

Федеральное государственное бюджетное учреждение науки

Институт общей генетики им.Н.И. Вавилова

Российской академии наук

*На правах рукописи*

Чухряева Марина Игоревна

**Сравнительный анализ популяций индоевропейской лингвистической  
семьи и генофондов других лингвистических групп в зонах их контакта**

03.02.07 Генетика

Диссертация

на соискание ученой степени  
кандидата биологических наук

Научный руководитель:  
доктор биологических наук  
Балановский О.П.

Москва - 2018

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ.....	12
1.1. Основные гипотезы локализации прародины индоевропейской лингвистической общности...12	
1.1.1. Степная гипотеза.....	17
1.1.2. Анатолийская гипотеза.....	18
1.1.3. Альтернативные гипотезы.....	20
1.2. Генетические маркеры в популяционно-генетических исследованиях различных лингвистических общностей.....	22
1.2.1. Маркеры Y-хромосомы.....	22
1.2.2. Аутосомные маркеры.....	26
1.2.3. Генетические маркеры разных типов в решении индоевропейской проблематики ...28	
1.3. Региональные исследования представителей индоевропейской языковой семьи.....	39
1.3.1 Изученность популяций балто-славянской лингвистической группы индоевропейской языковой семьи по разным типам генетических маркеров.....	42
1.3.2 Изученность популяций армянской лингвистической группы индоевропейской языковой семьи по разным типам генетических маркеров.....	47
ГЛАВА 2. Материалы и методы.....	54
2.1. Выбор объектов изучения и стратегия формирования выборок.....	54
2.2. Методы отбора образцов и проведения полногеномного генотипирования.....	65
2.3. Методы генотипирования полиморфизма Y-хромосомы.....	66
2.4. Общий статистический анализ полногеномных данных.....	67
2.5. Поиск маркеров, характерных для народов индоевропейской языковой общности.....	70
2.6. Статистический, филогеографический и картографический анализ данных полиморфизма Y-хромосомы.....	72
ГЛАВА 3. Положение генофондов народов индоевропейской языковой общности по полногеномным данным в контексте Евразии: поиск сходства и различий.....	76
3.1. Положение генофондов народов индоевропейской языковой общности в контексте современных популяций Евразии.....	76
3.2. Сравнение современных генофондов народов индоевропейской языковой общности с древним населением Евразии.....	81
3.3. Тестирование основных гипотез локализации индоевропейской прародины на соответствие полученным генетическим данным.....	88
3.4. Поиск маркеров, характерных для народов индоевропейской языковой общности по полногеномным данным.....	93
ГЛАВА 4. Генофонд народов балто-славянской языковой группы по маркерам Y-хромосомы....	103
4.1. Общая характеристика генофонда балто-славянских популяций.....	103
4.2. Генофонд населения Верхнего Поволжья.....	110
4.3. Генофонд донских казаков.....	116
ГЛАВА 5. Генофонд народов армянской языковой группы.....	124
5.1. Генетические взаимоотношения между группами армян.....	124
5.2. Филогенетический анализ гаплогрупп, наиболее распространенных среди армян.....	131
5.3. Генетический портрет армян в контексте окружающих популяций.....	138
Заключение.....	153
ВЫВОДЫ.....	155
Список литературы.....	156

## **ВВЕДЕНИЕ**

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 [Fedorova et al., 2013; Yunusbaev et al., 2015] и многие другие работы, 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 [Haak et al., 2015; Allentoft et al., 2015; 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 [Haak 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 испытавшие 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; Paganì et al., 2017], тогда как другие авторы обнаруживали четкую согласованность генетических и лингвистических границ [Kayzer et al., 2005].

Экспансия индоевропейцев – это не только события неолита 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; Mitnik 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 древе 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; Novhannisyan 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.**

В данном исследовании 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. Проведено полногеномное сканирование с целью проверки возможности выявления тех SNP, которые могут рассматриваться как маркирующие протогенофонд индоевропейцев и пути их миграций. Впервые на основе данных полногеномного генотипирования дана количественная оценка степени согласованности изменчивости индоевропейских популяций по их генофонду, их языковой близости и географическим расстояниям между ними. Также 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, выявлено структурирование генофонда Юго-Западной Азии на горные и

равнинные популяции и показано сохранение генофонда армян в диаспоре (по маркерам Y-хромосомы).

### **Практическая значимость.**

Научно-практическая значимость проведенного исследования состоит в следующем: во-первых, 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.

Во-вторых, созданная база данных по распределению панели 17 STR маркеров Y-хромосомы (коммерческий набор Y-filer, Applied Biosystems) 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.

В-третьих, 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.



### **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.

### **Степень достоверности и апробация результатов.**

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 в исследование.**

Автором лично 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 3130xl, 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 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.

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.

## **ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ**

### **1.1. Основные гипотезы локализации прародины индоевропейской лингвистической общности**

«Индоевропейская проблема», то есть вопрос о месте прародины индоевропейской языковой семьи (крупнейшей в мире по ареалу распространения и числу носителей в мире) и о ранних миграциях индоевропейцев, является одной из самых старых, сложных, и интенсивно разрабатывающихся проблем в междисциплинарной сфере изучения популяций человека. В XIX и XX веках в разработку этой проблемы основной вклад внесли лингвисты и археологи, выдвинув ряд конкурирующих гипотез о локализации индоевропейской прародины и путях миграций индоевропейцев, а в XXI веке к ним присоединились и генетики.

Индоевропейские языки входят в ностратическую макросемью языков. А собственно сам термин индоевропейские языки был впервые предложен английским исследователем Томасом Юнгом в 1813 году [Языки народов СССР, 1966]. Полная классификация индоевропейских языков заняла бы слишком много места, поэтому приведем здесь только основные ветви, в соответствии с классификацией Старостина [Ярцева В.Н. , 1990]:

1. славянскую;
2. германскую (включающую кроме английского, немецкого и скандинавских вымерший готский язык);
3. италийскую, которая в настоящее время представлена только производными от латинского языками: итальянским, испанским, французским, румынским;
4. индийскую;
5. иранскую, включающую персидский язык - фарси, курдский, балучи, пушту - язык афганцев, осетинский. В древности эта группа была одной

из самых распространенных во всем индоевропейском мире и кроме староперсидского (пехлеви) и языков других существовавших тогда древнеиранских народностей (парфян, согдийцев) включала еще и языки скифов, сарматов и саков, расширяя зону использования этих иранских диалектов и наречий до Причерноморья, Южной Сибири, Казахстана и Алтая;

6. несколько компактных обособленных групп носителей греческого, албанского и армянского языков;

7. сильно уменьшившиеся в настоящее время по числу носителей, ареалу и разнообразию кельтские и балтские языки;

8. а также еще целый ряд мертвых и не существующих уже языковых объединений, которые в прошлом играли заметную роль в отдельных регионах Евразии (тохарское, и хетто-лувийские), включая и не показанные здесь ветви, связанные с дако-фракийским, фригийским, иллирийским и некоторыми другими языками.

Поиск индоевропейской прародины ведется учеными уже около 150 лет. Данная научная проблема волнует специалистов различных специальностей: историков, лингвистов, археологов, антропологов, а в последние 10-20 лет и генетиков. За все это время было выдвинуто много гипотез локализации индоевропейской прародины, подкрепленных многочисленными доказательствами. Некоторые из этих гипотез имеют среди своих сторонников ученых с очень серьезной репутацией, поддерживаются большим количеством людей. Но, несмотря на это, до сих пор нет единой гипотезы, которая могла бы считаться окончательно доказанной. Как написал в своем труде Мэллори, один из известных ученых, работающих в этой области: «Нет сомнений в том, что прародина была в конце концов обнаружена, ведь ее искали повсюду – от Северного полюса до Южного и от Атлантического океана до Тихого. С точки зрения

хронологии прародину помещали как угодно – начиная с 100 000 г. до н.э. и до 1600 г. до н. э., приблизительного времени распространения боевых колесниц из Восточной Анатолии» [Мэллори, 1997].

Следует указать, что понимается под прародиной лингвистической общности. Общепринято считать, что в какое-то время существовал индоевропейский праязык (или праязыки) с диалектами, значительно более сходными между собой, чем различные группы индоевропейских языков периода их самой ранней письменной фиксации в эпоху поздней бронзы или в железном веке. Индоевропейская прародина – это территория распространения индоевропейского праязыка до его разделения на различные группы [Мэллори, 1997].

Среди всех теорий, ставящих своей целью найти индоевропейскую прародину, наряду с целым рядом неакадемических гипотез (таких как, теория исхода из Индии), существует несколько, имеющих большое количество доводов в свою пользу и серьезно рассматриваемых в научном сообществе гипотез. Это курганная, анатолийская, армянская и балканская гипотезы. Причем в последние годы наибольшее количество сторонников имеют первые две. Далее рассмотрим эти гипотезы несколько подробнее. Но прежде чем перейти к их подробному рассмотрению, следует остановиться на доводах и приемах лингвистов, используемых для доказательств выдвигаемых ими гипотез. Лингвисты по-прежнему резко расходятся относительно времени и места распада индоевропейской общности, а разные технические приемы, используемые ими, приводят доводы в пользу различных гипотез.

Одним из таких общепризнанных приемов является принцип «центра тяжести». Считается, что центр языковой дисперсии располагается в ареале, где отмечается наибольшее языковое разнообразие, тогда как периферийные области отличаются максимальной однородностью. Для

индоевропейских языков такая территория находится в Юго-Восточной Европе, на Балканах [Mallory J. P. 1989].

Другим надежным приемом считается принцип консерватизма. Считается, что центр языковой дисперсии располагался там, где отмечено наименьшее число языковых изменений по сравнению с праязыком, так как этот ареал менее других должен был испытывать воздействие иноязычных субстратов. К этому принципу чаще всего обращаются, желая поместить индоевропейскую прародину в Прибалтику на основании консерватизма литовского языка [Bender H., 1922]. Однако очевидно, что по другим критериям Прибалтика менее всего годится на роль прародины индоевропейцев

Еще одним приемом является анализ контактов между языковыми семьями [Бурлак, Старостин, 2001]. Согласно этому приему, считается, что прародина располагается в непосредственной близости от языковой семьи, в которой обнаружено наибольшее количество схождений с анализируемой (индоевропейской) семьей. В качестве такой языковой семьи, контактировавшей с индоевропейской, предполагались: семитская (при этом индоевропейская прародина помещалась в Анатолии/на Ближнем Востоке), картвельская (индоевропейская прародина – в Восточной Анатолии), северо-западнокавказская (индоевропейская прародина – в Северном Причерноморье), уральская (индоевропейская прародина локализовалась на территории между Северной/Центральной Европой и Уралом или даже еще восточнее) [Dolgopolsky, 1988]. Таким образом, использование подобного подхода также не дает однозначного ответа на вопрос локализации прародины.

В поисках локализации индоевропейской прародины могут помочь точные оценки времен образования праязыка индоевропейцев. Однако датировки, полученные с помощью различных методов, довольно существенно различаются. Так, применительно к проблеме индоевропейцев,

наиболее популярными являются датировки на основании лингвистических критериев, датировки по внешним контактам, глоттохронология и теоретическая оценка.

Как правило, лингвисты датируют начало расхождения индоевропейских языковых групп периодом около 5000—2500 гг. до н.э. Реконструируемый индоевропейский словарь отражает тип хозяйства, основанный на использовании культурных растений (зерновые), одомашненных животных (крупный рогатый скот, овцы, козы, свиньи, собаки) и соответствующей технологии (жернов, серп), и свидетельствует о том, что расхождение индоевропейских диалектных групп не могло ни в одном из предполагаемых ареалов начаться ранее VII тыс. до н.э., когда, как известно из археологии, появился такой тип хозяйства; насколько позднее это происходило, зависит от географической локализации прародины. Этот словарь содержит также ряд терминов, например, плуг, ярмо, колесный транспорт, шерсть, серебро, которые археологически засвидетельствованы не ранее 5000–3000 гг. до н.э.

При этом анализ культурной лексики дает широкие границы датировок индоевропейского праязыка – около 7000–2500 гг. до н. э. Наиболее поздние общеиндоевропейские термины позволяют отнести расхождение индоевропейских диалектных групп к 4000–3000 гг. до н. э., но это не обязательно отражает самые ранние индоевропейские миграции.

Таким образом, попытки датировать индоевропейский праязык дают широкий хронологический разброс от 7000 г. до н. э. (времени возникновения неолитического хозяйства, отраженного в реконструированном словаре) и до 2000 г. до н. э., когда появляются первые исторические свидетельства об индоевропейцах в Анатолии.



### 1.1.1. Степная гипотеза

Была предложена Гимбутас в 1965 году [Алексаха А.Г., 2012; Gimbutas M., 1977]. В основе курганной гипотезы лежат взгляды, высказанные ещё в конце XIX века Виктором Геном и Отто Шрадером. Она связывает индоевропейцев с ямной культурой и выдвигает в качестве их прародины причерноморские степи, откуда в дальнейшем и мигрировали различные ветви индоевропейцев. Временем образования индоевропейского языка указывает промежуток энеолит, около 3000 - 4500 тыс. лет до н.э. [Gimbutas M., 1977]. Исходя, из этой гипотезы, дольше всех изначальный ареал занимали предки славян и балтов. Предлагаемая локализация прародины устраивает тех, кто считает, что наиболее тесные внешние связи у индоевропейского праязыка обнаруживаются с уральским и северокавказским. В степях и лесостепях причерноморско-прикаспийского ареала возникли обширные культурные комплексы, которые могут рассматриваться как сферы взаимодействия в пределах единой языковой семьи.

Расширению ареала первых индоевропейцев способствовало одомашнивание лошади и позднее использование повозок. Археологически доказывается движение из этого региона в западном направлении – на Балканы и в восточном – в Казахстан. Стимулом к началу миграций могла стать очень мобильная экономическая стратегия при наличии социальной организации, способной ассимилировать и подчинить себе различные этноязыковые группы.

Данная модель объясняет локализацию (в историческое время) всех индоевропейских языков, как в Европе, так и в Азии. Данный ареал в доисторические времена не был заселен индоевропейцами. Датировка индоевропейской прародины согласно этой гипотезе не противоречит общественно известным фактам. Кроме того, археологические данные убедительно свидетельствуют о притоке населения из степных областей на

запад – до р. Тисы и на Балканы. Однако все последующие миграции на запад или на север, например, в кельтском, германском, балтийском и, возможно, славянском исторических ареалах, основаны на данных, которые не укладываются в общую картину и потому часто не принимаются в расчет. Например, распространение одомашненной лошади, колесного транспорта, оборонительной архитектуры, курганных погребений, боевых топоров, захоронений животных и т.д. Так же ненадежны свидетельства миграции на юг, в Анатолию и в Иран [Mallory J. P., 1995].

### **1.1.2. Анатолийская гипотеза**

Её сформулировал археолог Renfrew в 1987 г. Он предполагал, что индоевропейский язык существовал ещё в неолите, 7-6 тыс. лет до н.э. в Западной Анатолии (современная Турция), и только 6 тыс. лет до н.э. индоевропейцы пересекли Босфор и распространились по Юго-Восточной Европе [Renfrew, 1987]. Эта теория предполагает значительные миграции индоевропейского населения, включая «передовые отряды», благодаря которым носители индоевропейских языков распространились на обширных территориях Европы из ареала (Анатолии), в котором индоевропейцы известны с эпохи бронзы. Механизм этих миграций соотносят с распространением нового хозяйственного уклада, основанного на использовании домашних растений и животных. Это должно было привести к росту численности населения. В результате демографического взрыва массы населения устремились из Анатолии в Европу и Азию, при этом мигранты-земледельцы были носителями более высокой технологии, что, как полагают, позволило им ассимилировать местное население - охотников-собираателей. Эта гипотеза объединяет народы Балкан и Анатолии, между которыми в эпоху неолита отмечается сходство физического типа и культур, что могло бы свидетельствовать в пользу

общего языкового прошлого. Локализация прародины по этой гипотезе хорошо согласуется с данными внешних связей индоевропейской языковой семьи с семитской и картвельской [Сафронов, 1989].

В качестве критики теории выдвигаются положения о небольшом разнообразии индоевропейских языков на территории Анатолии, и, если принимать разделение индоевропейских языков столь древним, то становится необъяснимым факт существования в них хронологически информативных общеиндоевропейских терминов. Например, слова «колесо», «лошадь» не подтверждаются столь ранним археологическим материалом, однако представлены среди унаследованной лексики в анатолийских языках. [Кобычев В. Где находится родина индоевропейцев? - Scientific.ru]. Неолитические миграции, постулируемые, например, Renfrew, нарушают картину взаимосвязей индоевропейских языков: предполагаемое переселение народов не всегда согласуется с изоглоссами между индоевропейскими языками. Эта гипотеза хорошо согласуется с данными археологии для европейских народов, но сталкивается с серьезными проблемами, когда речь заходит об индоевропейском населении Азии.

В поддержку Анатолийской гипотезы выступили авторы одной из наиболее цитируемых работ в данном направлении, опубликованной в Nature - R. D. Gray and Q. D. Atkinson [Gray, Atkinson, 2003]. Ученые, опираясь на байесовский метод, провели статистический анализ 87 индоевропейских, как живых, так и мертвых языков, используя усовершенствованную базу Изидора Дайена. По результатам подсчетов было сгенерировано множество филогенетических деревьев. Из них было произвольно отобрано 10 000 и их уже и проверяли на соответствие реальным условиям. В результате был установлен возраст индоевропейского языка около 8700 лет, что было использовано авторами в качестве подтверждения анатолийской гипотезы [Gray, Atkinson, 2003]. Однако нашлось и немало критиков этой работы. Так подвергается

сомнению правомерность диагноза заимствования при удалении заимствованных слов, а также то, что близость схемы R. D. Gray и Q. D. Atkinson датировкам Renfrew может только косвенно подтвердить анатолийскую прародину индоевропейцев, но прямо об этом ничего не говорит [Клейн, 2007].

Интересно отметить, что в своей работе за 2011 г, R. D. Gray и Q. D. Atkinson [Gray, Atkinson, 2003] уже признают, время разделения большинства индоевропейских языков 6-7 тыс. лет до н.э., и предполагают, что обе гипотезы: курганная и анатолийская верны. Т.е. индоевропейский язык зародился 8700 лет назад на территории Анатолии, и затем отделение основных его групп произошло уже на территории современных России и Украины.

### **1.1.3. Альтернативные гипотезы**

Одной из альтернативных гипотез, заслуживающих внимания, является армянская гипотеза. Гипотеза о прародине индоевропейцев на территории Армянского нагорья и в степях Восточной Европы была сформулирована Миллером еще в 1873 году на основании близости индоевропейского праязыка с семито-хамитскими и кавказскими языками. Дальнейшее развитие эта гипотеза получила в трудах академиков Гамкрелидзе и В.В. Иванова (1980-1981 гг). Согласно гипотезе Т. В. Гамкрелидзе и В. В. Иванова ранее всего из индоевропейской общности выделился анатолийский праязык. Произошло это не позднее IV тыс. до н. э. где-то в районе Армянского нагорья. Отсюда предки анатолийцев начали движение на запад. Дальнейшему разделению индоевропейской общности предшествовало ее членение на группы диалектов. Армяне же остались единственным народом, который впоследствии вновь заселил Армянское нагорье – территорию гипотетической прародины.

Лингвистическая аргументация данной гипотезы основана на строгом использовании сравнительно-исторического метода и основных положений теории языковых заимствований, хотя и вызывает возражения оппонентов по некоторым частным вопросам. Очень важно подчеркнуть, что индоевропейские миграции рассматриваются согласно этой концепции не как тотальная этническая «экспансия», но как движение в первую очередь самих индоевропейских диалектов вместе с определенной частью населения, наслаивающегося на различные этносы и передающего им свой язык. Последнее положение методологически очень важно, так как показывает несостоятельность гипотез, опирающихся в первую очередь на антропологические критерии при этнолингвистической атрибуции археологических культур [Гамкрелидзе, Иванов, 1984).

Отдельного внимания стоит и гипотеза, помещающая прародину индоевропейцев в Центральную Европу – на Балканы. Она также позволяет предположить существование достаточно близко расположенной зоны контактов индоевропейских языков с уральскими или северокавказскими, постулируемой некоторыми лингвистическими моделями. С точки зрения археологии - это область культуры линейно-ленточной керамики, распространенной в обширном ареале Европы от атлантического побережья до Украины и демонстрирующей удивительную однородность. Эта общность могла бы свидетельствовать (по крайней мере, так считают некоторые исследователи) о наличии языковой общности [Idem, 1992). Слабость рассматриваемой модели заключается в том, что она не в состоянии описать индоевропейские языки Азии и вообще неудовлетворительно отражает продвижение индоевропейских народов на восток от Днепра.

Таким образом, лингвистические и археологические свидетельства не могут быть однозначно интерпретированы в пользу одной из гипотез

локализации прародины индоевропейских народов – в вопросах расселения этого населения до сих пор остается много нерешенных вопросов.

## **1.2. Генетические маркеры в популяционно-генетических исследованиях различных лингвистических общностей**

### **1.2.1. Маркеры Y-хромосомы**

Y-хромосома – самая маленькая в геноме человека (она занимает лишь около 1.6 % гаплоидного генома) и несет достаточно немного функционально активных генов [Lahn et al., 2001]. Она имеет размер около 60 Мб, 40 из которых приходится на гетерохроматин. Гетерохроматиновый регион подразделяют на три блока: первый располагается в районе центромеры, второй – в середине длинного плеча и третий – на дистальном участке длинного плеча Y-хромосомы. Все три блока содержат тандемные повторы, но именно третий блок отличается наиболее высокими показателями вариабельности. Так, он может различаться у двух родных братьев. Центромерный регион также демонстрирует достаточно высокую вариабельность, поэтому именно он часто используется в популяционно-генетических исследованиях [Carvalho et al., 2005].

Сейчас на Y-хромосоме известна локализация 156 транскрипционно активных единиц. 78 из них являются белок кодирующими генами, большая часть которых (60) – множественные копии 9 семейств. Остальные 18 генов представлены только одной копией [Степанов и др., 2006; Jobling et al., 2003].

Обычно все гены Y-хромосомы делят на следующие группы:

- PAR (Pseudoautosomal Region – гены псевдоаутосомных областей PAR1 и PAR2), локализованные в теломерных областях Y-хромосомы.

- X – Y гомологичные гены, локализованные в нерекombинирующих областях Y<sub>p</sub> и Y<sub>q</sub>.
- Y – специфичные гены, расположенные в нерекombинирующих областях Y<sub>p</sub> и Y<sub>q</sub>.

Благодаря наличию гомологичных PAR-регионов на X и Y-хромосомах, они регулярно конъюгируют и рекомбинируют участками этих регионов в зиготене и пахитене профазы I мейоза. Однако большая часть Y-хромосомы не принимает участия в рекомбинации, поэтому называется нерекombинирующей областью Y-хромосомы (NRY – Non Recombinant Region Y chromosome). NRY оценивается примерно в 50 млн. пар оснований и занимает примерно 95 % Y хромосомы [Hammer et al., 2002].

Присутствие NRY обуславливает самый низкий уровень полиморфизма Y-хромосомы среди всех 24 хромосом человека. [Hammer et al., 2003]. Генетическая вариабельность NRY определяется только мутационным процессом. Это значит, что отцовские линии представляют собой последовательную «запись» мутационных событий в продолжительном ряду поколений. В этом смысле Y-хромосома аналогична мтДНК. Но в отличие от мтДНК, размер которой лишь чуть больше 16 т.п.н. и где преобладают точечные мутации, Y-хромосома является хранилищем самого разнообразного полиморфизма, что делает её гораздо более информативной [Степанов и др., 2006; Thomas et al., 2000].

Генетические маркеры нерекombинантной части Y-хромосомы можно разделить на следующие классы: микросателлиты (STR), минисателлиты (локус MSY1(DYF155S1) единственный представитель этого класса, он содержит 50 – 150 фрагментов длиной 25 пар оснований каждый), [Jobling et al., 1999) инсерции\делеции (например, инсерция Alu - элемента в локусе DYS287 (YAP)) и SNP [Hammer et al., 2002].

Для изучения степени родства между древними популяциями, требуется маркер с низкой вероятностью возникновения параллельных и

обратных мутаций, а значит с низкой скоростью мутирования. Класс маркеров, отвечающих данным требованиям – это SNP (скорость мутирования различается по данным разных авторов и составляет в среднем  $0,8 \times 10^{-9}$  на локус на поколение) [Jobling et al., 2001, Xue et al., 2009; Mendez et al., 2012; Poznik et al., 2013, Francalacci et al., 2014; Karmin et al., 2015]. Поскольку этот тип маркера может существовать во всей популяции только в двух вариантах, SNP иначе называют бинарными маркерами, биаллельными маркерами или UEPs (unique event polymorphisms) [Thomas, 1998]. Комбинации UEP образуют «линиджи», которые и используются для выделения гаплогрупп [Jobling et al., 2001].

В настоящее время общепринятой филогенетически обоснованной классификацией гаплогрупп является номенклатура, принятая на консорциуме по Y-хромосоме в 2002 году [Y Chromosome Consortium, 2002]. Согласно этой классификации на филогенетическом древе Y-хромосомы современного человека выделено 18 основных кладов, обозначаемых буквами латинского алфавита от A до R, и эта классификация включала 243 маркера, по которым можно было выделить 153 конечных кластера, характеризующихся определенным аллельным состоянием группы последовательных по происхождению бинарных маркеров. По мере продвижения от корня дерева к ветвям в обозначениях линии используются арабские цифры и латинские буквы [Степанов В.А., 2006]. Благодаря гибкости этой системы, позволяющей последовательно расширять номенклатуру по мере обнаружения новых маркеров, не меняя топологию других ветвей дерева, к 2015 году уже было открыто около 15 000 маркеров [<http://www.isogg.org/>].

Не меньшую роль в популяционных исследованиях играет и другой тип маркеров – STR (short tandem repeats). У STR по сравнению с SNP темпы мутирования очень высоки. Высокое разнообразие STR вариантов делает их полезным инструментом в криминалистике, при выяснении



отцовства [Jobling et al., 1997; Ljubkovic et al., 2008] и для анализа разнообразия гаплотипов внутри гаплогрупп, определяемых по SNP [Mountain et al., 2002]. Также STR вариации помогают пролить свет на степень родства отдельных популяций [Srikumool et al., 2006]. Кроме того, STR широко используются в популяционной генетике в качестве «молекулярных часов» для определения возраста отдельных гаплогрупп [Gusmao et al., 2003].

Ещё одним достоинством Y-хромосомы является высокая степень кластеризации её вариантов. NRY фактически представляет собой один локус и при этом, естественно, подвергается отбору, как единое целое [Hammer et al., 2002]. Поскольку эффективная численность пула Y-хромосом в 4 раза меньше, чем для аутосом, и в 3, чем для X хромосомы, Y-хромосома в гораздо большей степени, чем другие генетические маркеры подвержена эффектам дрейфа и, как следствие, характеризуется большей степенью географической кластеризации своих вариантов [Karafet et al., 2002; Luca et al., 2007]. Структурированность распределения Y-хромосомы в пространстве ещё в большей степени усиливается за счет социальных особенностей человека: например патрилокальные браки в течение всей истории человечества гораздо чаще матрилокальных, а частота встречаемости полигамии выше по сравнению с полиандрией. [Cavalli-Sforza L.L., 2003] Так, значения таких показателей генетической дифференциации, как  $G_{ST}$  и  $F_{ST}$ , полученные для одних и тех же регионов, но по разным типам маркеров, для Y хромосомы оказались в 2-3 раза выше, чем для мтДНК и аутосом [Varzari et al., 2006]. А для восточнославянских популяций было показано, что межпопуляционное разнообразие Y-хромосомы в 7 раз выше, чем разнообразие по мтДНК и по аутосомным ДНК маркерам [Балановская и др., 2011].

Все, перечисленные выше особенности NRY, и делают его великолепным инструментом для детекции межгрупповых вариаций и для

реконструкции древних миграций человека [Rootsi et al., 2007; Shi et al., 2010; Underhill et al., 1997].

### **1.2.2. Аутомные маркеры**

В последние несколько лет в популяционно-генетических исследованиях все активнее используются полногеномные аутомные панели. Их использование позволяет получить данные о полиморфизме сотен тысяч маркеров. Первоначально исследования велись, основываясь на данных полногеномных панелей, предназначенных для медицинских целей. Наибольшее число популяционных исследований велось с использованием серии панелей Illumina (Illumina 550k, Illumina 660k и другие) [Ynusbayev et al., 2015; Fedorova et al., 2013]. Однако то, что данные панели были разработаны для медицинских целей, давало свой отпечаток на получаемые результаты: степень межпопуляционной дифференциации, получаемой посредством их использования, была значительно ниже, чем при использовании маркеров Y-хромосомы. Так в работе [Ynusbayev et al., 2011] при изучении кавказских популяций удалось выделить лишь два региона на Северном Кавказе: восточный и западный. Тогда как при исследовании полиморфизма Y-хромосомы в населении Кавказа было показано деление генофонда Кавказа на четыре этногеографических региона, для каждого из которых характерна собственная, мажорная гаплогруппа. И таких примеров, когда дорогостоящие полногеномные исследования не смогли превзойти результаты, полученные при изучении маркеров Y-хромосомы, существует множество.

Именно поэтому сразу несколько производителей занялись разработкой панелей предназначенных специально для популяционно-генетических исследований. На сегодняшний день мировыми лидерами являются две такие панели: GenoChip, созданная при участии Genographic

Project, и Axiom Human Origins, созданная при участии проф. D. Reich и его коллектива, Гарвард, США. Ниже мы опишем преимущества и недостатки этих двух панелей.

### Affymetrix Human Origins

Наиболее популярная полногеномная панель этой серии содержит 629 000 SNP, данные о которых были получены от следующих популяций: бушменов, йоруба, пигмеев, французов, сардинцев, китайцев, камбоджийцев, монголов, папуасов, шимпанзе и денисовцев. Помимо этого панель содержит дополнительные 87 000 маркеров митохондриальной ДНК и Y-хромосомы. Создатель этой панели D. Reich заявляет, что она представляет собой самую лучшую панель для изучения генофонда человечества и превосходит по своим качествам данные секвенирования. С использованием генотипирования по панели Human Origins уже опубликовано несколько масштабных исследований, в которых представлены оригинальные выводы [Naak et al., 2015; Lazaridis et al., 2014]. Но при этом при разработке панели не брались в расчет популяции Восточной Европы, Сибири, Дальнего Востока и многие другие.

### GenoChip

Данная полногеномная панель содержит 130 000 аутомных и X-хромосомных SNP, а также 13 000 SNP Y-хромосомы и 3 300 мтДНК. Информативно значимые маркеры для разработки данной панели брались из более чем 450 популяций мира, включая архаичных гоминид – неандертальцев и денисовцев (26 000 SNP). Оценки  $F_{ST}$  полученные при использовании панели GenoChip для тех же самых популяций оказались намного выше, чем при использовании коммерческих медицинских панелей [Elhaik et al., 2013]. Перекрытие панели GenoChip с Affimatrix Human Origins составляет 40%. Точность определения происхождения образцов при использовании данной панели была показана в работе [Elhaik et al., 2014], а эффективность её работы в целом показана в работе [ArunKumar et

al., 2015], где продемонстрировано, что генетическая подразделенность индусов коррелирует и с географией и с лингвистикой.

### **1.2.3 Генетические маркеры разных типов в решении индоевропейской проблематики**

Индоевропейская языковая семья в настоящее время является самой многочисленной, по количеству людей разговаривающих на её языках в мире. Неудивительно, что ученые, в том числе и генетики, с давних пор интересовались проблемой локализации прародины индоевропейских языков и отдельными этапами распространения её представителей.

Одно из наиболее известных исследований генофонда индоевропейских народов проведено L.L. Cavalli-Sforza ещё в начале 90-х годов. В его монографии, ставшей классикой популяционной генетики, обобщены проведенные им и его коллегами исследования изменчивости белковых генетических маркеров в Европе и других регионах мира. Полученная ими широко известная дендрограмма генетического сходства европейских популяций, согласно интерпретации самого L.L. Cavalli-Sforza, является отражением в первую очередь лингвистической классификации. А поскольку подавляющее большинство народов Европы говорит на языках индоевропейской семьи, выявленное сходство относится к связи структуры генофонда и лингвистической изменчивости именно индоевропейских популяций.

Сам L.L. Cavalli-Sforza всегда проявлял интерес к проблеме происхождения индоевропейцев и обладает оригинальной точкой зрения на этот счет, он, утверждает, что гипотезы Гимбутас и Renfrew не противоречат друг другу, а лишь подтверждают одна другую. Он полагает, что люди из курганной культуры, по крайней мере частично, являются

потомками населения мигрировавшего туда из Анатолии в неолите. В соответствии с этим экспансия индоевропейцев с территории Анатолии началась 9 500 лет назад. Дальнейшее их разделение на отдельные группы продолжилось в Волго-Уральском регионе 6 000 лет назад. Куда население попало, вероятно, через Балканы [Cavalli-Sforza, 1994].

Но основной массив работ по изучению генофонда народов индоевропейской языковой семьи велся преимущественно по данным полиморфизма Y-хромосомы и мтДНК, и уже по ним были получены выводы. Причем в большинстве статей доминировало мнение, что распространение индоевропейских языков шло путём языкового заимствования, а не вместе с распространением его носителей.

Так, в работе Rosser et al., 2000 показано, что структуру генофонда Европы формирует скорее географическая близость популяций, чем лингвистическое родство. Важно, что в этом исследовании были применены количественные лингвистические данные, хотя и несколько устаревшие с методологической точки зрения (при анализе лексико-статистического сходства языков не исключались заимствованные слова).

В работе Belledi et al., 2000 подтверждается вывод об отсутствии параллелизма языка и генетики на обеих системах маркеров: Y-хромосоме и мтДНК. Исследователи проводили сравнение лингвистически обособленных албанцев с другими популяциями индоевропейцев и выяснили, что генетически они очень сходны с остальными индоевропейцами. Особенно наглядно отсутствие дифференциации по маркерам Y-хромосомы, тогда как по мтДНК, авторы отмечают небольшую, но достоверную дифференциацию внутри лингвистических подгрупп индоевропейцев. Исследователи объясняют полученную картину относительно недавним разделением индоевропейских народов и последующими миграциями народов Европы.

С другой стороны, при исследовании изменчивости Y-хромосомы в отдельных странах Европы, нередко выявлялись и совпадения генетической и лингвистической изменчивости. Наиболее ярким примером является работа Kayzer et al., 2005, в которой показано совпадение генетической границы (резкие различия между генофондами) и лингвистической (а также политической) границы между германоязычными и славяноязычными популяциями.

Подтверждает согласованность данных лингвистики и генетики и картина, полученная для населения Индии по однородительским маркерам. В работе Cordaux et al., 2003 на основе данных полиморфизма мтДНК показано присутствие типично европейских гаплогрупп H, JT и W в северных и центральных племенах Индии, где разговаривают на индоевропейских языках. В противоположность этому, на юге Индии, где доминируют дравидские языки, обнаружить эти гаплогруппы не удалось. Не всегда при изучении генофондов популяций Индии удается установить параллели между данными разных дисциплин [Pagani et al., 2017].

Много работ посвящено и выявлению гаплогрупп, которые условно можно было бы назвать «индоевропейскими». Наиболее ярким кандидатом на эту роль является гаплогруппа R-M198. Поэтому особенно много дискуссий связано с обнаружением прародины этой гаплогруппы. Еще в одних из первых работ по географии Y-хромосомы [Passarino et al., 2001; Quintana-Murci et al., 2001; Wells et al., 2001] было высказано предположение о связи гаплогруппы R-M198 с распространением ямной культуры, которая, как считается, возникла в прикаспийских степях, и впоследствии ее носители мигрировали в Европу, Среднюю Азию и Индию в период 3000-1000 лет до н.э. Предполагается [Semino et al., 2000; Wells et al., 2001], что время возникновения мутации R-M198 более раннее (~ 5 900 лет назад, Karmin et al., 2015]. Было отмечено [Wells et al., 2001; Zerjal et al., 2002], что высокая частота гаплогруппы R-M198 сконцентрирована в

горных районах Центральной Азии и присутствует с относительно низкой частотой на Кавказе и Ближнем Востоке. В Средней Азии ее частота является самой высокой в горной местности (>50% среди таджиков, киргизов, алтайцев) и опускается <10 % на равнинах (среди народов Туркмении и казахов). По сравнению с другими более ранними исследованиями, в этих работах в Средней Азии наблюдали более высокую частоту гаплогруппы при более низком разнообразии STR-гаплотипов, чем в индийских кастах и племенах. На этом основании было сделано предположение о роли эффекта основателя: миграции из Западной Азии во время раннего голоцена. Эта миграция могла принести гаплогруппу R-M198 как в Среднюю Азию, так и в Индию еще задолго до появления на этих территориях индоевропейцев [Kivisild et al., 2003; Thanseem et al., 2006; Trivedi et al., 2006]. Однако, в работе Zerjal [Zerjal et al., 2002] приписывают низкое разнообразие Y-STR эффекту бутылочного горлышка в центральноазиатских популяциях.

Многие исследования установили значительную частоту гаплогруппы R1a в кастовых группах Индии по сравнению с племенными группами [Sahoo et al., 2005]. Это находит естественные аналогии в известном из данных физической антропологии преобладании в высших кастах потомков индоевропейской миграции из Средней Азии в Индию. Однако некоторые другие авторы считают, что такое влияние индоевропейского населения Средней Азии на автохтонный генофонд Индостана было незначительным [Sengupta et al., 2006]. Это заключение было основано на оценках возраста накопления микросателлитных мутаций в большинстве индийских гаплогрупп.

Если распространение гаплогруппы R-M198 в Индии естественно связывается с конкретной и хорошо известной для Индии миграцией индоиранских племен, то ее распространение в Восточной Европе и Сибири дает еще большую свободу для рассуждений. Действительно, поскольку

многие из существующих теорий происхождения индоевропейской семьи помещают ее прародину в Северное Причерноморье, то высокая частота гаплогруппы R-M198 у украинцев и других восточных славян может рассматриваться как сохранение высокой частоты гаплогруппы на этой территории в течение пяти тысячелетий. С другой стороны, значительно более древний возраст гаплогруппы R-M198 подразумевает ее прибытие на эту территорию намного раньше самого зарождения индоевропейской общности, и, следовательно, та же гаплогруппа с большой вероятностью должна присутствовать и в генофонде других (не индоевропейских) групп Восточной Европы и Кавказа. Наконец, высокая частота гаплогруппы R-M198 у населения Алтая снова вызывает ассоциации с хорошо известными скифо-сарматскими культурами (носители которых вероятно разговаривали на индоевропейских языках), распространявшимися от степей Причерноморья до Алтая. Такая трактовка наличия гаплогруппы R-M198 в Южной Сибири стала практически общепринятой в российской науке [Харьков, 2012; Балаганская и др., 2011].

Накопленные учеными данные указывают на две области с повышенными частотами R-M198: одна из них находится в Восточной Европе и включает в себя Польшу и часть России, а другая в Южной Азии и охватывает Северную Индию [Underhill, et al., 2009].

Ещё в 2009 году было проведено два исследования STR маркеров в пределах гаплогруппы R-M198, и они выявили наибольшее разнообразие её вариантов на территории Азии. А начальные этапы диверсификации R-M198 происходили на территории современных Ирана и Восточной Турции. Широта распространения данной гаплогруппы очень велика и лишь в последние годы удалось выявить её разделение не только по STR маркерам, но и по SNP заменам на две основные ветви: азиатскую и европейскую [Underhill et al., 2015]. Но в понимании проблем связанных с расселением народов индоевропейской семьи это открытие не помогло.



Таким образом, после ряда дискуссий, вопрос о путях и даже об основных направлениях распространения гаплогруппы R-M198 следует признать нерешенным.

Некоторые предположения о гаплогруппах, характерных для индоевропейцев, высказаны и в статье King et al., 2008. В ней изучалось влияние неолитических миграций из Анатолии на территорию Греции и Крита. В результате было выяснено, что на генофонд материковых греков, в отличие от греков Крита, они не оказали заметного воздействия. В заключение авторы рассуждают о возможности связать работы Gray, Atkinson и Renfrew со своими результатами: маркер M410, определяющий гаплогруппу J-M410, характерен для населения Анатолии и Крита, отделившихся от остальных индоевропейцев на ранних этапах развития, а субветвь J-M12, характерная для греков материковой части, знаменует собой дальнейшую дивергенцию индоевропейской общности.

С помощью полногеномных исследований ученые также активно пытались разрешить индоевропейскую проблематику. В том числе и с помощью панели GenoChip. В исследовании коллектива под руководством ArunKumar [ArunKumar et al., 2015] были исследованы 78 мужчин-индусов по полногеномной аутосомной панели GenoChip. В работе было показано преобладание компонента характерного для юго-западной Азии в генофонде населения Индии. Также ученые показали, что генетический ландшафт Индии хорошо согласуется с географическим и лингвистическим делением её населения. В работе продемонстрирована согласованность данных анализа полногеномных данных и Y-хромосомы. При этом с митохондриальными данными полногеномные не согласуются. Из этого авторы сделали вывод о том, что современный генофонд Индии сформировали миграции мужского населения. А женское население Индии генетически достаточно однородно из-за патрилокальной традиции, присущей этой стране.

Но главную возможность узнать о том, как же происходило расселение индоевропейцев, какой они имели генофонд, дает древняя ДНК. Изучая ДНК из древних захоронений можно последовательно восстановить, как формировался генофонд отдельных популяций и целых групп, в том числе и лингвистических.

В последние несколько лет, наибольший интерес у ученых-генетиков вызывает индоевропейская проблематика, и большое количество работ посвящено именно ей [Damgaard et al, 2018; Nielsen et al, 2017].

Так, исследование Haak et al., 2015 в некотором смысле и возродило интерес генетиков к этому вопросу. В работе был произведен анализ 69 древних геномов европейцев живших от 8000 до 3000 лет назад. Анализ производился по данным генотипирования указанных образцов по полногеномной панели Human Origin. Среди прочих изученных образцов, в работе анализировалось 9 представителей позднего медного/раннего бронзового века из ямной культуры (образцы датируются 3300-2700 лет до н.э.). Согласно курганной гипотезе Гимбутас, ямная культура связывалась с поздними протоиндоевропейцами. Именно поэтому интересно изучение ямников в исследовании индоевропейской проблематики. По результатам проведенного в статье анализа эта популяция расположилась между охотниками-собираателями Восточной Европы и некоторыми популяциями из круга Юго-Западной Азии и смежных регионов (в частности, Кавказа) на графике многомерного шкалирования. Результаты показали, что в позднем неолите и начале бронзового века население Западной Европы обнаруживает на 22-39% больше сходства с восточноевропейскими популяциями (ЕНG-компонент), чем их предки в среднем неолите. И это авторы статьи трактуют как результат миграции из Восточной Европы представителей ямной культуры. Исследование гаплогрупп Y-хромосомы образцов ямной культуры выявило у них доминантную гаплогруппу R-M269

– самую распространенную на нынешний день гаплогруппу в Западной Европе.

По представленным в статье результатам генетический вклад популяций ямной культуры в генофонд Центральной Европы оказался гораздо выше, чем считалось ранее. Все эти факты, по мнению авторов, говорят в пользу степной или курганной гипотезы распространения индоевропейских языков в Европе.

В другом масштабном исследовании древней ДНК было секвенировано еще больше образцов – 102. В данной работе изучались преимущественно образцы бронзового века - возрастом 3 тыс.-1 тыс. лет до н.э. [Allentoft et al., 2015]. Анализируя свои геномные данные в сравнении с уже опубликованными данными по древним и современным геномам, авторы пришли к заключению, что основные черты генетической структуры современной Европы сложились в бронзовом веке. Популяции в Северной и Южной Европе были сформированы смешением ранних охотников-собирателей и неолитических земледельцев, а в начале бронзового века получили приток генов из региона черноморско-каспийских степей. Эти данные хорошо соотносятся с археологическими данными по экспансии ямной культуры в Европу, что привело, по мнению авторов, к появлению популяции культуры шнуровой керамики. Анализ показал, что европейские культуры позднего неолита и бронзового века — шнуровой керамики, колоколовидных кубков, унетицкая и скандинавская – генетически близки друг другу. Все они проявляют сходство с ямной культурой. Наиболее близка к ней культура шнуровой керамики, наименее – культуры Венгрии, промежуточное положение занимает культура колоколовидных кубков Центральной Европы.

Но авторы не ограничились изучением образцов бронзового века в Европе, а включили в работу материалы из Азии. Их результаты показали, что образцы бронзового века, принадлежащие к афанасьевской культуре в

Алтае-Саянском регионе — генетически неотличимы от образцов ямной культуры. Это свидетельствует о том, что ямная культура из степей распространялась сразу в двух направлениях – как на запад, так и на восток. Благодаря этому в раннем бронзовом веке генетическое сходство появилось в таких отдаленных регионах, как Алтай и Скандинавия. Авторы статьи считают, что миграции в период бронзового века хорошо объясняют распространение индоевропейских языков. Они представляют процесс распространения индоевропейских языков так: экспансия ямников из черноморско-каспийских степей шла в двух направлениях - как в северную Европу, так и в Центральную Азию. Так, мигрирующие на восток ямники, вероятно говорившие на индоевропейском языке, дали начало афанасьевской культуре близ Алтая около 3000 лет до н.э. Представители афанасьевской культуры могли продвинуться далее на юг, вглубь Центральной Азии, и этим объясняется загадочное присутствие одного из старейших индоевропейских языков, тохарского, в бассейне реки Тарим (на стыке современных Казахстана, Монголии и Китая).

В продолжение своего исследования, тем же коллективом авторов была представлена еще одна статья [Matheison et al., 2016], но уже с большим количеством секвенированных древних образцов. Было изучено 230 древних геномов (с датировкой от 6500 до 1000 лет назад), в том числе 163 генома, данные о которых опубликованы впервые. В числе новых образцов – широкогеномные данные по неолитическому населению Анатолии. Анализ главных компонент показал, что неолитические образцы из Анатолии не похожи ни на одну из современных популяций Ближнего Востока, но смещены в генетическом пространстве в сторону Европы и входят в один кластер с ранними европейскими земледельцами по всей территории их распространения. Родство неолитических земледельцев Анатолии и первых европейских земледельцев показанное в статье не

вызывает сомнений - авторы доказали его с помощью многих надежных методов.

Авторы зафиксировали, что ямная культура из Самары и Калмыкии (5 300-4 700 лет назад), афанасьевская культура с Алтая (5 300-5 000 лет назад) и полтавкинская культура средней бронзы (4 900-4 700 лет назад) довольно близки генетически, на графике PCA они образуют единый «степной кластер». Во всех этих культурах доминирует Y-хромосомная гаплогруппа R-M269, как было показано и в работе Naak et al., 2015. Около половины предкового генофонда перечисленных выше групп составляет ближневосточный компонент, который отличается от генофонда ранних земледельцев Европы и Анатолии, зато похож на генофонд современных армян.

Но со временем, как показано в работе, генофонд носителей данных культур меняется - и в популяции следующей за срубной культурой поздней бронзы (3 900-3 600 лет назад) уже имелось около 17% предкового вклада от неолитических земледельцев Анатолии или первых европейских земледельцев.

В работе Allentoft с коллегами было показано, что этот предковый компонент обнаруживается даже к востоку от Урала, начиная, по крайней мере, со времени синташтинской культуры (4 300-4 100 лет назад). Высказывалось предположение, что он отражает миграцию на восток из культуры шнуровой керамики Центральной Европы. Однако, тот факт, что срубная культура также имеет этот предковый компонент (ранних земледельцев Европы) указывает, что он мог прийти в степь и из более восточных мест, чем Центральная Европа. Дополнительные свидетельства против миграций из Центральной Европы в степи дает Y-хромосома: изученные образцы срубной культуры (N=6) несут только гаплогруппу R-M198, и четыре из них (плюс один мужчина из полтавкинской культуры) принадлежат к ее ветви R-Z93, которая обычна в Центральной/Южной

Азии, очень редка у современных европейцев и отсутствует во всех изученных к настоящему времени древних образцах Центральной Европы.

Таким образом, результаты этих двух исследований показывают, что распространение индоевропейских языков шло не просто в виде культурного заимствования, а вместе с миграциями населения, и существенно уточняют отдельные этапы распространения индоевропейских языков.

В другой статье – Jones et.al., 2015 [Jones et.al., 2015] были обнаружены источники формирования генофонда популяций ямной культуры. Анализ трех древних геномов: верхнепалеолитического и мезолитического из Западной Грузии и верхнепалеолитического из Швейцарии показал, что популяции кавказских охотников-собирателей внесли вклад в генофонд ямной культуры и, таким образом, в миграционные потоки из понтийских степей на запад в Европу и на восток в Центральную Азию. Появление кавказского компонента в ямной культуре авторы связывают с майкопской культурой, которая возникла в IV веке до н.э. на Кавказе. Определяя место кавказских охотников-собирателей в формировании генофонда Европы, авторы работы приходят к выводу, что эта популяция составила четвертый источник генофонда (помимо трех, указанных в работе Lazaridis et al., 2014: охотников-собирателей Европы, земледельцев из стран плодородного полумесяца и северо-евразийского компонента). Четвертый компонент был перенесен в Европу миграцией скотоводов-ямников из евразийских степей, которые, в свою очередь, получили его из майкопской культуры, граничащей с Кавказом. По такому пути передачи кавказский предковый компонент был донесен до многих европейских популяций, в том числе и на северной части континента. С другой стороны, он распространился и на восток, достигнув северной Индии.

С выводами приведенной выше статьи согласуются и результаты работы Daniel E. Platt et al. 2017. В ней анализировались только данные по современной ДНК – по Y-хромосоме и полногеномным панелям. Суммируя результаты по всем видам анализа, авторы приходят к выводу о трех центрах, из которых произошло расселение людей по Юго-Западной Азии. Эти центры соответствуют трем климатическим рефугиумам Юго-Западной Азии, известным по археологическим данным. Первый центр находится в Восточном Причерноморье – в Грузии и восточной Турции; второй расположен в северном Леванте; третий – на юге Аравийского полуострова. В данной статье результаты указывают на роль Кавказа как источника наиболее древних постледниковых миграций. При этом кавказский регион определен как вероятный источник гаплогрупп J1 и J2, которые сейчас доминируют в Юго-Западной Азии, а ранее считались маркирующими распространение неолитической революции в Европе. С другой стороны гаплогрупп J-M267 и J-M172 не найдено в образцах ямной культуры – у них преобладает гаплогруппа R-M269.

Результаты, полученные в работе Broushaki et al., 2016 показывают, что в указанных нами выше работах, возможно, несколько упрощено рассматривается распространение индоевропейских языков. Ученые изучили образцы раннего неолита в Иране и пришли к следующим заключениям: образцы Ирана не могли послужить одной из предковых популяций для населения Европы. Зато для современного населения Пакистана, Афганистана, Ирана и Индии показано генетическое родство с древним населением Загроса (современный Иран). Такая картина согласуется с распространением индо-иранских языков на восток. Хотя распространение индоиранских языков – это не единственное объяснение полученных результатов, отмечают авторы. Распространение технологий сельского хозяйства либо дравидийских языков также могло быть причиной наблюдаемого сходства. Авторы статьи предполагают, что земледелие в

Юго-Западной Азии возникало независимо во множестве генетически различных популяций охотников-собирателей, откуда распространялось в окружающие регионы. Что касается популяции в иранском Загросе, то, по-видимому, она стала источником распространения культуры земледелия не в Европу, а на восток. И распространение земледелия шло совместно с миграциями населения и, возможно, распространением языков.

В работе Gallego-Llorente et al., 2016 так же показана генетическая обособленность древнего населения Загроса от Анатолии времен неолита, что подтверждает выводы, сделанные в работе Broushaki et al., 2016.

Проблема миграций в эпоху неолита и бронзового века поднята и в исследовании Goldberg et al., 2016. В ней рассматривалось, каков был половой состав эффективных миграций в генофонд Европы во время неолита из Анатолии и во время бронзового века из понто-каспийских степей. Исследование было проведено на основе данных полиморфизма X-хромосомы. Авторы изучили опубликованные образцы древней ДНК раннего и позднего неолита и бронзового века, проанализировав более 1,2 млн SNP. Исследуемые образцы относились к популяциям охотников-собирателей, земледельцев Анатолии и понто-каспийских степей. Анализ показал примерно одинаковое соотношение мужского и женского населения среди мигрантов для неолитической миграции, принесшей технологию земледелия в Европу. А вот миграция из понто-каспийских степей в Центральную Европу в течение позднего неолита и бронзового века, действительно, была преимущественно мужской: по подсчетам среди мигрантов на 5-14 мужчин приходилась одна женщина. Авторы показали, что эта миграция была растянута по времени на несколько поколений.

Таким образом, если в более ранних работах основанных на полиморфизме однородительских маркеров – Y-хромосомы и мтДНК, наличие общего компонента в генофонде всех групп популяций, разговаривающих на языках индоевропейской семьи, вызывало серьезные



сомнения, то применение данных секвенирования древних образцов изменило мнение ученых. Независимо, разными коллективами авторов был обнаружен генетический след миграций бронзового века в генофонде Европы. А поскольку большинство ученых склоняются к мнению, что эти мигранты из понто-каспийских степей были носителями индоевропейских языков, то напрашивается вывод о том, что в генофонде народов Европы, разговаривающих на индоевропейских языках, должен был сохраниться общий компонент. Однако, согласно исследованию Goldberg et al., 2016 среди мигрантов из степей преобладало мужское население, что не совсем согласуется с ранее накопленными данными по Y-хромосоме.

Для индо-иранской группы индоевропейской языковой семьи такой четкой миграции не было обнаружено, однако некие данные, свидетельствующие об общем генетическом компоненте в генофонде популяций этой группы все же имеются.

### **1.3. Региональные исследования представителей индоевропейской языковой семьи**

#### **1.3.1 Изученность популяций балто-славянской лингвистической группы индоевропейской языковой семьи по разным типам генетических маркеров**

Балто-славянская семья является довольно обширной и её исследованию посвящено достаточное количество работ: Balanovsky et al., 2008; Behar et al., 2010; Cruciani et al., 2007; Grzybowski et al., 2007; Hellenthal et al., 2014; Juras et al., 2014; Karachanak et al., 2012; Karlsson et al., 2006; Karmin et al., 2015; Kasperaviciū et al., 2002, 2004; Khrunin et al., 2013; Kushniarevich et al., 2013; Lappalainen et al., 2008; Lao et al., 2008; Malyarchuk et al., 2002, 2003, 2008; Mielnik-Sikorska et al., 2013; Mirabal et al., 2009; Mittnik et al., 2018; Morozova et al., 2012; Myres et al., 2011; Pericić et al., 2005; Ralph et al., 2013; Rootsi et al., 2004, 2007, 2012; Rosser et al., 2000; Semino et al., 2000; Shou et al., 2010; Wells et al., 2001; Underhill et al., 2007, 2010, 2015; Veeramah et al., 2011; Yunusbayev et al., 2012. В некоторых из этих работ изучена отдельная популяция балто-славянской языковой семьи, а в большинстве изучался обширный географический регион, либо распространение отдельной гаплогруппы.

Далее мы рассмотрим обобщающие работы по балто-славянам, дающие представление о генофонде её народов по различным типам маркеров. В литературе накоплен большой массив данных о генетическом разнообразии народов этой группы и, в частности, русских [Спицын, 2001; Балановская 2007].

Среди работ, основанных на полиморфизме Y-хромосомы особое положение занимает работа Kayzer et al., 2005. В ней показано совпадение лингвистической, географической и политической границ между славянами и германоязычными популяциями. Подобный же градиент обнаруживается

и при отдельном рассмотрении карт распространения ключевых для понимания генетической истории Европы гаплогрупп: R-M198 и R-M269. Так, гаплогруппа R-M269 встречается в Германии с частотой выше 50%, а в соседней Польше – ниже 20%. Гаплогруппа R-M198 демонстрирует противоположный тренд – она очень распространена в Польше (достигает 80% в некоторых выборках), несколько реже встречается в Чехии и Словакии (около 40%). В Западной Европе и Германии она встречается с частотой меньше 10%. Существование столь четкой генетической границы заставляет нас задуматься о том, что распространение славянских языков сопровождалось интенсивными миграциями населения.

Существенным для понимания общей структуры генофонда балто-славян является исследование распространения наиболее часто встречающихся среди них гаплогрупп.

Одной из мажорных гаплогрупп в генофонде славян является гаплогруппа R-M198 – она достигает максимальных частот среди русских Южной и Центральной России, составляя половину от всего пула Y-хромосом, постепенно, снижаясь к северу [Balanovsky et al., 2008; Балановский, Балановская, 2007]. В целом, эта гаплогруппа достаточно распространена среди восточных славян: русских, украинцев и белоруссов (около 40%). У балтов она встречается несколько реже (около 30%), как впрочем, и у русских Северной России. У поляков её частота составляет около 40%.

Особенно интересно для понимания истории славян, распространение варианта гаплогруппы R-M198 - R-M458. Данная гаплогруппа имеет высокую частоту в Европе с максимумом в Восточной Европе, где она составляет 70 % от всех Y-хромосом, несущих маркер M198. Следует заметить, что R-M458 отсутствует за пределами Европы, не распространяясь дальше Уральских гор на востоке и Турции на юге. Частоты выше 10 % характерны для западных и восточных славян, тогда как

у других европейцев, включая южных славян, частота R-M458 существенно меньше. Наивысшая же частота R-M458 наблюдается в Центральной и Южной Польше, где она превышает 30 %.

Также важна для понимания структуры генофонда балто-славян гаплогруппа I, определяемая маркером P37. Наибольшего распространения она достигает в Восточной Европе и на Балканах, присутствуя со значительной частотой и у всех славянских народов. Однако максимальная частота I-P37 замечена в Хорватии (31 %) и Боснии (40 %). На запад от Балкан частота I-P37 стремительно падает – данная гаплогруппа практически отсутствует у немцев, французов и итальянцев. Предполагают, что гаплогруппа I-P37 распространилась после периода последнего ледникового оледенения с Балкан [Primorac et al., 2011; Rootsi et al., 2004]

У русских, как и у других славянских народов, гаплогруппа I-P37 имеет весьма высокую частоту, являясь третьей по встречаемости гаплогруппой России [Балановский, 2015]. Внутри русского ареала распространение I-P37 закономерно уменьшается с удалением от Балкан – на юго-западе России частота приблизительно 15 %, в Центральной России – 10 %, на северо-востоке – 5%. У украинцев гаплогруппа I-P37 обнаружена с частотой 20 % [Balanovsky et al., 2008].

Большой интерес для понимания структуры генофонда балто-славян представляет распространение гаплогруппы N-M178 – как правило, связываемой с финно-угорскими популяциями. В Европе её распространение подчиняется строгому тренду – частота данной гаплогруппы начинает стремительно падать с востока на запад примерно на границе между Литвой и Польшей. Т.е. несмотря на свою принадлежность к балто-славянской группе, литовцы и латыши отличаются повышенными частотами гаплогруппы N-M178, совсем не характерной для славян. Гаплогруппа N1с занимает значительное место в русском генофонде, вбирая в себя пятую часть (19 %) фонда Y-хромосом. Примечательно, что в

пределах русского ареала частота N-M178 подчиняется широтной изменчивости, демонстрируя высокие (> 35 %) частоты на севере, низкие (< 10 %) – на юге [Балановская Е. В., Балановский О. П., 2007; Харьков и др., 2005, 2007]. Таким образом, за счет столь повышенных частот данной гаплогруппы, русские севера России генетически значительно отличаются от южных и центральных русских. Особый интерес представляет и изучение распространения этой гаплогруппы среди населения Верхнего Поволжья, поскольку на протяжении VIII-X веков н.э. Верхнее Поволжье входило в зону контактов населения, говорившего на финно-угорских, славянских и скандинавских языках. Понимание структуры генофонда этого региона может во многом объяснить ранние этапы расселения славян, помочь выявить их генетический след у современного населения.

Верхнее Поволжье было пограничным регионом между различными группами финно-угорских народов. Здесь обнаруживаются следы трех групп древних (реконструируемых только по данным топонимики) языков финно-волжской ветви: мерянских, севернофинских и верхневолжских [Альквист А., 2001]. Позднее обнаруживается след и прибалтийско-финской ветви, на языках которой говорят вепсы и карелы. По историческим данным регион был к X-XI в. заселен летописной мерей. Анализ субстратной топонимии показывает, что языки мери принадлежали в целом к западному (финно-волжскому) кругу финно-угорской лингвистической группы [Матвеев, 2007; Матвеев, 2015; Хелимский, 2006; Смирнов, 2015]. Миграция прибалтийских финнов - предков вепсов - в Ярославское Поволжье относится к более позднему времени (примерно середина IX века) [Захаров, 2012], а самая поздняя миграция прибалтийских финнов (карел) на верхнюю Волгу - к XVII веку. Скандинавский компонент фиксируется в Ярославском Поволжье не позднее IX века [Недошивина, Зозуля, 2012]. Славянское население в Ярославское Поволжье проникает в последние века I тыс. н.э. и складывается из двух колонизационных

потоков: ростово-суздальского (низового) и новгородского (верхового), соотношение которых в колонизации северо-востока Древней Руси было различным и требует дополнительного изучения [Третьяков, 1966]. Неполнота данных гуманитарных наук не позволяет надежно реконструировать эти миграции и оценить вклад финно-угорских народов (из которых наиболее значимым предполагается мерянский) в формирование современного русского населения Верхнего Поволжья.

В Поволжье по маркерам Y-хромосомы изучен генофонд коренного населения Тверской и Костромской областей [Balanovsky et al., 2008] (граничащих с Ярославской областью), в работах [Maluarchuk et al., 2008; Roewer L. et al., 2008] представлены данные о генофондах Нижегородской и Ивановской областей, хотя и по узкой панели маркеров. При этом генофонд Ярославской области, важный для понимания истории Верхнего Поволжья, по маркерам Y-хромосомы освещен недостаточно: ранее опубликованная [Maluarchuk et al., 2008] выборка мала ( $N=23$ ), а панель изученных маркеров слишком узка по современным меркам.

Отдельный интерес представляет изучение казачества – для понимания дальнейшего процесса этногенеза русского населения, его дифференциации на этно-социальные группы. В частности, вопрос о происхождении донского казачества является дискуссионным. Донские казаки ранее генетически не изучались, и в мировой науке нет сведений о степени вклада различных этносов в их генофонд. Анализ спектра гаплогрупп Y-хромосомы является одним из эффективных инструментов для решения этого вопроса.

Существуют две основные теории происхождения казачества на Дону: автохтонная и миграционная. И, несмотря на то, что миграционная теория поддерживается большинством авторитетных ученых-историков, на протяжении всего существования казачества на Дону вплоть до нашего времени возникали различные вариации автохтонных теорий, имеющие

большой успех в среде самого казачества. Они пытаются связать происхождение казаков с этносами, населявшими донские степи в разные исторические эпохи (с алано-яссами, бродниками), либо с народами Кавказа (черкесами) или же со степными тюркоязычными народами [Астапенко, 2000]. Сторонники миграционной теории, опираясь главным образом на русские письменные источники, полагают, что появление казачества на Дону произошло не ранее XVI в., и основная его часть была великорусской по происхождению [Волков, 1998]. Но даже миграционная теория оставляет много вопросов: в исторических документах зафиксировано, что миграционный поток в состав Войска Донского происходил не только из различных регионов России, но также и с территории Украины. Кроме того, в донское казачество вливались представители других народов: калмыки, греки, армяне, грузины, татары, черкесы, поляки, литовцы и другие [Проценко, 1999]. Так, в войсковом населении наряду с преобладающей численно и культурно русской группой существовали компактные группы казаков-мусульман и казаков-буддистов (калмыков). Таким образом, в донское казачество кроме великорусского и малорусского вошли и другие этнические элементы, и вопрос состоит в том, в какой мере они оказали влияние на генофонд казаков.

### **1.3.2 Изученность популяций армянской лингвистической группы индоевропейской языковой семьи по разным типам генетических маркеров**

Армянский народ уже становился объектом изучения генетиков, в том числе и по маркерам Y-хромосомы. Одним из наиболее масштабных исследований, посвященных данной проблематике стала работа Weale [Weale et al., 2001]. Исследование проведено на крупной выборке, включившей в себя четыре популяции из разных районов Армении и

представителей Лондонской диаспоры армян. Автором сформулированы выводы о высоком генетическом разнообразии армян, значительной их региональной стратификации и необходимости дальнейшего изучения популяций этого народа. Работа датируется 2001 годом, а в то время, к сожалению, генотипирование проводилось на гораздо менее глубоком уровне филогенетического разрешения, чем сейчас (в работе использовано 9 SNP и 6 STR маркеров). Таким образом, эти данные дают нам очень важные общие сведения о структуре генофонда армян, но для выяснения генетических взаимоотношения отдельных популяций армян друг с другом и с соседними народами, их недостаточно.

В работе Nasidze [Nasidze, 2003] армяне изучались в контексте всего Кавказа, и генотипирование проводилось также только на крупные ветви дерева Y-хромосомы. Эта работа показала, что генетически армяне ближе к кавказским народам, чем к народам индоевропейской языковой семьи, к которой относится и армянский язык. В связи с этим, авторами постулируется, что появление у армян языка, принадлежащего к индоевропейской языковой семье, вероятно, объясняется языковым замещением, без обмена генами.

Wells в своем исследовании [Wells et al., 2004], помимо остальных народов, проводил изучение армян по маркерам Y-хромосомы (размер выборки – 47 человек) в контексте всей Евразии. Но исследование не позволяет нам сделать никаких заключений о структуре генофонда собственно армян, степени его дифференциации и т.д. поскольку такой цели ученый перед собой не ставил.

В контексте народов Европы, фигурируют армяне и в статье Rosser [Rosser et al., 2000]. Но вновь генотипирование проводилось лишь по 11 SNP маркерам, что совершенно недостаточно для проведения анализа на современном уровне.



Проблема нахождения корреляций между данными лингвистики и генетики изучалась и в обширном исследовании Юнусбаева [Yunusbayev et al., 2011], где рассматриваются многочисленные народы Кавказского региона, в том числе и армяне. Роль географической структуры показана при использовании полногеномных панелей аутосомных маркеров, маркеров мтДНК и Y-хромосомы. Данные анализа аутосомных маркеров также указывают на генетическую общность армян с Кавказом, а не с народами индоевропейской языковой группы. Но вот выделить четкую структурированность кавказских популяций, как подчеркивают авторы, удалось только при использовании маркеров Y-хромосомы. Ни мтДНК, ни аутосомные маркеры для этой задачи не подходят. Однако, несмотря на то, что авторами использовалась большая панель SNP маркеров (52 маркера), она не была оптимизирована специально для кавказского генофонда и поэтому не включила, важные, дифференцирующие популяции Кавказа маркеры. Выборка армян составила 57 человек – достаточное количество для получения общего представления о положении армян в контексте окружающих популяций в генетическом пространстве, но недостаточное для изучения структуры генофонда самих армян.

В некоторых публикациях исследуются отдельные диаспоры армян. Так в работе Grugni [Grugni et al, 2012] автор изучал генофонд популяций Ирана по маркерам Y-хромосомы, в том числе и диаспору армян в Иране, в р-не Исфахан, прибывшую туда в 1600 году. Генотипирование проводилось на очень высоком по современным меркам уровне филогенетического разрешения. Но, к сожалению, выборка армян имеет достаточно невысокую численность – всего 34 человека и сравнение с другими группами армян в работе не проводилось.

Ещё одна работа, посвященная изучению генофонда армян в диаспоре, написана в 2012 году и исследует генетические особенности амшен [Margaryan et al, 2012]. В статье авторы пытаются найти генетические

доказательства одной из теорий происхождения амшен (существуют три основных теории их происхождения: из центральной Азии, из западной Армении, в частности древнего города Ани, или из восточной Армении) и приходят к выводу, что верной является теория, связывающая происхождения амшен с западной Арменией. Генотипирование в работе проведено лишь на 12 SNP и 6 STR маркеров. Так, доминирующей у амшен оказалась гаплогруппа ВТ\*(хDE,JK), дальнейшее генотипирование которой не проводилось. Но для выяснения происхождения амшен таких сведений явно недостаточно и требуется дополнительное, более глубокое исследование. Этот же коллектив авторов также провел обширное изучение генофонда армян по данным митохондриальной ДНК [Margaryan et al, 2017] и показали его консервативность на протяжении долгого времени.

Также в мировой науке имеются три статьи одного коллектива [Herrera, 2012; Lowery 2012; Novhannisyanyan A. et al., 2014] посвященные детальному изучению генофонда армян по достаточной панели маркеров Y-хромосомы и на обширной выборке. В статьях Herrera, 2012; Novhannisyanyan A. et al., 2014 была изучена обширные выборки образцов из центральной, восточной и западной исторической Армении. Авторы сделали заключение о том, что генофонд армян имеет неолитическое происхождение, и связан с распространением сельского хозяйства из Передней Азии, в частности Анатолии.

Другая статья того же коллектива [Lowery et al, 2012], по той же выборке представляет результаты анализа STR маркеров (17 локусов) и показывает специфичность Сасуна (одна из армянских популяций на территории Турции), его генетическую связь с Северным Кавказом. Однако в перечисленных выше трех работах в большей степени изучались генетические взаимоотношения внутри армянского народа и его положение в масштабе всего мира. Многие популяции Закавказья, Передней Азии просто не брались в анализ.

Обширная выборка армян (N=402) изучена в работе Platt et al., 2017. Выборка представлена следующими популяциями: армяне Юго-Восточной Турции (N=126), армяне Бахрейна (N=40), армяне Ирака (N=70), армяне Саудовской Аравии (N=7), армяне Кипра (N=38) и армяне Ливии (N=44). Однако сами исходные опубликованные данные работы не подразделены на субпопуляции, а представлены просто, как армяне. Субтипирование отдельных гаплогрупп (J-M267 и J-M172) проведено с достаточной глубиной. Тогда как гаплогруппы I, G субтипированы лишь до коровых вариантов, а R не глубже маркеров M198 и M269. Таким образом, без подразделения указанных гаплогрупп на более дробные варианты, а популяции армян на субпопуляции, теряется множество значимой информации.

У армян в работе Platt et al., 2017 было выявлено максимальное разнообразие гаплотипов в пределах гаплогруппы J-M172 и сделано предположение о том, что именно на территории Армении и возникла эта гаплогруппа впервые. На графиках многомерного шкалирования, полученных на основе  $R_{ST}$  дистанций, армяне кластеризуются вместе с популяциями Ирана и Турции. Согласно результатам PCA анализа по данным полногеномного генотипирования армяне объединяются в единый кластер с грузинами и турками, обособляясь от популяций Юго-Западной Азии.

Масштабное исследование генофонда армян было начато Теучеж И.Э. под руководством проф. Балановского О.П. В её диссертационной работе [Теучеж, 2013], представленной к защите в октябре 2013 г., изучались выборки армян Адыгеи и Краснодарского края и амшен. В исследовании Теучеж И.Э. армяне рассматривались в контексте всех кавказских популяций, и было показано их обособленное положение в генетическом пространстве.

Таким образом, мы видим, что в части ранних работ по изучению полиморфизма Y-хромосомы армяне не становились отдельным объектом изучения, а лишь исследовалось их положение в контексте других народов (Кавказа, Евразии). К тому же в большинстве работ генотипирование проведено по крайне узкой панели тестируемых маркеров, что делает перечисленные работы несопоставимыми с массивом современных данных. Однако в работах Herrera et al., 2012; Lowery et al., 2012; Novhannisyanyan et al., 2014; Grugni et al., 2012, Теучеж и др., 2013. популяции армян изучены с достаточным уровнем филогенетического разрешения и поэтому включены в суммарный анализ в нашем исследовании.

Помимо маркеров Y-хромосомы, армяне изучались и по полногеномным панелям во многих работах.

Единственным исследованием, где армяне стали главным объектом изучения стала работа коллектива Haber et al., 2015. По данным генотипирования 173 армян по полногеномной панели (панели Illumina 610k, 660k), учёные сделали вывод о том, что основные черты генофонда армян сформировались в бронзовом веке, между 3 000 -2 000 лет до н.э. Анализ результатов выявил: именно армяне являются наиболее вероятными потомками населения древней Анатолии. А сравнение с геномом тирольского человека показало, что среди всех ныне живущих популяций передней Азии, армяне генетически наиболее близки к неолитическому населению Европы.

Для анализа полученных данных учеными применен метод главных компонент. По его результатам авторы делают вывод о том, что армяне в мировом генетическом пространстве формируют обособленный кластер между европейцами, переднеазиатскими и кавказскими популяциями. Похожие выводы о положении армян в мировом генетическом пространстве сделаны в работе Novhannisyanyan et al., 2014. Однако, как было сказано выше, в работе Yunusbayev et al., 2012, где среди других популяций Кавказа и

Закавказья изучено и 16 армян по полногеномной панели, сделан вывод о большей генетической близости армян к народам Кавказа, чем к другим регионам.

В работе 2015 г. также приводится уточнение, что в Европе армяне наиболее близки к испанцам, итальянцам и румынам; в Передней Азии - ливанцам, евреям, друзам и киприотам; на Кавказе – абхазам и грузинам. Указанные популяции географически ближе располагаются территории исторической Армении. Вероятно поэтому они и генетически ближе: можно было бы ожидать в этом списке также и греков и азербайджанцев, но они просто не были включены в анализ.

Фигурируют армяне и в других исследованиях с использованием полногеномных систем маркеров. Но в них они не рассматриваются отдельно, а лишь в контексте изучаемого региона. Кроме того, в таких исследованиях армяне изучаются как одна популяция без подразделения на субпопуляции. Как правило, в таких работах показана близость армян к Переднеазиатским популяциям.

## **ГЛАВА 2. Материалы и методы**

### **2.1 Выбор объектов изучения и стратегия формирования выборок**

Корректное формирование выборок является необходимым этапом при исследовании. Только при его соблюдении можно полагаться на надежность последующих выводов. Все образцы, проанализированные в данной работе, принадлежат к Биобанку народонаселения Северной Евразии [Балановская и др., 2016]. Образцы для биобанка собираются по следующей стратегии:

1. Для охвата максимального генетического разнообразия анализируется широкий охват локальных субпопуляций.
2. Соблюдается правило трех поколений: в выборку включаются только те люди, у которых и оба его родителя, и все его бабушки и дедушки (т.е. предки до третьего поколения) родились в данной популяции и относили себя к данной этнической (или субэтнической) группе.
3. Проводится исключение родственников до третьего поколения из выборки. Наличие родственных индивидов не только уменьшает генетически эффективный размер выборки, но и смещает частоты.

Для анализа популяций по полногеномной панели GenoChip был применен оригинальный подход - метод сравнения пар популяций. С его помощью было проанализировано 245 собственных образцов (представлены в таблице 1) и 55 образцов, полученных благодаря сотрудничеству с зарубежными коллегами. 39 образцов испанцев, французов и басков были предоставлены Dr. Luis Quintano-Murci (Pasteur Institute, Париж, Франция), образцы эстонцев (N=16) были предоставлены Эстонским Биоцентром (Тарту, Эстония).

Все 300 образцов составили 7 пар сравнения: индоевропейские народы и соседние к ним народы, принадлежащие к иным лингвистическим семьям. Также в анализ включены данные из международного исследования [Elhaik et al., 2013] еще по двум комплексам сравнения: иранцы (N=16) и ливанцы (N=22) с кувейтцами (N=18); брахманы (N=12) и дравиды (N=14). Итого общее число комплексов сравнения составило 9, а число проанализированных образцов – 372.

Средний объем выборки, включаемой в анализ пар сравнения, составил 20 образцов. В ряде случаев эта выборка состояла из представителей двух или трех народов для более полного охвата генетического разнообразия. Например, русские (индоевропейская популяция, несущая в своем генофонде мощный финно-угорский субстрат) были сопоставлены со сборной выборкой окружающих финноугорских популяций (карелы, вепсы, мокша, эрзя), а баски (представители доиндоевропейского населения Европы) были сопоставлены со сборной выборкой двух окружающих индоевропейских народов (французов и испанцев).

На рисунке 1 представлены все девять комплексов сравнения: индоевропейские популяции с их географическими соседями, принадлежащими к другим лингвистическим семьям.

Таблица 1. Изученные по полногеномной панели GenoChip комплексы сравнения

	<b>индоевропейцы</b>	<b>не индоевропейцы</b>
1.	французы (N=9) и испанцы (N=22)	французские (N=10) и испанские баски (N=10)
2.	литовцы (N=20)	эстонцы (N=20)
3.	русские (N=20)	мордва (N=10), карелы (N=7), вепсы (N=4)

4.	таджики, памирские народы (N=31)	туркмены (N=20), узбеки (N=20)
5.	Брахманы (N=12), гуджарати (N=12)	дравиды (N=14)
6.	иранцы (N=16)	кувейтцы (N=18), ливанцы (N=22)
7.	армяне (N=20)	грузины (N=13)
8.	осетины (N=20)	кабардинцы (N=20), ингуши (N=4), чеченцы (N=4)
9.	украинцы (N=20)	ногайцы (N=12), крымские татары (N=8)
<b>ИТОГО: 9 комплексов сравнения, 416 образцов</b>		

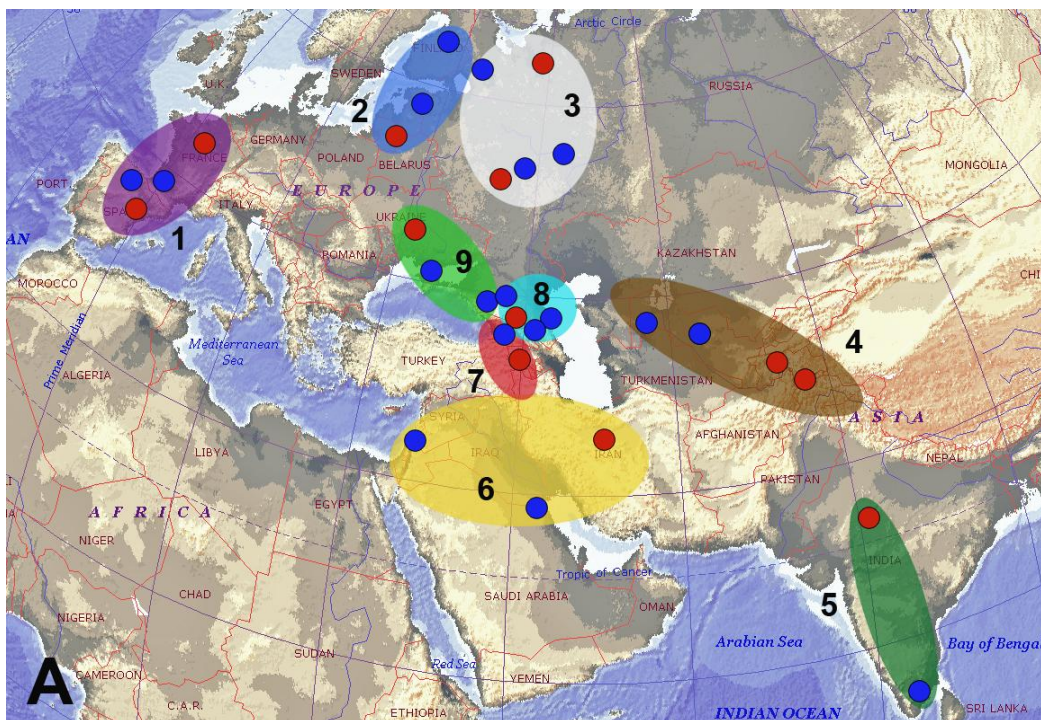


Рисунок 1. Карта популяций, изученных по полногеномной панели Genochip, объединенных в комплексы сравнения.

Обозначения:

**комплекс 1:** французы, испанцы (индоевропейская языковая семья) – французские и испанские баски (баскская семья);



**комплекс 2:** литовцы (индоевропейская семья) – эстонцы, финны (уральская семья);

**комплекс 3:** северные русские, центральные русские (индоевропейская семья) – мордва (мокша, эрзя), карелы, вепсы (уральская семья), казанские татары (алтайская семья);

**комплекс 4:** таджики, народы Памира (индоевропейская семья) – узбеки, туркмены (алтайская семья);

**комплекс 5:** каста брахманов (индоевропейская семья) – дравиды (дравидийская семья);

**комплекс 6:** иранцы (индоевропейская семья) – ливанцы, кувейтцы (афразийская семья);

**комплекс 7:** армяне (индоевропейская семья) – грузины (картвельская семья);

**комплекс 8:** осетины-иронцы, осетины-дигорцы (индоевропейская семья) – ингуши, кабардинцы, чеченцы (северо-кавказская семья);

**комплекс 9:** украинцы (индоевропейская семья) – ногайцы, степные крымские татары (тюркская семья).

Индоевропейские популяции показаны красными точками, а не индоевропейские – синими.

Образцы, для которых был изучен полиморфизм Y-хромосомы, также взяты из биобанка народонаселения Северной Евразии. В частности, изучено 599 образцов армян (таблица 2), в том числе 229 армян и 81 грузин исследованы в рамках данного диссертационного исследования (отмечены в таблице нижним подчеркиванием).

Недавние работы по популяционно-генетическому исследованию армян, проведенные уже на современном уровне филогенетического разрешения Y-хромосомы [Herrera et al., 2012], указали на насущную необходимость изучения широкого круга географически различных популяций армян для получения наиболее детальной и целостной картины генофонда народа. Для достижения этой цели нами изучены три ранее не исследованные армянские группы, представляющие современных потомков населения из различных регионов исторической Армении. Во все виды анализа включены все доступные данные по полиморфизму Y-хромосомы среди армян. Таким образом, с учетом локализации населенных пунктов, откуда произошли

миграции армян в диаспору, анализируемые в работе популяции, охватывают весь ареал исторической Армении (рисунок 2).



Рисунок 2. Карта изучаемых в работе выборок, с указанием путей миграций населения.

Таблица 2. Изученные популяции армян и грузин

<b>Популяций</b>	<b>N</b>	<b>Исторический регион проживания</b>	<b>Место сбора образцов</b>	<b>широта</b>	<b>Долгота</b>	<b>Источник данны</b>
<u>донские армяне*</u>	92	совпадает с местом сбора	Россия, Ростовская обл., Мясниковский р-н	47.17	39.3	данная работа
<u>армяне Эрзурума*</u>	99	Эрзурум. Историческая Армения	Грузия, Самцхе-Джавахетия	39.54	41.16	данная работа
<u>амшены*</u>	90	Трабзон. Историческая Армения	Россия, Краснодарский край	41.00	39.44	данная работа, Теучеж и др., 2013
армяне Краснодарского края*	117	совпадает с местом сбора	Россия, Краснодарский край	45.22	39.26	Теучеж и др., 2013
армяне Адыгеи*	49	совпадает с местом сбора	Россия, респ. Адыгея, Майкопский р-н	44.31	40.1	Теучеж и др., 2013
армяне Арарата	110	совпадает с местом сбора	Армения. Долина Арарат	40.04	44.33	Herrera et al. 2012
армяне	96	Гардман. Историческая	Армения	40.34	45.4	Herrera et al.

Гардмана		Армения				2012
армяне оз. Ван	10 3	Ван. Историческая Армения	Армения	38.37	42.52	Herrera et al. 2012
армяне Сасуна	10 4	Сасун. Историческая Армения	Армения	38..04	41..24	Herrera et al. 2012
армяне Ирана*	34	совпадает с местом сбора	Иран. Тегеран; Джодха, Исфахан	35.42	51.25	Grugni et al. 2012
армяне Салмаста	19 9	совпадает с местом сбора	Армения	38.11	44.45	Hovhannisyan et. al 2014
армяне Западной Армении	14 8	Историческая Западная Армения	Ливан	38.3	34.09	Hovhannisyan et. al 2014
армяне Алашкерта	20 0	Историческая Центральная Армения	Armenia	39.47	42.4	Hovhannisyan et. al 2014
армяне Карабаха	10 5	Историческая Восточная Армения	Карабах	39.48	46.45	Hovhannisyan et. al 2014
армяне Сюник	10 5	Историческая Восточная Армения	Армения	40.11	44.31	Hovhannisyan et. al 2014

<u>Восточные грузины</u> ( <u>пшавы, туши</u> <u>мтиулы,</u> <u>хевсуры</u> )	54	совпадает с местом сбора	Грузия. Душети. Гардабани. Тианети. Сагареджо	42.25	45.26	данная работа
<u>Имеретинцы</u>	62	совпадает с местом сбора	Грузия. Имеретия (Багдати. Чиатура. Харагаули. Хони. Кутаиси. Сачхере. Самтрedia. Тбилиси. Тержола. Ткибули. Вани. Зестафони)	42.10	42.59	Теучеж и др., 2013, данная работа
Лазы	36	совпадает с местом сбора	Турция (Ил Ризе. Фындыклы. Ардешен; ил Артвин (Хопа. Артвин). Грузия (Батуми)	41.08	41.01	Теучеж и др., 2013

\* - популяции армян, проживающие в диаспоре

Для народов балто-славянской языковой семьи, нами был проанализирован полиморфизм SNP маркеров Y-хромосомы более чем 6 000 образцов из 70 популяций. Изученные популяции представлены в таблице 3. Как указано в таблице, они объединяют данные из многочисленных источников литературы.

Таблица 3. Популяции балтов и славян взятые в исследование

<b>название популяции</b>	<b>Н, объем выборки</b>	<b>Источник</b>
Литовцы-аукштайты, восточные	<b>62</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Литовцы – аукштайты, южные	<b>38</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Литовцы-аукштайты, западные	<b>50</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Литовцы-жемайты, восточные	<b>65</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Литовцы-жемайты, южные	<b>48</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Литовцы-жемайты, западные	<b>38</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Латыши	<b>113</b>	Lappalainen et al., 2008; Underhill et al., 2010; Rootsi et al., 2012
Беларусы юго-восточные (Брест)	<b>96</b>	Kushniarevich et al., 2013
Беларусы северные (Витебск)	<b>100</b>	Kushniarevich et al., 2013
Беларусы центральные	<b>93</b>	Kushniarevich et al., 2013
Беларусы восточные	<b>85</b>	Kushniarevich et al., 2013
Беларусы, Восточное Полесье	<b>98</b>	Kushniarevich et al., 2013
Беларусы юго-западные	<b>121</b>	Kushniarevich et al., 2013

(Западное Полесье)		
Беларусы западные	<b>75</b>	Kushniarevich et al., 2013
Беларусы северные	<b>102</b>	Kushniarevich et al., 2013
Русские (Мезень)	<b>54</b>	Balanovsky et al., 2008
Русские (Пинега, Архангельская обл.)	<b>114</b>	Balanovsky et al., 2008
Русские (Красноборск, Архангельская обл)	<b>91</b>	Balanovsky et al., 2008
Русские (Вологда)	<b>121</b>	Balanovsky et al., 2008
Русские (Рязанская обл.)	<b>100</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Русские (Ливны, Орловская обл.)	<b>110</b>	Balanovsky et al., 2008
Русские (Пристенъ, Курская обл.)	<b>45</b>	Balanovsky et al., 2008
Русские (Репьевка, Воронежская обл)	<b>96</b>	uBalanovsky et al., 2008
Русские (Белгород)	<b>143</b>	Balanovsky et al., 2008
Русские (Унжа, Костромская обл.)	<b>52</b>	Balanovsky et al., 2008
Русские (Кашин, Тверская обл.)	<b>73</b>	Balanovsky et al., 2008
Русские (Порхов, Псковская обл.)	<b>57</b>	Balanovsky et al., 2008
Русские (Остров, Псковская обл.)	<b>75</b>	Balanovsky et al., 2008
Русские (Рославль, Смоленская обл)	<b>107</b>	Balanovsky et al., 2008
Украина (Ивано- Франковская обл.)	<b>57</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Украинцы (Львовская обл.)	<b>101</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Украинцы (Черкасская обл.)	<b>114</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Украинцы (Хмельницкая обл.)	<b>176</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Украинцы (Сумская обл.)	<b>101</b>	Данные лаборатории

		популяционной генетики человека (рук. Балановская Е.В.)
Украинцы (Харьковская обл.)	<b>56</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Украинцы (Белгородская обл.)	<b>56</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Украинцы (Закарпатская обл.)	<b>50</b>	Данные лаборатории популяционной генетики человека (рук. Балановская Е.В.)
Словенцы	<b>75</b>	Battaglia et al., 2008; Underhill et al., 2010
Хорваты 1	<b>89</b>	Battaglia et al., 2008; Underhill et al., 2010
Хорваты 2	<b>88</b>	Pericic et al., 2005; Underhill et al., 2010; Myres et al., 2010;
Боснийцы (Хорватия)	<b>90</b>	Battaglia et al., 2008; Underhill et al., 2010
Боснийцы (Герцеговина)	<b>133</b>	Pericic et al., 2005; Underhill et al., 2010; Myres et al., 2010;
Боснийцы (Сербия)	<b>81</b>	Battaglia et al., 2008; Underhill et al., 2010
Боснийцы 1	<b>65</b>	Pericic et al., 2005; Underhill et al., 2010; Myres et al., 2010;
Боснийцы 2	<b>84</b>	Battaglia et al., 2008; Underhill et al., 2010
Сербы 1	<b>110</b>	Pericic et al., 2005; Underhill et al., 2010; Myres et al., 2010;
Сербы 2	<b>103</b>	Regueiro et al., 2012
Македонцы 1	<b>73</b>	Pericic et al., 2005; Underhill et al., 2010; Myres et al., 2010;
Македонцы 2	<b>211</b>	Noveski et al., 2009
Болгары (Бургас)	<b>45</b>	Karachanak et al., 2013
Болгары (Ловеч)	<b>62</b>	Karachanak et al., 2013
Болгары (Монтана)	<b>80</b>	Karachanak et al., 2013
Болгары (Пловдив)	<b>159</b>	Karachanak et al., 2013
Болгары (София)	<b>257</b>	Karachanak et al., 2013
Болгары (Хасково)	<b>41</b>	Karachanak et al., 2013
Чехи (Брно)	<b>46</b>	Luca et al., 2006; Underhill et al., 2010; Myres et al., 2010;
Чехи (Градец-Карлове)	<b>49</b>	Luca et al., 2006; Underhill et al.,



		2010; Myres et al., 2010;
Чехи (Клатови)	<b>48</b>	Luca et al., 2006; Underhill et al., 2010; Myres et al., 2010;
Чехи (Писек)	<b>65</b>	Luca et al., 2006; Underhill et al., 2010; Myres et al., 2010;
Чехи (Требик)	<b>49</b>	Luca et al., 2006; Underhill et al., 2010; Myres et al., 2010;
Поляки Кашубский регион (центр)	<b>93</b>	Rebala et al., 2012
Поляки Кашубский регион (север)	<b>70</b>	Rebala et al., 2012
Поляки Кашубский регион (юг)	<b>41</b>	Rebala et al., 2012
Поляки (Кашубы)	<b>158</b>	Rebala et al., 2012
Поляки	<b>158</b>	Rebala et al., 2012
Поляки (Вроцлав)	<b>99</b>	Battaglia et al., 2008; Underhill et al., 2010
Сорбы (Германия)	<b>123</b>	Rebala et al., 2012
Словаки	<b>164</b>	Rebala et al., 2012

## 2.2 Методы отбора образцов и проведения полногеномного генотипирования

Для успешного проведения генотипирования образцов ДНК по полногеномной панели Genochip предварительно был проведен трудоемкий этап отбора образцов.

Для того, чтобы охватить максимальное генетическое разнообразие, для каждого народа отбирались образцы из разных субэтносов или разных географических районов.

Далее отбор велся по качеству ДНК. Этот фактор важен, чтобы избежать технических проблем при проведении генотипирования и для получения полных результатов. Мы использовали трехступенчатую систему проверки качества ДНК.

- Промер концентрации ДНК на спектрофотометре Nanodrop 2000C. Определялась общая концентрация ДНК в образце и присутствие там белкового

либо фенольного загрязнения. Образцы с соотношением поглощаемых волн  $260/280 < 1.8$  или  $260/230 > 2.3$  не брались в дальнейшую работу.

- Постановка электрофореза отобранных образцов. Этот пункт может показаться избыточным, однако реальный опыт проведения полногеномного генотипирования показывает, что это не так. Результаты электрофореза позволяют убедиться, что отобранные образцы не содержат деградированной ДНК. Образцы с деградированной ДНК исключались.
- Промер концентрации ДНК с помощью флуориметра Qubit. Данный прибор измеряет лишь концентрацию двухцепочечной ДНК, поэтому эти значения точнее, чем при измерении с помощью Nanodrop. На этом этапе нами были отобраны образцы с концентрацией не менее 80 нг/мкл и приведены к единой стандартной концентрации 75 нг/мкл.

По результатам многоэтапного отбора был сформирован единый набор образцов из 245 собственных образцов, удовлетворяющих всем целям нашего исследования. Генотипирование проводилось на платформе Illumina с использованием полногеномной панели GenoChip [Elhaik et al., 2013]. По результатам генотипирования для 95% наших образцов показатель «call rate» составил 99% и выше, что значит, что использованная нами стратегия отбора образцов оказалась эффективной.

### **2.3 Методы генотипирования полиморфизма Y-хромосомы**

Материалом для генотипирования служила тотальная ДНК, выделенная из слюны фенол-хлороформным методом. Все образцы были исследованы по двум системам генетических маркеров - SNP и STR маркерам Y-хромосомы. Генотипирование SNP маркеров проводилось методом ПЦР в реальном времени на приборах StepOnePlus и 9700HT (Applied Biosystems) с использованием

технологии Taqman (Applied Biosystems). Анализировался полиморфизм 50 SNP маркеров. Фрагментный анализ 17 локусов STR маркеров Y-хромосомы проведен с использованием коммерчески доступного набора Y-filer PCR Amplification Kit (Applied Biosystems) на генетическом анализаторе ABI 3130xl (Applied Biosystems) с последующей обработкой данных в программе GeneMapper (Applied Biosystems). Исследованы локусы: DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATAH4.

## **2.4 Общий статистический анализ полногеномных данных**

### **Первичная биоинформационная обработка.**

Данные полногеномного генотипирования были получены в формате .csv и затем объединены в единый массив данных в форматах, пригодных для работы в программе Plink [PLINK (1.07), author: Shaun Purcell, URL: <http://pngu.mgh.harvard.edu/purcell/plink/>; Purcell et al., 2007]: .map и .ped. (которые были затем преобразованы в троичную систему файлов: .bed, .bim, .fam).

### **РСА (анализ главных компонент).**

Этот классический метод многомерной статистики широко используется при популяционно-генетическом анализе полногеномных данных. Нами был проведен анализ главных компонент для всех изучаемых в данной работе популяций, а также привлечены данные из работы [Elhaik et al., 2013] и данные полученные благодаря сотрудничеству с иностранными коллегами. Всего был проведен анализ главных компонент для 2 наборов данных по образцам современной ДНК и график сравнения данных по современной и древней ДНК. Для графика в масштабе Евразии количество проанализированных образцов составило 546 индивидов.

Популяции, использованные для основных видов анализа, показаны на рисунке 3.

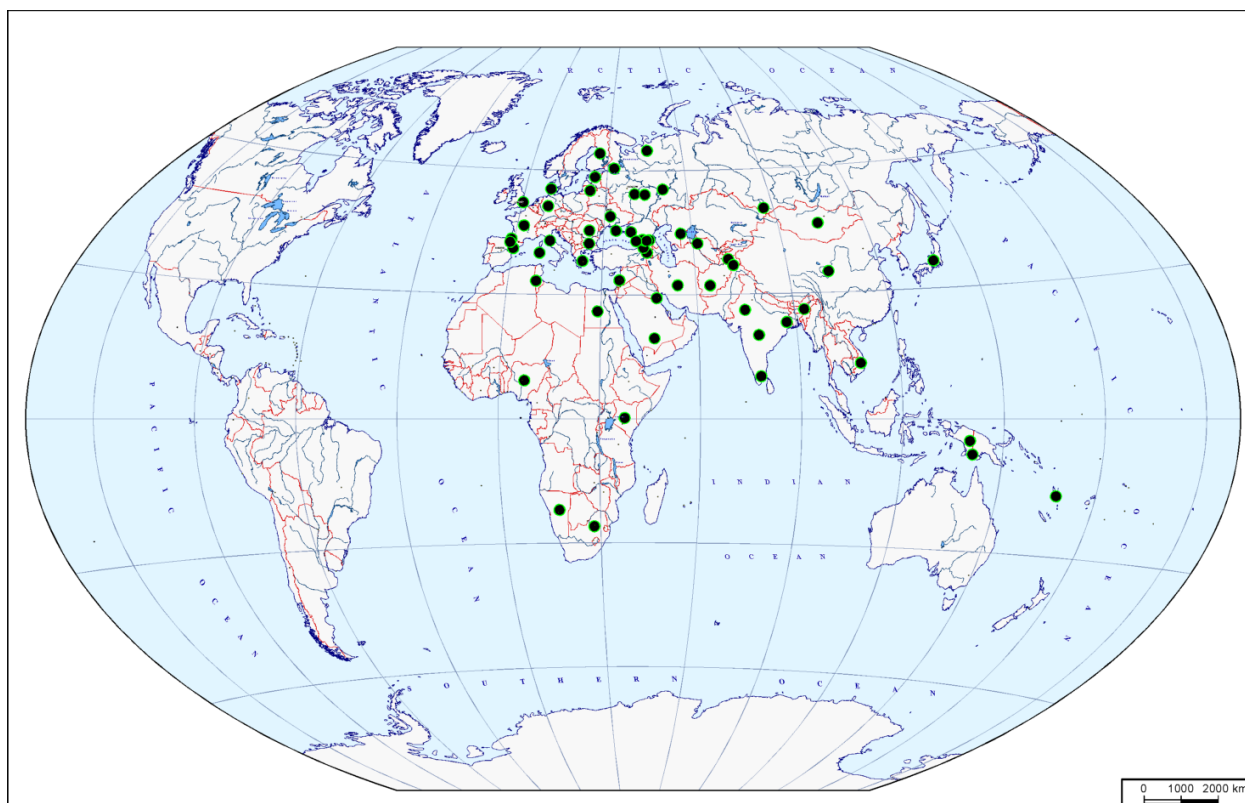


Рисунок 3. Карта популяций, изученных по полногеномной панели GenoChip и взятых в анализ в данной работе

Для анализа современных популяций в контексте данных по древней ДНК использован массив данных по 644 современным образцам, данные по древней ДНК взяты из работ [Naak et al., 2015; Allentoft et al., 2015; Matheison et al., 2016].

Перед проведением анализа главных компонент все данные были обработаны в программе plink по следующим параметрам: mind=0,1 (исключает индивидов в большом количестве (0,1%) пропущенных, не прошедших генотипирование замен), maf=0,05 (исключаются SNP с общей частотой меньше 0.05); geno=0,1 (исключаются SNP, по которым не были получены результаты для большого количества образцов). Затем проводится «отсев» сцепленных однонуклеотидных замен – LD pruning. Отсев также проводился в программе plink по следующим параметрам: попарная корреляция генотипов  $r^2 > 0.2$  с окном в 1500 SNP, с шагом в 150 SNP. В анализ для построения графиков главных компонент брались только аутосомные SNP. Таким образом, для проведения PCA

анализа был получен единый массив данных с общим количеством SNP около 46 000.

Непосредственно анализ главных компонент проводился и визуализировался в программной среде R studio 3.2.1.

Построение графика главных компонент с данными по древней ДНК и современными данными проводилось методом Прокрустова анализа. Для начала нами с использованием опции программы smartpca1 (версия: 10210) в программе EIGENSOFT2,3 5.0.1 был проведен анализ главных компонент для современных образцов. Затем был проведен Прокрустов анализ для совмещения графика главных компонент по современным образцам с данными генотипирования древних образцов, имеющих не слишком высокое покрытие. Анализ проводился как описано в [Lazaridis et al., 2014].

#### **ADMIXTURE анализ.**

Анализ, основанный на методе кластеризации, реализован нами на массиве данных из 767 образцов (собственные данные и образцы из работы Elhaik et al., 2013] в программе ADMIXTURE. Перед проведением анализа данные были обработаны в программе plink по следующим параметрам: maf=0.01, geno=0.03, mind=0.03, также был проведен «отсев» сцепленных однонуклеотидных замен – LD pruning. Анализ проведен при K=2-16. ADMIXTURE дает оценку дающих наиболее достоверную картину K путем вычисления индекса кросс-валидации (CV), который оценивает точность предсказания [Yunusbaev et al., 2015] модели при заданном K. В нашем анализе наилучший показатель достоверности получен при K=9. Поэтому для дальнейшего анализа были выбраны данные при K=9. Визуализация результатов ADMIXTURE проводилась в Microsoft Excel 2003.

#### **Расчет матрицы $F_{ST}$ расстояний.**

Расчет проводился согласно Weir and Cockerham [Weir and Cocker, 1984] с помощью скрипта для работы в программной среде R studio 3.2.1. [Metspalu et al., 2011] с последующей визуализацией матрицы в Microsoft Excel 2003. В анализ были включены 43 популяции.

### **Расчет евклидовых расстояний.**

Поскольку из-за проблем с качеством секвенирования древней ДНК, рассчитать матрицу генетических расстояний между современными популяциями и древними образцами ямной культуры (N=4) из работ [Haak et al., 2015; Allentoft et al., 2015] и древними образцами времен неолита из Анатолии (N=3) [Matheison et al., 2015] не представляется возможным, нами были рассчитаны евклидовы расстояния от каждого из этих образцов до всех анализируемых популяций. Визуализация полученной матрицы расстояний проведена в Microsoft Excel 2003.

### **Картографирование.**

Был проведен картографический анализ по матрице  $F_{ST}$  расстояний и по частотам маркеров, обнаруженных методами поиска, описанными в пункте 2.5 построены серии карт в программном пакете GeneGeo, созданном под руководством О.П. Балановского [Haak W. et al., 2010; Кошель С.М. и др., 2011; Balanovsky O. et al., 2015]. Для построения карт генетических расстояний, брались следующие показатели: весовая функция 3 и радиус влияния – 10 000 км.

## **2.5 Поиск маркеров, характерных для народов индоевропейской языковой общности.**

Все имеющиеся данные разбиты на комплексы сравнения: популяция, принадлежащая к индоевропейской языковой семье и её иноязычные географические соседи. Всего было проанализировано 557 образцов. Комплексы сравнения представлены на рисунке 1 и в таблице 1. Были рассчитаны частоты маркеров в каждой из анализируемых субпопуляций. Поиск проводился по трем критериям:

1. В каждом комплексе сравнения (например, русские - казанские татары, мордва, карел, вепсов) отбирались маркеры, присутствующие у индоевропейцев (в данном примере у русских), но отсутствующие у не индоевропейцев (в данном примере у татар, мордвы, карел, вепсов). При

этом можно требовать, чтобы маркер удовлетворял этому критерию в каждой паре, или только в пяти парах и так далее.

2. В каждом комплексе сравнения отбирались маркеры, присутствующие у индоевропейцев с частотой, в два раза чаще, чем у не индоевропейцев из данного комплекса сравнения. Как и для предыдущего критерия, возможны варианты, различающиеся числом пар, в которых это условие должно соблюдаться.
3. Подход, аналогичный предыдущему, только маркер должен был встречаться у индоевропейцев хотя бы немного чаще, чем среди их «парных» не индоевропейских соседей.

Поиск проводился в программе MicrosoftOffice 2013. Для проверки надежности количества обнаруженных маркеров, нами проводилось случайное объединение исследуемых популяций в комплексы сравнения. Был сгенерирован ряд случайных чисел, каждое число было присвоено одной популяции. Затем числа были отфильтрованы по возрастанию, после чего популяции оказались расположены в случайном порядке. Для всех полученных таким случайным образом пар популяций проводился поиск маркеров по критериям описанным выше.

Проверка проводилась три раза, при этом популяции были объединены в указанные ниже комплексы сравнения.

1. При первой случайной проверке популяции были объединены в следующие комплексы сравнения: армяне и казанские татары; испанцы и кувейтцы; осетины и испанские баски; карелы, вепсы, узбеки и дравиды, русские; иранцы и финны; французы, эстонцы и французские баски; таджики, ингуши и брахманы, чеченцы; ногайцы, украинцы и литовцы, кабардинцы; мордва, ливанцы и крымские татары, грузины.
2. При второй случайной проверке популяции были объединены в следующие комплексы сравнения: кувейтцы и осетины; финны и

армяне; крымские татары и чеченцы; ногайцы, грузины и татары, мордва; украинцы, карелы, вепсы и иранцы; испанские и французские баски и испанцы; кабардинцы, эстонцы и испанцы, узбеки; французы, ингуши и литовцы, брахманы; ливанцы, дравиды и таджики, русские.

3. При третьей случайной проверке популяции были объединены в следующие комплексы сравнения: литовцы и узбеки; русские и финны; мордва и таджики; испанцы, ливанцы и испанские баски, карелы, вепсы; ингуши, ногайцы и кувейтцы, французы; чеченцы, грузины и французские баски; дравиды, крымские татары и брахманы, иранцы; осетины и кабардинцы; украинцы, татары и эстонцы, армяне.

## **2.6 Статистический, филогеографический и картографический анализ данных полиморфизма Y-хромосомы**

Генетические расстояния между анализируемыми популяциями рассчитаны в программе DJ [Balanovsky et al., 2008] и визуализированы на графиках многомерного шкалирования в программе Statistica 6.0 [StatSoft. Inc., 2001] [StatSoft, Inc. STATISTICA (data analysis software system). version 62001. [www.statsoft.com](http://www.statsoft.com)). Для изучения генофонда балто-славянских популяций использовался весь массив информации указанный в таблице 3. Для исследования представителей армянской языковой группы помимо данных из работ, указанных в таблице 2, в анализ были взяты результаты следующих исследований: Abu-Amero et al., 2009; Balanovsky et al., 2011; Battaglia et al., 2008; Cadenas et al., 2007; Cinnioglu et al., 2003; Di Cristofaro et al., 2013; El Sibai et al., 2009; Flores et al., 2005; Grugni et al., 2012; Haber et al., 2011; Karachanak et al., 2013; King et al., 2008;



Pericic et al., 2005; Sanchez et al., 2005; Varzary et al., 2009; Zalloua et al., 2008a; Zalloua et al., 2008b.

Филогенетические сети по STR гаплотипам Y-хромосомы созданы в программах Network 4.1.1.2 [FluxusTechnologyLtd., www.fluxus-engineering.com] и Network Publisher [Fluxus Engineering, Clare, U.K.] на основе алгоритма reduced median. Для построения сетей использованы данные из работ [Balanovsky et al., 2011; Cristofaro et al., 2013; Haber et al., 2011; Herrera et al., 2012; Lowery et al., 2013; Myres et al., 2010;]. Возраст кластеров гаплотипов определялся с помощью показателя  $\rho$  согласно публикациям [Forster et al., 1996; Saillard et al., 2000]. Использовались две скорости мутирования STR маркеров: «эволюционная» скорость [Zhivotovsky et al., 2004], и «генеалогическая» скорость, полученная при прямых сравнениях пар отец-сын [Gusmao et al., 2005; Sánchez-Diz et al., 2008; Ge et al., 2009]. Время смены поколения оценивалось до 25 лет для эволюционной скорости, и 30 лет [Почешхова, 2008; Fenner, 2005] для генеалогической скорости мутации.

Поиск совпадающих гаплотипов для популяции донских казаков проведен в программе Haplomatch.

Расчет  $R_{ST}$  дистанций от донских казаков проведен в программе Arlequin 3.5.

Карты генетических расстояний и распространения отдельных гаплогрупп построены с помощью программного пакета GeneGeo, созданного под руководством О.П. Балановского [Naak et al., 2010; Кошель и др., 2011; Balanovsky et al., 2015]. Для построения карт генетических расстояний, брались следующие показатели: весовая функция 3 и радиус влияния – 10 000 км.

Тест Мантеля и анализ межпопуляционной изменчивости (AMOVA) проведены в программе Arlequin 3.5. Для анализа межпопуляционной изменчивости при изучении народов армянской лингвистической группы использовались данные представленные в таблице 2 и из работ: Abu-Amero et al., 2009; Balanovsky et al., 2011; Battaglia et al., 2008; Cadenas et al., 2007; Cinnioglu et al., 2003; Di Cristofaro et al., 2013; El Sibai et al., 2009; Flores et al., 2005; Grugni et

al, 2012; Haber et al., 2011; Karachanak et al., 2013; King et al., 2008; Pericic et al., 2005; Sanchez et al., 2005; Varzary et al., 2009; Zalloua et al., 2008a; Zalloua et al., 2008b.

Для расчета теста Мантеля на первом этапе рассчитывается степень сходства изучаемых популяций по признакам генетики, лингвистики и географии (создаются матрицы попарных генетических, лингвистических и географических расстояний между изучаемыми популяциями). На втором этапе вычисляются коэффициенты парных корреляций, а также коэффициенты частных корреляций, позволяющие выявить связь двух анализируемых параметров при исключении влияния третьего параметра.

Но для сравнительного анализа генетического и лингвистического разнообразия требуются количественные лингвистические данные о степени сходства и времени разделения языков, на которых говорят изучаемые популяции. Эти данные для индоевропейской семьи к настоящему времени уже получены методами лексикостатистики и глоттохронологии [Swadesh, 1955; Старостин, 1989]. Значительная часть результатов и исходных баз данных представлена на сайте «Вавилонская башня» [[www.starling.runet.ru](http://www.starling.runet.ru)]. Матрица для языков индоевропейской языковой семьи составлена и любезно предоставлена представителем московской научной школы глоттохронологии член-корр. РАН А.В. Дыбо.

Для расчета теста Мантеля для балто-славянской лингвистической группы использовался весь массив информации указанный в таблице 3. Для исследования представителей армянской языковой группы помимо данных из работ, указанных в таблице 2, в анализ были взяты результаты следующих исследований: Abu-Amero et al., 2009; Balanovsky et al., 2011; Battaglia et al., 2008; Cadenas et al., 2007; Cinnioglu et al., 2003; Di Cristofaro et al., 2013; El Sibai et al., 2009; Flores et al., 2005; Grugni et al, 2012; Haber et al., 2011; Karachanak et al., 2013; King et al., 2008; Pericic et al., 2005; Sanchez et al., 2005; Varzary et al., 2009; Zalloua et al., 2008a; Zalloua et al., 2008b.

Для анализа генофонда донских казаков использована программа Harpmatch. При сравнении с донскими казаками использовались данные о генофонде популяций, которые по историческим данным имели или могли иметь отношение к становлению донских казаков. Из соображений сопоставимости данных, в анализ брались только народы, по которым имелись характеристики разнообразия панели из тех же 17 STR-маркеров, которые были изучены у казаков Дона. Сравнение производилось с 15 популяциями: абхазами, азовскими греками, армянами, восточными украинцами, грузинами, западными украинцами, казанскими татарами, калмыками, литовцами, ногайцами, северными русскими, черкесами, центральными русскими, шапсугами, южными рускевклидовыми. Данные по полиморфизму STR локусов Y-хромосомы в этих популяциях были взяты из работ [Balanovsky et al., 2008; Roewer et al., 2008]. Вероятность ошибки (p-value) при сравнении указанных популяций рассчитывалась в программе Statistica 6.0.

## **ГЛАВА 3. Положение генофондов народов индоевропейской языковой общности по полногеномным данным в контексте Евразии: поиск сходства и различий**

### **3.1 Положение генофондов народов индоевропейской языковой общности в контексте современных популяций Евразии**

Нами был проведен анализ главных компонент для всех собственных данных, а также и для других популяций Западной Евразии, ранее изученных в мировой науке по полногеномной панели Genochip.

На построенном графике (рисунок 4) изученные популяции четко распределились на европейский и азиатский кластеры. Между двумя кластерами расположились лишь популяции татар, выступая своеобразным мостом. Уже по обнаруженной картине видно, что между популяциями, разговаривающими на индоевропейских языках восточных и западных ветвей мало генетического сходства. Поэтому далее рассмотрим закономерности для этих двух частей света по отдельности.

В «европейской» части графика проявляется несколько большее сходство друг с другом индоевропейских популяций, чем всех популяций Европы в целом. Обширное облако индоевропейских популяций ограничено слева облаком финно-угорских и тюркских популяций, а сверху – компактным баскским кластером. В этом контексте особенно важны различия между взаимным положением популяций в географическом пространстве и их положением в генетическом пространстве.

Так, французы и испанцы географически находятся по разные стороны ареала басков. Но на генетическом графике французы и испанцы находятся по одну, «индоевропейскую» сторону относительно басков, что хорошо согласуется с идеей, что баски представляют собой доиндоевропейское население Европы.

Несмотря на выраженное генетическое сходство басков с французами, их этнические облака хотя и примыкают друг к другу, но не совпадают. Сходным образом, русские и украинцы достаточно четко отделены от своих, не принадлежащих к индоевропейской семье, соседей (за исключением северных русских, для которых неоднократно была показана генетическая близость с финно-угорскими популяциями). Аналогично и литовцы, несмотря на географическую близость к эстонцам, на графике образуют своё «облако», чётко отделенное от их финно-угорских соседей.

В «азиатской» части графика также проявляются множественные сходства с лингвистической принадлежностью популяций. Удивительно генетическое положение осетин: географически расположенные между популяциями ингушей и кабардинцев и очень далеко от популяций Ирана, на генетическом графике осетины располагаются не между популяциями Северного Кавказа, а значительно сдвинуты в сторону образцов из Ирана. Этнические облака осетин и народов Ирана даже перекрываются, что согласуется с принадлежностью осетин к иранской группе индоевропейской семьи.

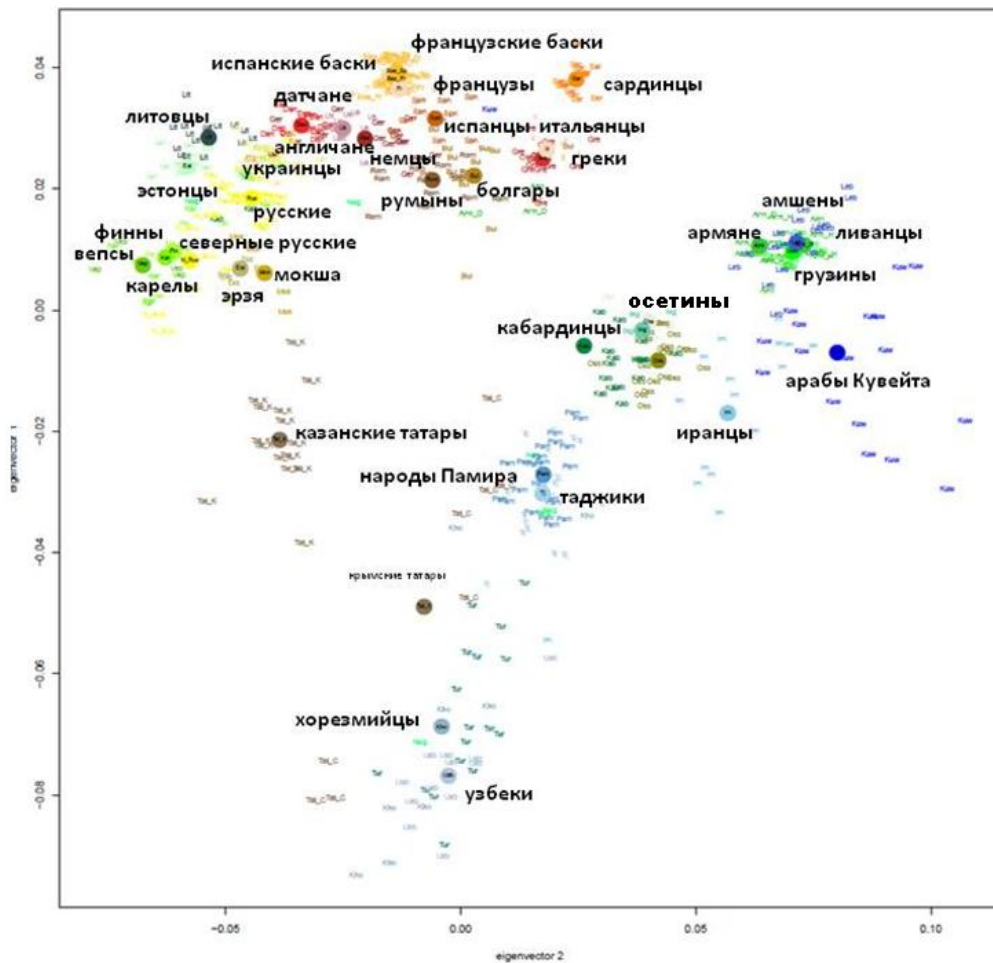


Рисунок 4. График главных компонент (данные по полногеномной панели Genochip).

На графике главных компонент, говорящие на языках индоевропейской языковой семьи народы Памира и таджики образуют отдельный кластер от соседних тюркоязычных народов (узбеков и туркмен). Важно отметить, что и жители Хорезма, чьими предками по историческим сведениям были хорезмийцы, разговаривавшие на языках иранской языковой группы, на графике располагаются рядом с узбеками. Таким образом, вероятно, современное население Хорезма не сохранило в своем генофонде следов древних хорезмийцев.

В целом, по проведенному нами анализу главных компонент для широкого круга популяций индоевропейской языковой семьи, видно, что генетические взаимоотношения изучаемых популяций значительно лучше объясняются

сочетанием географического и лингвистического факторов, чем одной лишь географией.

Далее мы провели Admixture анализ уже для более широкого набора популяций (рисунок 5). На данном графике мы специально разместили анализируемые нами популяции парами: популяция, разговаривающая на одном из языков индоевропейской языковой семьи, и её лингвистически контрастные географические соседи.

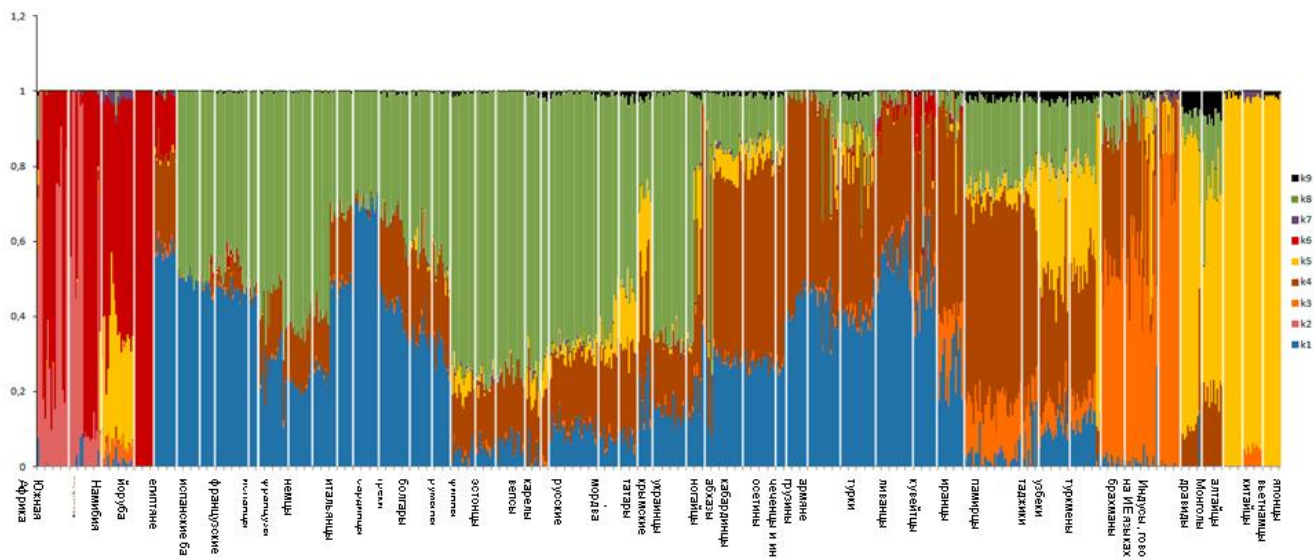


Рисунок 5. График Admixture для популяций Северной Евразии.

Анализ ADMIXTURE (выявление предковых компонентов) был проведен при разных  $k$ , (на рисунке 5 приведены результаты при  $k=9$ ). Как и в случае с анализом главных компонент, на данном графике можно видеть, что генетическую близость изучаемых популяций определяет не только их географическая близость, но также и лингвистическая.

Баски отличны от соседних испанцев и французов отсутствием коричневого переднеазиатско-кавказского компонента. Однако этот компонент присутствует не только у популяций, разговаривающих на языках индоевропейской языковой семьи, а также у финнов и тюркоязычных популяций.

Литовцы по графику Admixture очень похожи на своих географических соседей – эстонцев. Единственное отличие заключается в несколько меньшем присутствии у литовцев желтого, восточноазиатского компонента.

Несколько меньше, чем у окружающих популяций, указанный желтый компонент представлен и в генофонде русских. У окружающих их мордвы, татар, карел и вепсов он выражен значительно сильнее.

Для украинцев обнаружены достаточно сильные отличия от тюркоязычных крымских татар и ногайцев – у которых практически полностью отсутствует зеленый европейский компонент. У украинцев в свою очередь он составляет больше половины от всего пула компонент.

Единственное отличие армян от грузин, разговаривающих на языках картвельской языковой семьи, заключается в наличии у армян небольшой процентной доли зеленой европейской компоненты, отсутствующей у грузин. Интересно отметить, что турки очень похожи генетически по результатам данного анализа на армян. Что подтверждает сохранение в их генофонде мощного пласта предшествовавшего тюркской инвазии населения.

Памирские народы и таджики отличаются от своих географических тюркоязычных популяций, как это показали и результаты анализа главных компонент. У узбеков и туркмен ярко выражен желтый компонент, отсутствующий у памирцев и таджиков, по соотношению других компонент они довольно сильно отличаются.

Очевидно отличие и между народами Индии, разговаривающими на языках индоевропейской языковой семьи и дравидами: у дравидов почти 100% составляет оранжевый компонент, практически не встреченный у других популяций.

Однако обнаружить генетические отличия между изучаемыми нами комплексами сравнения оказалось возможно не для всех изучаемых пар: для осетин не обнаружено никаких заметных отличий от их географических соседей, принадлежащих к иным языковым семьям. Но это является скорее исключением –



как видно из рисунка 5 для большинства пар (популяция, разговаривающая на индоевропейском языке и лингвистически контрастная географически соседняя популяция) обнаружено отличие в соответствии с лингвистической принадлежностью.

### **3.2 Сравнение современных генофондов народов индоевропейской языковой общности с древним населением Евразии**

Особый интерес представляет сравнение генофонда современных популяций индоевропейской языковой семьи с древними образцами, собранными с территорий предположительной локализации прародины индоевропейцев и совпадающие по временным рамкам со временем существования индоевропейской прародины.

Ведь, как показано во многих масштабных исследованиях последних лет, генофонд народонаселения менялся на протяжении времени [Lazaridis et al., 2014].

Особый интерес в связи с этой проблемой представляет изучение генетических взаимоотношений между современными популяциями, разговаривающими на индоевропейских языках, населением ямной культуры и населением Анатолии, проживавшем там 6-7 тыс. лет назад.

Для анализа положения современных образцов относительно массива секвенированных с достаточным покрытием образцов древней ДНК нами был проведен анализ главных компонент (рисунок 6). На графике заметны четкие кластеры, образованные как только современными, так и совместно древними и современными образцами. Так, на графике четко выделяется Кавказско-переднеазиатский кластер, обозначенный сиренево-розовыми цветами. Между этим кластером и «облаком» образцов сицилийцев располагаются образцы из Европы (Испании) времен неолита. Подобный факт не вызывает удивления, поскольку во многих ранних исследованиях показано, что сицилийцы лучше

других популяций отражают генофонд переселенцев из стран плодородного полумесяца, принесших в Европу технологии земледелия.

Отдельные кластеры формируют и таджики с памирцами и туркмены с узбеками. Все древние образцы располагаются обособленно от них, поскольку нами не были включены в анализ древние образцы из этого региона. Образцы из Индии также располагаются в стороне от остальных, как древних, так и современных образцов. Причем на графике заметны два субкластера: образованный дравидийскими популяциями Индии и популяциями (показаны темно серыми квадратами), принадлежащими к индоевропейской языковой семье (показаны оранжевыми квадратами).

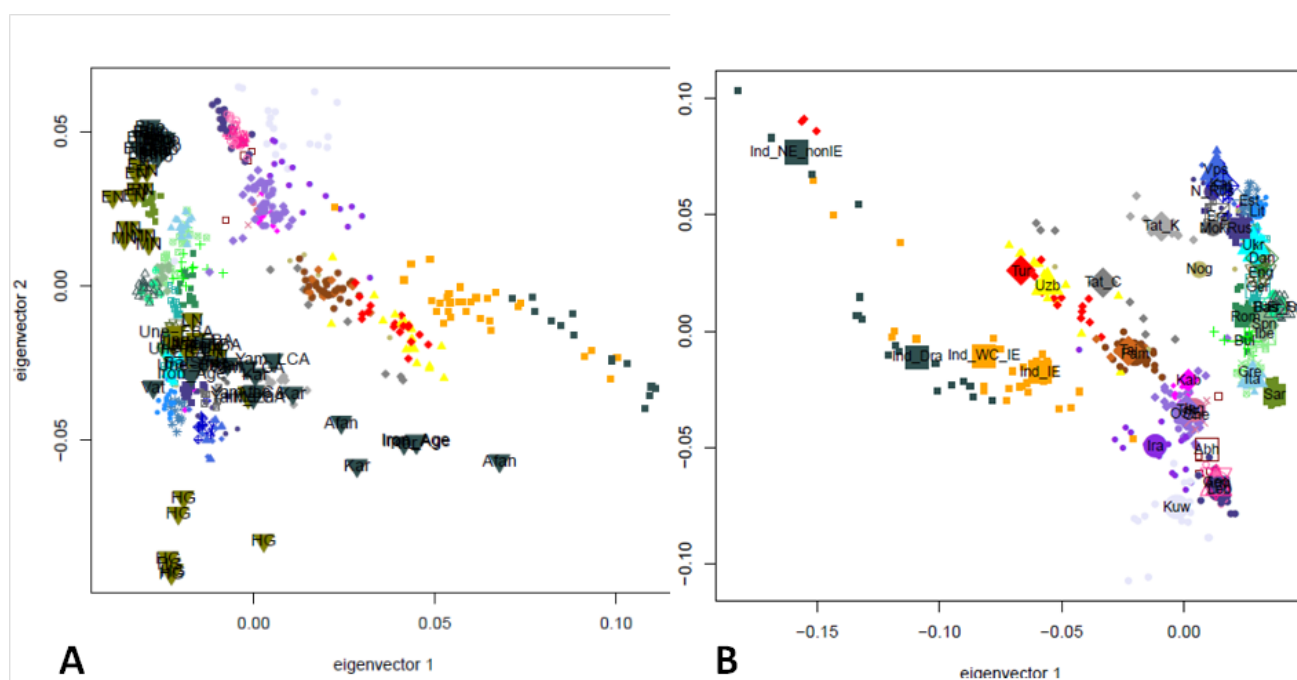


Рисунок 6. График главных компонент: сравнение современных популяций Евразии с древним населением этого региона

*Обозначения: А – график главных компонент, где древние образцы даны крупными темно-серыми и болотного цвета фигурами с подписями, а современные образцы представлены разноцветными геометрическими фигурами без подписей. Цвет и форма фигур соответствуют цвету и форме фигур на рисунке В, где даны только современные популяции с подписями. Где:*

*Abh* – абхазы, *Arm* – армяне, *Bas\_Fr* – французские баски, *Bas\_Sp* – испанские баски, *Bul* – болгары, *Che* – чеченцы, *Dan* – датчане, *Eng* – англичане, *Erz* – эрзя, *Est* – эстонцы, *Fin* – финны, *Fr* – французы, *Geo* – грузины, *Ger* – немцы, *Gre* – греки, *Ibe* – испанцы, *Ind\_Dra* – индусы разговаривающие на дравидийских языках, *Ind\_IE* – индусы, разговаривающие на языках индоевропейской языковой семьи, *Ind\_NE\_nonIE* – индусы северо-восточной Индии, не принадлежат к индоевропейской языковой семье, *Ind\_WC\_IE* – индусы центральной и западной Индии, не принадлежат к индоевропейской языковой семье, *Ing* – ингуши, *Ira* – иранцы, *Ita* – итальянцы, *Kab* – кабардинцы, *Kar* – карелы, *Kiw* – кувейтцы, *Leb* – ливанцы, *Lit* – литовцы, *Mok* – мокша, *N\_Rus* – северные русские, *Nog* – ногайцы, *Ose* – осетины, *Pat* – памирцы, *Rom* – румыны, *Rus* – русские, *Sar* – сардинцы, *Spn* – испанцы, *Taj* – таджики, *Tat\_C* – крымские татары, *Tat\_K* – казанские татары, *Trk* – турки, *Tur* – туркмены, *Ukr* – украинцы, *Uzb* – узбеки, *Vps* – венсы.

Древние популяции обозначены: *Afan* – Afanasievo [Allentoft et al., 2015]; *Andr* – Andronovo [Allentoft et al., 2015], *BA* – BA [Allentoft et al., 2015]; *Bat\_Axe* – Battle\_Axe/Nordic\_LN+ Battle\_Axe [Allentoft et al., 2015]; *Bell\_B* – Bell\_Beaker [Allentoft et al., 2015]; *Bho* – Barcin Höyük [Matheison et al., 2016]; *Cord\_W* – Corded\_Ware [Allentoft et al., 2015]; *EBA* – Halberstadt\_LBA+ EBA [Allentoft et al., 2015; Haak et al., 2015]; *EN* – LBK\_EN+ LBKT\_EN + Starcevo\_EN+ Spain\_EN [Haak et al., 2015]; *HG* – Motala\_HG+ Karelia\_HG+ Samara\_HG [Haak et al., 2015]; *Iron\_Age* – Iron\_Age [Allentoft et al., 2015]; *Kar* – Karasuk [Allentoft et al., 2015]; *LBA* – LBA [Allentoft et al., 2015]; *LN* – Corded\_Ware\_LN+ BenzigerodeHeimburg\_LN+ Bell\_Beaker\_LN+ Corded\_Ware\_LN+ Alberstedt\_LN+ Karsdorf\_LN [Haak et al., 2015]; *Mar* – Maros [Allentoft et al., 2015]; *MBA* – MBA [Allentoft et al., 2015]; *Mezh* – Mezhovskaya [Allentoft et al., 2015]; *Mho* – Menteşe Höyük; *MN* – Esperstedt\_MN+ Spain\_MN+ Baalberge\_MN [Haak et al., 2015]; *Nor\_BA* – Nordic\_BA [Allentoft et al., 2015]; *Nor\_LBA* – Nordic\_LBA; *Nor\_LN* – Nordic\_LN [Allentoft et al., 2015]; *Nor\_MN\_B* – Nordic\_MN\_B [Allentoft et al., 2015]; *Oku* – Okunevo [Allentoft et al., 2015]; *Une-EBA* – Unetice\_EBA+ Unetice [Allentoft et al., 2015]; *Vat* – Vatyа [Allentoft et al., 2015]; *Yam\_LCA* – Yamnaya [Allentoft et al., 2015; Haak et al., 2015].

В целом, как и графике главных компонент, построенном без учета данных по древней ДНК (рисунок 4), генетические взаимоотношения популяций определяются главным образом географией – или иными словами географической близостью проживания популяций друг относительно друга. Основная масса древних образцов располагается вокруг европейского кластера. В соответствии с результатами ранних

работ [Lazaridis et al., 2014; Haak et al., 2015], генофонд современной Европы сформировался из трех источников, сложение которых привело к его современному виду. На рисунке 6 все современные популяции Европы расположены между этими тремя «источниками»: между охотниками-собираателями Европы, земледельцами неолита и населением причерноморских степей - представителями ямной культуры. При этом население Южной Европы ближе всего к земледельцам времен неолита, население Северной Европы - к охотникам-собираателям, а население причерноморских степей – к представителям ямной культуры.

Но больше всего нас интересует расположение образцов представителей ямной культуры на графике главных компонент – видно, что они размещаются близко к образцам казанских татар, русских Центральной и Южной России, украинцам и ногайцам. Из чего можно сделать заключение о том, что население современных причерноморских степей и более отдаленных территорий сохранило генофонд своих предшественников. Однако это еще не дает нам никаких доводов в пользу степной гипотезы локализации прародины индоевропейцев. Поскольку сходство с образцами ямников демонстрируют не только представители индоевропейской языковой семьи, но и тюркоязычные популяции: ногайцы, татары. Правда, в работе Ynusbaev et al., 2015 было показано, что распространение тюркских языков шло с небольшим изменением генофонда, по преимуществу по модели элиты. Т.е. обнаруженное сходство может отражать сохранение в генофонде указанных популяций «дотюркского» пласта.

Однако обнаруженная нами закономерность, указывает на то, что использование ногайцев и татар в качестве пары для украинцев и русских может не позволить идентифицировать маркеры, присущие лишь народам, разговаривающим на языках индоевропейской языковой семьи.

Далее нами был проведен расчет евклидовых расстояний от древних образцов представителей ямной культуры (3000-2400 до н.э.), и образцов из Анатолии (6700-6600 до н.э.), из культурного комплекса Барсин-Хуююк, расположенного в районе Енишехир (провинция Бурса) в северо-западной Анатолии (рисунок 7-8). Конкретные

образцы для расчета расстояний были отобраны по параметрам максимального покрытия, пересечения надежно отсеквенированных участков древних образцов с полногеномной панелью GenoChip.

Идея этого анализа заключалась в том, что минимальные генетические расстояния должны отделять популяции индоевропейской языковой семьи от образцов людей, проживавших на территории гипотетической прародины индоевропейцев.

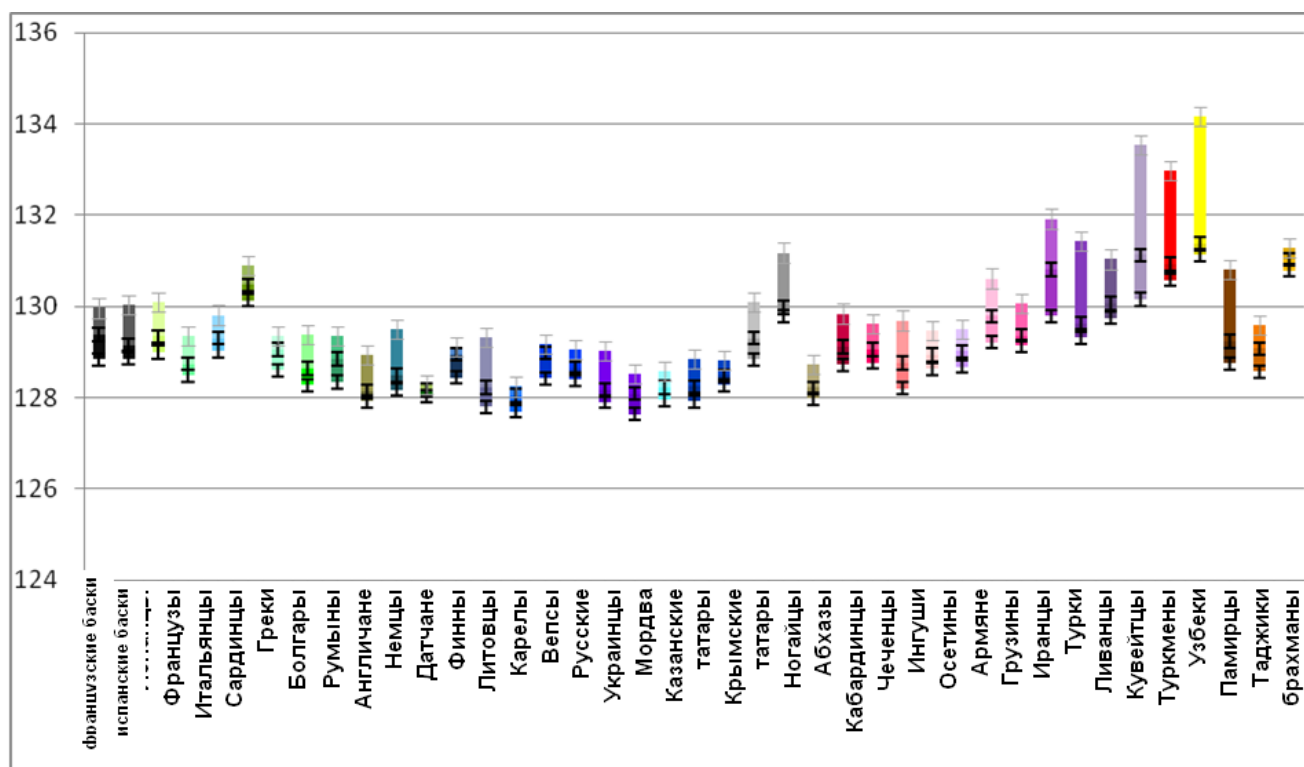


Рисунок 7. Евклидовы расстояния от образца ERR844324 древней ДНК ямной культуры из Allentoft et al., 2015 до современных популяций, генотипированных по полногеномной панели GenoChip

Образцы из Анатолии рассматривались нами в качестве населения наиболее близкого к гипотетической индоевропейской прапопуляции согласно Анатолийской гипотезе. А образцы представителей ямной культуры рассматривались нами в качестве населения наиболее близкого к гипотетической индоевропейской прапопуляции согласно степной гипотезе. Соответственно от народов, не принадлежащих к индоевропейской языковой семье, мы ожидали большие расстояния от образцов с предполагаемых прародин.

На рисунке 7 не выделяются популяции близкие к анализируемому образцу, однако на уровне общих тенденций видно, что минимальными являются расстояния от образца ямной культуры до русских, украинцев, литовцев, популяций Волго-Уральского региона. А максимальными – до народов Азии и Кавказа. Однако больших расстояний от народов финно-угорской языковой семьи нами также не обнаружено – финны, эстонцы, карелы ведут себя в целом схоже с указанными выше народами индоевропейской языковой семьи. Сходным образом ведут себя и другие проанализированные нами образцы ямников.

Для образцов времен неолита из Анатолии картина не столь однозначна, как для ямников. Образцы из Центральной Азии по-прежнему генетически далеки от анализируемых образцов. А наибольшее сходство с ними демонстрируют сардинцы, ливанцы, население Южной Европы и армяне – т.е. население в наибольшей степени, сохранившее следы неолитических миграций связанных с распространением технологии земледелия. Однако грузин отделяют почти такие же расстояния, как и армян от древних образцов из Анатолии. Что заставляет предположить, что обнаруженное сходство связано с миграциями населения, связанными с распространением технологий земледелия.

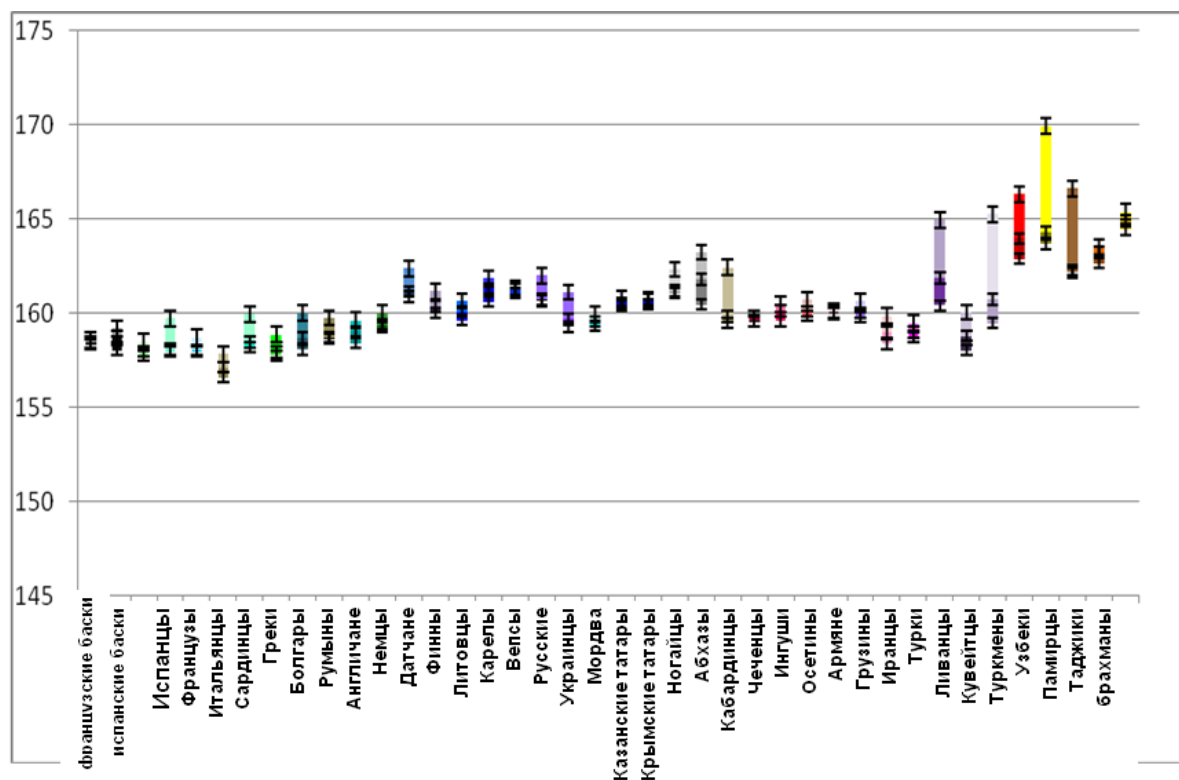


Рисунок 8. Евклидовы расстояния от образца I0708 древней ДНК, из Анатолии, культурный комплекс *Varcin Höyük* расположен в районе Енишехир (провинция Бурса) в северо-западной Анатолии из Matheison et al., 2015 до современных популяций, генотипированных по полногеномной панели GenoChip.

Таким образом, сравнение с древними геномами также не может дать однозначного ответа, откуда шло расселение народов индоевропейской языковой семьи. Однако в целом, полученные нами данные о сравнении с древней ДНК свидетельствуют о том, что миграции с территории, предполагаемой Анатолийской гипотезой в качестве прародины индоевропейских языков, были связаны в большей степени с распространением техники земледелия. Они не привели к появлению изменений в генофондах народов индоевропейской семьи не то, что на всем ареале её распространения, но даже и только у представителей западных её ветвей. Что касается подтверждения степной гипотезы данными сравнения с древней ДНК, то находится большее количество её подтверждений – картина сравнения современных образцов с образцами древней ДНК ямников

объясняет распространение части западных языков индоевропейской языковой семьи.

### **3.3 Тестирование основных гипотез локализации индоевропейской прародины на соответствие полученным генетическим данным**

Нами была рассчитана по полногеномным данным матрица  $F_{ST}$  расстояний для всех популяций Евразии, информация по генотипированию которых по панели Genochip, доступна на сегодняшний день. На основе полученной матрицы мы построили серию карт генетических расстояний от популяций с мест локализации предполагаемой прародины индоевропейцев (рисунок 9).

Так, мы рассмотрели в качестве возможной прародины причерноморские степи [Gimbutas, 1977]. Для рассмотрения этой гипотезы, генетические расстояния рассчитывались от украинцев – современного населения причерноморских степей. На рисунке 9А видно, что зона наибольшего генетического сходства с населением гипотетической прародины ограничивается Европой и Анатолией. С продвижением на восток заметно все меньше сходства с населением причерноморских степей. В целом эти результаты согласуются и с проведенным нами тестированием степной гипотезы с помощью образцов древней ДНК из ямной культуры. С применением степной гипотезы можно объяснить лишь распространение носителей индоевропейских языков на запад. В генофонде представителей индоевропейской языковой семьи на востоке не обнаружено близости с представителями ямной культуры.

Следующей мы рассмотрели Анатолийскую гипотезу [Renfrew, 1987] – в этом случае генетические расстояния считались усреднено от армян и турок, как наиболее вероятных современных популяций - потомков древнего населения Анатолии. Полученная карта (рисунок 9 Б) похожа на предыдущую, однако зона максимального сходства уже не распространяется на прибалтийские страны и север России, зато занимает практически все Средиземноморье, достигая

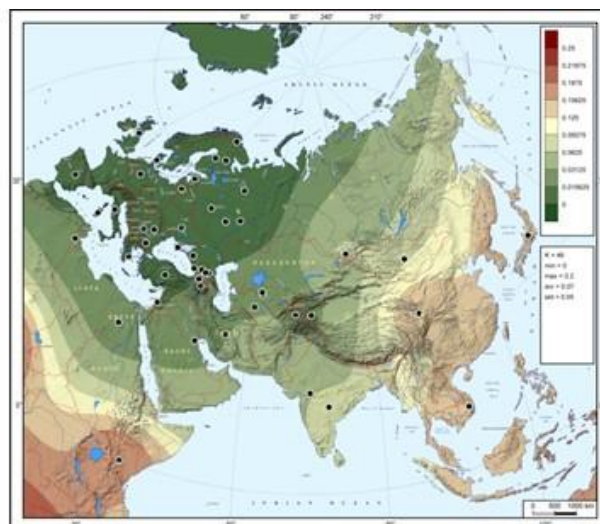


наиболее высоких значений в Леванте. На восток зона сходства так же не простирается – условной границей можно считать Каспийское море и Уральские горы.

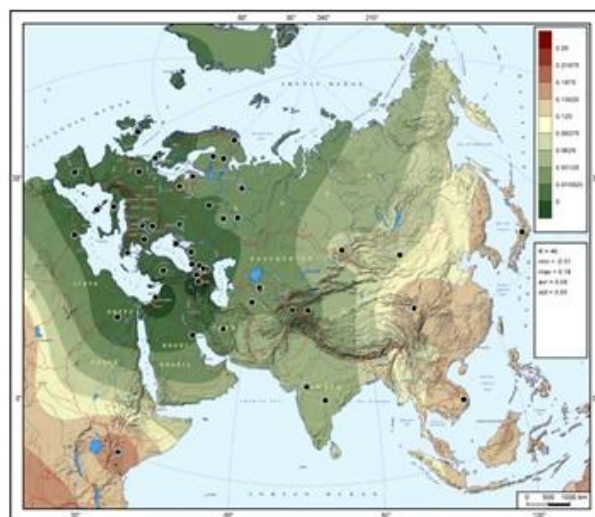
Отсутствие генетического сходства населения Анатолии с севером Европы может найти свое объяснение в разной лингвистической принадлежности населения этих регионов. Однако на карте генетических расстояний сходство с населением гипотетической прародины заметно и с населением Аравийского полуострова – можно сказать, что на карте видно два максимума сходства – на Закавказье и на Аравийском полуострове. А поскольку население Аравийского полуострова по преимуществу семито язычное, то обнаруживаемый на карте (рисунок 9 Б) паттерн сходства невозможно объяснить одним лишь расселением народов индоевропейской языковой семьи.

Третьей мы рассмотрели балканскую гипотезу локализации прародины индоевропейцев. Зона генетического сходства наиболее обширна, и включает всю Европу и Ближний Восток (рисунок 9 В). Любопытно отметить, что данная карта в наибольшей степени гомогенна и не позволяет выделить отдельные зоны максимального сходства.

**А.** Карта генетических расстояний от популяций причерноморских степей (полногеномные данные, картографированы  $F_{ST}$ -дистанции)



**Б.** Карта генетических расстояний от популяций Анатолии (полногеномные данные, картографированы  $F_{ST}$ -дистанции)



**В.** Карта генетических расстояний от популяций Балкан (полногеномные данные, картографированы  $F_{ST}$ -дистанции)

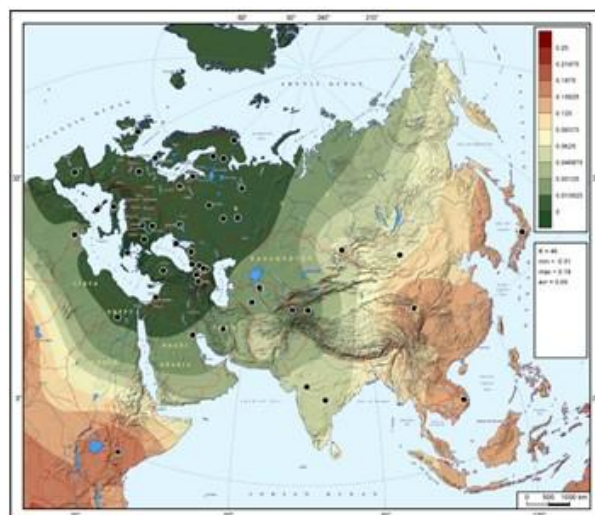


Рисунок 9. Карты генетических расстояний от современного населения проживающего на территории современных прародины индоевропейцев (предполагаемых согласно разным гипотезам локализации прародины).

В целом все три карты наглядно показывают, что ни одна из вероятных прародин не может претендовать на роль источника основного компонента в генофонде всех индоевропейских популяций. И какую бы прародину мы ни взяли, значительная часть индоевропейских популяций не вписывается в постулируемую ею схему миграций.

Для проверки того, насколько все же в количественном отношении лингвистическая подразделенность индоевропейской языковой семьи оказала влияние на структуру генофонда её представителей, нами был рассчитан тест Мантеля на основе данных полногеномного генотипирования по панели GenoChip (таблица 4).

Как и предполагалось, самый высокий коэффициент корреляции был обнаружен для генетики с географией – для данных о близости популяций индоевропейской языковой семьи, согласно этим двум системам, показатель корреляции составил 0,72. Однако данные генетики и лингвистики для популяций индоевропейской языковой семьи имеют достаточно высокий коэффициент корреляции – 0,56. Важно отметить, что показатели частной корреляции подтверждают эти данные – они равны 0,65 для генетики с географией (лингвистика остается постоянной) и 0,4 для генетики с лингвистикой (география остается постоянной).

Таблица 4. Тест Мантеля для определения корреляции матриц генетических и географических расстояний и матриц лингвистического сходства для популяций индоевропейской лингвистической семьи

<b>Тип корреляции</b>	<b>Параметры корреляции</b>	<b>Значения показателя</b>
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

*Примечание:* p-value<0,001

Таким образом, полученные нами результаты свидетельствуют о том, что генетическую структуру популяций индоевропейской языковой семьи определяет главным образом географический фактор. Однако лингвистическая близость также оказала значительное влияние на их генетическую структуру.

### **3.4 Поиск маркеров, характерных для народов индоевропейской языковой общности по полногеномным данным**

Современные индоевропейские популяции могли унаследовать от популяции, говорившей на прото-индоевропейском языке, не только язык, но и хотя бы небольшую часть генофонда, при этом у их не индоевропейских соседей такого генофонда должно быть меньше.

Отличительной особенностью проведенного нами поиска является то, что все имеющиеся данные разбиты на комплексы сравнения: популяция, принадлежащая к индоевропейской языковой семье и её иноязычные географические соседи. Всего для выявления маркеров индоевропейского протогенофонда использованы три критерия, описанных в главе 2.

Самому строгому критерию (присутствие/отсутствие) для всех пар не соответствовал ни один маркер, а для уменьшенного числа пар – лишь считанные маркеры из более ста тысяч исследованных маркеров. Критерию значительного преобладания по частоте соответствовали уже 639 маркеров (при условии соответствия критерию хотя бы в пяти парах). Самому мягкому третьему критерию соответствовали 5065 маркеров (для всех популяционных пар, кроме одной), при еще большем смягчении критерия число маркеров стремительно росло (таблица 5).

Также нами была проведена проверка полученных результатов (таблицы 6-9) – было проведено случайное объединение популяций и проведены аналогичные расчеты, как и при объединении популяций в комплексы сравнения.

Результаты получились аналогичными, как и при «осмысленном» объединении популяций. Что заставляет с осторожностью относиться к маркерам, которые могли бы быть связанными с распространением носителей индоевропейских языков. Результаты проверки представлены в таблице 5 в столбце «Количество обнаруженных маркеров при случайном объединении пар»

Таблица 5. Маркеры-кандидаты на роль «характерных для индоевропейских генофондов».

Количество комплексов сравнения, удовлетворяющих условию	<b>Присутствие/отсутствие.</b> Маркеры, присутствующие у индоевропейцев с любой частотой, отличной от 0, и отсутствующие у не-индоевропейцев	<b>Преобладание по частоте на 50% минимум.</b> Маркеры, присутствующие у индоевропейцев с частотой на половину превышающей частоту у не-индоевропейцев.	<b>Небольшое преобладание по частоте.</b> Маркеры, присутствующие у индоевропейцев с частотой, превышающей (не важно на сколько) частоту среди не-индоевропейцев из их комплекса.
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

Таблица 6. Проверка № 1 маркеров-кандидатов на роль «характерных для индоевропейских генофондов».

Количество комплексов сравнения, удовлетворяющих условию	<b>Присутствие/отсутствие.</b> Маркеры, присутствующие в группе 1 с любой частотой, отличной от 0, и отсутствующие у группы 2	<b>Преобладание по частоте в два раза.</b> Маркеры, присутствующие у группы 1 с частотой в два раза превышающей частоту у группы 2.	<b>Небольшое преобладание по частоте.</b> Маркеры, присутствующие у группы 1 с частотой, превышающей (не важно на сколько) частоту среди группы 2 из их комплекса.
Все комплексы сравнения	0	0	221
8 комплексов сравнения	0	3	3404
7 комплексов сравнения	0	12	17943
6 комплексов сравнения	11	170	56094
5 комплексов сравнения	88	966	117924
4 комплекса сравнения	493	3894	183080

Таблица 7. Проверка № 2 маркеров-кандидатов на роль «характерных для индоевропейских генофондов».

<b>Присутствие/отсутствие.</b> Маркеры, присутствующие в группе 1 с любой частотой, отличной от 0, и отсутствующие у группы 2	<b>Небольшое преобладание по частоте.</b> Маркеры, присутствующие у группы 1 с частотой, превышающей (не важно на сколько) частоту среди группы 2 из их комплекса.	<b>Присутствие/отсутствие.</b> Маркеры, присутствующие в группе 1 с любой частотой, отличной от 0, и отсутствующие у группы 2	<b>Преобладание по частоте в два раза.</b> Маркеры, присутствующие у группы 1 с частотой в два раза превышающей частоту у группы 2.
Все комплексы сравнения	0	0	209
8 комплексов сравнения	0	1	2869
7 комплексов сравнения	0	20	16509
6 комплексов сравнения	6	143	53713
5 комплексов сравнения	34	871	116169
4 комплекса сравнения	198	3912	182576



Таблица 8. Проверка № 3 маркеров-кандидатов на роль «характерных для индоевропейских генофондов».

Количество комплексов сравнения, удовлетворяющих условию	<b>Присутствие/отсутствие.</b> Маркеры, присутствующие в группе 1 с любой частотой, отличной от 0, и отсутствующие у группы 2	<b>Преобладание по частоте в два раза.</b> Маркеры, присутствующие у группы 1 с частотой в два раза превышающей частоту у группы 2.	<b>Небольшое преобладание по частоте.</b> Маркеры, присутствующие у группы 1 с частотой, превышающей (не важно на сколько) частоту среди группы 2 из их комплекса.
Все комплексы сравнения	0	1	374
8 комплексов сравнения	0	6	4015
7 комплексов сравнения	1	26	19920
6 комплексов сравнения	6	193	59869
5 комплексов сравнения	60	1092	122442
4 комплекса сравнения	372	4030	186282

Таблица 9. Усредненные показатели проверки маркеров-кандидатов на роль «характерных для индоевропейских генофондов».

Количество комплексов сравнения, удовлетворяющих условию	Присутствие/отсутствие. Маркеры, присутствующие в группе 1 с любой частотой, отличной от 0, и отсутствующие у группы 2	Преобладание по частоте в два раза. Маркеры, присутствующие у группы 1 с частотой в два раза превышающей частоту у группы 2.	Небольшое преобладание по частоте. Маркеры, присутствующие у группы 1 с частотой, превышающей (не важно на сколько) частоту среди группы 2 из их комплекса.
Все комплексы сравнения	0	0	268
8 комплексов сравнения	0	3	3429
7 комплексов сравнения	0	19	18124
6 комплексов сравнения	8	169	56559
5 комплексов сравнения	61	976	118845
4 комплекса сравнения	354	3945	183979

Далее мы построили серию карт распространения маркеров, обнаруженных по каждому из указанных в таблице вариантов. Наибольший интерес представляют карты, показанные на рисунках 10 и 11.

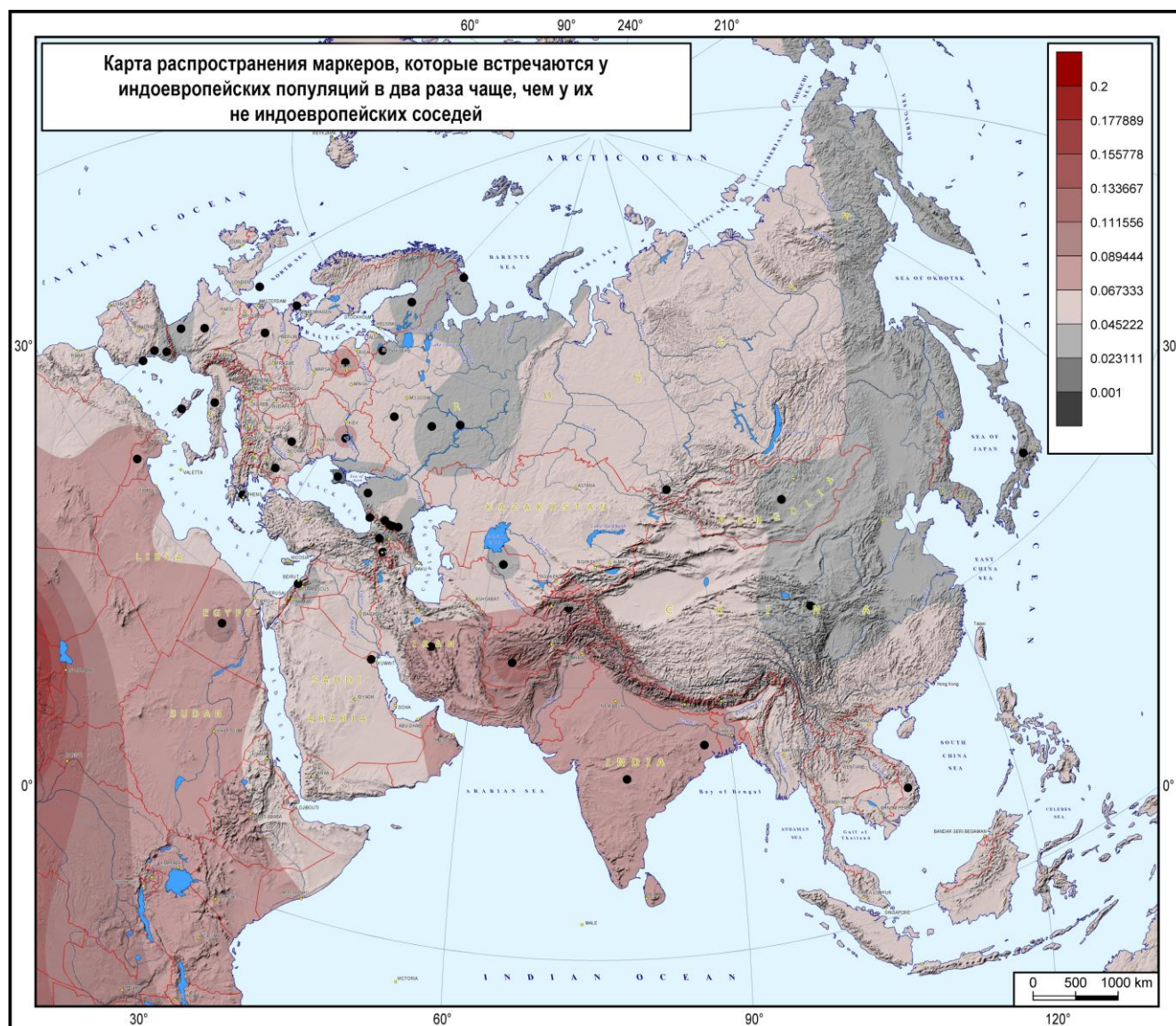


Рисунок 10. Карта распространения маркеров, которые встречаются у индоевропейских популяций в два раза чаще, чем у их не индоевропейских соседей (данное условие соблюдается не менее, чем во всех парах, кроме одной, картографирована средняя частота отобранных по критерию маркеров).

На рисунке 10 представлена карта средней частоты маркеров, удовлетворяющих критерию значительного различия по частоте. Карта показывает, что в целом отобранные маркеры действительно приурочены к индоевропейским популяциям: характерно снижение частоты почти до нуля у басков, финно-угорских народов Северо-Восточной Европы, у монголов, китайцев, японцев. В то же время максимальные частоты наблюдаются в Индии,

что не согласуется ни с одной из основных теорий происхождения индоевропейцев. Это может быть связано со сравнительно небольшим числом маркеров, удовлетворяющих такому критерию, среди которых случайно оказалось много приуроченных к народам Индии.

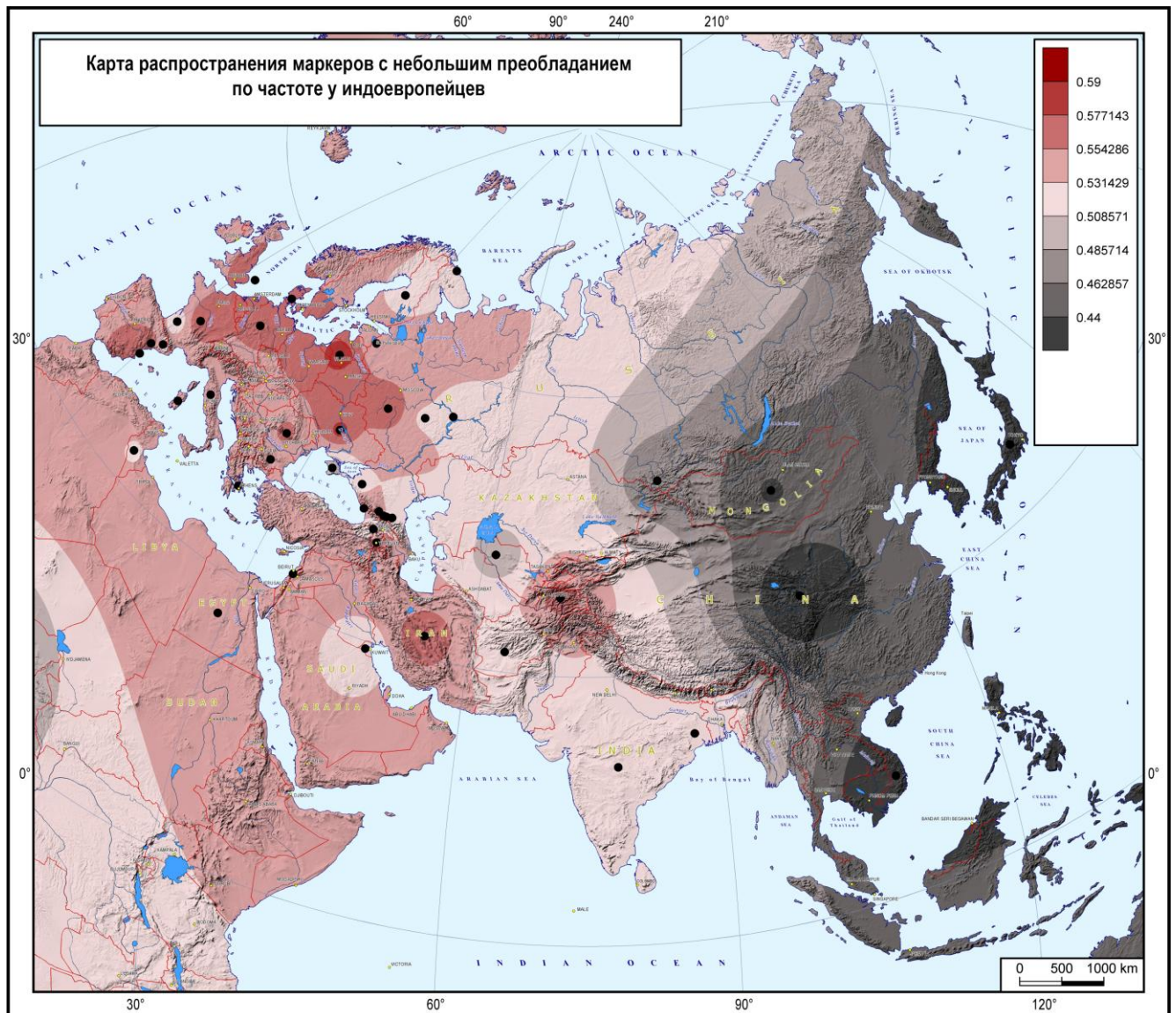


Рисунок 11. Карта распространения маркеров с небольшим преобладанием по частоте у индоевропейцев (данное условие соблюдается не менее, чем в семи парах, картографирована средняя частота отобранных по критерию маркеров).

На рисунке 11 показано распространение маркеров, удовлетворяющих критерию хотя бы небольшого преобладания по частоте в индоевропейских, чем в

парных к ним не индоевропейских популяциях. За счет большого числа маркеров различия выражаются не в контрасте нулевых и ненулевых частот, а в контрасте высоких и низких частот, но это не делает паттерн карты менее выразительным. Рельефно выступают те же закономерности: снижение частоты у не индоевропейских народов Европы (басков на западе, финно-угров на северо-востоке). Характерно также полное отсутствие данной группы маркеров в Восточной Евразии. И действительно, «индоевропейские» генетические маркеры могут быть только приуроченными к европеоидным, западноевразийским популяциям. Максимальные частоты потенциальных «индоевропейских» маркеров при данном критерии приурочены к Центральной и Восточной Европе. Такая локализация, хотя также не соответствует ни одной из двух основных гипотез индоевропейской прародины («степной» и «анатолийской»), но согласуется с рядом других гипотез.

Таким образом, самый строгий критерий из более чем ста тысяч изученных генетических маркеров не выявил ни одного чисто «индоевропейского» маркера. Более мягкие критерии, разрешающие маркеру быть характерным для индоевропейцев только в нескольких комплексах сравнения, выявили большое число маркеров, которые действительно более свойственны индоевропейским популяциям. Однако при разных критериях географическая зона накопления «индоевропейских» маркеров варьирует от Центральной Европы до Северной Индии, не позволяя использовать эти данные для верификации гипотез о месте прародины индоевропейцев.

В целом надо признать, что предложенный нами подход к выявлению кандидатных «индоевропейских» маркеров действительно позволяет отбирать из множества генетических маркеров те, которые в наибольшей степени приурочены к современным индоевропейским популяциям. Но различные критерии отбора выявляют группы маркеров с разными паттернами географического распределения, которое к тому же может отражать не локализацию исходного протогенофонда индоевропейцев, а «центр тяжести» генофондов более поздних

эпох, например, генетических компонентов, свойственных индоевропейским популяциям только Европы или только Азии.

То есть наиболее вероятно, что нет таких 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 результаты исследования свидетельствуют, что даже в этом случае их доля будет крайне мала.

## **ГЛАВА 4. Генофонд народов балто-славянской языковой группы по маркерам Y-хромосомы**

### **4.1. Общая характеристика генофонда балто-славянских популяций**

В раннем средневековье произошло быстрое распространение представителей славянской ветви балто-славянской языковой семьи по территории Центральной и Восточной Европы. Степень генетического вклада славянского населения в автохтонный генофонд, существовавший на этих территориях, до сих пор не известна. Именно это мы и намеревались исследовать в своей работе, изучив опубликованные данные по полиморфизму Y-хромосомы у 6 079 образцов из популяций всех славянских народов.

Для того, чтобы проследить положение славян относительно их географических соседей, принадлежащих к другим лингвистическим группам, нами была построена дендрограмма на основе генетических расстояний между изученными популяциями. Так, для народов Северного Кавказа в работе [Balanovsky et al., 2011] было показано четкое совпадение филогенетического дерева с лингвистическим. Для славянских народов получить столь четкую картину не удалось, лингвистическое родство у них меньше связано с генетической близостью (рисунок 12).

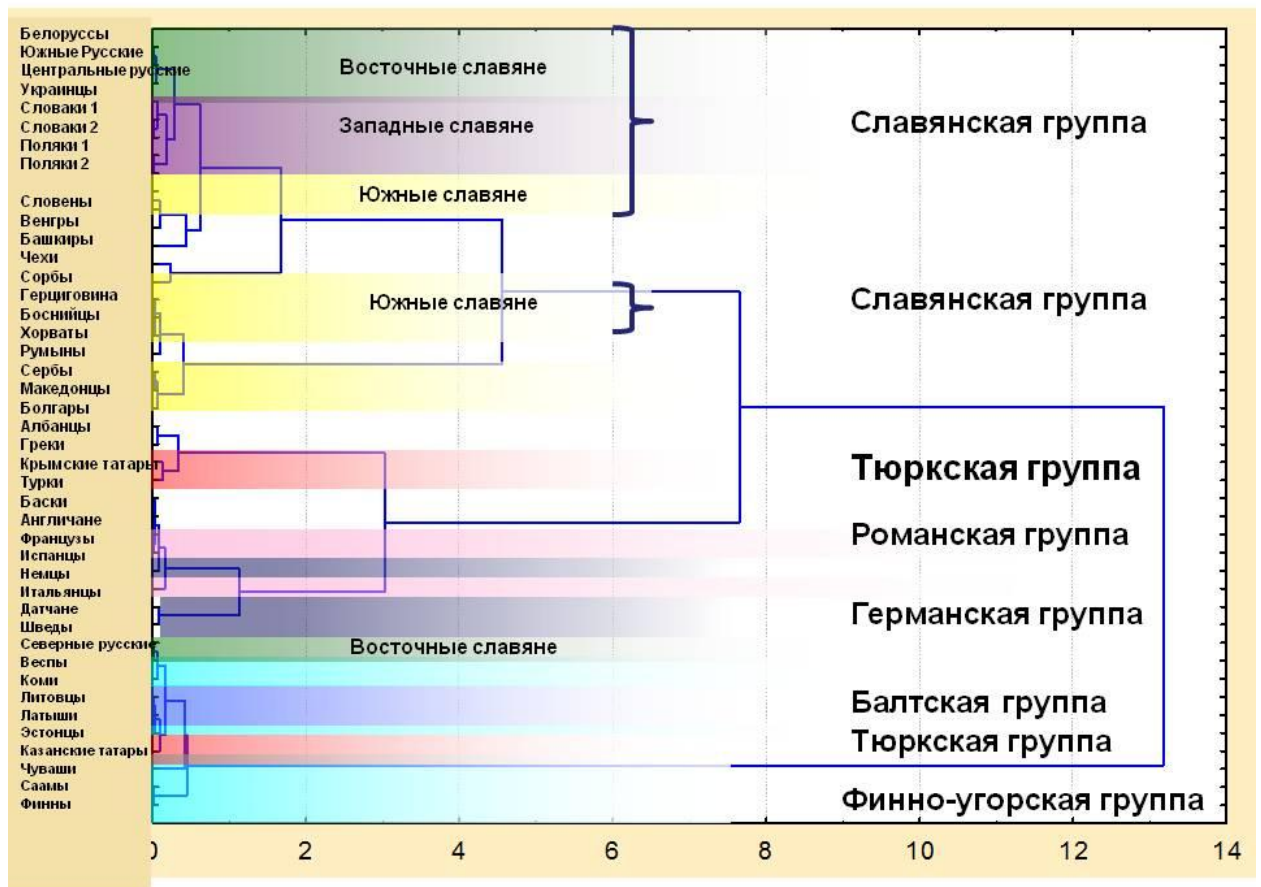


Рисунок 12. Дендрограмма генетических взаимоотношений славян в контексте окружающих лингвистически контрастных популяций.

Дендрограмма генетических взаимоотношений западных индоевропейских популяций и соседних народов, принадлежащих к иным лингвистическим семьям, позволяет выделить определенные кластеры, приуроченные к отдельным лингвистическим группам: так восточные и западные славяне, а также популяции, говорящие на языках балтской группы, образуют обособленные генетические кластеры. Но многочисленны и некоторые несовпадения лингвистических и генетических данных: северные русские, лингвистически являющиеся славянами, генетически ближе к балтам. Неоднороден по лингвистической принадлежности и кластер балканских популяций: в общий кластер вместе с южными славянами попали венгры, чехи и румыны. Но наиболее наглядно такие несовпадения видны на графике многомерного шкалирования (рисунок 13), построенного нами для тех же популяций, вновь на основе данных о генетических расстояниях между ними.



Анализ графика показал, что славяне образуют отдельный кластер, хотя даже в пределах этого кластера можно заметить отдельные исключения (несовпадения лингвистической и генетической реконструкций). Народы, принадлежащие к германской ветви, и народы романской группы также образуют отдельные кластеры среди окружающих популяций. Интересно отметить, что столь географически разбросанные тюркские народы сформировали на графике отдельную группу. Но как уже отмечалось ранее, несовпадения генетического и лингвистического объединения народов также нередки: давно известна ситуация с венграми, принадлежащими к финно-угорской языковой семье, но генетически близкими к южным славянам (своим географическим соседям). Но, несмотря на ряд этих несовпадений, вырисовывается достаточно четкая картина генетической структурированности западных индоевропейцев, которая может быть предсказана по лингвистическим данным. Это достаточно четко проявляется на графике (рисунок 20), который оказывается возможным подразделить на зоны (показаны красным пунктиром) в пределах каждой из которых преобладают популяции определенной лингвистической группы.

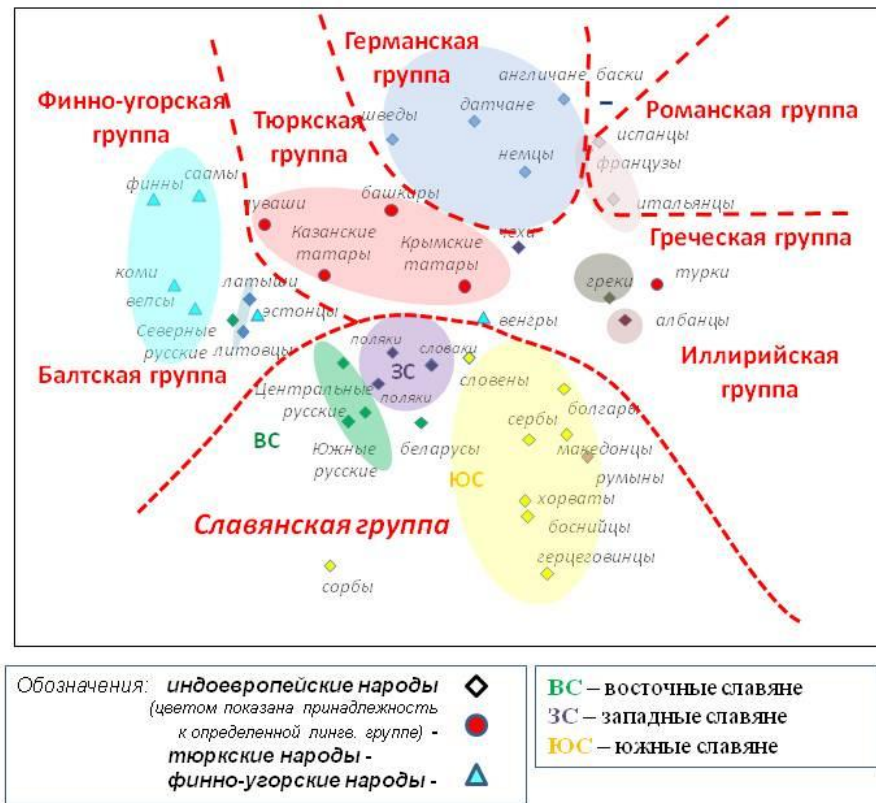


Рисунок 13. График многомерного шкалирования генетических взаимоотношений славян в контексте окружающих лингвистически контрастных популяций.

Далее нами была подробно разобрана генетическая структура славянских популяций в узком кругу окружающих популяций. Для этого нами были рассчитаны генетические расстояния  $D_{NEI}$  по данным полиморфизма гаплогрупп Y-хромосомы. По генетическим расстояниям четко прослеживается широтное распределение популяций, в соответствие с их происхождением. Восточные славяне – русские Центральной и Южной России близки с украинцами и белорусами. При этом генетические расстояния между ними являются довольно незначительными: ( $D_{Nei} = 0.125$ ). Северные русские, напротив, резко отличны от остальных восточных славян и тяготеют к финноязычным популяциям. Соответственно генетические расстояния между северными русскими и

остальными восточными славянами существенно больше и составляют  $D_{Nei}=0.5844$ .

В сравнении с восточными славянами, западные славяне не столь схожи друг с другом. Например, чехи и словаки тяготеют к германским популяциям, в то время как поляки крайне близки генетически к восточным славянам. Таким образом, средние попарные расстояния между западными славянами в два раза выше для западных славян, относительно восточных ( $D_{Nei} = 0.241$ ).

Таблица 10. Генетические расстояния ( $D_{NEI}$ ) для балто-славянских и окружающих популяций на основе данных полиморфизма Y-хромосомы

	Рус	С.рус	укр	бел	болг	макед	Босн., хорв и сербы	пол	чех	слов	лит	лат
Русские	0,00	0,22	0,06	0,00	0,52	0,62	0,52	0,04	0,18	0,05	0,17	0,20
Сев.рус.		0,00	0,39	0,27	0,95	1,07	0,98	0,37	0,52	0,39	0,01	0,01
украинцы			0,00	0,04	0,26	0,29	0,20	0,10	0,18	0,04	0,35	0,39
белорусы				0,00	0,49	0,57	0,46	0,04	0,17	0,03	0,22	0,25
болгары					0,00	0,02	0,16	0,50	0,42	0,37	0,96	1,01
Македонцы						0,00	0,09	0,65	0,55	0,45	1,12	1,18
боснийцы, хорваты и сербы							0,00	0,64	0,61	0,39	1,06	1,11
поляки								0,00	0,07	0,03	0,30	0,31
Чехи									0,00	0,08	0,48	0,44
словаки										0,00	0,34	0,36
литовцы											0,00	0,01
латыши												0,00

Примечательно то, что добавив поляков к расчету средних попарных генетических расстояний между русскими центральной и южной России, украинцами и белорусами показатель почти не меняется -  $D_{Nei}=0.14389$ . Это означает то, что, несмотря на принадлежность к разным лингвистическим

группам славян, на обширном географическом ареале славянские популяции имеют невысокий уровень дифференциации.

Большинство южных славян отделяется от остальных балто-славянских популяций. Средние попарные генетические расстояния для южных славян составили ( $D_{Nei}=0.2387$ ), что сопоставимо и даже выше аналогичных значений для восточных славян. И это, несмотря на то, что южные славяне занимают достаточно географически обособленный Балканский полуостров.

Обе популяции балтов – латыши и литовцы, генетически близки к эстонцам, разговаривающим на финно-угорском языке.

Нами был проведен анализ AMOVA для того чтобы определить, какой уровень лингвистической подразделенности в наибольшей степени близок реальной генетической структуре изученных популяций. Наименьшее генетическое разнообразие отмечено нами для локальных популяций, разговаривающих на одном и том же языке. Он варьирует от 0 для чехов и македонцев до 0.05 для северных русских. Дифференциация между популяциями, относящимися к одной и той же лингвистической ветви, составила 0,03. Вариация среди ветвей – 0,06.

Также нами был проведен тест Мантеля для выяснения степени влияния лингвистических и географических расстояний на структуру генофонда популяций балто-славянской лингвистической группы (таблица 11). Максимальная корреляция нами обнаружена для генетики с географией - 0,924. При этом лингвистический фактор повлиял на структуру генофонда славян гораздо меньше. Это видно уже на парных корреляциях, тогда как при анализе частных корреляций роль лингвистического фактора становится практически нулевой.

Таблица 11. Тест Мантеля для оценки 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 ( $p < 0,05$ )

Результаты теста Мантеля и другие проведенные виды анализа единогласно указывают, 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 преимущественно культурным и не оставило заметных следов в генофонде коренного населения Европы. Однако для отдельных групп славян показано, что генетическая близость совпадает с лингвистической подразделенностью. Это позволяет все-таки сделать заключение о том, что распространение славянских языков шло не только путем культурной передачи языка, но также сопровождалось и миграциями населения, пусть и оставившими слабый след в генофонде [Kushniarevich et al., 2015].

## 4.2. Генофонд населения Верхнего Поволжья

Верхнее Поволжье некогда являлся регионом пограничным для населения, говорившего на финно-угорских, славянских и скандинавских языках. В связи с этим особенно интересно изучить современный генофонд населения этого региона, с целью обнаружения вклада разных групп населения. Мы сосредоточились на изучении генофонда популяций Ярославской области – наименее изученного региона Верхнего Поволжья по маркерам Y-хромосомы. Поскольку русское население Ярославской области подразделялось на ряд этнографических групп, мы изучили этнографические группы сицкарей, кацкарей и бывшее население г. Мологи, которое согласно некоторым версиям археологов могло сохранить древние мерянские пласты. Четвертая популяция представляет сборную выборку из разных районов Ярославской области и дает точку отсчета - общую характеристику ее современного коренного русского населения, не подразделенного на этнографические группы. А все четыре популяции вместе позволяют охватить значительную часть генетического разнообразия Ярославского Поволжья.

Анализ суммарной выборки (N=132) позволяет создать общий генетический портрет русских Ярославской области. Он представлен тремя наиболее частыми гаплогруппами: R-M198(xM458) (34%); R-M458 (24%); N-M178 (11%), включая два субварианта этой гаплогруппы - N3a3-CTS10760 и N3a4-Z1936. Еще десять гаплогрупп встречены с частотой менее 10%: I-M253 (8%); R-L23 (7%); I-P37.2 (5%); E-M35 (2%); E-M78 (2%); J-P58 (2%); G-M406 (1%); I-M170(xM253, M223, P37.2) (1%); J-M172 (1%); J-M92 (1%) [Чухряева и др., 2017]. Такой спектр гаплогрупп типичен для русских средней полосы России.

На графике многомерного шкалирования, построенном по панели из 16 гаплогрупп Y-хромосомы (рисунок 14), две из четырех популяций Ярославской

области – кацкари и сборная выборка – вошли в один тесный кластер вместе с большинством восточнославянских популяций. К этому кластеру близки популяции Мордовии (мокша и эрзя). Третья популяция Ярославской области - сицкари - отличается генетическим своеобразием и несколько обособлена от других и славянских, и финно-угорских популяций. Четвертая популяция Ярославской области - мологжане – вошла в финно-угорский кластер, включивший также популяции Русского Севера (по данной панели маркеров изучено население Псковской и Новгородской областей).



Рисунок 14. Население Ярославской области в контексте окружающих популяций: график многомерного шкалирования по данным субтипирования 16 гаплогрупп: A-M91, B-M60, C-M130, D-M174, E-M35, G-M201, I-M170(xM253, M223, P37.2), I-M253, I-M223, I-P37.2, J-M267, J-M172, N-LLYG, R-M198(xM458), R-M458, R-M269 (показатель стресса = 0.086; алиенации = 0.102).

Поскольку для большинства финно-угорских популяций имеются данные лишь по узкой панели из 13 гаплогрупп Y-хромосомы, без субтипирования вариантов гаплогрупп R и I, важных для понимания генетической истории региона, эти популяции отражены на графике многомерного шкалирования по узкой панели маркеров (рисунок 15). Из четырех четких кластеров графика -

угорского, финно-волжского, восточнославянского и северо-восточной Европы (включившего и прибалтийских финнов, и балтов, и популяции Русского Севера) наибольшей специфичностью обладает угорский кластер - в него вошли лишь популяции угорской языковой группы: ханты и манси.

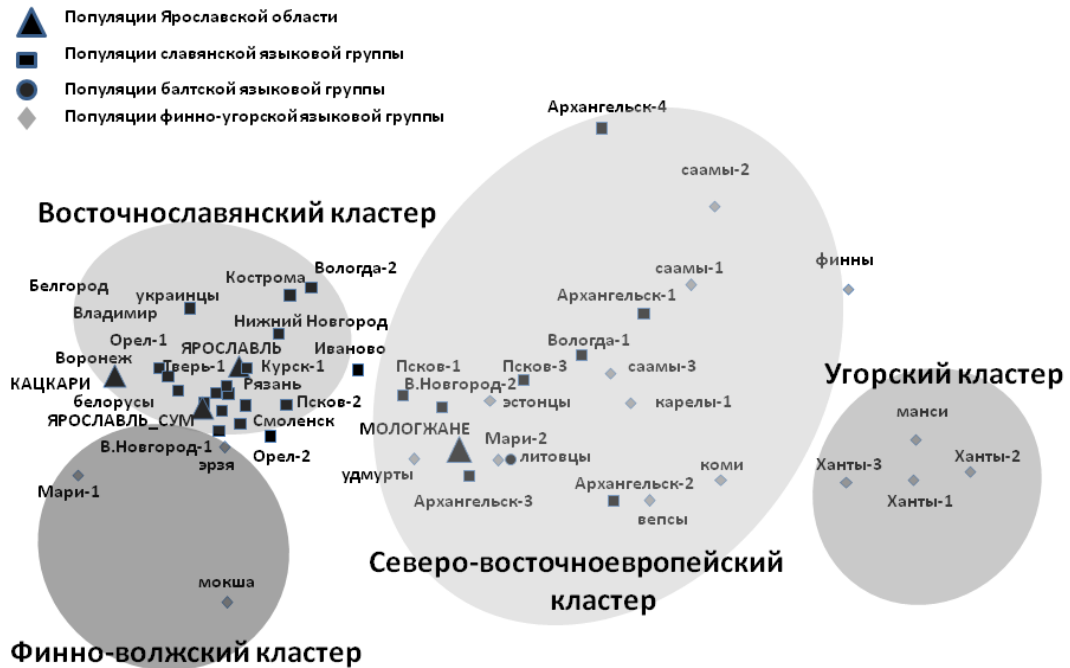


Рисунок 15. Население Ярославской области в широком контексте окружающих популяций. График многомерного шкалирования по панели 13 крупных гаплогрупп: С-M130, DE-P205, G-L116, H-M69, I-M170, J-M304, L-M11, N-M231, O-M175, Q-M242, R1-M173, R2-L261, T-L206 (показатель стресса = 0.034; алиенации = 0,038).

Популяции Ярославской области расположились так же, как и на предыдущем графике: мOLOGЖАНЕ кластеризуются с популяциями русского Севера, Новгородской области и финно-язычными народами, что может указывать на наличие в генофонде мOLOGЖАН довольно мощного пласта дославянского населения [Балановская и др., 2011].



Все три другие популяции Ярославской области – кацкари, сборная и суммарная - входят в общий кластер русских популяций Центральной и Южной России, украинцев и белорусов. Отличие их генофонда от северных русских популяций и финно-угорских народов может свидетельствовать в пользу того, что территория современной Ярославской области заселялась в ходе низовой, ростово-суздальской миграции. Эта гипотеза подтверждается генетической близостью современного населения Новгородской области к популяциям русского Севера и финно-угорским народам, но не к суммарной и сборной выборкам русских Ярославской области. Если бы при заселении славянами Ярославского Поволжья преобладала верховая волна миграций, то можно было ожидать, что ее современное население будет генетически ближе к генофонду Русского Севера, а не к популяциям областей Центральной России.

Картографирование генетических расстояний от каждой из выборок Ярославской области (рисунок 16) подтверждает и уточняет эти выводы. Зона максимального сходства с суммарной выборкой русского населения Ярославской области охватывает соседние области Центральной России (рисунок 16 А) - Тверскую, Владимирскую, Костромскую, Московскую, Ивановскую, и лишь отчасти заходит на окраины Новгородской и Вологодской областей. Иными словами, зона наибольшего генетического сходства с генофондом населения Ярославской области охватывает Волго-Окское междуречье и Верхнее Поволжье и не простирается выше на север. Если рассматривать только сборную выборку русских Ярославской области, то картина становится еще наглядней (рисунок 16 Б): зона генетического сходства простирается не только на все Верхнее Поволжье (Ярославскую, Тверскую, Московскую, Костромскую и Ивановскую области), но и идет дальше на юг вплоть до Ростовской области. При этом с популяциями Русского Севера сборная выборка почти не обнаруживает генетического сходства, хотя и присутствует сходство с популяциями Волго-Уральского региона (марийцами, удмуртами, чувашами).

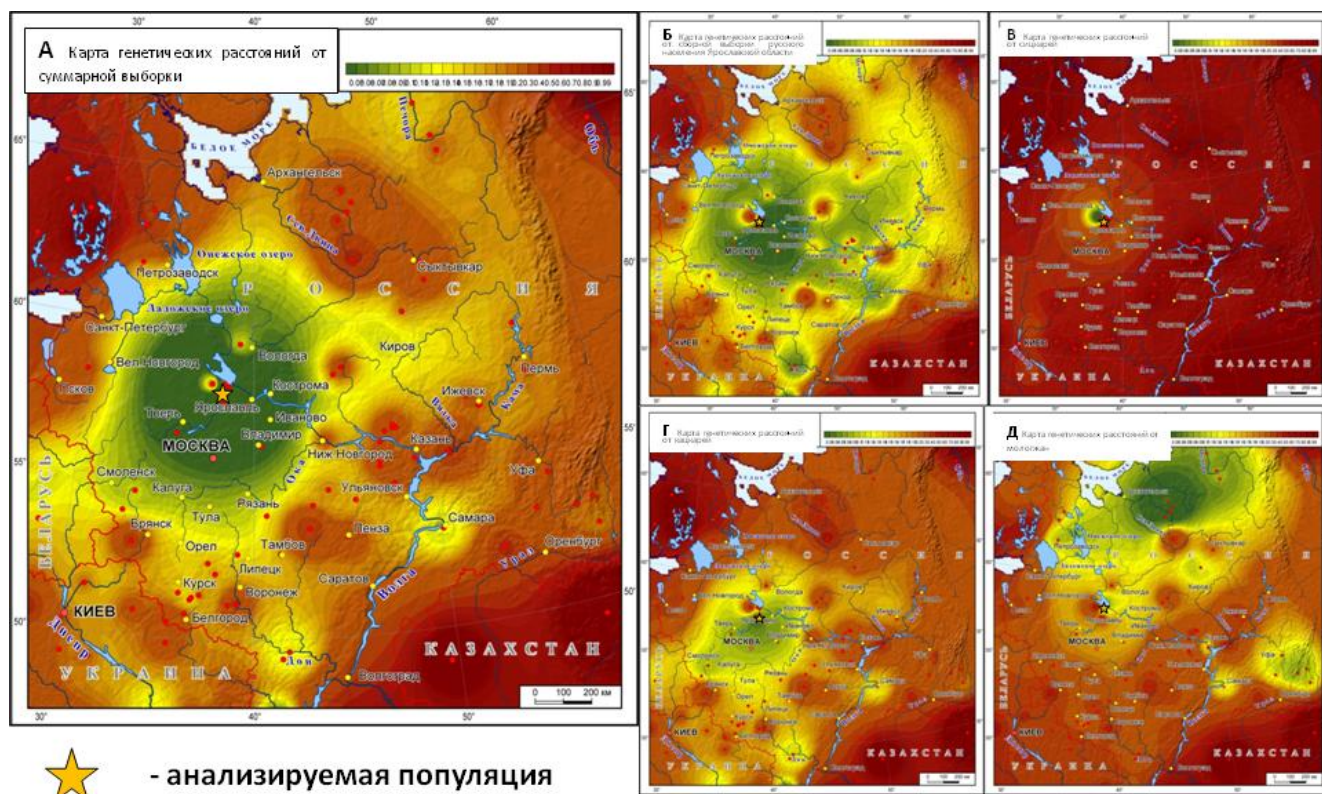


Рисунок 16. Карты генетических расстояний от популяций Ярославской области до окружающих популяций Евразии.

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

Все эти факты, собранные вместе, показывают, что основные черты генофонда Ярославской области сближают его с генофондом Центральной России и других восточных славян (украинцев и белорусов), что может говорить в пользу гипотезы формирования генофонда Ярославского Поволжья при преобладании низовой (Ростово-Суздальской) славянской экспансии, а не верховой (Новгородской) волны миграций. Конечно, надежно реконструировать генофонд новгородцев по современному населению проблематично из-за высылки

новгородцев после падения Новгородской республики в 1478 г. Только данные палеодНК смогут внести ясность в вопросы направления и интенсивности генетических потоков славянской колонизации. Финно-угорский вклад в генофонд современного населения Ярославского Поволжья был или весьма ограниченным, или же снизился в результате более поздних массовых миграций славянского населения. Вместе с тем, нельзя исключить, что летописная меря сформировалась на древнем балто-славянском субстрате и представляла собой в значительной мере центрально-европейское по происхождению население (потомков создателей верхневолжских культур шнуровой керамики и боевых топоров), перешедшее на финно-угорский язык [Напольских, 2015]. В этом случае их генофонд должен быть близок к наиболее древним балто-славянским племенам Восточной Европы. Проверка этой гипотезы методами генетики крайне затруднена (даже анализ древней ДНК может оказаться мало информативен): изучение балто-славянских народов показало, что основную часть их современного генофонда в Восточной Европе составляет генофонд дославянского населения [Чухряева и др., 2017].

Отдельно, с целью обнаружения вклада финно-угорских народов, нами были рассмотрены генетические особенности популяции мологжан. Хотя спектр основных гаплогрупп в генофонде мологжан такой же, что и в суммарной выборке Ярославской области - R-M198(xM458), R-M458, N-M178, однако наиболее частой (мажорной) гаплогруппой является N-M178 (35%). Столь высокая частота гаплогруппы N-M178 приближается к значениям популяций русского Севера (35-50%) [Balanovsky et al., 2008]. Это может указывать либо на след верховой (Новгородской) славянской колонизации, либо, что более вероятно, на сохранение в генофонде жителей Мологи следа финно-угорских племен - предположительно, мерян - в соответствии с данными археологии и топонимики.

Для решения этого вопроса было проведено субтипирование гаплогруппы N-M178 по маркерам, выявленным в результате полногеномного анализа Y-хромосомы [Pumae et al., 2016]. Оказалось, что в большинстве русских популяций

преобладает субвариант N3a3-CTS10760, который доминирует и в обследованной нами сборной выборке Ярославской области. Однако в генофонде мологжан доминирует (с частотой 26%) другой субвариант N3a4-Z1936, который характерен для Волго-Уральского региона и севера Восточной Европы, в основном её финно-угорского населения – карел, вепсов, эстонцев. Таким образом, обнаружение именно этого варианта гаплогруппы N-M178 у мологжан может свидетельствовать о мощном финно-угорском пласте в их генофонде. Положение мологжан на графике многомерного шкалирования в общем кластере с марийцами, литовцами и северно-русскими популяциями также указывает на сохранение в их генофонде этого пласта. Этот вывод подтверждается и картой генетических расстояний от мологжан (рисунок 16 Е), указывая на их генетическое сходство с севером Восточной Европы и несколько меньшее сходство с народами Волго-Уральского региона. Полученные данные позволили выдвинуть гипотезу, что в генофонде летописных мерян преобладал именно этот субвариант гаплогруппы N: N3a4-Z1936 [Чухряева и др., 2017].

### **4.3. Генофонд донских казаков**

Для понимания дальнейшего развития генофонда славян, после этапов массовых миграций интересно рассмотреть процессы образования отдельных субэтнических групп. В частности нами были изучены донские казаки, как субэтническая группа русского населения.

Среди 314 генотипированных казаков Дона обнаружено 18 различных гаплогрупп. Причем среди них встречено лишь шесть гаплогрупп превышающих по частоте 5% уровень полиморфизма: R-M198, I-P37.2, R-M458, N-M178, R-M269 и I-M253. Остальные 12 гаплогрупп имеют, соответственно, частоту ниже 5%, а именно гаплогруппы: E-M78, G-M406, G-P303, J-M267, J-P58, J-M172, J-M92, J-M12, N-LLY22, O-M242, R-M297 и T-M70.

Почти половину от всего спектра гаплогрупп, обнаруженных среди донских казаков, составляет гаплогруппа R-M198. Подобный факт не вызывает удивления, поскольку если посмотреть на общемировое распространение R-M198, то наиболее часто она встречается у славян и у балтов. Да и остальные 5 наиболее распространенных среди казаков гаплогрупп характерны для славян в целом и для русских в частности.

Проведенный нами многомерный анализ положения донских казаков в контексте окружающих популяций (рисунок 17) показал генетическую близость донского казачества с популяциями русских и украинцев. При этом турки вместе с абхазами, шапсугами и черкесами выделяются в отдельный кластер, генетически далекий от казачества Дона. Калмыки также занимают обособленное положение, весьма далекое от казаков. Таким образом, нами показано, что этносы абхазо-адыгской языковой группы, калмыки и турки генетически весьма далеки от казаков Дона [Чухряева и др., 2016].

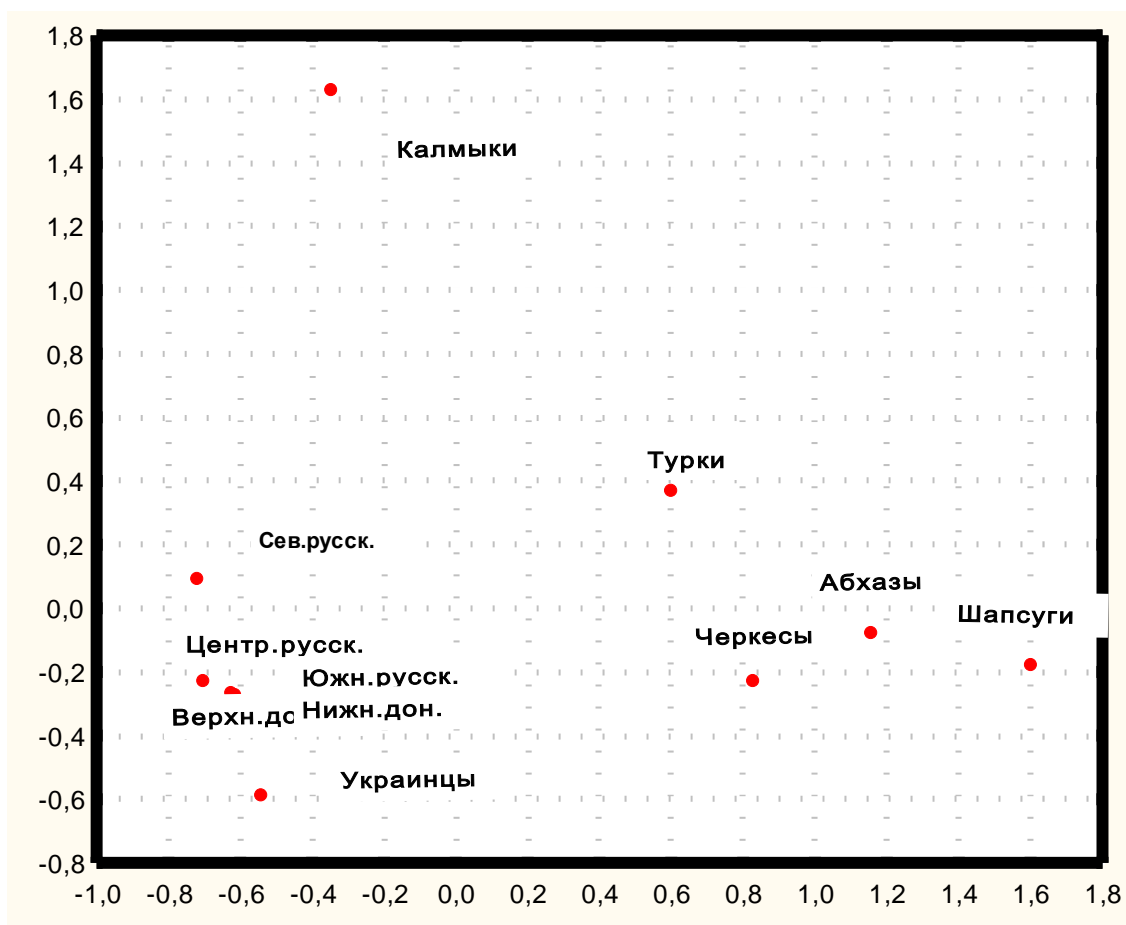


Рисунок 17. Положение донских казаков в генетическом пространстве окружающих популяций. График многомерного шкалирования по маркерам Y-хромосомы

Также нами проведен и анализ STR гаплотипов донских казаков. Уровень гаплотипического разнообразия составил  $1.0000 \pm 0.0009$ , столь высокое разнообразие характерно для многих европейских популяций.

Нами были подсчитаны  $R_{ST}$  дистанции между казаками и окружающими популяциями. На основе полученных значений построена карта генетических расстояний от популяции донских казаков до остальных рассматриваемых популяций Восточной Европы и Кавказа (рисунок 18). Карта демонстрирует, что популяции, генетически сходные с казаками (темно-зеленые тона) расположены в южных областях России и в Украине, тогда как степные тюркоязычные популяции демонстрируют умеренные отличия (салатовый цвет), а кавказские популяции – выраженные отличия от генофонда казаков (желтый цвет).

Полученная картина свидетельствует о близости генофонда донского казачества к генофонду восточных славян в целом. Полученные результаты подтверждает и анализ гаплотипов казаков, проведенный в программе Haplomatch (Таблица 12).



Рисунок 18. Карта генетических расстояний от донских казаков до окружающих популяций. Картографированы значения  $R_{ST}$  расстояний между популяцией донских казаков и окружающими популяциями.

Таблица 12. Показатели сходства генофонда казаков с окружающими популяциями, рассчитанные в режиме сравнения популяций программы Haplomatch.

Название популяции	N	Число точных совпадений	Число различий в один шаг	Число различий в два шага	Число различий в три шага	Отношение точных совпадений к размеру выборки*	Отношение совпадений с разницей в один шаг к размеру выборки	Отношение совпадений с разницей в два шага к размеру выборки	Отношение совпадений с разницей в три шага к размеру выборки
Южные русские	<b>115</b>	9	93	317	605	<b>0.08</b> ±0.03	0.81	2.76	5.26
Восточные украинцы	<b>397</b>	22	136	485	1129	<b>0.06</b> ±0.01	0.34	1.22	2.84
Центральные русские	<b>442</b>	28	186	702	1662	<b>0.06</b> ±0.01	0.42	0.58	3.76
Ногайцы	<b>57</b>	3	36	97	221	<b>0.05</b> ±0.03	0.63	1.70	3.88
Литовцы	<b>298</b>	14	121	423	996	<b>0.05</b> ±0.01	0.41	1.42	3.34
Западные украинцы	<b>586</b>	22	226	813	1763	<b>0.04</b> ±0.01	0.39	1.39	3.01
Северные русские	<b>121</b>	4	64	188	352	<b>0.03</b> ±0.02	0.53	1.55	2.9
Казанские татары	<b>138</b>	2	33	146	343	<b>0.01</b> ±0.01	0.24	1.06	2.49
Черкесы	<b>129</b>	1	5	35	135	<b>0.01</b> ±0.01	0.04	0.27	1.05
Калмыки	<b>156</b>	1	6	21	52	<b>0.01</b> ±0.01	0.04	0.13	0.33
Азовские греки	<b>95</b>	0	7	48	100	<b>0</b>	0.07	0.51	1.05
Грузины	<b>169</b>	0	2	13	47	<b>0</b>	0.01	0.08	0.28
Абхазы	<b>51</b>	0	1	9	32	<b>0</b>	0.02	0.18	0.63
Армяне	<b>168</b>	0	0	2	11	<b>0</b>	0.00	0.01	0.07
Шапсуги	<b>97</b>	0	0	5	10	<b>0</b>	0.00	0.05	0.10

\* Показана статистическая ошибка выборочной доли.



Проведенный нами сравнительный анализ в программе Haplomatch (таблица 12) гаплотипов донских казаков с окружающими популяциями показал наибольшее сходство донских казаков с восточно-славянскими популяциями. Наибольшая доля точных совпадений с гаплотипами казаков обнаружена в южнорусских популяциях (8%). Несколько меньшее (однако, отличие недостоверно) генетическое сходство донские казаки показывают с русскими Центральной России (показатель совпадений равен 6%). Подобные результаты хорошо согласуются с историческими сведениями об истории заселения донских степей: наиболее мощный поток мигрантов шел из южных губерний России (современные Воронежская, Курская, Орловская, Белгородская области). Также значимые миграции в казачество происходили и из Центральной России - например, в летописях зафиксированы массовые миграции из Рязани на Верхний Дон. Полных совпадений гаплотипов казаков с северными русскими популяциями обнаружено гораздо меньше – это свидетельствует о том, что русское население удаленных от донских земель территорий слабо принимало участие в формировании казачества на Дону. Кроме южных русских и русских Центральных регионов России, ощутимый вклад в генофонд казачества внесли и украинцы: они являются вторым по значимости компонентом, участвовавшим в формировании генофонда донского казачества. Достаточно высокий показатель сходства с казачьим населением показали и литовцы – 5%. Мы полагаем, что сходство с литовцами скорее всего объясняется их долгим существованием в одном государстве с украинцами (Великое княжество Литовское, позднее Речь Посполитая), непосредственные же массовые миграции из Литвы на Дон представляются маловероятными [Чухряева и др., 2016].

Таким образом, на уровне тенденции можно отметить, что генофонд казаков проявляет больше сходства с южными русскими, чем с остальными наиболее сходными популяциями (восточными украинцами, русскими Центральных регионов России, ногайцами, литовцами), но эти различия не достигают порога

достоверности, и все перечисленные популяции в целом высоко сходны с генофондом верхнедонского казачества.

Что касается степных народов, то все они (за исключением ногайцев) показали невысокое генетическое сходство с донскими казаками. Среди казачьих станиц на Дону присутствовали и татарские станицы, причем термин «донские татары» - обобщающий и подразумевает не только этнических татар. Консолидация в этой локальной группе происходила на основе ислама, близости обычаев, языка, а также принадлежности к казачьему сословию. Предположительно в этой группе преобладали ногайцы. Таким образом, благодаря столь активным историческим взаимоотношениям казаков Дона с ногайцами мы и наблюдаем сегодня 5% совпадающих гаплотипов между их генофондами. Генетическое же сходство донских казаков с татарами в современном понимании этого этнонима невелико: показатель схождения с казанскими татарами равен всего 1%. Вероятно, это означает, что в группу «донские татары» действительно преимущественно входили ногайцы, а не этнические татары. Калмыки также не оказали на генофонд верхнедонских казаков большого влияния: наш анализ показал всего 1% общих гаплотипов с ними. И, хотя в Войске Донском существовали отдельные группы казаков-калмыков, эти сообщества, по всей видимости, были достаточно замкнуты, и поэтому сколько-нибудь значимый их генетический вклад в основную группу казаков не обнаруживается.

Народы Кавказа все оказались генетически далеки от верхнедонских казаков – за исключением черкесов, ни в одной из кавказских популяций вообще не обнаружено полных совпадений гаплотипов с казачьими. Ненулевое (1%) сходство с черкесами может быть сформировано и общим потоком генов между Восточной Европой и Кавказом, не обязательно связанным с прямыми миграциями между сравниваемыми популяциями.

Таким образом, генофонд казаков Верхнего Дона (по данным о STR и SNP маркерах Y-хромосомы) сформировался преимущественно за счет

восточнославянского компонента, из степных 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.

## ГЛАВА 5. Генофонд народов армянской языковой группы

### 5.1. Генетические взаимоотношения между группами армян

По маркерам Y-хромосомы нами был изучен генофонд армян – единственных представителей армянской языковой группы индоевропейской языковой семьи. В генофонде изученных нами популяций армян выявлено 27 гаплогрупп (рисунок 19). В целом для их генофонда характерно большое разнообразие вариантов Y-хромосомы и отсутствие какой-либо одной доминантной гаплогруппы, в отличие от народов Северного Кавказа, где в каждом регионе решительно преобладает какая-либо один вариант, определяющий региональный генетический портрет [Balanovsky et al, 2011]. Наиболее распространены в генофонде армян гаплогруппы R-L23, J-M67 и G-M285. В суммарной выборке армян эти три варианта представлены примерно в равных долях (20-10%). Но генофонды разных популяций армян значительно различаются по их частоте. Так, гаплогруппа G-M285 составляет почти половину генофонда амшенских армян, однако у донских она полностью отсутствует. В распределении J-M67 обнаруживаются столь же значительные различия между популяциями: J-M67 составляет почти треть генофонда амшенских армян, с высокой частотой встречена она у армян Краснодарского края, но в других трех изученных нами популяциях она составляет всего 5-7%. Частоты гаплогруппы R-L23 также варьируют от 24% у армян Адыгеи до полного отсутствия у амшенских армян [Balanovsky et al., 2017].

Для выяснения соотношений между генофондами разных популяций армян нами были рассчитаны генетические расстояния между ними. Матрица генетических расстояний с иерархически кластеризованными популяциями армян и график многомерного шкалирования приведены на рисунке 20.

Полученные результаты демонстрируют генетическую близость изученных популяций армян вне зависимости от того, представляют ли они армян диаспоры

либо самой Армении. Так в основной кластер вошли шесть популяций армян: Эрзурума, Адыгеи, Ирана, Гардмана, Ван и долины Арарат – это наглядно видно по результатам всех проведенных нами видов анализа (рисунок 20 А, В). Все популяции армян, вошедшие в «основной» кластер, как указывалось выше, объединяет общий генетический портрет высоких частот гаплогрупп R-L23, J-M67 и J-M172\*.

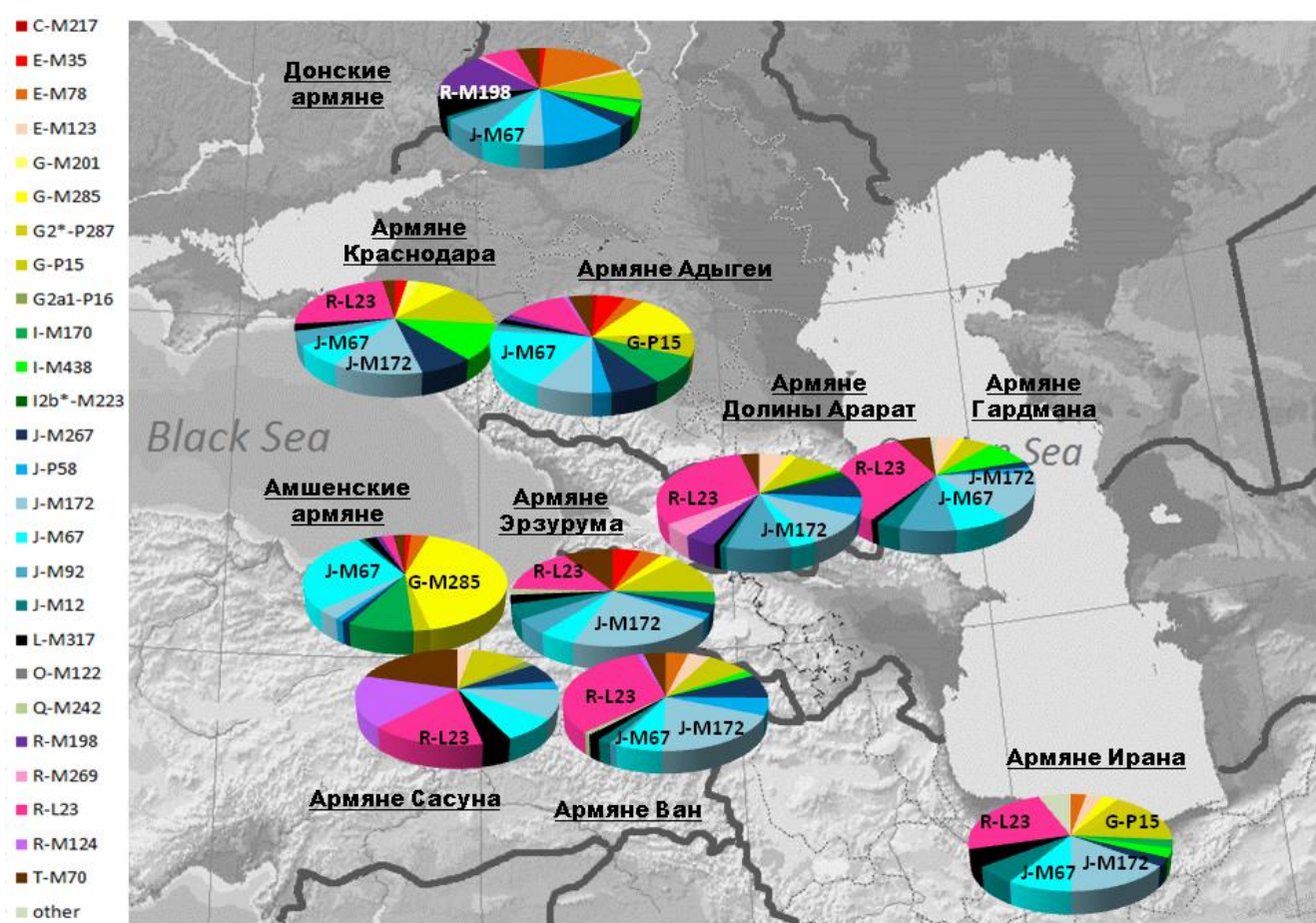


Рисунок 19. Частоты гаплогрупп Y-хромосомы у изученных групп армян

При оценке генетических расстояний от усредненной популяции армян (ее параметры получены путем усреднения частот всех анализируемых популяций армян), наиболее близкими к ней оказались армяне Адыгеи и Ирана. Вероятно, это

произошло, из-за того, что эти группы формировались за счет мигрантов из разных городов Армении. Видимо этот процесс миграции и воссоздал в этих популяциях своеобразный «среднеармянский спектр гаплогрупп».

Армяне Краснодарского края генетически наиболее близки к амшенам: самыми распространенными среди них являются гаплогруппы G-M285 (13,9%), J-M67 (19,1%). Однако по спектру остальных, немажорных гаплогрупп, они близки к основному массиву армянских популяций. Вероятно, это объясняется долгим совместным проживанием указанных групп армян в тесном соседстве на территории Краснодарского края.

Обособленное положение армян Сасуна связано с высокими частотами гаплогрупп T-M70 и R-M124. Подобное явление можно объяснить влиянием курдских племен на армян Сасуна в XIV, что подтверждено историческими данными. В современной популяции курдов частота T-M70 составляет 8,5% [Grugni et al., 2011], выше, чем у всех окружающих иранских популяций и армян в целом. Вероятно, курды и принесли этот вариант Y-хромосомы в генофонд армян Сасуна. А за счет действия дрейфа генов частота T и достигла столь высоких частот у них.

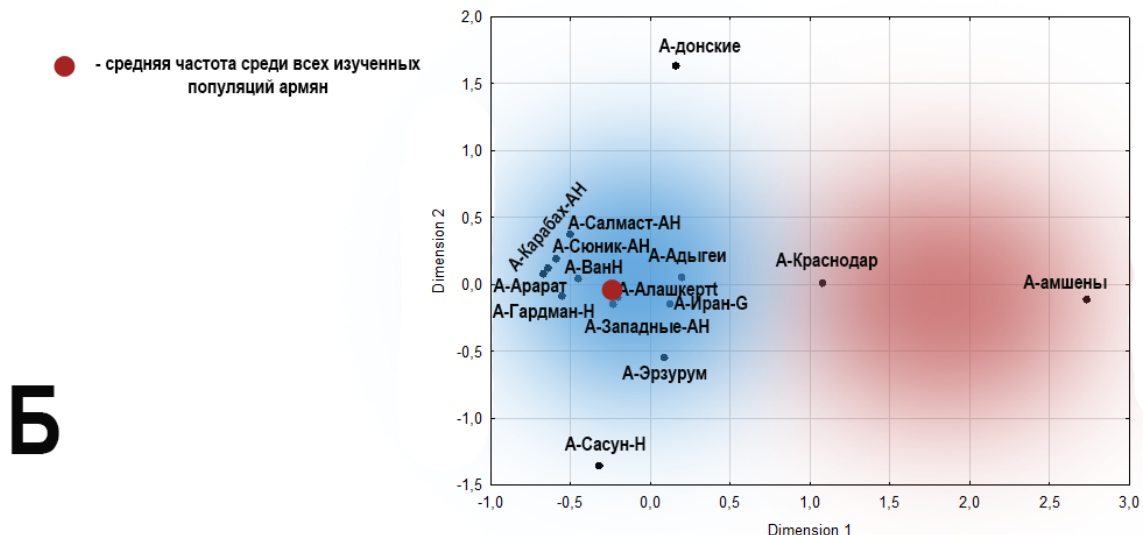
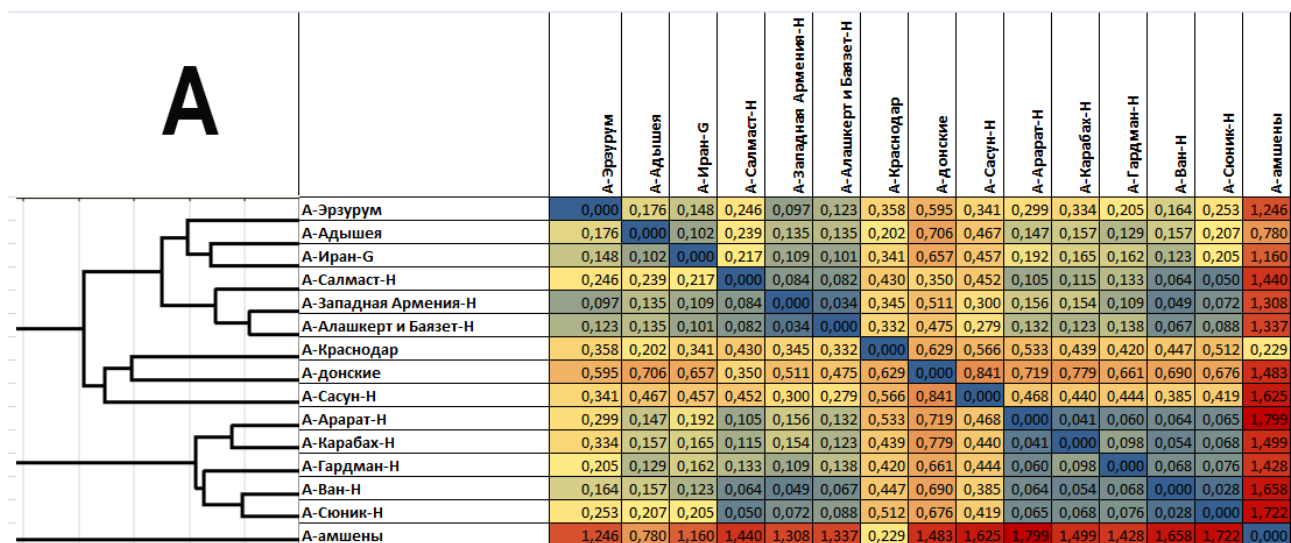


Рисунок 20. Анализ генетических взаимоотношений между армянами

А) Heatmap, построенная на основе генетических расстояний ( $D_{NEI}$ ) между изученными популяциями армян

Б) График многомерного шкалирования генетических взаимоотношений между изученными популяциями армян

На основе проведенного анализа, мы можем сделать общее заключение, что проживание в диаспоре, при условии сохранения прежней брачной структуры, позволило популяциям армян сохранить свой генофонд. Таким образом, тщательное исследование родословных позволяет использовать население

диаспоры для изучения и реконструкции генофонда уже не существующих по тем или иным причинам государств.

Предположение о том, что армяне в диаспоре сохранили черты исходного протогенофонда Исторической Армении подтверждает проведенный нами картографический анализ: на рисунке 21 представлена карта генетических расстояний от усредненной армянской популяции до всех окружающих народов. Картографический анализ выявляет удивительное совпадение современного ареала популяций, наиболее близких к среднеармянскому генофонду, с ареалом Исторической Армении. При этом географически соседние популяции равнинной Юго-Западной Азии генетически далеки от армян.

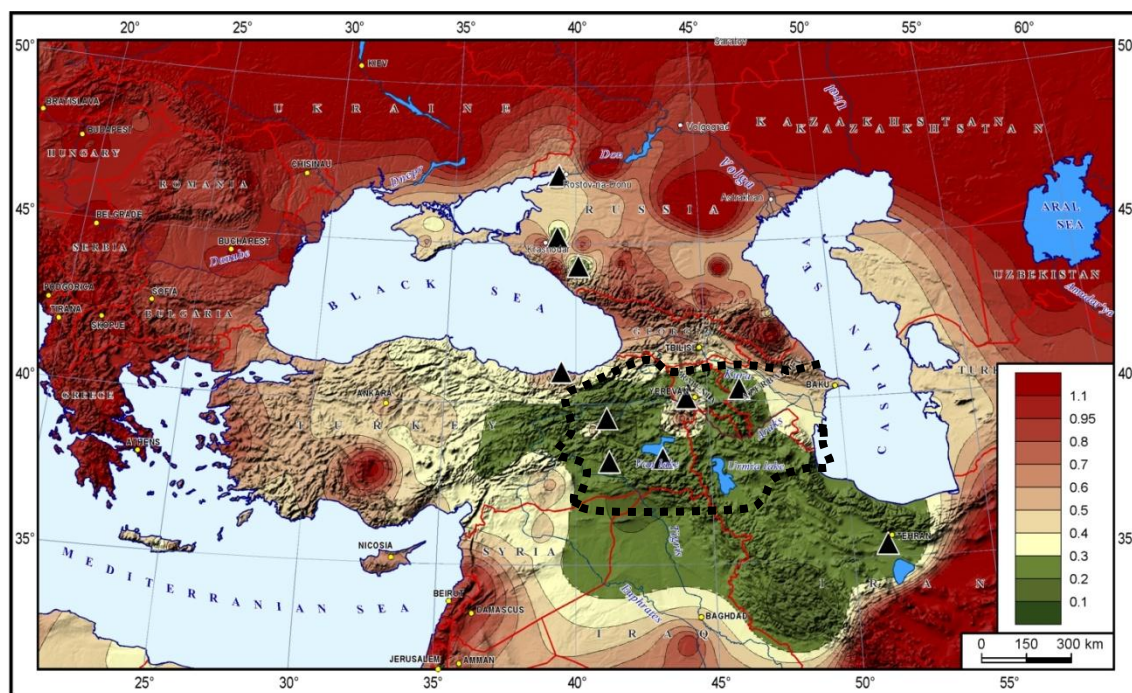


Рисунок 21. Карта генетических расстояний от анализируемых групп армян до окружающих популяций.

*Примечания: пунктирной линией показаны границы исторической Армении, темно зеленые цвета означают максимальное генетическое сходство, небольшие генетические расстояния; темно красные – указывают на высокие генетические расстояния, показывают на отсутствие сходства.*



Для проверки гипотезы о сохранении армянами генофонда в диаспоре, нами также был проведен анализ генетической дифференциации AMOVA (таблица 13). Нами было выбрано несколько способов подразделения популяций: на амшен и остальных популяций армян, на армян диаспоры и армян не диаспоры, и случайное объединение изучаемых популяций в 2 группы. Если подразделение согласуется с реальными генетическими отличиями между популяциями, то при анализе мы увидим высокий показатель межгрупповых вариаций. В нашем случае показатель оказался высоким для подразделения амшены\не амшены. Что указывает на обособленность амшен от остальных групп армян, связанную с особенностью образования этой группы. А при подразделении диаспора\не диаспора показатель дифференциации не высокий, что подтверждает, гипотезу о том, что армяне в диаспоре сохранили черты исходного протогенофонда Исторической Армении и не имеют специфических черт в своем генофонде.

Таблица 13. Оценка генетической дифференциации популяций армян методом AMOVA.

Процент вариаций	межгрупповых	внутригрупповых	Внутрипопуляционных
Амшены\не амшены	11.8*	2.2	86.0
диаспора\не диаспора	2.6*	2.7	94.7
Случайное объединение в 2 группы	-0.3	4.2	96.1

\*p-value<0,05

Далее нами был проведен анализ лингвистической подразделенности армян (таблица 14). Мы хотели узнать, насколько принадлежность армян к разным языковым подгруппам отразилась на структуре их генофонда. Нами был проведен анализ при подразделении всех популяций на говорящих на западной и восточной группах армянских языков и анализ с отдельным рассмотрением представителей крымско-анинского диалекта (донских армян). Однако, и в одном, и в другом случае показатель дифференциации крайне невысокий. В случае подразделения на западную и восточную группу армян он составил всего 0,07. А в случае подразделения на западный, восточный и крымско-анинский диалект показатель межгрупповых вариаций составил всего 0,8. Это указывает на отсутствие влияния лингвистической подразделенности на структуру генофонда армян [Balanovsky et al., 2017].

Таблица 14. Оценка генетической дифференциации популяций армян методом AMOVA.

Процент вариаций	Межгрупповых (%)	Внутригрупповых (%)	Внутрипопуляционных (%)
Западные и восточные	0,07	2,38	97,55
Западные /восточные/крымско-анинский диалект	0,8	1,94	97,26

## **5.2. Филогенетический анализ гаплогрупп, наиболее распространенных среди армян**

### Гаплогруппа R-L23

Первоначальное распространение гаплогруппы R-L23 связывают с процессом неолитизации [Herrera et al, 2012]. Картографический анализ современного распространения R-L23 (рисунок 22) показывает, что максимальные частоты R-L23 характерны для армян (20% в среднем) и окружающих их «горных» юго-западных популяций, не распространяясь дальше ни на юг, ни на Кавказ, ни на ареалы русского и степных народов. По сути, ареал её наибольшего распространения совпадает с территорией Исторической Армении. Филогенетический анализ гаплогруппы R-L23, проведенный по 15 локусам Y-хромосомы, выявил кластер, который можно назвать «армянским» (рисунок 22). Его «генеалогическая» датировка ( $1900 \pm 500$  лет) близка к исторической датировке возникновения армянского государства.

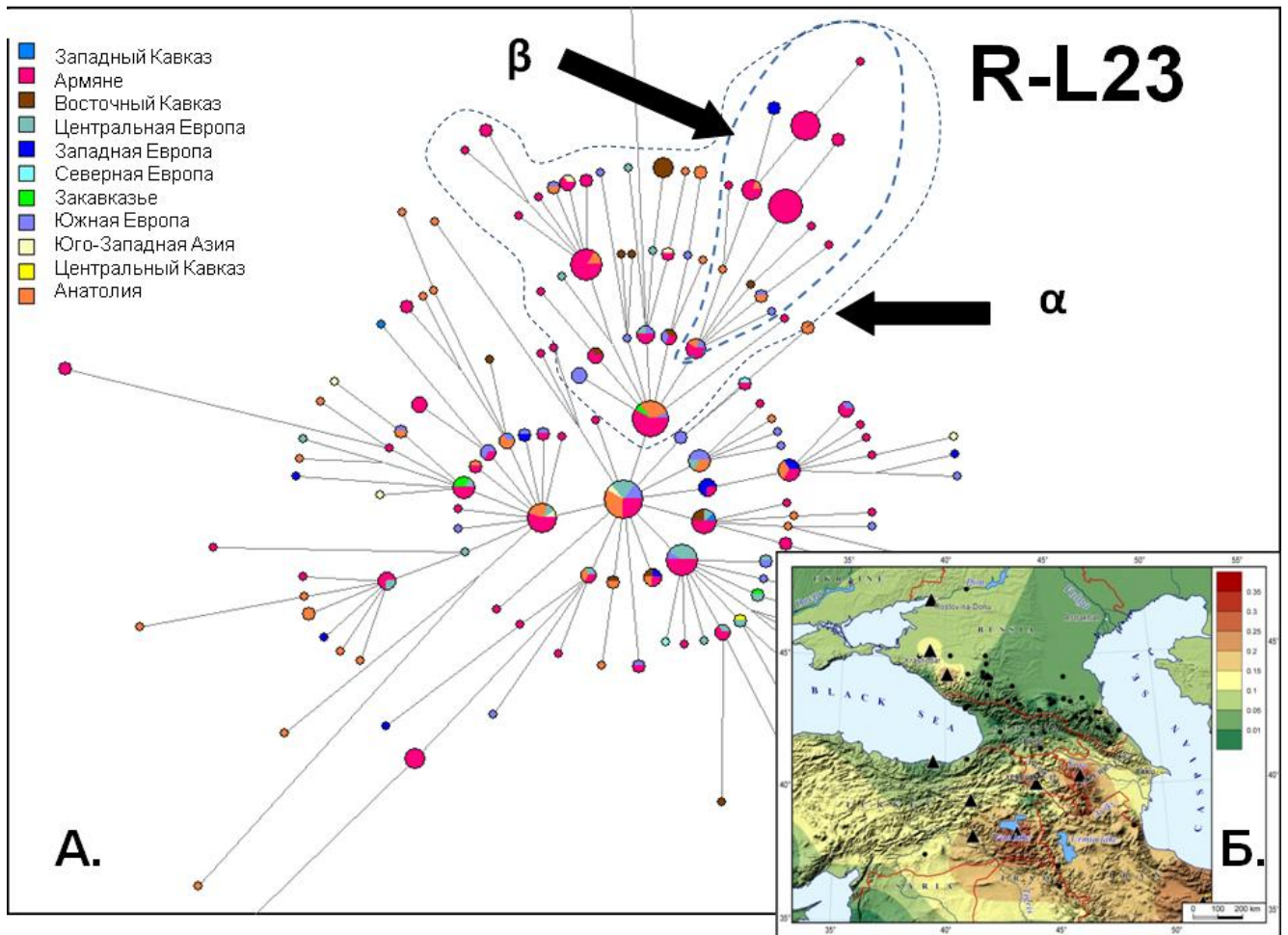


Рисунок 22. А) Филогенетическая сеть гаплогруппы R-L23

*Генеалогический возраст кластера  $b$ :  $3000 \pm 1200$  лет*

В) Карта распространения гаплогруппы R-L23

**Западный Кавказ:** Abkhazes and Cherkessians of Myres 2010; **Армяне:** A-Krasnodar, A-Adygei, A-Don, эта работа; A-Ararat, A-Gardman, A-Van, A-Sasun of Hererra 2012; Armenians of Myres 2010; **Восточный Кавказ:** Avars, Bagvalals, Balkars, Darginians, Lezgis, Tabasarans of Myres 2010; **Центральная Европа:** Czech, Hungarians, Poles of Myres 2010; **Западная Европа:** Danes, Germans, Swiss of Myres 2010; **Северная Европа:** Estonians, Swades South (Malmo) of Myres 2010; **Закавказье:** Georgians, Azerbaijanians - неопубликованные данные лаборатории популяционной генетики человека, Georgians Megrels of Myres 2010; **Южная Европа:** Greeks, Italians, Romanians of Myres 2010; **Южная Азия:** Jordanians, Pakistani, Palestinians of Myres 2010; **Центральный Кавказ:** Osetins North of Myres 2010; **Анатолия:** Turkey of Myres 2010.

### Гаплогруппа J-M67.

Для гаплогруппы J-M67 также предполагается переднеазиатское происхождение. Карта ее современного распространения (рисунок 23) демонстрирует, что наибольшей частоты J-M67 достигает у народов нахской языковой группы (у чеченцев и ингушей - до 79%), со средними частотами она распространена среди армян, а для других популяций Юго-Западной Азии относится к редким. Филогенетический анализ выявляет, что армяне формируют свой собственный кластер гаплотипов, обособленный от кластера народов нахско-дагестанской языковой семьи. Примечательно, что в «армянский» кластер входят представители всех изученных нами популяций армян. «Генеалогическая» датировка кластера (2000±500 лет) близка к датировке «армянского» кластера на филогенетической сети гаплогруппы R-L23. Это еще раз свидетельствует, что основные черты генофонда армян сложились в период формирования армян как единого этноса.

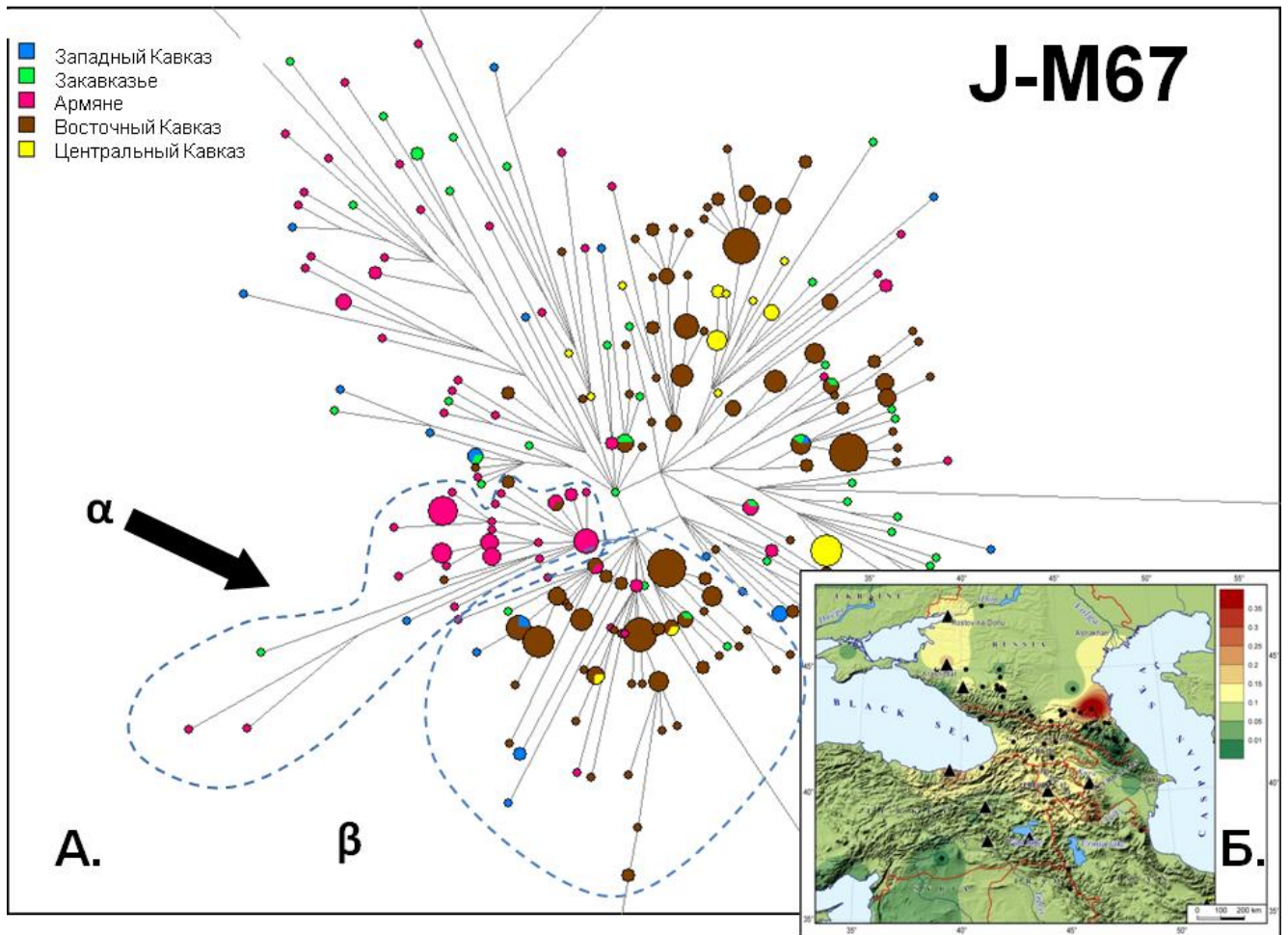


Рисунок 23 . Анализ гаплогруппы J-M67

А) Филогенетическая сеть гаплогруппы J-M67

Б) Карта распространения гаплогруппы J-M67

West Caucasus: Adyghes, Abkhazes Balanovsky et al. 2011; Transcaucasia: Georgians, Azerbaijanians our data unpublished; Armenians: A-Krasnodar, A-Adygei, A-Don, A-Hemshin, A-Erzurum this study; A-Ararat, A-Gardman, A-Van, A-Sasun of Hererra 2012; East Caucasus: Ingush, Chechen Balanovsky et al. 2011; Central Caucasus: Ossetians Balanovsky et al. 2011.

Генеалогический возраст кластера:  $2000 \pm 500$ .

Гаплогруппа G-M285

На карте распространения гаплогруппы G-M285 (рисунок 24) видно, что эта гаплогруппа не является распространенной в целом, но при этом она имеет достаточно широкий ареал распространения: от Италии до Монголии, но с наивысшими частотами обнаружена в центрально-азиатских степях и на Переднеазиатском нагорье. У амшенских частота G-M285 равна 42,6% и является одной из самых высоких среди изученных популяций. Ранее считалось, что высокие частоты G-M285 присущи только населению Казахстана - до 87% у мадьяров Казахстана [Viro et al, 2009]. Позднее, нашим коллективом было выяснено, что повышенные частоты гаплогруппы G1 обнаружены также среди родоплеменной группы казахов аргын и среди амшенских армян.

В целом, можно сделать предположение, что данный вариант Y-хромосомы впервые возник в Иране, где частота G1-M285 не превышает 5%, но обнаружено максимальное гаплотипическое разнообразие. Дальнейшее распространение гаплогруппы шло на запад и на восток, откуда она вероятно и проникла в Армению. [Cinnioglu et al, 2004 Rootsi et al, 2012 Regueiro 2005]. Возникновение гаплогруппы на территории современного Ирана и её генетический ландшафт в целом, могут быть использованы в качестве свидетельств Анатолийской гипотезы: её распространение из Передней Азии в степную зону Евразии и далее могло произойти благодаря миграциям степных кочевников.

Филогенетический анализ гаплогруппы G-M285 (рисунок 25) среди армян обнаружил в их генофонде действие дрейфа генов. Амшенские армяне и армяне, впоследствии мигрировавшие в Краснодарский край, формируют свой, обособленный от казахов кластер. Он включает в себя помимо образцов амшенских армян и армян Краснодарского края так же армян Адыгеи.

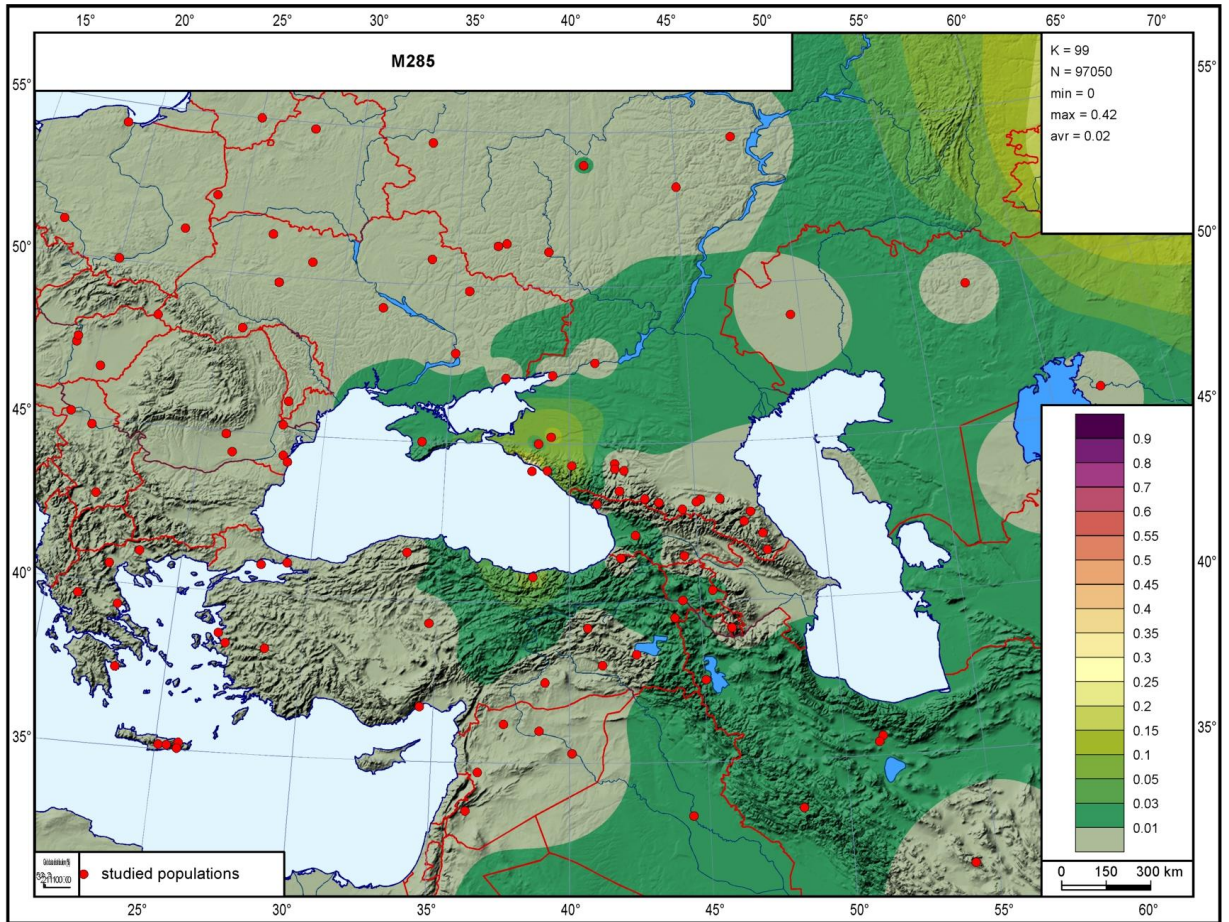


Рисунок 24 . Карта распространения гаплогруппы G-M285



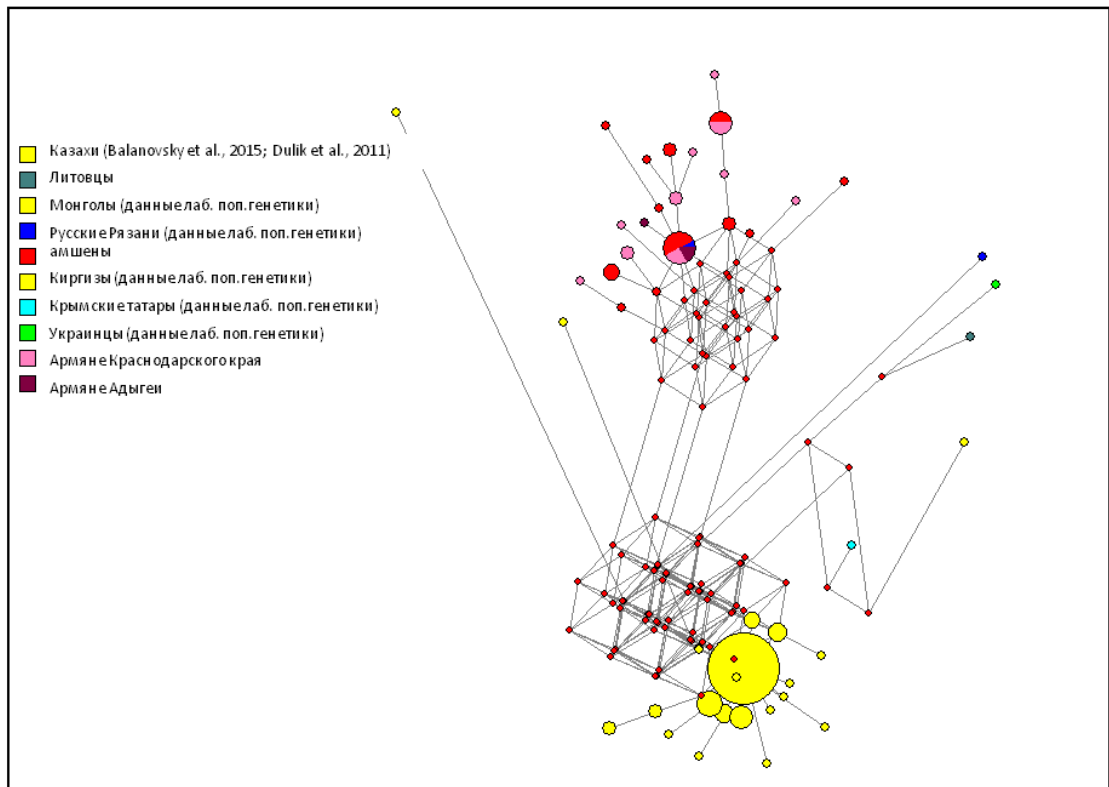


Рисунок 25 . Филогенетическая сеть гаплогруппы G-M285

Изучение нами причин наблюдаемой картины показало, что эффект основателя привел к накоплению в генофонде амшен гаплогруппы G-M285. Об этом свидетельствуют исторические данные – в XX в. численность амшенских армян была не меньше 150 000 человек. Таким образом, наблюдаемая нами сейчас картина, не является следствием действия дрейфа генов – популяция амшен была достаточно многочисленной. Скорее всего мы сейчас действительно наблюдаем эффект основателя – по летописным данным группа амшенских армян произошла от принца Аматуни из династии Аббасидов и его 12 тысяч сподвижников, мигрировавших в 791 в Трапезунд из Персии. Эта дата согласуется с полученной нами оценкой возраста кластера гаплогруппы G-M285. Она составила 1150 лет назад.

Таким образом, исследование нами гаплогруппы G-M285 позволило верифицировать летописные данные об образовании амшен, как единой субэтнической группы.

### 5.3. Генетический портрет армян в контексте окружающих популяций

Для определения генетического положения армян с широким контексте окружающих популяций, нами был проведен многомерный анализ по данным полиморфизма 21 гаплогруппы Y-хромосомы (рисунок 26).



Рисунок 26. График многомерного шкалирования: генетический портрет народов Юго-Западной Азии.

График построен по данным полиморфизма 21 гаплогруппы Y-хромосомы (С-M130, E-M35.1, G1-M285, G2-P287, H-M69, I\*-M170, I-P37.2, I-M253, J-M267, J\*-M172, J-M67, J-M92, L-M11, N-M231, O-M175, Q-M242, R-M448, R-M73, R-M269, R-L261, T-L206)

В анализ нами были включены популяции Европы, Восточного, Западного и Центрального Кавказа, Передней Азии. Проведенный нами анализ показал, что армяне входят в Переднеазиатский кластер, обособленный от народов Европы и Кавказа. В пределах Переднеазиатского кластера армяне формируют свой собственный субкластер. Существование этого субкластера еще раз свидетельствует об общем генофонде всех изученных нами групп армян.

Удивительно, что не только население Северного Кавказа, абхазы, но и представители Закавказья – грузины генетически далеки от армян. Генетические расстояния между ними ( $d=0,589259$ ) намного больше, чем между армянами и куда более географически удаленными популяциями турок ( $d= 0,080364$ ).

Результаты многомерного шкалирования иллюстрируют, что кавказские популяции образуют отдельные кластеры от Переднеазиатского. А именно: кластер, куда вошли популяции Восточного Кавказа и кластер, объединивший в себя популяции Центрального и Западного Кавказа. Филогенетический анализ гаплогруппы J-M67, распространенной как среди армян, так и среди чеченцев и ингушей, обнаружил разную генетическую историю формирования этих популяций. Таким образом, пути проникновения гаплогруппы J-M67 в генофонд армян и чеченцев с ингушами различны.

Зато горные популяции Ирана генетически достаточно близки к армянам (от  $d=0,307622$  до  $d=0,145995$ ). Таким образом, генофонд горного пояса Юго-Западной Азии (куда мы относим и армян) обособлен от генофонда Кавказа. Исключение составляет лишь генофонд азербайджанцев. Он изучен крайне слабо. Однако даже изученная нами популяция, в основном представленная азербайджанцами Восточного Дагестана (т.е. географически наиболее далекая от Армении), генетически оказалась близка к армянам и другим народам Юго-Западной Азии. Это можно объяснить тем, что азербайджанцы лишь относительно недавно подверглись инвазии тюркских племен и переняли их тюркский язык (XI в.) - до этих событий их историческое развитие было тесно

связано с Ираном. Они разговаривали на одном из иранских языков и их происхождение связано с древнеиранскими племенами. Вероятно, тюркизация азербайджанцев не сильно изменила их генофонд и генетически они остались близки к азиатским народам - носителям индоевропейских языков.

Также нами были рассчитаны показатели генетического разнообразия для изучаемых групп армян и соседних с ними популяций: грузин и народов Северного Кавказа (таблица 15). Оказалось, что для армян и грузин характерны сопоставимые значения генетического разнообразия. Лишь амшены среди всех армян выделяются сниженными показателями разнообразия, но про них нами было сказано отдельно. При этом для северокавказских популяций характерны на порядок меньшие значения показателей генетического разнообразия. Таким образом, выявлено, что генетическое разнообразие в популяциях Закавказья, среди армян и грузин, существенно выше, чем у соседних северокавказских народов [Balanovsky et al., 2017].

Таблица 15. Генетическое разнообразие в популяциях Западной Азии на основе полиморфизма гаплогрупп Y-хромосомы

Популяция	N	Генетическое разнообразие	источник
Армяне			
А-Адыгей	49	0,88 +/- 0,02	Эта работа
А-Дона	92	0,91 +/- 0,01	Эта работа
А-Эрзурума	99	0,90 +/- 0,01	Эта работа
А-амшены	89	0,74 +/- 0,03	Эта работа
А-Краснодара	117	0,90 +/- 0,01	Эта работа
А-Арарата	110	0,83 +/- 0,03	Herrera et al, 2012
А-Гардмана	96	0,86 +/- 0,02	Herrera et al, 2012
А-озера Ван	103	0,84 +/- 0,02	Herrera et al, 2012
А-Сасуна	104	0,89 +/- 0,03	Herrera et al, 2012
А-Ирана	34	0,89 +/- 0,03	Grugni et al, 2012
А-Сюника	105	0,81 +/- 0,03	Hovhannisyan et al,

			2014
А-Карабах	105	0,81+/- 0,03	Hovhannisyan et al, 2014
А-Алашкерт и Баязет	200	0,89 +/- 0,01	Hovhannisyan et al 2014
А-Западная Армения	148	0,88 +/- 0,01	Hovhannisyan et al 2014
А-Салмаст	199	0,89 +/- 0,01	Hovhannisyan et al 2014
Грузины			
Имеретинцы	62	0,92 +/- 0,02	Эта работа
Лазы	36	0,79 +/- 0,06	Эта работа
Восточные грузины	54	0,82 +/- 0,03	Эта работа
Северокавказские популяции			
Шапсуги	100	0,24 +/- 0,06	Balanovsky et al., 2011
Абхазы	58	0,65 +/- 0,06	Balanovsky et al., 2011

Таблица 15. Генетическое разнообразие в популяциях Западной Азии на основе полиморфизма гаплогрупп Y-хромосомы

<b>Популяция</b>	<b>N</b>	<b>Генетическое разнообразие</b>	<b>источник</b>
Черкесы	142	0,78 +/- 0,03	Balanovsky et al., 2011
Осетины	362	0,51 +/- 0,03	Balanovsky et al., 2011
Ингуши	29	0,37 +/- 0,11	Balanovsky et al., 2011
Чеченцы	330	0,64 +/- 0,02	Balanovsky et al., 2011
Аварцы	115	0,63 +/- 0,05	Balanovsky et al., 2011
Лезгины	81	0,70 +/- 0,03	Balanovsky et al., 2011

После выяснения генетических взаимоотношений армян с окружающими популяциями в широком географическом масштабе, нами был проведен анализ генофонда Юго-Западной Азии. Большая часть изученных популяций этого региона, информация по которым имеется в литературе, генотипирована по более узкой панели гаплогрупп Y-хромосомы. Поэтому обобщающий анализ для Юго-Западной Азии проведен нами по панели из 12 гаплогрупп Y-хромосомы.

Результаты многомерного анализа генофонда Юго-Западной Азии оказались довольно неожиданными (рисунок 27) – нами было выделено четкое подразделение всех изученных популяций на два кластера. В первый кластер вошли иорданцы, ливанцы, иракцы, палестинцы, сирийцы и популяции Аравийского полуострова. Второй кластер включил в себя следующие популяции: субпопуляции армян, азербайджанцев, курдов, иранцев и турок. Внимательное рассмотрение популяций, вошедших в кластеры, позволило нам сделать наблюдение, что первый кластер включил в себя народы, проживающие на равнине, а второй объединяет народы, проживающие в гористой местности. Естественно, это несколько обобщенное наблюдение, поскольку небольшие горы имеются и в Сирии. Однако в анализе мы учитывали только крупные горные системы Передней Азии: Переднеазиатское нагорье, состоящее из Иранского, Армянского и Малоазиатского нагорий.



Рисунок 27. График многомерного шкалирования: положение армян в контексте популяций Юго-Западной Азии.

График построен по данным полиморфизма 12 гаплогрупп Y-хромосомы (С-M130, E-M35, G-L116, H-M69, I-M170, J-M267, J-M172, L-M11, NO-M214, Q-M242, R-M173, T-L206).

Для того, чтобы подтвердить либо опровергнуть свое наблюдение, о том, что именно географический рельеф повлиял на структуру генофонда Передней Азии мы провели анализ генетической дифференциации популяций Передней Азии методом AMOVA (таблица 16).

Таблица 16. Оценка генетической дифференциации популяций Передней Азии методом AMOVA.

Процент вариаций	Межгрупповых	внутригрупповых	Внутрипопуляционных
горные\равнинные популяции	5,1*	3,0*	91,9*
западные\восточные популяции	0,55*	5,7*	93,8*
Популяции из разных лингвистических семей (Индоевропейская, Семитская, Тюркская)	4,2*	3,1*	92,7*

\*p-value<0,001

Анализ был проведен тремя разными способами: с подразделением всех популяций Передней Азии на народы, проживающие в горах и на равнине; с подразделением популяций согласно их лингвистической принадлежности (разговаривающих на индоевропейских, тюркских и семитских языках). Для проверки полученных результатов, нами также был проведен тест AMOVA, с подразделением на восточные и западные популяции. Деление было проведено по 43,5° в.д.

При группировке популяций согласно рельефу местности, где они проживают, нами был получен наиболее высокий процент межгрупповых вариаций, равный 5,17% (таблица 16). Для тех же популяций подразделенных согласно долготе, показатель дифференциации оказался в десять раз меньше –

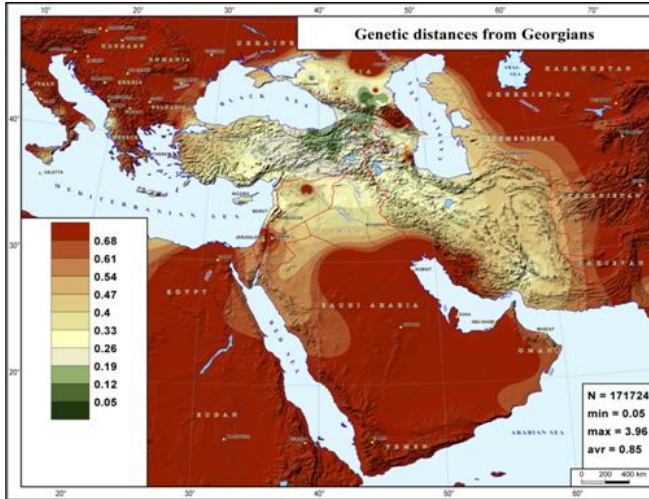


всего 0,56%. Зато при группировке популяций согласно их лингвистической принадлежности показатель вариаций составил 4,2%. Что значительно выше, чем при подразделении на восточные и западные, но несколько меньше, чем подразделение на горные и равнинные популяции.

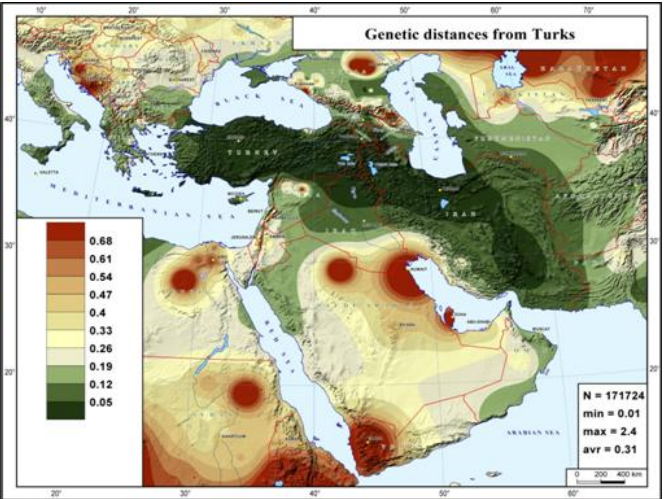
Чтобы увидеть насколько обнаруженные закономерности справедливы для основных из представленных на графике многомерного шкалирования популяций, нами была построена серия карт генетических расстояний. Итого было создано 12 карт генетических расстояний (рисунок 28). Как и ожидалось для каждой из исследуемых популяций зона максимального сходства (показана на картах зеленым цветом) совпала с географическим местом проживания исследуемой популяции. Однако зона распространения этого сходства существенно различается для разных популяций. Так для грузин обнаружена крайне небольшая зона генетического сходства с окружающими популяциями (рис.28 А). Возможно, это объясняется их принадлежностью к Картвельской семье, единственными представителями которой они являются. В отличие от них для турок наоборот присуща широкая зона генетического сходства (рисунок 28 Б) совпадающая с территориями, принадлежащими Османской империи. Армяне также показывают достаточно обширную зону генетического сходства с окружающими популяциями, совпадающую очертаниями с границами исторической Армении (рисунок 28 Д). Но и в целом, карты генетических расстояний от популяций, проживающих в горной местности (турки, армяне, иранцы, курды и грузины) демонстрируют тренд, показанный и предыдущими видами анализа: горные популяции проявляют максимальное генетическое сходство с другими популяциями, проживающими в данном типе рельефа, и демонстрируют значительные отличия от популяций, проживающих на равнине. Карты генетических расстояний от «равнинных» популяций (арабы Саудовской Аравии, арабы ОАЭ, сирийцы, иорданцы, ливанцы и палестинцы) демонстрируют противоположный тренд. Карта генетических расстояний от иракцев показывает некий промежуточный тренд, но все же в большей степени тяготеет к равнинным

популяциям (рисунок 28 Ж). Также мы заметили, что сирийцы (рисунок 28 И), проживая главным образом в равнинном ландшафте, генетически ближе к географически далеко проживающим от них арабам, чем своим непосредственным географическим соседям.

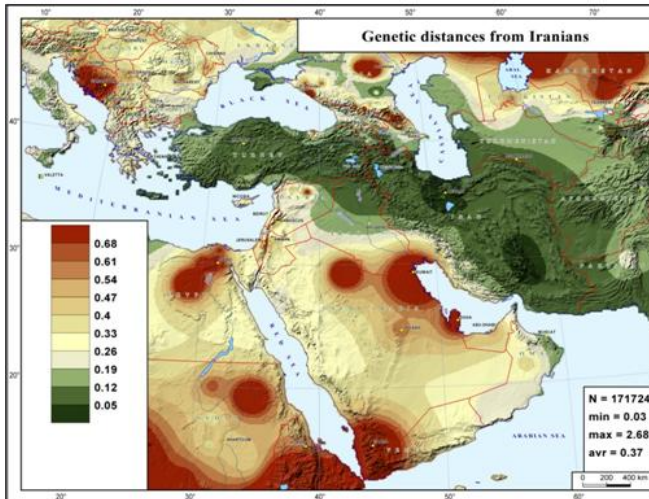
Но в целом, проведенный нами картографический анализ подтверждает закономерности, обнаруженные методом многомерного шкалирования: существование в генофонде народов Юго-Западной Азии двух паттернов сходства. Один из них приурочен к равнинному ландшафту, второй – к горной местности.



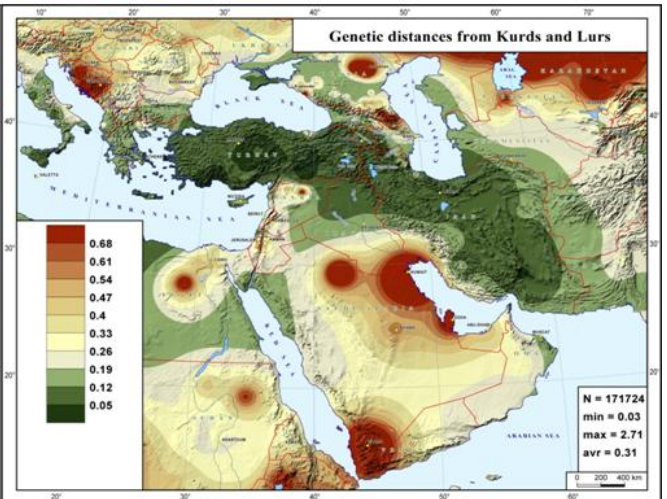
**A**



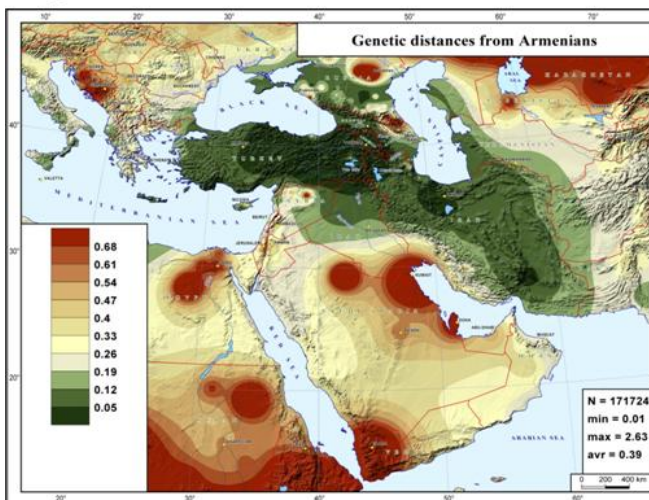
**B**



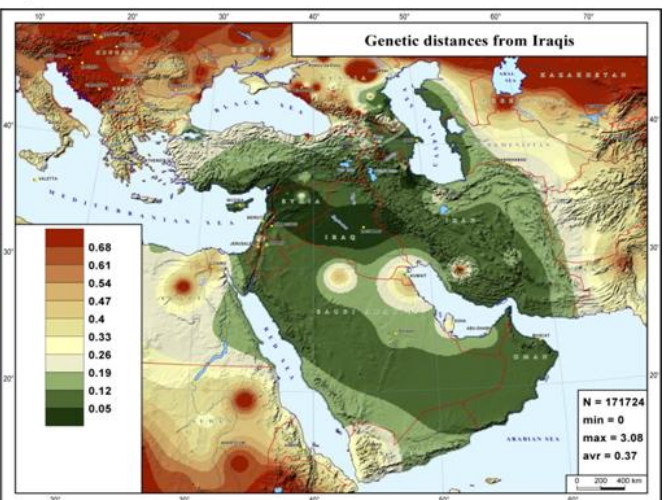
**B**



**Г**



**Д**



**E**

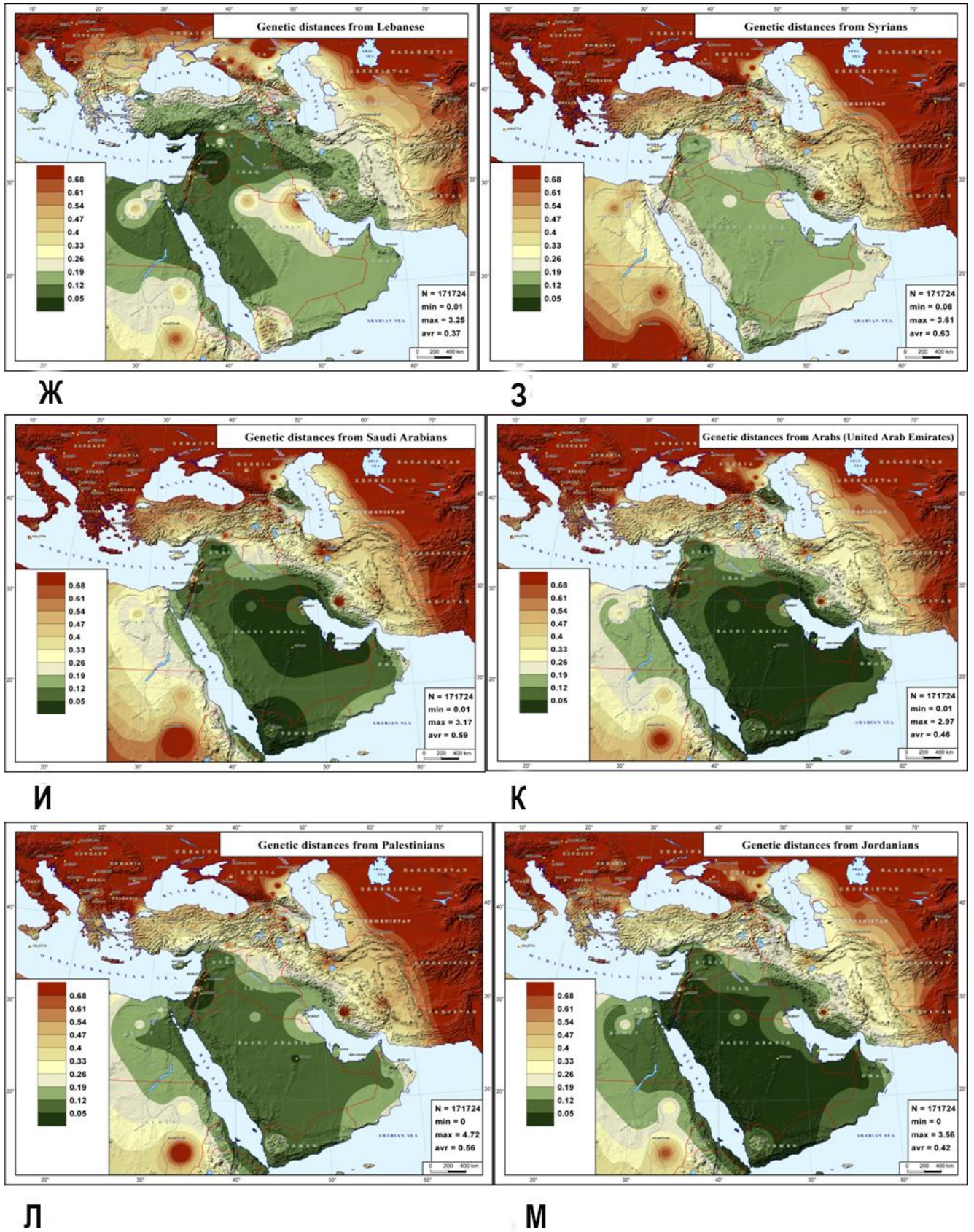


Рисунок 28. Карты генетических расстояний от популяций Юго-Западной Азии.

Рисунок 28. Карты генетических расстояний от популяций Юго-Западной Азии.

(Продолжение)

А – карта генетических расстояний от грузин;

Б – карта генетических расстояний от турок;

В – карта генетических расстояний от иранцев;

Г – карта генетических расстояний от курдов и луров;

Д – карта генетических расстояний от армян;

Е – карта генетических расстояний от иракцев;

Ж – карта генетических расстояний от ливанцев;

З – карта генетических расстояний от сирийцев;

И – карта генетических расстояний от арабов Саудовской Аравии;

К – карта генетических расстояний от арабов Объединенных Арабских Эмиратов;

Л – карта генетических расстояний от палестинцев;

М – карта генетических расстояний от иорданцев.

По всей видимости, основные черты генофонда армян сложились приблизительно 2000-2500 лет назад, согласно полученным нами филогенетическим датировкам. То есть генетические датировки образования армян, как единого этноса, приблизительно совпадают с историческими. А проживание в горной, труднодоступной для проживания местности, обеспечило сохранение их генофонда, несмотря на все последующие исторические миграции. Так, зона максимального генетического сходства со среднеармянской популяцией удивительно схожа с границами Исторической Армении. Получение такого результата, несмотря на то, что в настоящее время на этой территории по преимуществу проживают хорошо генетически изученные турки, свидетельствует о реально существующей генетической общности этих народов. Вероятно, это сходство обусловлено тем, что этногенез турок изначально шел при большой роли анатолийского субстрата, которым и послужило население западной Исторической Армении.

В завершение этого раздела рассмотрим одну из доминирующих теорий проникновения индоевропейского языка в Историческую Армению, объясняющую его появление миграцией фрако-фригийских народов на Армянское нагорье [Novannisian, 1997].

В поисках генетических свидетельств в пользу или против этой гипотезы мы обнаружили в некоторых популяциях армян ряд вариантов гаплогруппы I-M170 (в частности I\*-M438 и I-P37.2), с частотами выше, чем в соседних популяциях. Этот факт можно было бы связать с древними миграциями фрако-фригийских народов. Однако преобладание вариантов I-M170 у армян по сравнению с окружающими популяциями является лишь тенденцией и статистически недостоверно. Проведенный нами филогенетический анализ гаплогруппы I-P37.2 (рисунок 29) подтверждает это.

При этом спектр наиболее распространенных среди армян гаплогрупп – R-L23 и J-M67 и их филогенетические датировки (около 2000 лет) также слабо вписываются в концепцию данной гипотезы – согласно ей основные черты армянского генофонда должны были сложиться раньше. Историки считают, что фрако-фригийская миграция произошла приблизительно 1200-800 лет назад до н.э. Т.е. значительно раньше образования армянского этноса. Поэтому мы считаем, что явных генетических свидетельств в пользу этой гипотезы мы не обнаружили.

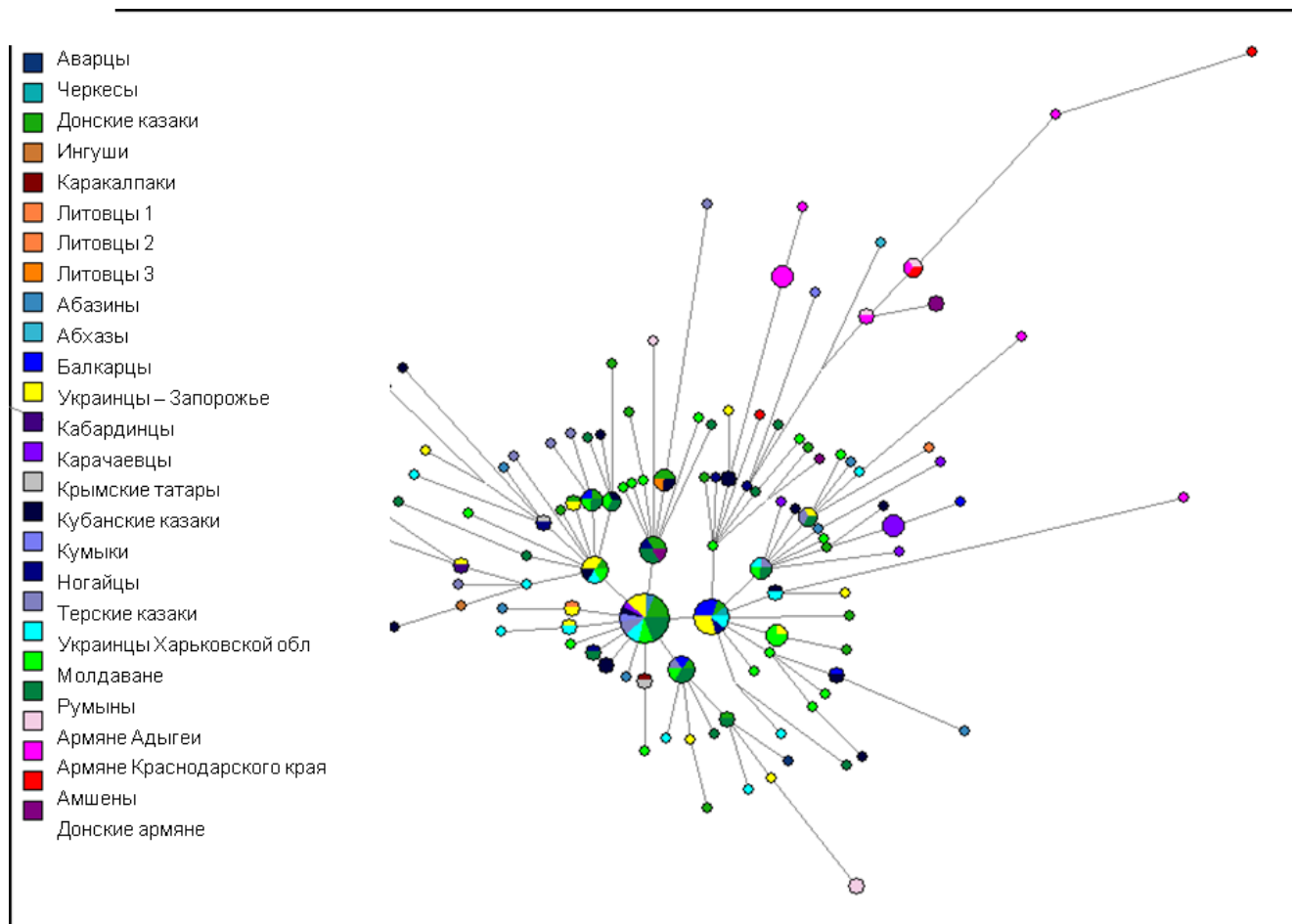


Рисунок 29. Филогенетическая сеть гаплогруппы I-P37.2

Принадлежность армян к индоевропейской языковой семье, все же нашла некоторое отражение и в структуре их генофонда – генетическая близость с принадлежащими к индоевропейской языковой семье народами Ирана, в совокупности с достаточно сильными генетическими различиями с соседними представителями иных языковых семей: картвельской (грузины), семитской (сирийцы, иракцы и др. народы Леванта) указывают на это. Проведенный нами анализ иерархической изменчивости с подразделением популяций Юго-Западной Азии в соответствии с их лингвистической принадлежностью показал, достаточно высокий уровень межгрупповой изменчивости. Многомерный анализ проведенный нами, также выявил этот тренд. Таким образом, индоевропейские популяции Юго-Западной Азии формируют в генетическом пространстве

окружающих популяций единый паттерн, географически приуроченный к Переднеазиатскому нагорью.



## Заключение

Изучение народов индоевропейской языковой семьи по разным системам маркеров показало, что популяции, разговаривающие на языках индоевропейской языковой семьи и соседние с ними иноязычные популяции генетически похожи. Их генетические взаимоотношения лучше объясняются сочетанием географического и лингвистического факторов, чем одним лишь географическим ареалом.

Общих маркеров для всех народов индоевропейской языковой семьи не обнаружено. Анализ пар популяций показал, что распространение индоевропейских языков чаще всего шло по модели культурного заимствования, а не демической экспансии.

Единый общий компонент генофонда, характерный для всех народов индоевропейской языковой семьи не обнаружен по полногеномным данным. При этом на формирование генетической структуры популяций, разговаривающих на языках индоевропейской языковой семьи, оказали влияние как географический (показатель общей корреляции между генетической и географической близостью популяций при проведении теста Мантеля – 0,72), так и лингвистический факторы (показатель корреляции - 0,56).

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, мы сосредоточились на подробном анализе отдельных лингвистических групп в пределах индоевропейской семьи.

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 преимущественно культурным и не оставило заметных следов в генофонде коренного населения Европы.

В генофонде армян нами обнаружено большое разнообразие гаплогрупп и отсутствие какой-либо одной доминантной гаплогруппы. Наиболее распространены в генофонде армян гаплогруппы R-L23, J-M67 и G-M285.

Предположение о том, что армяне в диаспоре сохранили черты исходного протогенофонда исторической Армении подтверждает и проведенный нами анализ.

Анализ положения армян в контексте всей Юго-Западной Азии выявил 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.

## **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 литературы

1. Abu-Amero, K.K. [et al.]. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions / K.K. Abu-Amero [et al.] // BMC Genetics. – 2009. – P.10-59.
2. Allentoft, M.E. [et al.]. Population genomics of Bronze Age Eurasia / M.E. Allentoft [et al.] // Nature. – 2015. – V.522(7555). – P.167-172.
3. ArunKumar, G. [et al.]. Genome-wide signatures of male-mediated migration shaping the Indian gene pool / G. ArunKumar [et al.] // J Hum Genet. – 2015. – V. 60(9). P.493-499.
4. Balanovsky, O. [et al.]. Two sources of the Russian patrilineal heritage in their Eurasian context / O. Balanovsky [et al.] // Am. J. Hum. Gen. – 2008. – V. 82 (1). – P. 236–250.
5. Balanovsky, O. [et al.]. Parallel Evolution of Genes and Languages in the Caucasus Region / O. Balanovsky [et al.] // Molecular biology and evolution. – 2011. – V.28(10). – P.2905-2920.
6. Balanovsky, O. [et al.]. Geographical Relief and Prehistoric Migrations Shape the Y-Chromosomal Landscape of West Asian Populations/ O. Balanovsky [et al.] // Human Genetics.- 2017. – P. 136:437-450.
7. Battaglia, V. [et al.]. Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe / V. Battaglia [et al.] // European Journal of Human Genetics. – 2009. – V.17(6). – P.820-830.
8. Behar, D.M. [et al.]. The genome-wide structure of the Jewish people / D.M. Behar [et al.] // Nature. – 2010. – V.466. – P.238–242.
9. Belledi, M. [et al.]. Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations / M. Belledi [et al.] // Eur. J. of Hum. Gen. – V. 8. – 2000. – P. 480-486.
10. Bender, H. The Home of the Indo-Europeans. - Princeton, 1922. P. 53.

11. Biro, A.Z. [et al.]. A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary) / A.Z. Biro [et al.] // *American Journal of Physical Anthropology*. – 2009. – V.139(3). – P.305-10.
12. Broushaki, F. [et al.]. Early Neolithic genomes from the eastern Fertile Crescent / F. Broushaki [et al.] // *Science*. – 2016. – V. 29, 353(6298). – P.499-503.
13. Cadenas, A.M. [et al.]. Y-chromosome diversity characterizes the Gulf of Oman / A.M. Cadenas [et al.] // *European journal of human genetics*. - . 2008. – V.16(3). – P.374–386.
14. Carvalho, B.M., Santos F.R. Human Y-chromosome variation and male dysfunction / B.M. Carvalho, F.R. Santos // *Journal of Molecular and Genetic Medicine*. – 2005. – V.1 (2). – P. 63–75.
15. Cavalli-Sforza, L. L. *The History and Geography of Human Genes*. - Princeton, NJ. - Princeton University Press. – 1994. P.358.
16. Cavalli-Sforza, L. L., Luca L. *Genes, Peoples, and Languages*. - New York: North Point Press . – 2000. – P. 453.
17. Cinnioglu, C. [et al.]. Excavating Y-chromosome haplotype strata in Anatolia / C Cinnioglu [et al.] // *Human genetics*. – 2004. – V.114(2). -P.127–48.
18. Cordaux, R., Saha, N., Bentley, G.R. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India / R. Cordaux, N. Saha, G.R. Bentley // *Eur. J. of Hum. Gen.* – V.11. – 2003. – P. 253-264.
19. Cruciani, F. [et al.]. Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12 / F. Cruciani [et al.] // *Mol Biol Evol.* - 2007. - V. 24. - № 6. P.1300-1311.
20. Damgaard, P. [et al.]. The first horse herders and the impact of early Bronze Age steppe expansions into Asia / Damgaard P. [et al.] // *Science*. – 2018. – V.360 – P.1422-1433.

21. Di Cristofaro, J. [et al.]. Afghan Hindu Kush: Where Eurasian Sub-Continent Gene Flows Converge / J. Di Cristofaro [et al.] // PLoS ONE. -2013. V/8(10).
22. Dolgopolsky, A. The Indo-European Homeland and Lexical Contacts of Proto-Indo-European with other Languages/ A. Dolgopolsky // Mediterranean Language Review. - 1988. - №3. - P. 12.
23. Elhaik, E. [et al.]. The GenoChip: a new tool for genetic anthropology / E. Elhaik [et al.] // Genome Biol Evol. – 2013. – V.5(5). – P.1021-31.
24. El-Sibai, M. [et al.]. Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast / M. El-Sibai [et al.] // Annals of human genetics. – 2009. - V.73. – P.568-581.
25. Fedorova, S.A. [et al.]. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia / S.A Fedorova [et al.] // BMC Evol Biol. – 2013. – V.19. – P.113-127.
26. Fenner, J.N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies / J.N. Fenner // Am J Phys Anthropol. – 2005. V.128. – P.415–423.
27. Flores, C. N. [et al.]. Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan / C. N. Flores [et al.] // J Hum Genet .- 2005. – V.50(9). – P. 435-441.
28. Forster, P. [et al.]. Origin and evolution of Native American mtDNA variation: a reappraisal / P. Forster [et al.] // American Journal of Human Genetics. – 1996. V.59(4). – P.935-945.
29. Francalacci, P. [et al.]. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny / P. Francalacci [et al.]. // Science. - 2013. – 341. – P.565–569.
30. Gallego-Llorente, M. [et al.]. The genetics of an early Neolithic pastoralist from the Zagros, Iran / M. Gallego-Llorente [et al.] // Sci Rep. - 2016. – V.9. – P.6:31326.

31. Ge, J. [et al.]. Mutation rates at Y chromosome short tandem repeats in Texas populations / J. Ge [et al.] // *Forensic science international Genetics*. – 2009. V. 3(3). P.179–84.
32. Gimbutas, M. The First Wave of Eurasian Steppe Pastoralists into Copper Age Europe / M. Gimbutas // *JIES*. - 1977.- № 5. - P. 277-338.
33. Goldberg, A., Mychajliw A.M., Hadly E.A. Post-invasion demography of prehistoric humans in South America / A. Goldberg, A.M. Mychajliw, E.A. Hadly // *Nature*. – 2016. – V.14;532(7598). P.232-5.
34. Gray, R. D. [et al.]. Language evolution and human history: what a difference a date makes / Gray R. D. [et al.] // *Phil. Trans. R. Soc. B*. –2011.- 366. – P.1090–1100.
35. Gray, R. D., Atkinson Q. D. Language-tree divergence times support the Anatolian theory of Indo-European origin / R. D. Gray, Q. D. Atkinson // *Nature*. - Vol. 426 – 2003. – P. 435-439.
36. Grugni, V. [et al.]. Ancient Migratory Events in the Middle East: New Clues from the Y-Chromosome Variation of Modern Iranians / V. Grugni [et al.] // *PLoS ONE*. – 2012. V.7(7):e41252.
37. Grzybowski, T. [et al.]. Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis/ T. Grzybowski [et al.] // *Forensic Sci Int Genet*. – 2007. V.1. P. 141–147.
38. Gusmao, L. [et al.] Grouping of Y-STR haplotypes discloses European geographic clines / L. Gusmao [et al.] // *Forensic Science International*. – 2003. – V. 134. – P. 172–179.
39. Haak, W. [et al.]. Massive migration from the steppe was a source for Indo-European languages in Europe / W. Haak [et al.] // *Nature*. – 2015. – V.7555. P.207-11.
40. Haak, W. [et al.]. Ancient DNA from European early neolithic farmers reveals their near eastern affinities / W. Haak [et al.] // *PLoS Biol*.- 2010. - V. 8. - № 11.

41. Haber, M. [et al.]. Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon / M. Haber [et al.] // *European Journal of Human Genetics*. – 2011. – V.19(3). – P.334-340.
42. Haber, M. [et al.]. Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations / M. Haber [et al.] // *European Journal of Human Genetics*. – 2016. – V. 24. - P.931-936.
43. Hammer, M.F. [et al.] Human population structure and its effects on sampling Y chromosome sequence variation / Hammer M.F. [et al.] // *Genetics*. – 2003. - V.164. – P. 1495–1509.
44. Hammer, M.F., Zegura S.L. The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions / M.F. Hammer, S.L. Zegura // *The Annual Review of Anthropology*. – 2002. – V.31. – P. 303–321.
45. Hellenthal, G. [et al.]. A Genetic Atlas of Human Admixture History / G. Hellenthal [et al.] // *Science*. – 2014. – V.343. – P.747–751.
46. Herrera K.J., [et al.]. Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists / K.J. Herrera [et al.]. // *European Journal of Human Genetics*. – 2012. - 20(3). P.313-320.
47. Hovannisian, R.G. *The Armenian people*. - New York: St Martin's, 1997. – 228 p.
48. Hovhannisyán, A. [et al.]. Different waves and directions of Neolithic migrations in the Armenian Highland / A. Hovhannisyán [et al.] // *Investigative Genetics*. – 2014. P.5-15.
49. Idem, A. A Neolithic Model of Indo-European Prehistory/ A. Idem // *JIES*. 1992. - 20. - P. 193-238.
50. Ilumäe, A-M. [et al.]. Human Y-chromosomal haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families / A-M. Ilumäe [et al.] // *Am J Hum Genet*. - 2016. – V.7, 99(1). – P.163-73.



51. International Society of Genetic Genealogy. Y-DNA Haplogroup Tree 2017 / International Society of Genetic Genealogy // [isogg.org](http://isogg.org) . - Version: [12.98]. - Date: [14 April 2017]. – Режим доступа: <http://www.isogg.org/tree/> [15 April 2017].
52. Jobling, M.A., Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age / M.A. Jobling, C. Tyler-Smith // *Nat. Rev. Genet.* – 2003. – V.4. – P. 598–610.
53. Jobling, M.A. Y-chromosomal SNP haplotype diversity in forensic analysis/ M.A. Jobling // *Forensic Science International.* – 2001. – V.118. – P. 158–162.
54. Jobling, M.A. [et al.]. Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, MSY1 / M.A. Jobling [et al.] // *Hum. Mol. Genet.* – 1999. – V.8, №11. – P. 2117– 2120.
55. Jobling, M.A. The Y chromosome in forensic analysis and paternity testing / M.A. Jobling, A. Pandya, C. Tyler-Smith et al. // *Int. J. Legal Med.* – 1997. – V.110. – P. 118–24.
56. Jones, E. [et al.]. Upper Palaeolithic genomes reveal deep roots of modern Eurasians / E Jones [et al.] // *Nature Communications.* – 2015. – V.9. – P.66-69.
57. Juras, A. [et al.]. Ancient DNA Reveals Matrilineal Continuity in Present-Day Poland over the Last Two Millennia / A. Juras [et al.] // *PLoS ONE.* – 2014. – V.9. – P. 8912-8920.
58. Karachanak, S. [et al.]. Bulgarians vs the other European populations: a mitochondrial DNA perspective / S. Karachanak [et al.] // *Int J Legal Med.* -2012. – V.126. –P. 497–503.
59. Karafet, T.M. High levels of Y-chromosome differentiation among native siberian populations and the genetic signature of a boreal hunter-gatherer way of life / T.M. Karafet [et al.] // *Human Biology.* – 2002. – V. 74, №6. – P. 761–789.
60. Karlsson, A.O. [et al.]. Y-chromosome diversity in Sweden - a long-time perspective / A.O. Karlsson [et al.]. // *Eur J Hum Genet.* - 2006. - V. 14. - № 8. - P. 963-70.

61. Karmin, M. [et al.]. A recent bottleneck of Y chromosome diversity coincides with a global change in culture / M. Karmin [et al.] // *Genome Research*. – 2015. - 25(4). – P.459-466.
62. Kasperaviciūte, D., Kucinskas, V., Stoneking, M. Y chromosome and mitochondrial DNA variation in Lithuanians / D. Kasperaviciūte, V. Kucinskas, M. Stoneking // *Ann Hum Genet*. – 2004. - V68. – P. 438–452.
63. Kayser, M. [et al.]. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis / M. Kayser [et al.] // *Hum Genet*. – 2005. – V.117(5). – P.428-43.
64. Khrunin, A.V. [et al.]. A genome-wide analysis of populations from European Russia reveals a new pole of genetic diversity in northern Europe / A.V. Khrunin [et al.]. // *PloS One*. – 2013. – V.8.
65. King, R.J., Ozcan S.S., Carter T. Differential Y-chromosome Anatolian Influences on the Greek and Cretan Neolithic / King R.J., Ozcan S.S., Carter T. // *Annals of Human Genetics*. – 2008. – V.72. – P.205–214.
66. Kivisild, T., Rootsi, S., Metspalu M. The Genetic Heritage of the Earliest Settlers Persists Both in Indian Tribal and Caste Populations / T. Kivisild, S. Rootsi, M. Metspalu // *Am. J. Hum. Genet*. - V. 72. – 2003. – P.313–332.
67. Kushniarevich, A. [et al.]. Uniparental genetic heritage of belarusians: encounter of rare middle eastern matrilineages with a central European mitochondrial DNA pool / A. Kushniarevich [et al.] // *PLoS One*.- 2013. – V.13;8(6).
68. Kushniarevich, A. [et al.]. Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data / A. Kushniarevich [et al.] // *PLoS One*. - 2015. - V. 10(9):e0122968.
69. Lahn, B. T. [et al.]. The human Y chromosome, in the light of evolution / B. T. Lahn [et al.] // *Nat. Rev. Genet*. . – 2001. – V.2. – P. 207–216.
70. Lao, O. [et al.]. Correlation between genetic and geographic structure in Europe / O. Lao [et al.] // *Curr Biol CB*. – 2008. – V.18. – P.1241–1248.

71. Lappalainen, T. [et al.]. Migration waves to the Baltic Sea region / T. Lappalainen [et al.] // *Ann Hum Genet.* - 2008. - V. 72. - № 3. - P. 337-348.
72. Lazaridis, I. [et al.]. Ancient human genomes suggest three ancestral populations for present-day Europeans / I. Lazaridis [et al.]. // *Nature.* - V.713. - P.409-13.
73. Lowery, R.K. [et al.]. Regionalized autosomal STR profiles among Armenian groups suggest disparate genetic influences / R.K. Lowery [et al.] // *Am J Phys Anthropol.* – 2011. P.146:171–178.
74. Luca, F. Y-chromosomal variation in the Czech Republic / F. Luca [et al.] // *Am J Phys Anthropol.* -2007. - V.132. - P.132–139.
75. Mallory, J. P. *In Search of the Indo-Europeans.* - L., 1989. P. 152.
76. Mallory, J. P. *The Indo-European Homeland: an Asian Perspective* / J. P. Mallory // *Bulletin of the Deccan College Post Graduate and Research Institute.* - 1995. P.154.
77. Malyarchuk, B. [et al.]. Mitochondrial DNA Phylogeny in Eastern and Western Slavs / B. Malyarchuk [et al.] // *Mol Biol Evol.* – 2008. – V.25. – P.1651–1658.
78. Malyarchuk, B.A. [et al.]. Mitochondrial DNA variability in Bosnians and Slovenians / B.A. Malyarchuk [et al.] // *Ann Hum Genet.* - 2003. – V.67. P.412–425.
79. Malyarchuk, B.A. [et al.]. Mitochondrial DNA variability in Poles and Russians / B.A. Malyarchuk [et al.]. // *Ann Hum Genet.* – 2002. – V.66. – P.261–283.
80. Margaryan, A. [et al.]. Paternal lineage analysis supports an Armenian rather than a Central Asian genetic origin of the Hamshenis / A. Margaryan [et al.] // *Hum Biol.* – 2012. – V.84(4). – P.405-22.
81. Margaryan, A. [et al.]. Eight Millennia of Matrilineal Genetic Continuity in the South Caucasus / A. Margaryan [et al.] // *Curr Biol.* – 2017. – V.27(13). – P.2023-2028.

82. Mathieson, I.[et al.]. Genome-wide patterns of selection in 230 ancient Eurasians / I. Mathieson [et al.] // *Nature*. – 2015. – V.528(7583). P.499-503.
83. Mendez, F.L. [et al.]. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree / F.L. Mendez [et al.] // *Am J Hum Genet*. – 2013. – 92. – P.454–459.
84. Mielnik-Sikorska, M. [et al.]. The history of Slavs inferred from complete mitochondrial genome sequences // M. Mielnik-Sikorska [et al.] / *PloS One*. – 2013. –V.8: e54360.
85. Mirabal, S. [et al.]. Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia / S. Mirabal [et al.]. // *Eur J Hum Genet*. 2009. – V. 17. – № 10. – P. 1260-73.
86. Mitnik, A. [et al.]. The genetic prehistory of the Baltic Sea region / A. Mitnik [et al.]. // *Nat Commun*. 2018. –V. 9(1). – P. 442.
87. Morozova, I. [et al.]. Russian ethnic history inferred from mitochondrial DNA diversity / I. Morozova [et al.] // *Am J Phys Anthropol*. – 2012. – V.147. – P. 341–351.
88. Mountain, J.L. [et al.]. SNPSTRs: Empirically Derived, Rapidly Typed, Autosomal Haplotypes for Inference of Population History and Mutational Processes / J.L. Mountain [et al.] // *Genome Research*. - 2002. - V.12(11). - P.1766-1772.
89. Myres, N.M. [et al.]. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe / N.M. Myres N.M. [et al.] // *Eur J Hum Genet*. – 2011. – V. 19. – № 1. – P. 95-101.
90. Nasidze, I. [et al.]. Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome / I. Nasidze [et al.] // *Hum Genet*.- 2003. – V.112(3). – P. 255-261.
91. Network 4.1.1.2 (FluxusTechnologyLtd., [www.fluxus-engineering.com](http://www.fluxus-engineering.com)) Network Publisher (Fluxus Engineering, Clare, U.K.).

92. Nielsen, R. [et al]. Tracing the peopling of the world through genomics / R. Nielsen [et al] // Nature. – 2017. – V.541(7637). – P.302–310.
93. Passarino, G. [et al.]. The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea / G. Passarino [et al.] // Hum Immunol. 2001. – V.62(9). – P.922-32.
94. Pagani, L. [et al.]. An Ethnolinguistic and Genetic Perspective on the Origins of the Dravidian-Speaking Brahui in Pakistan / L. Pagani [et al.] // Man India. 2017. – V.97(1). – P.267-278.
95. Perićić, M. [et al.]. High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations / M. Perićić [et al.] // Mol Biol Evol. - 2005. - V.22. - P.1964–1975.
96. Platt, D.E. [et al.]. Mapping Post-Glacial expansions: The Peopling of Southwest Asia / D.E. Platt [et al.] // Sci Rep. – 2017. V. 6. – P. 7:40338.
97. Poznik, G.D. [et al.]. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females/ G.D. Poznik [et al.] // Science. 2013. – 2. – V.341. – P.562-5.
98. Primorac, D. [et al.]. Croatian genetic heritage: Y-chromosome story. Croat Med J. – 2011. - V.52(3). – P.225-34.
99. [Purcell, S. Plink \(PLINK \(1.07\)\) \(электронный ресурс\) / S. Purcell. – Режим доступа: http://pngu.mgh.harvard.edu/purcell/plink/.](http://pngu.mgh.harvard.edu/purcell/plink/)
100. Purcell, S. [et al.]. PLINK: a tool set for whole-genome association and population-based linkage analyses / S. Purcell [et al.] // Am J Hum. - Genet.- 2007. V. 81(3). – P.559-75.
101. Quintana-Murci, L. [et al.]. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia / L. Quintana-Murci [et al.] // Am J Hum Genet. – 2001. – V.68(2). P.537-42.
102. Ralph, P., Coop, G. The Geography of Recent Genetic Ancestry across Europe / P. Ralph, G. Coop // PLoS Biol. – 2013ю – V.11. – P:e1001555.

103. Rębała, K. [et al.]. Contemporary paternal genetic landscape of Polish and German populations: from early medieval Slavic expansion to post-World War II resettlements / K. Rębała [et al.] // *Eur J Hum Genet.* – 2013. – V.21(4). – P.415-22.
104. Regueiro, M. [et al.]. Iran: tricontinental nexus for Y-chromosome driven migration / M. Regueiro [et al.] // *Hum Hered.* -2006. – V.61(3). - P.132-43.
105. Regueiro, M. [et al.]. High levels of Paleolithic Y-chromosome lineages characterize Serbia / M. Regueiro [et al.] // *Gene.* – 2012. – V.498(1). P.59-67.
106. Renfrew, A. *Archaeology and Language.* L.: Jonathan Cape - 1987. – P.386.
107. Rootsi, S. [et al.]. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe / S. Rootsi [et al.] // *Am J Hum Genet.* - 2004. - V. 75. - № 1. - P.128-137.
108. Rootsi, S. [et al.]. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe / S. Rootsi [et al.] // *Eur J Hum Genet.* – 2007. - V.15(2). – P.204-11.
109. Rootsi, S. [et al.]. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus / S. Rootsi [et al.] // *Eur J Hum Genet.* - 2012. - V. 20. - № 12. - P. 1275-1282.
110. Rosser, Z. [et al.]. Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language / Z. Rosser, T. Zerjal, M. Hurles // *Am. J. Hum. Genet.* – V. 67. - . 2000. - . P.196-199.
111. Roewer, L., [et al.]. Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations / Roewer, L., [et al.] // *Int. J. Legal Med.* - 2008. - V. 122(3). - P. 219-223.
112. Sahoo S., Slngh A., Hlmablndu G. A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios / Sahoo S., Slngh A., Hlmablndu G. // *PNAS* . – 24. – Vol. 26. – 2003. – P. 843–848.

113. Saillard, J. [et al.]. mtDNA Variation among Greenland Eskimos: The Edge of the Beringian Expansion / J. Saillard [et al.] // *American Journal of Human Genetics*. – 2000. V.67(3). – P.718-726.
114. Sanchez, J. J. [et al.]. High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males / J. J. Sanchez, [et al.] // *Eur J Hum Genet*. – 2005. V.13(7). – P. 856-866.
115. Sanchez-Diz, P. [et al.]. Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study / P. Sanchez-Diz [et al.] // *International journal of legal medicine*. – 2008. – V.122(6). - P.529–33.
116. Semino, O. [et al.]. The Genetic Legacy of Paleolithic Homo sapiens sapiens in Extant Europeans: A Y Chromosome Perspective / O. Semino [et al.] // *Science*. – Vol.290. – 2000. – P. 1155-1159.
117. Sengupta, S. [et al.]. Polarity and Temporality of High-Resolution Y-Chromosome Distributions in India Identify Both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists / S Sengupta [et al.] // *Am J Hum Genet*. 2006. –V.78. – P.203-21.
118. Shi, W. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP–CEPH populations / Shi W. [et al.] // *Mol. Biol. Evol*. – 2010. – V. 27. – P. 385–393.
119. Shou, W.H. [et al.]. Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians / W.H. Shou [et al.] // *J Hum Genet*. - 2010. - V. 55. - № 5.-P. 314-322.
120. Srikummool, M. [et al.]. Minimal Y-chromosomal haplotype selection for phylogenetic study using the bootstrapped DTI method / M. Srikummool [et al.]. // *ScienceAsia*. - 2006. – V.32. – P.361-364.
121. Statistica 6.0 (StatSoft. Inc., 2001) (StatSoft, Inc. STATISTICA (data analysis software system). version 62001. – 2001. – Режим доступа: [www.statsoft.com](http://www.statsoft.com).

122. Swadesh, M. Towards greater accuracy in lexicostatistic dating / M. Swadesh // *International Journal of American Linguistics*. -1955. – V.21.- P.121-137.
123. Thanseem, I. [et al.]. Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA / I. Thanseem [et al.]. // *BMC Genet.* – 2006. – V.7. – P.7:42.
124. Thomas, M. G. [et al.]. Origins of old testament priests / M. G. Thomas [et al.] // *Nature*. – 1998. – V.394. – P. 138–140.
125. Thomas, M.G. [et al.]. Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba--the "Black Jews of Southern Africa / M. G. Thomas [et al.] // *Am. J. Hum. Genet.* – 2000. – V.66. – P. 674–686.
126. Trivedi, R. [et al.]. Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago / R. Trivedi [et al.]. // *J Hum Genet.* - 2006. – V.51. – P. 217-26.
127. Underhill, P.A. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography / P.A. Underhill [et al.] // *Genome Research*. – 1997. – V.7. – P. 996–1005.
128. Underhill, P.A. [et al.]. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a / P.A. Underhill [et al.] // *European Journal of Human Genetics*. - 2010. – V.18(4). – P.479-484.
129. Underhill, P.A. [et al.]. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a / P.A. Underhill [et al.] // *Eur J Hum Genet.* – 2015. – V.23(1). P.124-31.
130. Varzari, A. [et al.]. Population history of the Dniester-Carpathians: evidence from Alu markers / A. Varzari [et al.]. // *J Hum Genet.* - 2007. - V.52(4). - P.308-16.
131. Varzari, A. [et al.]. Searching for the origin of Gagauzes: inferences from Y-chromosome analysis / A. Varzari [et al.] // *Am J Hum Biol.* – 2009. - V.21(3). – P.326-36.



132. Veeramah, K.R. [et al.]. Genetic variation in the Sorbs of eastern Germany in the context of broader European genetic diversity / K.R. Veeramah [et al.] // *Eur J Hum Genet EJHG*. 2011. – V.19. – P. 995–1001.
133. Weale, M. E. [et al.]. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group / M. E. Weale [et al.] // *Hum Genet*. - 2001. - V.109(6). - P.659-674.
134. Weir, B. S., Cockerham C. C. Estimating F-statistics for the analysis of population structure / B. S. Weir, C. C. Cockerham // *Evolution*. – 1984. – V.38. – P.1358-1370.
135. Wells, R.S. [et al.]. The Eurasian heartland: a continental perspective on Y-chromosome diversity / R.S. Wells [et al.] // *Proc. Natl. Acad. Sci. USA* - V.98. - 2001. - P.10244-10249.
136. Xue, Y. [et al.]. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree / Y. Xue [et al.] // *Curr Biol*. – 2009. - 19(17). – P.1453-7.
137. Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups / Y Chromosome Consortium // *Genome Res*. – 2002. – V.12. – P. 339–348.
138. Yunusbayev, B. [et al.]. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations / B. Yunusbayev [et al.] // *Molecular biology and evolution*. – 2012. – V.29(1). P.359–65.
139. Yunusbayev, B. [et al.]. The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia / B. Yunusbayev [et al.] // *PLoS Genet*. – 2015. – V.21. - P. 4-11.
140. Zalloua, P.A. [et al.]. Identifying Genetic Traces of Historical Expansions: Phoenician Footprints in the Mediterranean / P.A. Zalloua [et al.] // *American Journal of Human Genetics*. – 2008. V.83(5). – P.633-642.

141. Zalloua, P.A. [et al.]. Y-Chromosomal Diversity in Lebanon Is Structured by Recent Historical Events / P.A. Zalloua [et al.] // *American Journal of Human Genetics*. – 2008. V.82(4). P.873-882.
142. Zerjal, T. [et al.]. The Genetic Legacy of the Mongols / T. Zerjal [et al.] // *Am J Hum Genet*. – 2003. – V.72. – P.717-21.
143. Zhivotovsky, L.A. [et al.]. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time / L.A. Zhivotovsky [et al.] // *Am J Hum. Genet.* - 2004. V.74. – P.50–61.
144. Алексаха, А.Г. Происхождение славян. Прогрессологическая реконструкция / А.Г. Алексаха // *Гуманитарный журнал* . - № 2-3. – 2012. - с. 155-181.
145. Альквист, А. Субстратная топонимия Ярославского Поволжья. Очерки исторической географии: Северо-Запад России: Славяне и финны / А. Альквист; Ред. А. С. Герд, Г. С. Лебедев. - Санкт-Петербург: СПбГУ, 2001. - С. 436-467
146. Астапенко, М.П. История донского казачества в повествованиях и рассказах / М.П. Астапенко. - в 5 т. - Ростов\_н/Д:Терра, 2000. - Т. 1. - 700 с.
147. Балаганская, О.А. [и др.]. Полиморфизм Y хромосомы у тюркоязычного населения Алтае-Саян, Тянь-Шаня и Памира в контексте взаимодействия генофондов Западной и Восточной Евразии / О.А. Балаганская [и др.] // *Медицинская генетика*. - 2011. - Т10. - №3. - С. 12-22.
148. Балановская, Е. В., Балановский О. П. Русский генофонд на Русской равнине. — М: Луч, 2007. – 416 С.
149. Балановская, Е.В. [и др.] Генофонд Русского Севера: славяне? Финны? Палеоевропейцы? / Е.В. Балановская [и др.] // *Вестник Московского Университета*. - Серия XXIII "Антропология". - 2011. - №3. - С. 27-58.
150. Балановская, Е. В. [и др.]. Популяционные биобанки: принципы организации и перспективы применения в геногеографии и

- персонализированной медицине / Е. В. Балановская [и др.] // Генетика. – 2016. - № 12. – С.1371-1378.
151. Балановский, О.П. Генофонд Европы / Балановский О.П. – Москва Товарищество научных изданий КМК. - 2015. С.338
152. Бурлак, С.А., Старостин С.А. Введение в лингвистическую компаративистику / С.А. Бурлак, С.А. Старостин. – Москва - Эдиториал УРСС. – 2001. С. 270.
153. Бурлак, С. А. Languages, dna, relationship and contacts / С. А. Бурлак // Вопросы языкового родства. — 2013. — №. 9. — С. 55–67.
154. Волков, Ю.Г. Донские казаки в прошлом и настоящем / Ю.Г. Волков. - Ростов\_н/Д: ГинГо, 1998. - 540 с.
155. Гамкрелидзе, Т.В., Иванов В.В. Индоевропейский язык и индоевропейцы. – Тбилиси. – 1984. – С.184.
156. Захаров, С.Д. Белоозеро. Русь в IX-X веках. Археологическая панорама / С.Д. Захаров, ред. Н. А. Макаров. - Москва – Вологда: «Древности Севера», 2012. С. 212-241.]
157. Индоевропейские языки: лингвистический энциклопедический словарь / гл.р. Ярцева В.Н. – М.: "Советская энциклопедия" - 1990. – с. 566.
158. Клейн, Л.С.. Древние миграции и происхождение индоевропейских народов. – С.-П. – 2007. – С. 226.
159. Кобычев, В. Где находится родина индоевропейцев?/ В. Кобычев // Scientific.ru. – 2004 - Режим доступа:  
<http://www.scientific.ru/journal/translations/vlad.html>.
160. Кошель, С.М. Геоинформационные технологии в геногеографии. Современная географическая картография / С.М. Кошель; под ред. И.К. Лурье и В.И. Кравцовой. М.: Дата+, 2012. - С. 158–166.
161. Лакин, Г.Ф. Биометрия / Г.Ф. Лакин. - М.: Издательство: Высшая школа, 1990. – 350 с.

162. Макаров, Н.А. Колонизация северных окраин Древней Руси в XI-XIII вв / Н.А. Макаров. - Москва: «Скрипторий», 1997. 358 с.
163. Матвеев, А.К. Субстратная топонимия Русского Севера / А.К. Матвеев III. - Екатеринбург: Изд-во Уральского университета, 2007. -215 с.
164. Матвеев, А.К. Субстратная топонимия Русского Севера. IV. Топонимия мерянского типа / А.К. Матвеев. - Екатеринбург: Изд-во Уральского университета, 2015. - 313 с.
165. Мэллори, Дж. П. Индоевропейские прародины / Дж. П. Мэллори // . Вестник древней истории. - 1997. - №1. - С. 61.
166. Недошивина, Н.Г., Зозуля С.С. Курганы Ярославского Поволжья // Русь в IX-X веках. Археологическая панорама / Н.Г. Недошивина, С.С. Зозуля; ред. Н. А. Макаров. - Москва – Вологда: «Древности Севера», 2012. - С. 179-193.
167. Почешхова, Э.А. [и др.]. Динамика генофонда во времени по данным о фамилиях и родословных / Э.А. Почешхова [и др.] // Медицинская генетика. - 2008. -Т.7. - №8. - С.25-29.
168. Проценко, Б.Н. Оппозиция “свой–чужой” и менталитет донских казаков. История и культура народов степного Предкавказья и Северного Кавказа: проблемы межэтнических отношений / Б.Н. Проценко. - сб. научн. статей. - Ростов\_на\_Дону, 1999. - С. 211–221.
169. Сафронов, В. Индоевропейские прародины. Горький: волговятское-Вятское кн. Издательство. - 1989. – с. 398 с.,
170. Смирнов, О.В. «Марийская» гипотеза в исследовании топонимии Оки и Унжи и западные границы древнемарийской топонимии. Вопросы ономастики / О.В. Смирнов. - Екатеринбург, 2015. - №2(19). - С. 7-61.
171. Спицын, В.А. [и др.]. Русский генофонд. Частоты генетических маркеров / В.А. Спицын [и др.] // Генетика. – 2001. – Т.37, № 3. – С.386–401.

172. Старостин, С. А. Сравнительно-историческое языкознание и лексикостатистика // Лингвистическая реконструкция и древнейшая история Востока. М.: Наука, 1989, стр.3-39.
173. Степанов, В.А., Харьков В.Н., Пузырев В.П. Эволюция и филогеография линий У хромосомы человека/ В.А. Степанов, В.Н. Харьков, В.П. Пузырев // Вестник ВОГиС. – 2006. – Т.10, № 1. – С.57–73.
174. Теучеж, И.Э. [и др.]. Генофонды абхазо-адыгских народов, грузин и армян в евразийском контексте / И.Э. Теучеж [и др.] // Вестник Московского Университета. Серия XXIII "Антропология". - 2013. - №2. - С. 49-62.
175. Теучеж, И.Э. Генофонд абхазо-адыгских народов, грузин и армян по данным о полиморфизме Y-хромосомы и фамилий: автореф. дис. на соискание ученой степени канд. биол. наук: 03.02.07 / Теучеж Ирина Эдуардовна. - М., 2013. — 24 с.
176. Третьяков, П.Н. Финно-угры, балты и славяне на Днепре и Волге. Москва/ П.Н. Третьяков. – Ленинград: «Наука», 1966. 294 с.
177. Харьков, В.Н. [и др.]. Частоты диаллельных гаплогрупп Y-хромосомы у белорусов / В.Н. Харьков [и др.] // Генетика. – 2005. – т.41, №8. – С. 1132–1136.
178. Харьков, В.Н. [и др.]. Различия структуры генофондов северных и южных алтайцев по гаплогруппам Y-хромосомы/ В.Н. Харьков [и др.] // Генетика человека. – 2007. – Т.43, №5. – С. 1–13.
179. Харьков, В.Н. Структура и филогеография генофонда коренного населения Сибири по маркерам Y-хромосомы: автореф. дис. д.б.н.: 03.02.07 / Харьков Владимир. – Томск, 2012. – 43 с.
180. Хелимский, Е.А. Северо-западная группа финно-угорских языков и её субстратное наследие. Вопросы ономастики/ Е.А. Хелимский. - №3. - Екатеринбург, 2006. - С. 38-51.

181. Чухряева, М.И. [и др.]. Программа HAPLOMATCH для сравнения STR-гаплотипов Y-хромосомы и её применение к вопросу происхождения донских казаков // Генетика. - 2016. Т.52., №5. - С.595-604.
182. Чухряева, М.И. [и др.]. Сохранились ли следы финно-угорского влияния в генофонде русского населения Ярославской области? Свидетельства Y-хромосомы // Генетика. - 2017. № 3. - С 388-399.
183. Языки народов СССР: т. 1, Индоевропейские языки. – М.: Изд-во АН СССР, 1966. – 433 с.