличить до 9.

Полученные данные обработаны в программном пакете plink 1.07 [<http://pngu.mgh.harvard.edu/~purcell/plink/>]. Проведен анализ главных компонент (PCA) в программном пакете plink 1.07 [<http://pngu.mgh.harvard.edu/~purcell/plink/>]. Визуализация PCA, Admixture анализ и расчет F_{ST} дистанций осуществлены в программном пакете R studio 3.2.1, картографический анализ проведен с помощью программного пакета GeneGeo, тест Мантеля - в программе Arlequin 3.5., поиск общих маркеров осуществлялся в стандартном пакете Microsoft Excel.

Таблица 1. Изученные по полногеномной панели GenoChip комплексы сравнения

популяция	изученные группы	языковая семья	N
русские	центральные, южные, северные русские	индоевропейская	20
мордва	мокша, эрзя	уральская	10
карелы		уральская	7
вепсы		Уральская	4
украинцы	западные и восточные украинские популяции	индоевропейская	20
ногайцы	кубанские ногайцы	алтайская	12
крымские татары	степные крымские татары	алтайская	8
армяне	донские, амшенские, армяне Адыгеи и Эрзурума	индоевропейская	20
грузины	имеретинцы	картвельская	13
осетины	иронцы и дигорцы	индоевропейская	20
кабардинцы	-	северокавказская	20
таджики, памирские народы	ишкашимцы, рушанцы, шугнанцы, бартангцы, ваханцы, язгулемцы, таджики	индоевропейская	31
туркмены	-	алтайская	20
узбеки	ташкентские, хорезмские узбеки	алтайская	20
литовцы*	литовцы-аукштайты, литовцы-жемайты	индоевропейская	20
ИТОГО:	6 комплексов сравнения, 245 образцов		

Примечание: темной заливкой выделены индоевропейцы из каждого комплекса сравнения

**Составляют комплекс сравнения с эстонцами, данные по которым предоставлены Эстонским биоцентром.*

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

ГЕНЕТИЧЕСКИЙ ЛАНДШАФТ НАРОДОВ, ГОВОРЯЩИХ НА ЯЗЫКАХ ИНДОЕВРОПЕЙСКОЙ ЯЗЫКОВОЙ СЕМЬИ

Анализ данных полногеномного генотипирования по панели GenoChip показал, что географическая близость народов является решающим фактором, повлиявшим на генофонд народов индоевропейской языковой семьи и их соседей. Однако лингвистический фактор в ряде случаев также оказал влияние на формирование структуры генофонда Западной Евразии. Анализ главных компонент (Рисунок 1) подтверждает это. Так, французы и испанцы географически находятся по разные стороны ареала басков. Но на генетическом графике (Рисунок 1) французы и испанцы находятся по одну сторону относительно басков, в соответствии с их лингвистическим сходством. Подобным образом, русские (кроме северных) и украинцы достаточно четко отделены от своих не принадлежащих к индоевропейской семье соседей.

Для всех народов индоевропейской языковой семьи нами были сопоставлены три матрицы сходства: генетических (рассчитаны по полногеномным данным), лингвистических (матрица лингвистических расстояний была предоставлена член-корр. РАН А.В. Дыбо) и географических. Результаты представлены в Таблице 2.

Таблица 2. Результаты теста Мантеля для индоевропейских популяций

Тип корреляции	Параметры корреляции	Значения показателя
Парная корреляция	Генетика и география	0,72*
Парная корреляция	Генетика и лингвистика	0,56*
Частная корреляция	Генетика и география (лингвистика остается постоянной)	0,65*
Частная корреляция	Генетика и лингвистика (география остается постоянной)	0,40**

*- p-value < 0,001; **- p-value = 0,002

Из полученной таблицы видно, что самые высокие значения корреляции обнаружены между генетикой и географией (0,72), но между лингвистикой и генетикой также выявлена достоверная корреляция (0,56), сохраняющаяся и после исключения влияния географии (0,40).

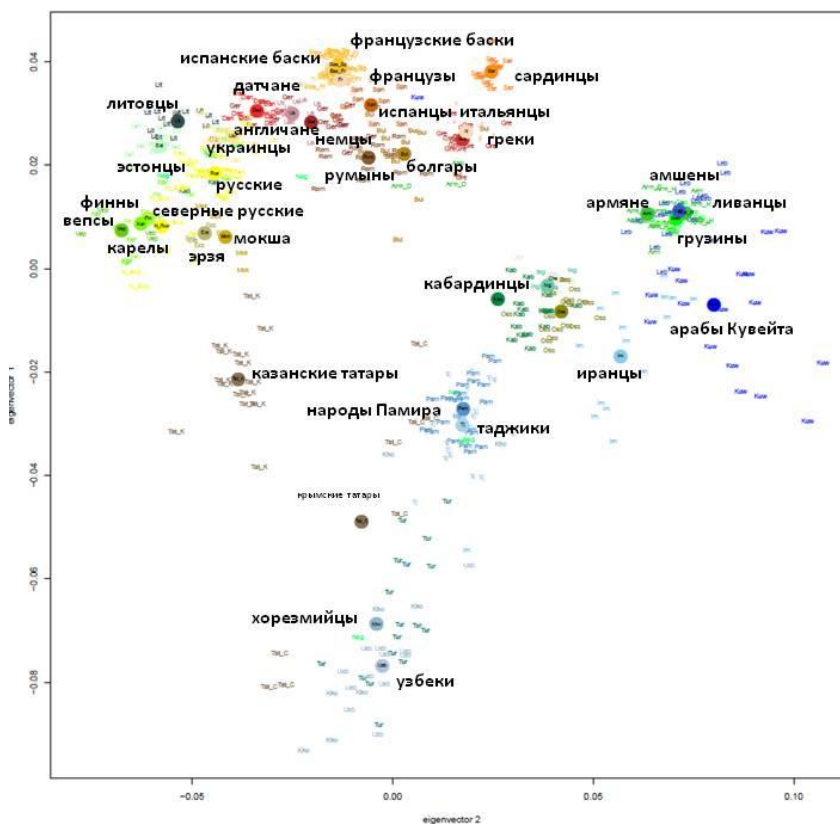


Рисунок 1. График главных компонент популяций Западной Евразии, изученных по полногеномной панели GenoChip.

ПОИСК МАРКЕРОВ, ХАРАКТЕРНЫХ ДЛЯ НАРОДОВ ИНДОЕВРОПЕЙСКОЙ ЯЗЫКОВОЙ ГРУППЫ

Мы исходили из предположения, что современные индоевропейские популяции унаследовали от популяции, говорившей на прото-индоевропейском языке, не только язык, но и небольшую часть генофонда, при этом у их неиндоевропейских соседей генофонд должен быть отличен, либо содержать меньшую часть этого индоевропейского протогенофонда. В связи с чем, мы провели поиск маркеров в популяциях индоевропейцев, используя принцип парного сравнения. Для выявления маркеров использованы три критерия:

Критерий 1. В каждом комплексе сравнения отбирались маркеры, присутствующие у индоевропейцев, но отсутствующие у неиндоевропейцев.

Критерий 2. В каждом комплексе сравнения отбирались маркеры, присутствующие у индоевропейцев с частотой в два раза превышающей частоту у неиндоевропейцев из данного комплекса сравнения.

Критерий 3. Подход, аналогичный предыдущему, только маркер должен был встречаться у индоевропейцев хотя бы немного чаще, чем среди их неиндоевропейских соседей из анализируемого комплекса сравнения.

Для каждого критерия можно требовать, чтобы маркер удовлетворял ему в каждом комплексе сравнения, или только в пяти комплексах и так далее.

Самому строгому критерию (присутствие/отсутствие) для всех комплексов сравнения не соответствовал ни один маркер, а для уменьшенного числа комплексов – лишь считанные маркеры. Критерию значительного преобладания по частоте соответствовали уже 639 маркеров (при условии соответствия критерию хотя бы в пяти комплексах сравнения). Самому мягкому третьему критерию соответствовали 5713 маркеров, при снижении требований к числу комплексов число маркеров стремительно росло (Таблица 3).

Таблица 3. Результаты поиска маркеров, характерных для народов индоевропейской языковой семьи.

Количество комплексов сравнения, удовлетворяющих условию	Присутствие/отсутствие.	Преобладание по частоте в два раза.	Небольшое преобладание по частоте.
Все комплексы сравнения	0	0	648
8 комплексов сравнения	0	3	5713
7 комплексов сравнения	0	22	24356
6 комплексов сравнения	2	104	64816
5 комплексов сравнения	11	639	122715
4 комплекса сравнения	83	2637	181315

Также нами была проведена проверка полученных результатов – три раза популяции были объединены в случайные комплексы сравнения и проведены расчеты, как и при объединении популяций в индоевропейцев и неиндоевропейцев. Результаты всех трех проверок были усреднены. Они получились аналогичными (Таблица 4), как и при «осмысленном» объединении популяций (Таблица 3). Это заставляет с осторожностью относиться к маркерам, которые могли бы быть связанными с распространением носителей индоевропейских языков.

Таблица 4. Результаты проверки поиска «индо-европейских» маркеров (поиск для случайной группы популяций).

Количество комплексов сравнения, удовлетворяющих условию	Присутствие /отсутствие	Небольшое преобладание по частоте	Преобладание по частоте в два раза
Все комплексы сравнения	0	268	0
8 комплексов сравнения	0	3429	3
7 комплексов сравнения	0	18124	19
6 комплексов сравнения	8	56559	169
5 комплексов сравнения	61	118845	976
4 комплекса сравнения	354	183979	3945

То есть наиболее вероятно, что нет таких SNP-маркеров, которые однозначно можно было бы связать с распространением носителей индоевропейских языков. Хотя теоретически и можно предположить, что такие маркеры существуют, но не попали в используемую полногеномную панель, но наши результаты свидетельствуют, что даже в этом случае их доля будет крайне мала.

Все полученные нами данные подводят к выводу, что распространение индоевропейских языков чаще всего шло по модели культурного заимствования, а не демической экспансии. При этом доля исходного индоевропейского компонента могла стать крайне небольшой в случае цепи заимствований: популяция, получившая язык лишь при небольшом генетическом вкладе его носителей, передает этот язык в другую популяцию, в свою очередь, совершая в нее лишь небольшой генетический вклад.

Таким образом, для понимания отдельных этапов расселения индоевропейцев с помощью данных об их генофонде, необходимо сосредоточиться на подробном анализе отдельных лингвистических групп, в пределах индоевропейской семьи.

АНАЛИЗ ГЕНОФОНДА ПОПУЛЯЦИЙ БАЛТО-СЛАВЯНСКОЙ ГРУППЫ ИНДОЕВРОПЕЙСКОЙ ЯЗЫКОВОЙ СЕМЬИ

Для народов балто-славянской лингвистической группы нами проведено обобщающее исследование по маркерам Y-хромосомы с привлечением всего массива накопленных в мировой науке данных по данному региону (проанализировано N=6 079 образцов).

Нами был проведен тест Мантеля для выяснения степени влияния лингвистических и географических расстояний на структуру генофонда популяций балто-славянской лингвистической группы (Таблица 5). Максимальная корреляция нами обнаружена для генетики с географией - 0,924. При этом лингвистический фактор повлиял на структуру генофонда славян гораздо меньше, а частная корреляция с лингвистической недостоверно отличается от нуля.

Данный результат подтверждается и аутосомными данными – также показана максимальная корреляция для генетики с географией [Kushniarevich et al., 2015].

Таким образом, проведенный нами на большом массиве данных анализ балто-славянских популяций четко указывает на то, что распространение славянских языков по обширным территориям не привело к кардинальным изменениям генофонда населения этих территорий. Следовательно, генофонд современных славян был сформирован преимущественно за счет автохтонного субстрата, а влияние древних славян было в большей мере культурным и лингвистическим, чем генетическим.

Таблица 5. Тест Мантеля по связи данных генетики, лингвистики и географии для балто-славянских популяций.

Вид корреляции	Параметры корреляции	Данные по Y-хромосоме
Парная корреляция	Генетика и лингвистика	0.76*
Парная корреляция	Генетика и география	0.924*
Частная корреляция	Генетика и лингвистика (география остается постоянной)	0.137
Частная корреляция	Генетика и география (лингвистика остается постоянной)	0.812*

* статистическая значимость коэффициента корреляции ($p < 0.05$)

Анализ генофонда популяций Верхнего Поволжья

Анализ суммарной выборки (N=132) русского населения Ярославской области в контексте окружающих популяций по маркерам Y-хромосомы позволил получить генетический портрет исследуемого региона. Основные черты генофонда Ярославской области сближают его с генофондом Центральной России и других восточных славян (Рисунок 2). Поэтому можно предполагать, что финно-угорский вклад в генофонд современного населения Ярославского Поволжья был или весьма ограниченным, или же снизился в результате более поздних массовых миграций славянского населения. Лишь в генофонде населения ныне затопленного района долины Мологи был выявлен значительный финно-угорский пласт. Это проявляется, например, в своеобразии карты генетических расстояний от этой популяции: зона генетического сходства (показанная зелеными оттенками) находится в ареале финно-угорских популяций, тогда как на остальных картах зоны сходства располагаются в ареале центральных русских групп (Рисунок 3).



Рисунок 2. Население Ярославской области в контексте окружающих популяций.

График многомерного шкалирования по частотам 16 гаплогрупп.

В целом, анализ генофонда Верхнего Поволжья показывает, что хотя расселение славян мало повлияло на генофонд больших территорий (Восточно-европейская равнина), при анализе меньших регионов генетический эффект славянских миграций начинает проявляться сильнее.

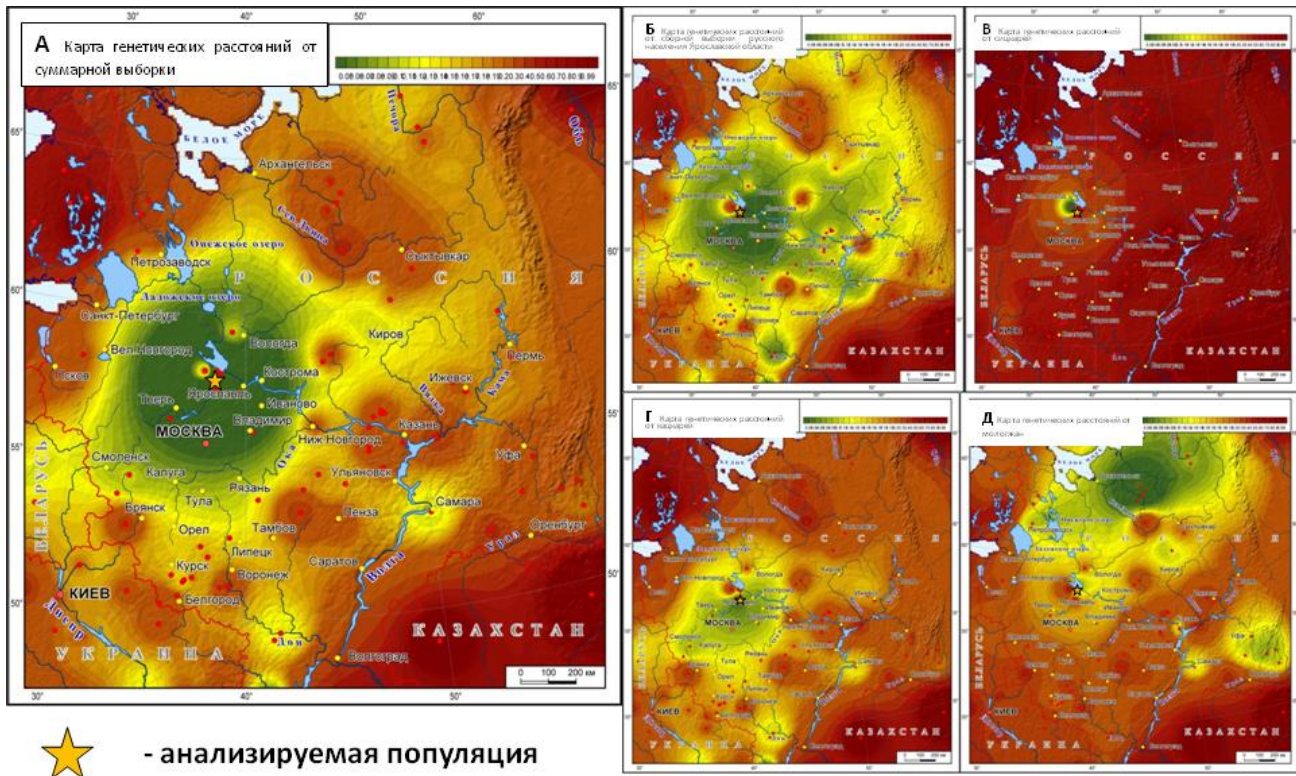


Рисунок 3. Карты генетических расстояний от популяций Ярославской области до окружающих популяций Евразии.

А. Карта генетических расстояний от суммарной выборки.

Б. Карта генетических расстояний от сборной выборки русского населения Ярославской области

В. Карта генетических расстояний от сицкарей.

Г. Карта генетических расстояний от кацкарей.

Д. Карта генетических расстояний от мологжан

Анализ генофонда донских казаков

Исследование 314 образцов донских казаков по STR и SNP маркерам Y-хромосомы показало, что их генофонд сформировался преимущественно за счет восточнославянского компонента. Из степных тюркоязычных популяций заметное влияние оказали только ногайцы, а влияние народов Кавказа не прослеживается. Эти генетические результаты находятся в соответствии с миграционной теорией происхождения казачества, хотя не отрицают и возможность ограниченного влияния степных популяций в лице ногайцев. А поскольку предшествующее казакам население этого ареала было тюркоязычным, формирование казачества является примером кардинального изменения генофонда на определенной территории в результате славянских миграций.

КОМПЛЕКСНЫЙ АНАЛИЗ ГЕНОФОНДА ПОПУЛЯЦИЙ АРМЯНСКОЙ ГРУППЫ ИНДОЕВРОПЕЙСКОЙ ЯЗЫКОВОЙ СЕМЬИ

Нами было проведено комплексное исследование генофонда армян. В итоговый анализ были включены собственные данные по 5 популяциям армян, N=437. А так же данные по 10 популяциям армян из работ Herrera et al. 2012; Grugni et al., 2012; Novhannisyan et. al 2014, N=1204.

Нами показано, что в целом для генофонда армян характерно большое разнообразие гаплогрупп и отсутствие какой-либо одной доминантной гаплогруппы, в отличие от народов Северного Кавказа, где в каждом регионе решительно преобладает какая-либо одна гаплогруппа, определяющая региональный генетический портрет [Balanovsky et al, 2011; Yunusbaev et al., 2012]. Наиболее распространены в генофонде армян гаплогруппы R-L23, J-M67 и G-M285. В суммарной выборке армян эти три гаплогруппы представлены примерно в равных долях (10-20%).

Предположение о том, что армяне в диаспоре сохранили черты исходного протогенофонда исторической Армении, подтверждает проведенный нами многомерный анализ: популяции армян объединяются в общий кластер на графике вне зависимости от того, представляют ли они армян диаспоры либо самой Армении. Сохранение генофонда в диаспоре наглядно видно и по результатам картографического анализа. Он выявляет совпадение современного ареала популяций, наиболее близких к средне-армянскому генофонду, с ареалом исторической Армении. Генетического сходства с современными соседями армян в диаспоре картографический анализ не выявил.

Нами был проведен филогенетический анализ наиболее распространенных среди армян гаплогрупп J-M67 и R-L23. В результате анализа обнаружено существование специфических армянских кластеров. Для гаплогруппы R-L23 «генеалогическая» датировка кластера равна 1900 ± 500 лет. Для гаплогруппы J-M67 датировка кластера равна 2000 ± 500 лет.

Примечательно, что в «армянские» кластеры входят представители всех изученных нами популяций армян. Это подтверждает сложение основных черт генофонда армян в период их формирования как единого этноса.

Для анализа положения популяций армян в широком географическом масштабе нами был привлечен большой объем литературных данных. Сравнение с ближайшими географическими соседями выявило генетические отличия генофонда армян от Северного Кавказа и сходство с другими группами Закавказья и Юго-Западной Азии. Исследование армян в контексте всей Юго-Западной Азии (Рисунок 4) выявило генетическую дифференциацию регионального

генофонда на два основных кластера популяций: «равнинный» и «горный» кластеры. «Горный» кластер включил в себя исключительно население Переднеазиатского нагорья (и, соответственно, входящего в него Армянского нагорья): основной массив армянских популяций, объединившихся с популяциями западного Ирана и турками. В равнинный кластер вошли популяции Леванта.

Результаты анализа иерархической межпопуляционной изменчивости (AMOVA) подтверждают эту дифференциацию. Таким образом, нами впервые показана своеобразная «генетическая граница» подразделяющая генофонд Юго-Западной Азии на две части, приуроченные к разному типу рельефа – горному и равнинному. Корреляция генофонда региона с лингвистическим делением лишь немного уступает корреляции с географическим делением на нагорья и равнины. Это связано с тем, что популяции армянской, иранской и тюркской лингвистических групп населяют главным образом нагорья Передней Азии, а популяции семитской группы – равнины.

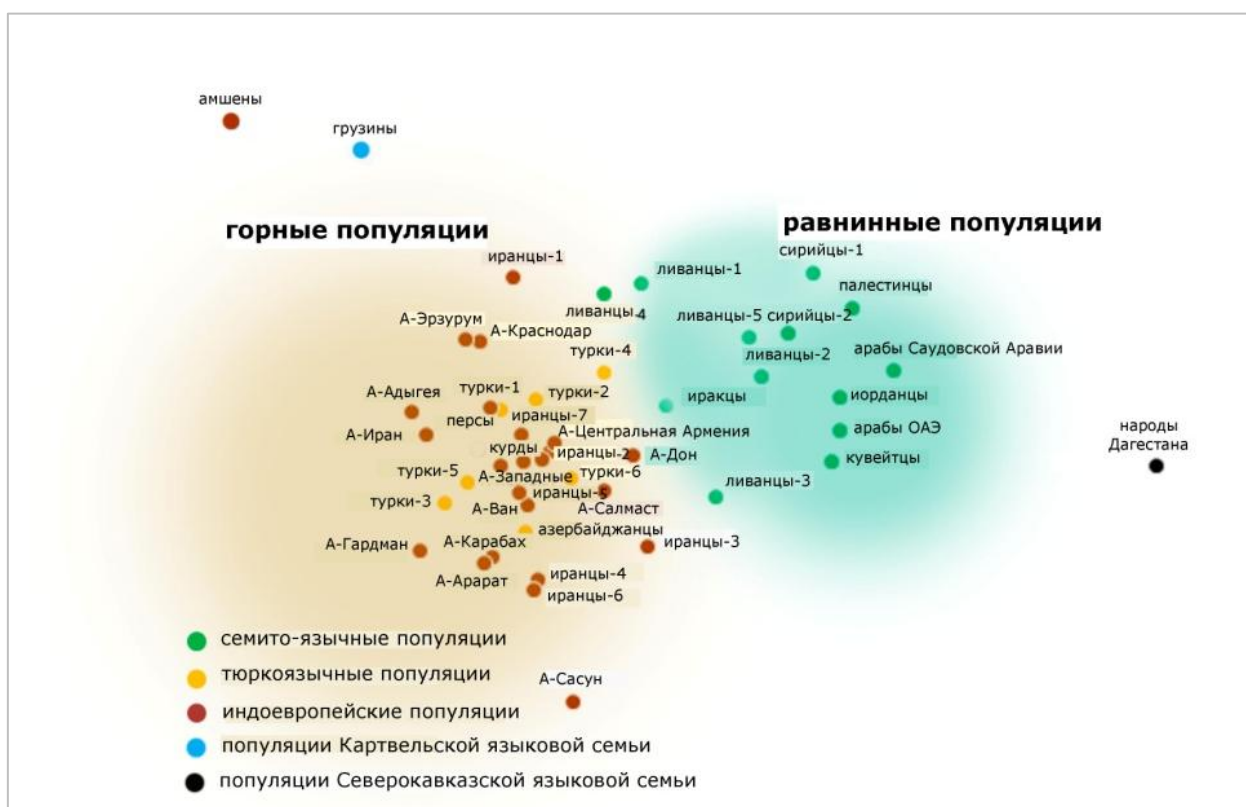


Рисунок 4. График многомерного шкалирования: армяне в контексте Юго-Западной Азии.

ВЫВОДЫ

1. Анализ девяти популяций, говорящих на языках индоевропейской семьи, в сравнении с их не-индоевропейскими соседями по полногеномной панели GenoChip не выявил маркеров, надежно отличающих индоевропейские популяции, что указывает на распространение индоевропейских языков по модели культурного заимствования, а не демической экспансии.

2. Генетические взаимоотношения популяций Западной Евразии лучше объясняются сочетанием географического и лингвистического факторов, чем одним лишь географическим; генетическая структура индоевропейцев тоже определяется прежде всего географическим соседством популяций (корреляция 0,72), хотя при исключении этого фактора выявлена и связь с лингвистическим родством популяций (корреляция 0,40).

3. Распространение славянских языков не изменило значительно Y-хромосомный генофонд на всей территории их исторического расселения (частная корреляция генетических и лингвистических расстояний между славянскими популяциями нулевая, при этом корреляция генетических и географических расстояний составляет 0,8). Но на формирование генофондов регионального масштаба - Верхнего Поволжья и донских казаков - миграция славян оказала решающее влияние.

4. Генофонд армян по маркерам Y-хромосомы близок к переднеазиатским популяциям, но генетически удален от генофондов Северного Кавказа.

5. Большинство популяций армян диаспоры генетически сходны с популяциями на исходной территории расселения, а не со своими географическими соседями, указывая на возможность сохранения генофонда в диаспоре.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Kushniarevich A., Utevska O., **Chuhryaeva M.**, Agdzhoyan A., Dibirova K., Uktveryte I., Möls M., Mulahasanovic L., Pshenichnov A., Frolova S., Shanko A., Metspalu E., Reidla M., Tambets K., Tamm E., Koshel S., Zaporozhchenko V., Atramentova L., Kučinskas V., Davydenko O., Tegako L., Evseeva I., Churnosov M., Pocheshchova E., Yunusbaev B., Khusnutdinova E., Marjanović D., Rudan P., Rootsi S., Yankovsky N., Endicott Ph., Kassian A., Dybo A., Tyler-Smith C., Balanovska E., Metspalu M., Kivisild T., Villems R., Balanovsky O., The Genographic Consortium. Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data // *PLoS One*. 2015. V. 10(9):e0122968. doi 10.1371/journal.pone.0135820.
2. Balanovsky O., **Chukhryaeva M.**, Zaporozhchenko V., Urasin V., Zhabagin M., Hovhannisyan A., Agdzhoyan A., Koshel S., Pocheshkhova E., Alborova I., Shalyakho R., Utevska O., The Genographic Consortium, Mustafin Kh., Episkoposyan L., Tyler-Smith C., Balanovska E. Geographical Relief and Prehistoric Migrations Shape the Y-Chromosomal Landscape of West Asian Populations // *Human Genetics*. 2017. 136:437-450. DOI 10.1007/s00439-017-1770-2.
3. **Чухряева М.И.**, Иванов И.О., Фролова С.А., Кошель С.М., Утевская О.М., Схляхо Р.А., Агджоян А.Т., Богунов Ю.В., Балановская Е.В., Балановский О.П. Программа HAPLOMATCH для сравнения STR-гаплотипов Y-хромосомы и её применение к вопросу происхождения донских казаков // *Генетика*. 2016. Т.52. №5. С.595-604.
4. **Чухряева М.И.**, Павлова Е.С., Напольских В.В., Гарин Э.В., Темняткин С.Н., Клопов А.В., Запорожченко В.В., Романов А.Г., Агджоян А.Т., Утевская О.М., Маркина Н.В., Кошель С.М., Балановский О.П., Балановская Е.В. Сохранились ли следы финно-угорского влияния в генофонде русского населения Ярославской области? Свидетельства Y-хромосомы // *Генетика*. 2017. № 3. С 388-399.
5. Утевская О.М., **Чухряева М.И.**, Агджоян А.Т., Атраментова Л.А., Балановская Е.В., Балановский О.П. Популяции Закарпатья и Буковины на генетическом фоне окружающих территорий // *Вісник Дніпропетровського університету. Біологія, медицина*. 2015. 6(2) С. 133–140.
6. **Чухряева М.И.**, Дибирова Х.Д., Епископосян Л.М., Балановский О.П., Балановская Е.В. Генофонд донских армян: столетия в диаспоре // *Армяне Юга России: история, культура, общее будущее*. 26–28 мая 2015 г. Ростов-на-Дону. С. 81-84.
7. **Chukhryaeva M.**, Lukyanova E., Quintana-Murci L., Comas D., Balanovska E., Balanovsky O. Applying the GenoChip genotyping array for tracing Indo-European expansion // *Abstracts of Papers European Human Genetics Conference (Glasgow, Scotland, United Kingdom, 2015, June 6 – 9)*. 2015. Vol. 23(Suppl. 1). P. 473.
8. **Chukhryaeva M.**, Dibirova Kh., Teuchezh I., Kuznetsova M., Agdzhoyan A., Yepiskoposyan L., Pocheshkhova E., Balanovsky O. Gene pool of the Southwest Asia: reflection of geographical relief and linguistic stratification // *Материалы конференции The 19th Congress of the European Anthropological Association “Anthropology: Unity in Diversity”*. August 2014. Moscow. *Вестник Московского университета. Серия XXIII "Антропология"*. 2014. №3. С. 116.
9. **Чухряева М.И.**, Дибирова Х.Д., Теучеж И.Э., Кузнецова М.А., Епископосян Л.М., Почешхова Э.А., Балановский О.П. Генофонд носителей армянских и индоиранских языков в контексте индоевропейской проблематики // *Сборник «Языковая политика и языковые конфликты в современном мире. Доклады и сообщения»*. М.: Тезаурус. Языки Народов Мира. 2014. С. 579-584.
10. **Чухряева М.И.**, Альборова И.Э., Дибирова Х.Д., Схляхо Р.А., Кагазежева Ж.А., Романов А.Г., Епископосян Л.М., Балановская Е.В. Исследование генофонда народов Передней Азии по маркерам Y-хромосомы: мост из настоящего в прошлое // *"50 лет ВОГиС: успехи и перспективы"*, Москва 8-10 ноября 2016. С.327.
11. **Чухряева М.И.**, Агджоян А.Т., Схляхо Р.А., Запорожченко В.В., Степанов Г., Кузнецова М.А., Почешхова Э.А., Урасин В., Епископосян Л.М., Мустафин Х.Х.,

- Балановская Е.В., Балановский О.П. Применение данных полногеномного секвенирования Y-хромосомы к изучению генофонда Передней Азии в контексте индоевропейской проблематики // International Scientific Conference on "Anthropology and Ethnology of Caucasus". – 2016. – Tbilisi, (October, 24 – 27). С. 211-216.
12. **Чухряева М.И.**, Теучеж И.Э., Дибирова Х.Д., Почешхова Э.А., Балановский О.П. Анализ генофонда армян как представителей индоевропейской языковой семьи в контексте окружающих популяций Кавказа // Конференция «Проблемы генетики населения и этнической антропологии», памяти Ю.Г. Рычкова. 19-21 ноября 2013. С. 25.
 13. **Чухряева М.И.**, Теучеж И.Э., Дибирова Х.Д., Епископосян Л.М., Почешхова Э.А., Балановский О.П. Сравнительное изучение генофонда армян и соседних народов в контексте проблемы «прародины индоевропейцев» (по данным о полиморфизме Y-хромосомы) // VI Съезд Вавиловского Общества Генетиков и Селекционеров. Ростов-на-Дону. 15-20 июня 2014. С. 25.
 14. **Чухряева М.И.**, Лукьянова Е., Quintano-Murci L., Балановский О.П. Генофонд популяций индоевропейской лингвистической семьи: полногеномный анализ на основе использования биобанков // Материалы VII Съезда Российского общества медицинских генетиков. 9-23 мая 2015. Санкт-Петербург. Медицинская генетика. 2015. Т. 14. № 2(152). С. 40.
 15. **Чухряева М.И.**, Дибирова Х.Д., Теучеж И.Э., Агджоян А.Т., Почешхова Э.А., Епископосян Л.М., Балановский О.П. Геногеография армян диаспоры: донских, амшен, Адыгеи и Краснодарского края (по данным о полиморфизме Y-хромосомы) // Международная конференция «Алексеевские чтения» памяти академиков Алексеевой Т.И. и Алексеева В.П. «Человек в окружающей среде: этапы взаимодействия». 6-8 ноября 2013 г. Москва. С. 107.
 16. Утевская О.М., Дибирова Х.Д., Чухряева М.И., Агджоян А.Т., Атраментова Л.А., Балановская Е.В., Балановский О.П. Западноевропейские и западноазиатские варианты гаплогруппы R1b в украинских популяциях // Факторы экспериментальной эволюции организмов. Сборник научных трудов. Киев. 2015. Логос. Т.16. С. 255-259.
 17. **Чухряева М.И.**, Агджоян А.Т., Схалыхо Р.А. Создание программы HAPLOMATCH: выявление сходных гаплотипов Y-хромосомы и происхождение донских казаков // Материалы конференции молодых ученых ФГБНУ «МГНЦ». 2015. Москва. С. 12.
 18. Епископосян Л.М., **Чухряева М.И.**, Дибирова Х.Д., Оганесян А.А., Теучеж И.Э., Агджоян А.Т., Грештакян П.А., Балановский О.П. Из Армении на Дон: генетическая летопись донских армян (по данным о полиморфизме Y-хромосомы) // Сборник трудов Международная научная конференция "Донские антропологические чтения «Население юга России с древнейших времен до наших дней»" 26–30 августа 2013 г., Ростов-на-Дону, Россия. стр. 119.