

ՀՀ ԿՐԹՈՒԹՅԱՆ ԵՎ ԳԻՏՈՒԹՅԱՆ ՆԱԽԱՐԱՐՈՒԹՅՈՒՆ

ԵՐԵՎԱՆԻ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ

ՀՈՎՀԱՆՆԻՍՅԱՆ ՀՐԱՆՏ ՀԵՆՐԻԻ

ՀԱՅԵՐԻ ՄՈԼԵԿՈՒԼԱՅԻՆ ՖԻԼՈՍՇԽԱՐՀԱԳՐՈՒԹՅՈՒՆԸ  
ԸՍՏ ԱՄԲՈՂՋԱԿԱՆ ՄԻՏՈՔՈՆԴՐԻՈՒՄՍԱՅԻՆ ԳԵՆՈՄՆԵՐԻ

Գ.00.02 – Կենսաֆիզիկա և կենսաինֆորմատիկա մասնագիտությամբ  
կենսաբանական գիտությունների թեկնածուի  
գիտական աստիճանի հայցման ատենախոսության

Ս Ե Ղ Մ Ա Գ Ի Դ

ԵՐԵՎԱՆ 2017

---

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РА  
ЕРЕВАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ОГАНЕСЯН ГРАНТ ГЕНРИЕВИЧ

МОЛЕКУЛЯРНАЯ ФИЛОГЕОГРАФИЯ АРМЯН  
НА ОСНОВЕ ПОЛНЫХ МИТОХОНДРИАЛЬНЫХ ГЕНОМОВ

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени  
кандидата биологических наук по специальности  
03.00.02 – Биофизика и биоинформатика

ЕРЕВАН 2017

Ատենախոսության թեման հաստատվել է Երևանի պետական համալսարանում:

Գիտական ղեկավար՝ կենսաբանական գիտությունների դոկտոր,  
պրոֆեսոր Լ.Մ.Եպիսկոպոսյան

Պաշտոնական ընդդիմախոսներ՝ կենսաբանական գիտությունների դոկտոր  
Ն.Մ. Այվազյան  
կենսաբանական գիտությունների թեկնածու,  
դոցենտ Ս.Գ. Տիրացույան

Առաջատար կազմակերպություն՝ ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման  
պրոբլեմների ինստիտուտ

Ատենախոսության պաշտպանությունը կայանալու է 2017թ. մարտի 17-ին, ժամը 14<sup>00</sup>-ին, Երևանի պետական համալսարանում գործող ՀՀ ԲՈՂ-ի Կենսաֆիզիկայի 051 մասնագիտական խորհրդի նիստում (0025, Երևան, Ալեք Մանուկյան փ. 1, ԵՊՀ, կենսաբանության ֆակուլտետ):

Ատենախոսությանը կարելի է ծանոթանալ Երևանի պետական համալսարանի գրադարանում:

Ատենախոսության սեղմագիրն առաքված է 2017թ. փետրվարի 14-ին:

051 մասնագիտական խորհրդի գիտական քարտուղար,  
կենս. գիտ. թեկնածու, դոցենտ՝  Մ.Ա. Փարսադանյան

---

Тема диссертации утверждена в Ереванском государственном университете.

Научный руководитель: доктор биологических наук,  
профессор Л.М. Епископосян

Официальные оппоненты: доктор биологических наук Н.М. Айвазян  
кандидат биологических наук,  
доцент С.Г. Тирацуюн

Ведущая организация: Институт проблем информатики и автоматизации  
НАН РА

Защита диссертации состоится 17-го марта 2017 г., в 14<sup>00</sup> часов, на заседании специализированного совета 051 по Биофизике ВАК РА при Ереванском государственном университете (0025, Ереван, ул. Алека Манукяна 1, ЕГУ, биологический факультет).

С диссертацией можно ознакомиться в библиотеке Ереванского государственного университета.

Автореферат диссертации разослан 14-го февраля 2017 г.

Ученый секретарь специализированного совета 051  
кандидат биологических наук, доцент  М.А. Парсаданян

## ВВЕДЕНИЕ

### Актуальность темы

Будучи естественным коридором, связывающим Европу и Ближний Восток, Армянское нагорье играло ключевую роль в неолитических миграциях, которые впоследствии привели к колонизации Европы поселенцами из Ближнего Востока. Несмотря на то, что территория Армянского плато в настоящее время активно исследуется археологами, до сих пор нет единого мнения о последовательности демографических процессов в этом регионе. Данная картина может быть в значительной степени прояснена при изучении генетической структуры популяций, автохтонных для Армянского нагорья, к которым относятся и армяне. К настоящему времени большинство исследований генетической истории армян основано на патрилинейно наследуемых маркерах Y хромосомы, в то время как материнская генетическая система, которая характеризуется митохондриальной ДНК (мтДНК), изучена недостаточно. Обобщая результаты многочисленных генетических исследований армян по маркерам мтДНК, необходимо выделить ряд все еще существующих проблем и недостатков в данной области. Во-первых, в большинстве работ армяне рассматриваются как единая группа, т.е. без учета географического фактора расселения популяции. Лишь в ограниченном числе публикаций [Weale et al., 2001; Harutyunyan et al., 2009; Herrera et al., 2012] сбор образцов ДНК проведен на основе территориального принципа, и результаты этих исследований свидетельствуют о том, что исследование армян как единой группы может привести к ошибочным выводам [Weale et al., 2001]. Другим недостатком является малый объем выборок, в большинстве случаев не превышающий 50 образцов. Кроме того, указанные исследования проводились с использованием неодинаковых маркеров мтДНК, что делает практически невозможным мета-анализ агрегированных данных для преодоления проблемы малых выборок. И наконец, сбор образцов в большинстве случаев проводился в Ереване, столице Армении, население которого преимущественно представлено потомками армян из различных областей исторической Армении.

За последнее десятилетие было проведено немало популяционно-генетических исследований армян, однако до сих пор все еще существует целый ряд неразрешенных вопросов, касающихся ключевых аспектов этногенеза, направлений доисторических миграций и генетических контактов армян с другими этническими группами. Только с помощью крупномасштабных, хорошо структурированных исследований с использованием географически различных групп армян и сравнительных выборок аналогичного свойства возможно получить ответы на вышеуказанные вопросы.

В рамках диссертационного исследования было собрана, просеквенирована и проанализирована представительная выборка образцов (n~1900) митохондриальной ДНК, включающая как полные митогеномы, так и HVR1 (Hyper Variable Regions) последовательности. Эти образцы представляют различные географические регионы, охватывающие практически всю территорию Армянского нагорья. Более того, помимо использования современных образцов мтДНК, мы впервые также проанализировали большой набор образцов (n=52) древней мтДНК (aDNA), собранных в результате археологических раскопок захоронений на территории Республики Армения и Арцаха и охватывающих временные рамки от 6-го тысячелетия до н.э. до средневековья.

Учитывая, что крупномасштабные митогеномные исследования требуют сложного биоинформатического анализа большого массива генетических данных, нами были разработаны соответствующие инструменты, которые позволяют быстро и

эффективно проанализировать митохондриальную информацию, а также облегчают надежное хранение и подробный анализ подобного рода информации.

### Цель и задачи исследования

Основная цель диссертационного исследования заключалась в восстановлении матрилинейной генетической истории армян на основе анализа частичных и полных древних и современных митохондриальных геномов в географически различных армянских популяциях.

Для достижения поставленной цели нами были сформулированы следующие задачи:

1. Описать и проанализировать филогенетическую изменчивость полных и частичных современных и древних митохондриальных геномов в девяти армянских региональных группах, используя различные биоинформатические методы.
2. Оценить уровень генетического разнообразия во всех рассмотренных образцах на основе анализа различных параметров последовательности HVRI, кодирующих и некодирующих участков полных митохондриальных геномов.
3. Определить уровень генетического родства между различными армянскими группами, армянами и соседними популяциями, а также идентифицировать локализацию армян на генетическом ландшафте юго-западной Азии.
4. Восстановить динамику демографических процессов в восточных регионах Армянского нагорья начиная с эпохи неолита, используя подходы симуляционного моделирования сгенерированных генетических данных.
5. Построить наиболее вероятную демографическую модель этногенеза армян на основе анализа данных об изменчивости современных и древних митохондриальных геномов методами симуляционного коалесцентного моделирования и приближенных Байесовых вычислений (ABC).
6. Создать новую базу данных о полных митохондриальных геномах человека, а также разработать биоинформатические инструменты для повышения эффективности крупномасштабных митохондриальных исследований человека.

### Научная новизна и научно-практическая значимость работы

В нашем исследовании впервые армянский митохондриальный генофонд был исследован с использованием высококачественных полных и частичных митохондриальных геномов из различных региональных групп армян. Более того, мы также впервые проанализировали древние митохондриальные геномы из Республики Армения и Арцах и, с использованием методов биоинформатики, сравнили современные и древние образцы между собой, а также с популяциями из различных регионов мира. Основываясь на полученных результатах, было показано, что вариабельность гаплогрупп во всех армянских популяциях очень высока при сохранении в целом сходной матрилинейной генетической структуры для всех территориальных групп. Кроме того, было установлено, что армянский матрилинейный генофонд не содержит следов центрально- или восточноазиатского происхождения, несмотря на многочисленные нашествия тюрко- и монголоязычных завоевателей на территорию Армянского нагорья. С помощью методов многомерного анализа продемонстрирована тесная генетическая близость армянских групп, а также показано, что армянский кластер занимает промежуточное положение между народами Леванта, Кавказа и Европы, что может свидетельствовать об автохтонном происхождении армян на территории Армянского нагорья.

Кроме того, в данной работе было выявлено, что по структуре исследуемых маркеров современные армянские популяции из различных географических регионов статистически не отличаются между собой. Данный результат свидетельствует о генетической однородности армянского матрилинейного генофонда, что может быть объяснено культурной практикой патрилокальности.

Мы реконструировали наиболее вероятную демографическую историю всех армянских групп (по отдельности и в совокупности) с использованием симуляционного моделирования эффективного размера популяции  $N_e$ . Полученные результаты позволили отследить флуктуацию численности населения на протяжении по меньшей мере 40 тысяч лет и установить в целом сходную популяционную историю всех армянских групп. Общая картина изменения численности населения свидетельствует о том, что Армянское нагорье могло служить рефугиальной зоной для анатомически современного человека во время последнего ледникового периода. Поскольку все результаты сравнения исследованных наборов данных указывали на то, что современные армяне генетически очень близки к древнему населению региона, перед нами была поставлена задача определить, существует ли генетическая преемственность между вышеупомянутыми наборами данных. Для ответа на данный вопрос мы использовали недавно разработанные, но уже широко применяемые методы симуляционного моделирования и приближенных Байесовских вычислений (ABC). Полученные результаты показали, что древние образцы являются прямыми генетическими предками современных популяций и что армянская матрилинейная генетическая история имеет возраст по крайней мере около 3,5-4 тысяч лет.

В целом, в настоящем исследовании были получены ответы на ряд фундаментальных вопросов этногенеза армян, которые были слабо освещены в предшествующих исследованиях по генетической истории армянской популяции. Эти результаты могут быть широко использованы представителями различных дисциплин, занимающихся проблемами происхождения армян и других народов юго-западной Азии.

Практическая ценность полученных в данной работе результатов связана также с разработкой новой базы данных для полных митохондриальных геномов человека, которая была снабжена биоинформатическими инструментами для быстрого и эффективного анализа крупномасштабных митохондриальных данных.

### Апробация работы

Материалы диссертации были доложены на международных и республиканских конференциях: Девятая и десятая годовые научные конференции РАУ (Ереван, 2014, 2015), Международная конференция «Computer Science and Information Technologies» (г. Ереван, 2015), Международная конференция «Применение современных научных методов и технологий в области экспертиз» (Цахкадзор, 2015), Международная конференция «DNA polymorphism in human populations» (Париж, 2016).

### Публикации

Основные результаты диссертации отражены в восьми научных публикациях.

### Объем и структура работы

Работа изложена на 113 страницах, содержит 12 таблиц и 20 рисунков. Библиография включает 167 наименований литературных источников. Рукопись состоит из введения, обзора литературы, описания методов исследования, изложения результатов и их обсуждения, заключения, выводов и списка литературы.

## ОБЗОР ЛИТЕРАТУРЫ

В разделе «Обзор литературы» обсуждаются основные генетические исследования армян, проведённые до сих пор, и проанализированы их недостатки. В данной главе также дано описание современных биоинформатических методов и подходов, которые успешно применяются в исследовании вопросов этногенеза, исторической демографии и популяционной генетики человека. В конце раздела представлен подробный анализ различных баз данных для полных митохондриальных геномов человека.

## МАТЕРИАЛЫ И МЕТОДЫ

### Анализируемые образцы

В нашей работе мы использовали 284 неопубликованных современных армянских митогенома, представляющих 4 различных географических региона исторической Армении – Араратская долина, Арцах, Центральная Армения и Западная Армения, 52 образца древней мтДНК, 8 из которых были ранее опубликованы, 208 опубликованных современных армянских митохондриальных геномов и 1307 неопубликованных последовательностей HVRI, 734 из которых представляют 5 других армянских региональных групп и 572 описывают общую армянскую популяцию.

Таким образом, в общей сложности, в данном исследовании мы описали и проанализировали 1851 образец мтДНК этнических армян, что делает нашу популяцию одной из наиболее изученных этнических групп по маркерам митохондриальной ДНК. Для сравнения армянского митохондриального генетического профиля на основе частот гаплогрупп мтДНК с таковыми у других популяций мы собрали большое количество сравнительных данных, представляющих этнические группы со всего мира. Эта база данных включает в себя около 23 тысяч образцов, охватывающих 113 популяций. Для выполнения различных видов сравнительного анализа между армянами и вышеупомянутыми группами на основе частот гаплогрупп мы унифицировали разрешение гаплогрупп всех образцов до 40 основных мтДНК родословных – A, B, C, D, F, G, H, HV, HV0 & V, HV1, HV2, I, J, K, L, M, M1, N, N1a, N1b, N1c, N1D, R, R0a, T1, T\* (xT1), U, U1, U2, U3, U4, U5, U6, U7, U8, U9, W, X, Y и Z.

### Биоинформатический анализ

Гаплогруппы целых митогеномов были определены с помощью Haplofind [Vianello et al., 2013]. Для оценки гаплогрупп на основании результатов генотипирования HVRI региона были использованы программы Haplosearch [Fregel et al., 2011], Mitotool [Fan et al., 2013], mitohap (<https://dna.jameslick.com/mthap/>), haplogrep [Kloss-Brandstätter et al., 2011] и EMMA [Rock et al., 2013]. Образцы с неоднозначно определенными гаплогруппами были исключены из дальнейшего анализа. Частотный анализ распределения гаплогрупп в изученных популяциях был сделан с помощью MS Excel. На основании частот гаплогрупп были также рассчитаны матрицы различных генетических расстояний. Для оценки статистической значимости генетических расстояний был привлечен точный тест Фишера [Raymond et al., 1995] с использованием пермутационного подхода (n=1000). Значения  $p < 0,05$  считались статистически достоверными. Тепловые матрицы были построены с помощью программы GENE [<http://www.broadinstitute.org/cancer/software/GENE>]. Анализ главных компонент (PCA) был проведён на основе пакета программ Genstat [Trust, 1995]. Анализ соответствия (correspondence analysis – CorA) проводился с использованием SPSS v. 19 [Brosius, 2011]. Для визуализации полученных данных использован написанный нами скрипт на языке R

на основе библиотеки ggplot2 [Wickham, 2009], свободно доступный на нашей странице GitHub [https://github.com/GrantHov/My\\_R\\_codes](https://github.com/GrantHov/My_R_codes).

Множественное выравнивание последовательностей мтДНК было проведено с помощью MAFFT [Kato et al., 2013]. Для оценки динамики демографических изменений, происшедших в ходе этногенеза армянской популяции, мы применили метод Bayesian Skyline Plots (BSP) [Drummond et al., 2005], реализованный в пакете BEAST [Drummond et al., 2007]. Для выполнения такого анализа мы разбили митогеномные выравнивания на несколько областей – в зависимости от частот мутаций. Такое разбиение было проведено при помощи скрипта, написанного на языке программирования Python (доступен через нашу страницу в GitHub [https://github.com/GrantHov/My\\_Python\\_codes](https://github.com/GrantHov/My_Python_codes)). Для генерации входного файла для BEAST V. 1.8.1 мы использовали программу Beauti [Drummond et al., 2012]. Симуляции в 10000000 шагов проводились с отбором параметров на каждом 10000-м шагу. Расчёты проводились с использованием открытого сервера для филогенетических задач CIPRES [Miller et al., 2010]. Для вычисления основных параметров генетического разнообразия последовательностей мтДНК мы использовали программу DnaSP v. 5 [Librado et al., 2009]. Географические карты, иллюстрирующие пространственное распределение частот гаплогрупп и генетического разнообразия, были построены при помощи Surfer v. 11 (Golden Software).

#### Коалесцентное симуляционное моделирование

Для симуляции митогеномных данных мы вначале построили альтернативные модели, описывающие различные версии происхождения армянской популяции. Моделирование данных ДНК проводилось при помощи Fastsimcoal v. 2.5 (fsc) [Excoffier et al., 2013]). Для всех моделей симуляции проводились в 10000000 шагов. Для расчёта и сравнения суммарных статистик полученных симуляций использовался пакет Arlsumstat программы Arlequin. В дальнейшем для проведения ABC и оценки наиболее правдоподобной модели мы использовали пакет abc языка программирования R [Csillery et al., 2012].

#### Создание базы данных

Для облегчения организации, управления и дальнейшего анализа большого массива информации о митогеномах человека мы разработали mtMART – новую базу данных о полных митохондриальных геномах человека. Для создания самой базы данных использованы язык программирования и платформа MySQL. Функциональные характеристики mtMart реализованы на языках PHP и JavaScript. Информация о полных митогеномах человека извлекается из базы данных NCBI Genbank, используя интерфейс программирования приложений (API) NCBI.

## **РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ**

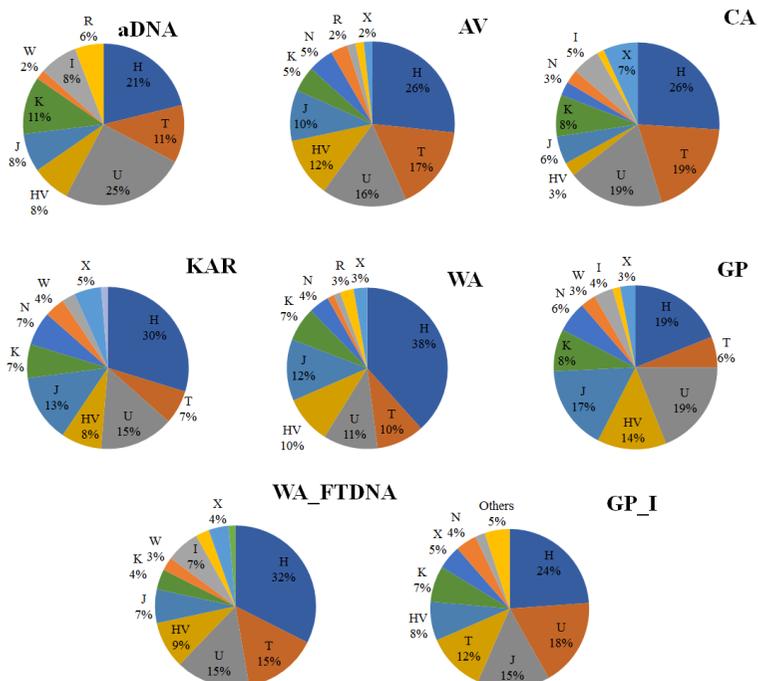
#### Распределение гаплогрупп в армянских популяциях.

Рисунок 1 демонстрирует, что общее разнообразие матрицейной генофонда армян очень высокое и представлено несколькими крупными родословными, которые, в свою очередь, подразделяются на многочисленные субгаплогруппы. Например, гаплогруппа H, которая по своему географическому распределению считается европейской, представлена более чем 20-ю ответвлениями. Аналогичное разнообразие наблюдается и для других крупных гаплогрупп, таких как HV и U.

Рисунок 1 показывает также, что армянский матрицейной генофонд включает в себя несколько крупных гаплогрупп – H, U, J, T, HV и др. – в различных количественных соотношениях, с модальной гаплогруппой H во всех популяциях. Согласно особенностям распределения этих гаплогрупп в различных этно-

территориальных группах, они считаются западно-европейскими, т.к. в больших пропорциях встречаются в популяциях данного региона.

Однако вопрос о географическом происхождении этих родословных все еще остается открытым. Анализ частот гаплогрупп выявил, что в армянском матрилинейном генофонде практически отсутствуют центрально- и восточноазиатские гаплогруппы, такие как A, B, C, D и G, которые являются наиболее распространёнными в большинстве азиатских популяций и достигают 70-80% в генофондах этих этнических групп.



**Рисунок 1.** Распределение гаплогрупп в исследованных армянских образцах.

Подобная структура матрилинейного генофонда свидетельствует о том, что несмотря на многочисленные нашествия центральноазиатских завоевателей (тюрков и монголов) на Армянское нагорье и Ближний Восток, в армянской популяции практически не регистрируются сигналы о генетических контактах с указанными популяциями. Более того, аналогичное отсутствие этих же гаплогрупп в образцах древней ДНК указывает на то, что армянский генофонд не был подвержен воздействию центральноазиатского субстрата по крайней мере в течение последних 3.5-4 тысяч лет.

Для вычисления параметров генетического разнообразия мы провели множественное выравнивание мтДНК и вычленили участок HVRI (16024-16400), а также перевели гаплотипы HVRI в последовательности ДНК. Параметры генетического разнообразия были вычислены с помощью программы DNASP5 и представлены в Таблице 1.

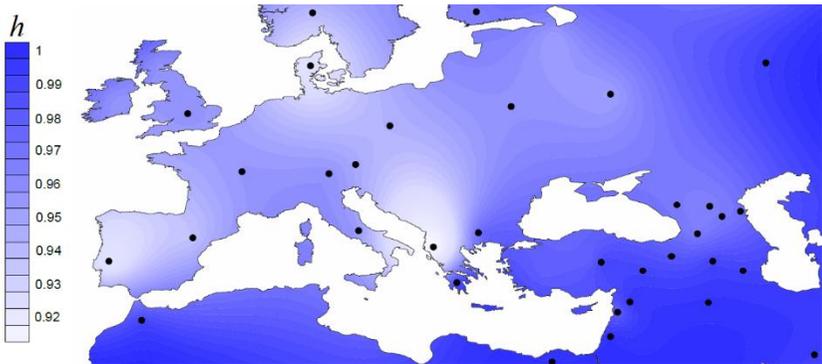
Отрицательные значения параметра Tajima's D, выявленные во всех армянских группах, свидетельствуют о недавней популяционной экспансии или о негативном отборе некоторых полиморфизмов в армянской популяции.

**Таблица 1.** Параметры генетического разнообразия армянских популяций. Аббревиатуры: N – размер выборки, S – количество вариабельных локусов, Eta – количество полиморфизмов, h – разнообразие гаплотипов, Pi – количество мутаций в локусе, SigD – статистическая значимость параметра Tajima D, \*p < 0.05, \*\*p < 0.01.

Популяция	N	S	Eta	hap	h	Pi	AvNu mDif	Tajima D	SigD
<b>SYN</b>	140	81	86	77	0.973	0.015	5.634	-2.025	*
<b>AV</b>	61	65	66	40	0.975	0.016	5.915	-1.982	*
<b>WA</b>	65	69	72	51	0.982	0.014	5.378	-2.196	**
<b>KAR_I</b>	226	102	109	128	0.984	0.016	5.846	-2.088	*
<b>WA_FTDN A</b>	74	62	62	47	0.985	0.015	5.581	-1.867	*
<b>Iran_Salma st</b>	200	107	112	131	0.986	0.015	5.651	-2.189	**
<b>GP_I</b>	396	125	138	241	0.989	0.015	5.527	-2.201	**
<b>WA</b>	73	65	66	55	0.99	0.015	5.543	-1.978	*
<b>CA</b>	73	73	75	57	0.99	0.018	6.610	-1.922	*
<b>A_UNK</b>	176	99	102	117	0.991	0.015	5.720	-2.122	*
<b>Iran</b>	46	54	55	38	0.991	0.016	5.856	-1.867	*
<b>CA</b>	58	69	71	49	0.993	0.017	6.299	-2.031	*
<b>KAR</b>	74	76	79	60	0.993	0.016	6.023	-2.115	*
<b>GP</b>	134	99	104	107	0.996	0.017	6.259	-2.151	**

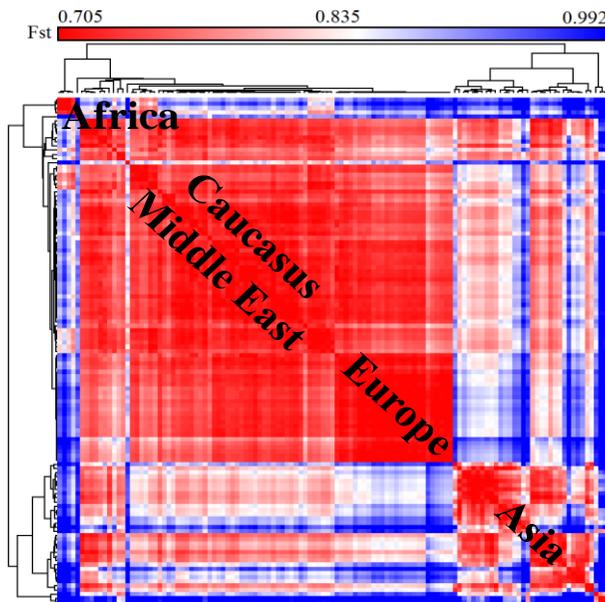
Для визуализации показателя разнообразия гаплотипов, *h*, было проведено географическое картирование распределения величины данного параметра в популяциях Европы, Кавказа, Ближнего Востока и Северной Африки (рис. 2). Построенная карта показывает, что народы Северной Африки и Ближнего Востока характеризуются относительно высоким уровнем генетического разнообразия матрилинейного генофонда, что отражает более ранний возраст этих популяций.

Для оценки взаимоотношений между армянскими группами и другими популяциями было собрано большое количество сравнительных данных, охватывающих более ста популяций. Мы определили положение армянских групп среди остальных популяций, используя ряд статистико-генетических подходов. Во-первых, на основе частот 40 гаплогрупп были вычислены генетические расстояния *Fst* между популяциями, которые затем были использованы для построения тепловой матрицы с ее последующей иерархической кластеризацией (рис. 3).



**Рисунок 2.** Пространственное распределение параметра  $h$  в исследованных популяциях.

Полученная карта отчётливо демонстрирует, что популяции группируются друг с другом в соответствии с их географической близостью, что в очередной раз свидетельствует о положительной корреляции между географическим и генетическими расстояниями. В соответствии с данным принципом, армянские группы кластеризуются с популяциями Ближнего Востока (Леванта) и Кавказа.

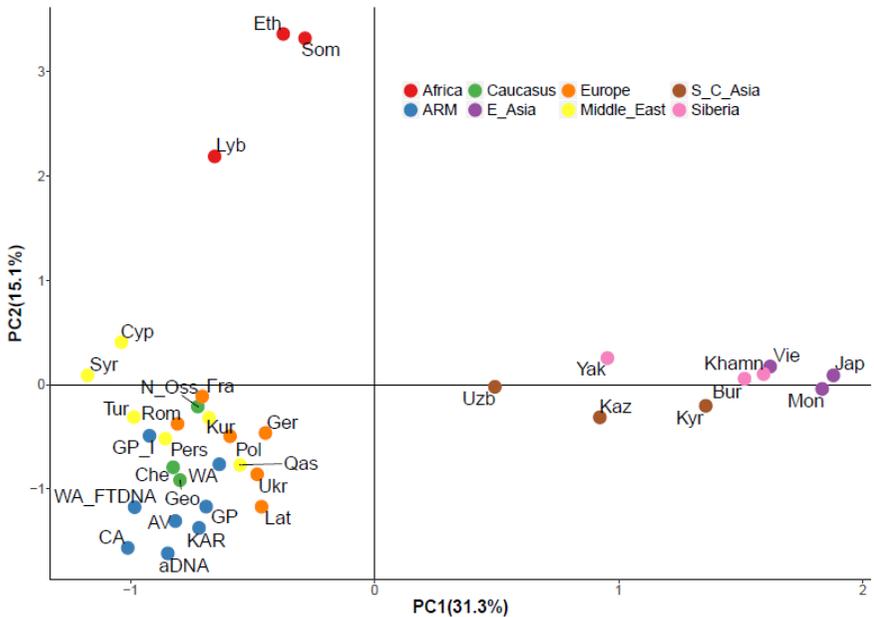


**Рисунок 3.** Тепловая матрица исследованных популяций на основе генетических дистанций  $F_{st}$ . Синие оттенки отражают большие генетические расстояния, красные – меньшие.

Для более детального описания местоположения армян на генетической карте мира мы применили статистически более достоверный подход – анализ соответствий

(Correspondence analysis) на основе частот гаплогрупп (рис. 4), который демонстрирует сходную кластеризацию популяций, выявленную на тепловой матрице расстояний Fst.

Рисунок показывает, что армянские региональные выборки имеют тенденцию группироваться друг с другом, за исключением выборки WA, которая располагается ближе к европейским и кавказским группам. Необходимо отметить, что образцы древней ДНК также локализованы в армянском кластере и находятся ближе к KAR и AV, что указывает на то, что эти популяции генетически близки между собой.



**Рисунок 4.** Результаты анализа соответствий исследованных популяций на основе частот 40 гаплогрупп.

Для дальнейшего выяснения структуры глобальной армянской популяции мы оценили статистическую значимость генетических расстояний Fst между географическими различными группами армян (табл. 2).

Как видно из таблицы, большинство армянских групп статистически незначимо отличаются друг от друга, что свидетельствует о том, что структура матрилинейных генофондов популяций из разных географических регионов Армянского нагорья являются в целом одинаковыми. Важно отметить, что образцы древней ДНК также следуют этой тенденции, подтверждая тем самым исходную гипотезу о близкой генетической связи древних образцов с современными армянскими группами. Единственным исключением являются опубликованные иными авторами результаты исследования армянских популяций. Данное несовпадение может объясняться тем, что эти обобщенные выборки могут содержать образцы ДНК из географических областей, которые не рассматриваются в нашем исследовании.

**Таблица 2.** Показатели теста дифференциации между различными территориальными группами армян на основе генетических расстояний *F<sub>st</sub>*. В нижней треугольной матрице знак «-» означает отсутствие статистической значимости ( $p > 0.05$ ), знак «+» - значимые отличия, верхняя треугольная матрица – фактические значения *r*.

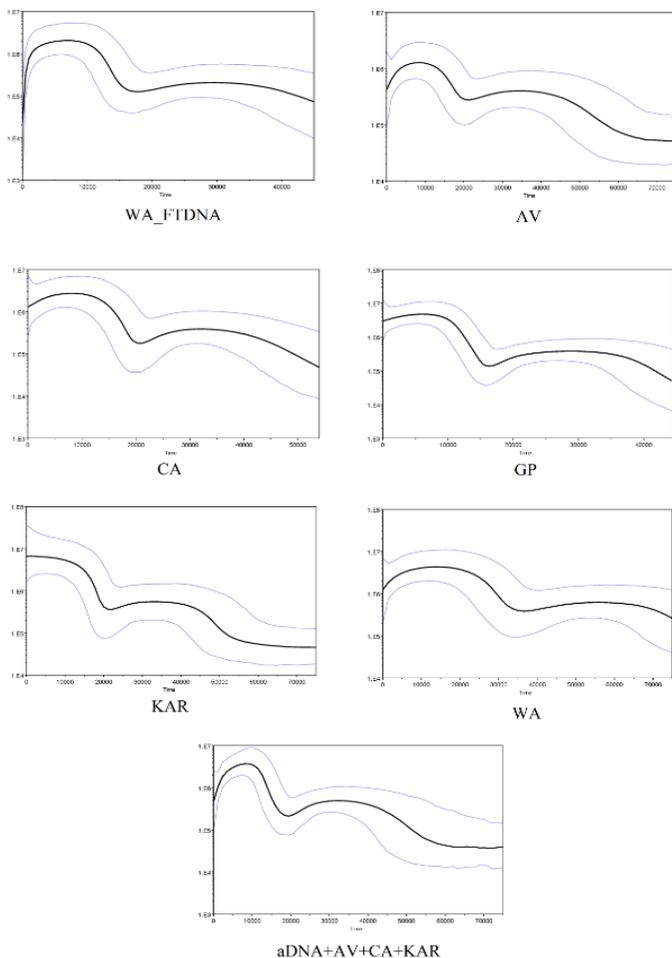
Группы	aDNA	AV	CA	GP	KAR	WA	WA_FTDNA	GP_I
aDNA	*	0.3889	0.4130	0.0401	0.0823	0.0175	0.3768	0.0001
AV	-	*	0.2975	0.3096	0.6716	0.8148	0.2607	0.0113
CA	-	-	*	0.0016	0.2626	0.1351	0.5451	0.0033
GP	+	-	+	*	0.9552	0.1770	0.0040	0.0000
KAR	-	-	-	-	*	0.7618	0.2057	0.0005
WA	+	-	-	-	-	*	0.2122	0.0008
WA_FTDNA	-	-	-	+	-	-	*	0.0000
GP_I	+	+	+	+	+	+	+	*

Этот важный результат отличается от данных предыдущих исследований армян, проведенных на основе анализа маркеров Y хромосомы [Weale et al., 2001, Yeriskorosyan et al., 2001]. Наиболее вероятным объяснением этого расхождения может быть социальная традиция патрилокальности в армянских популяциях. Патрилокальность, способствующая более широкой пространственной дисперсии женской части населения, является широко распространенным культурным обычаем, согласно которому женщина после замужества переходит жить в местообитание мужа. Тем самым внутрипопуляционный уровень миграций женщин увеличивается, что приводит к пространственной однородности распределения частот гаплогрупп митохондриальной ДНК [Seielstad et al., 1998 ; Perez-Lesaun et al., 1999; Jorde et al., 2000; Willson et al., 2001]. Результаты нашего исследования показывают, что традиция патрилокальности оказала сильное влияние на формирование определенной структуры матрилинейного генофонда армян.

Для восстановления демографической истории армянского народа мы использовали метод симуляционного моделирования эффективного размера популяции  $N_e$ , используя BEAST v. 1.8.3. В анализ были включены как современные образцы, так и древние молекулы ДНК, которые являются калибровочными пунктами для симуляции, что увеличивает достоверность и точность полученных результатов. Нами были сгенерированы кривые Bayesian skyline plots, описывающие флуктуации  $N_e$  в течение наблюдаемого времени для всех армянских популяций, представленных целыми митогеномами, а также для комбинированной армянской популяции вместе с древними образцами (рис. 5).

Результаты показывают, что демографические изменения во всех группах очень похожи, выявляя четыре основных демографических события: (1) популяционная экспансия около 40 тыс. лет назад и относительно постоянный размер популяции до 20 тыс. лет назад, когда наблюдается небольшой спад численности; (2) после этого

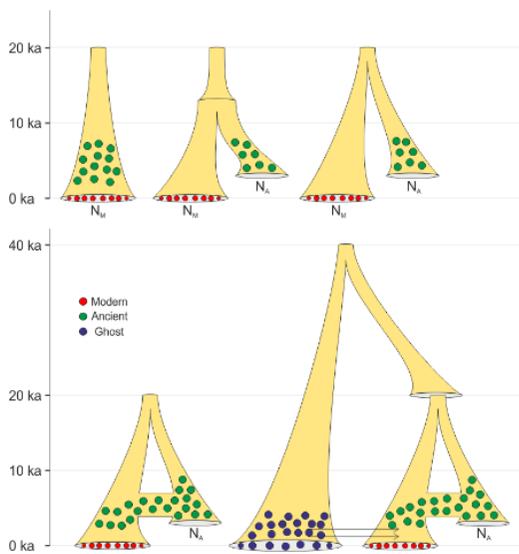
обнаруживается резковыраженная популяционная экспансия около 15-18 тысяч лет назад, что совпадает с последним ледниковым максимумом и указывает на то, что в данный период Армянское нагорье являлось рефугиальной зоной; (3) затем для всех групп, за исключением KAR, около 7-10 тысяч лет назад наблюдается небольшой популяционный спад и (4) достигается современный эффективный размер популяции. Необходимо отметить, что аналогичная демографическая история была получена и для агрегированной популяции, что может свидетельствовать о том, что исследованные популяции представляют собой единую генетически изолированную этническую группу.



**Рисунок 5.** Bayesian skyline plots, отражающие изменение в  $N_e$  в течение времени. Черная линия соответствует среднему значению  $N_e$ , синие линии охватывают 95%-й доверительный интервал.

Используя различные методы популяционной генетики и биоинформатики, в нашей работе мы несколько раз продемонстрировали, что в большинстве случаев образцы древней ДНК генетически близки к современным образцам ДНК, представляющих армянские популяции из центральных и восточных областей Армянского нагорья. Исходя из этого, можно предположить, что между древними и современными популяциями региона существует выраженная генетическая преемственность и что корни армянского матрилинейного генофонда могут восходить к эпохе неолита. Однако для проверки данной гипотезы необходимы более достоверные и чувствительные методы анализа. Именно поэтому для выяснения степени генетической связи между популяциями различных исторических эпох мы применили метод симуляционного моделирования сгенерированных генетических данных при различных демографических сценариях и с дальнейшим использованием аппроксимированных байесовых вычислений для выделения наиболее правдоподобной модели.

Нами были сконструированы пять альтернативных моделей (рис. 6) этногенеза армянской популяции и для каждой модели просимулированы результаты анализа ДНК при помощи *fastsimcoal2*. Для полученных данных на основе *arlsimstat* рассчитаны суммарные статистики, которые были сравнены между различными моделями с использованием пакета *abc* языка программирования R.



**Рисунок 6.** Схематическое представление пяти демографических сценариев использованных для симуляционного моделирования.

Результаты моделирования выявили, что наиболее правдоподобной является первая модель, подразумевающая прямую генетическую преемственность между современными и древними образцами. Постериорные вероятности моделей и факторы Байеса указаны в таблице 3.

**Таблица 3.** Постериорные вероятности моделей (в скобках) и попарные факторы Байеса.

<b>Модель</b>	<b>model-1 (0.6471)</b>	<b>model-2 (0.0983)</b>	<b>model-3 (0.042)</b>	<b>model-4 (0.1596)</b>	<b>model-5 (0.053)</b>
<b>model-1</b>	1	6.5804	15.4242	4.0539	12.2104
<b>model-2</b>	0.152	1	2.344	0.6161	1.8556
<b>model-3</b>	0.0648	0.4266	1	0.2628	0.7916
<b>model-4</b>	0.2467	1.6232	3.8047	1	3.012
<b>model-5</b>	0.0819	0.5389	1.2632	0.332	1

### **База данных**

В нашей работе мы разработали mtMart (по аналогии с поисковой системой ENSEMBL BioMart [Yates et al., 2016]) – новую базу данных для полных митогеномов человека, которая значительно облегчает управление и дальнейший анализ крупных массивов митохондриальной ДНК человека.

База данных разработана в системе MySQL. Его функциональные характеристики реализованы в PHP и JavaScript (jQuery). mtMart извлекает информацию о митохондриальных геномах человека (однако, в принципе, она может быть использована для получения любого запроса) непосредственно из Национального центра биотехнологической информации (NCBI, <http://www.ncbi.nlm.nih.gov/>), который предоставляет API (интерфейс программирования приложений), используя произвольный набор или диапазон входных номеров (accession numbers). Затем разработанные нами скрипты обрабатывают INSDSeq XML-файл из GenBank и переводят все необходимые данные во внутреннюю память mtMART.

mtMart полуавтоматическим образом работает с программой Haplofind для определения гаплогрупп митогеномов и мутаций, ассоциированных с различными болезнями. Информация о популяции и географическом регионе для каждого образца добавляется вручную, что обеспечивает высокое качество хранимых данных. Для того, чтобы избежать возможных нежелательных манипуляций, функции добавления и удаления данных ограничены для всех пользователей, за исключением администратора. Другой важной особенностью mtMart является возможность поиска, сортировки и загрузки данных. Можно сортировать данные по гаплогруппам (в любом разрешении филогенетического дерева), популяциям и географическому региону или объединить все эти варианты, чтобы получить нужный результат. После того как данные получены, пользователь может скачать их в различных форматах (fasta, tsv и др.). Кроме того, mtMart может также генерировать входные файлы для широкоиспользуемой популяционно-генетической программы Arlequin. Наша база данных будет постоянно развиваться, позволяя исследователям использовать новые функционально удобные инструменты для быстрого и эффективного управления данными. В настоящее время развитие mtMart продолжается; она находится в открытом доступе по ссылке <http://mtmart.org/>.

## ЗАКЛЮЧЕНИЕ

В данной работе нами впервые была изучена матрилинейная генетическая структура географически различных армянских популяций с использованием большого количества высококачественных как современных, так и древних митохондриальных данных, совокупное количество которых составило около 1900 образцов.

Для выяснения генетической и демографической истории армян мы применили как уже существующие биоинформатические методы, так и разработали новые скрипты для эффективной обработки данных. Кроме того, в рамках нашего проекта мы разработали новую базу митохондриальных данных человека, которая позволяет повысить эффективность крупномасштабных исследований в этой области генетики и биоинформатики.

На основе проанализированных данных мы показали, что армянские группы, представляющие различные географические регионы Армянского нагорья, имеют очень сложный состав гаплогрупп. Исследованные генофонды преимущественно представлены несколькими крупными родословными, такими как H, U, J, T и др., которые считаются европейскими гаплогруппами по причине их высоких частот в Центральной и Западной Европе. Гаплогруппа H является наиболее распространенной практически во всех наборах данных. Анализ частот гаплогрупп показал также, что армянские популяции не содержат центральноазиатских гаплогрупп A, B, C, D и G, указывая тем самым, что несмотря на многочисленные вторжения центральноазиатских племен и этнических групп на Армянском нагорье, которые происходили с начала прошлого тысячелетия, армянский генофонд не был подвергнут их генетическому воздействию.

При сравнении географически различных групп армян нами не выявлено заметной генетической дифференциации между ними. Наиболее вероятным объяснением этого факта может служить культурная традиция патрилокальности – этнокультурного обычая, согласно которому после замужества женщина переезжает на постоянное жительство к мужу. На популяционном уровне, в результате многовекового следования данной практике, происходит гомогенизация матрилинейного генофонда на просторах Армянского нагорья – географического региона исторического обитания армян на протяжении тысячелетий.

Для определения местоположения армян на генетическом ландшафте Среднего Востока было использовано большое количество сравнительных выборок при проведении различных методов анализа для установления степени генетической связи между армянами и другими популяциями региона. Результаты многомерного статистического анализа показали, что армянские группы имеют тесную генетическую близость друг с другом и армянский кластер занимает промежуточное положение между популяциями Леванта, Кавказа и Европы, что свидетельствует в пользу автохтонного происхождения армян на Армянском нагорье.

С помощью Bayesian skyline графиков мы восстановили демографическую историю армянских групп, которые были представлены полными митохондриальными геномами. Полученные результаты показали, что закономерности демографических изменений во всех армянских группах очень похожи в отношении четырех основных демографических событий: рост численности населения около 40 тыс. лет назад; постоянный размер популяции до 20 тыс. лет назад; интенсивная экспансия населения около 15 тыс. лет назад; медленное снижение численности до настоящего времени. Полученные результаты свидетельствуют о том, что Армянское нагорье могло служить рефугиальной зоной во время последнего ледникового периода и что последующий рост населения

около мог быть результатом миграции в этот регион выходцев из Ближнего Востока вслед за отступлением ледникового щита на север. Примечательно, что при выполнении анализа BSP для объединённой выборки армян мы получили аналогичный результат, указывающий на то, что на территории Армянского нагорья демографическая история популяций в целом одинакова, и, как было показано в результатах при использовании иных подходов, общие закономерности свидетельствуют о том, что рассматриваемые популяции принадлежат к одному репродуктивно изолированному (в течение нескольких тысячелетий) эническому образованию.

Для более подробного исследования генетической и демографической истории армян мы применили методы коалесцентного симуляционного моделирования и приближённых байесовых вычислений. Результаты симуляционного моделирования показали, что наиболее правдоподобной моделью этногенеза армян является вариант, подразумевающий генетическую преемственность между современными армянами и образцами древней ДНК, что свидетельствует о том что корни армянского матрилинейного генофонда восходят к эпохе неолита.

В рамках проведенного диссертационного исследования разработана и создана новая база данных полных митогеномов человека mtMART ([www.mtmart.org](http://www.mtmart.org)), которая в значительной мере облегчает эффективное управление и последующий анализ крупномасштабных массивов информации в популяционно-генетических исследованиях.

## ВЫВОДЫ

1. Матрилинейные генофонды географически различных армянских популяций имеют комплексную и разнообразную структуру. Они включают в себя многочисленные крупные гаплогруппы и их разветвления, с высокой частотой встречающиеся у народов Европы и Ближнего Востока. Кроме того, в армянском матрилинейном генофонде практически отсутствуют гаплогруппы центральноазиатского происхождения, несмотря на многочисленные вторжения на Армянское нагорье монголов и тюркоязычных племен в начале прошлого тысячелетия.
2. Все исследованные армянские географические группы имеют высокий уровень генетического разнообразия, сопоставимый с таковым для популяций Леванта, что свидетельствует в пользу автохтонного происхождения армян на территории Армянского нагорья.
3. Географически различные группы армян не отличаются друг от друга по структуре митохондриального генофонда, что может быть объяснено культурной традицией патрилокальности.
4. На генетической карте региона армяне занимают промежуточное положение между популяциями Ближнего Востока и Кавказа, что также указывает на коренное происхождение популяции в границах Армянской возвышенности.
5. Все исследованные популяции обладают сходной динамикой демографических показателей, указывающих на то, что Армянское нагорье могло служить рефугиальной зоной на протяжении последнего ледникового периода.
6. На основе методов симуляционного моделирования результатов генотипирования современной и древней мтДНК построена наиболее вероятная модель этногенеза армян, свидетельствующая о генетической преемственности населения восточных регионов Армянского нагорья с эпохи неолита.
7. Для эффективного управления и анализа крупномасштабных массивов информации в популяционно-генетических исследованиях разработана и создана новая база данных полных митогеномов человека mtMART ([www.mtmart.org](http://www.mtmart.org)).

## ПУБЛИКАЦИИ

1. Епископосян Л., Маргарян А., Хачатрян З., Худоян А., Оганесян А., Оганесян Г., Погосян А.С. Популяционная генетика в решении вопросов этногенеза армян. Этнос и среда обитания // Сборник статей по этноэкологии, вып. 4. М.: Старый сад, С. 198-207, 2014.
2. Hovhannisyan H. Databases for complete human mitogenomes: a review // Herald of Russian-Armenian University, (2), p. 95-103, 2015.
3. Hovhannisyan H., Yesayan A., Yepiskoposyan L. mtMART – a Novel Database for Complete Human Mitochondrial Genomes // International Conference “Application of Modern Scientific Methods and Technologies in Expertise Sphere», p. 388-391, June 16-17, Tsakhkadzor, Armenia, 2015.
4. Hovhannisyan H., Yesayan A. and Yepiskoposyan L. Designing and implementing a novel database of human mitogenomes // International Conference Computer Science and Information Technologies, p. 277-279, Yerevan, Armenia, 2015.
5. Hovhannisyan H., Yesayan A., Antonosyan M., Yepiskoposyan L. mtMART – A New Human Mitogenomic Database // Herald of Russian-Armenian University, (1), p. 142-146, 2015.
6. Hovhannisyan H., Yepiskoposyan L. Mitochondrial genetic diversity in the Armenians: a review // International Journal of Anthropology, (31), p. 137-144, 2016.
7. Hovhannisyan H, Antonosyan M., Yepiskoposyan L. Human mtDNA DBome – a Review on Databases for Human mitochondrial DNA // International Journal of Advanced Biotechnology and Research, 7(3), p.888-896, 2016.
8. Margaryan A., Derenko M., Hovhannisyan H., Malyarchuk B., Heller R., Yepiskoposyan L., Willerslev E., Allentoft M. 7800 Years of Mitochondrial Genetic Continuity in Armenia // International Conference “DNA Polymorphism in Human Populations”, p.49, Paris, France, 2016.

**ՀԱՅԵՐԻ ՄՈՒԵԿՈՒԼԱՅԻՆ ՖԻԼՈՍՓԻԱԸՀԱԳՐՈՒԹՅՈՒՆԸ  
ԸՍՏ ԱՄԲՈՂՋԱԿԱՆ ՄԻՏՈՔՈՆԴԻՌԻՄՍՅԻՆ ԳԵՆՈՄՆԵՐԻ**

**ԱՄՓՈՓՈՒՄ**

**Հանգուցային բառեր՝** հայեր, միտոքոնդրիումային ԴՆԹ, հնագույն ԴՆԹ, ժողովրդագրական պատմություն, տվյալների բազա

Հայերը հանդիսանում են Հայկական բարձրավանդակի բնիկ էթնիկ խումբ՝ ընդգծված սոցիալական և էթնիկական առանձնահատկություններով: Գտնվելով Եվրոպան և Ասիան միացնող խաչմերուկում՝ Հայկական լեռնաշխարհը հանդիսացել է բազմաթիվ ցեղախմբերի և էթնիկ հանրությունների տեղաշարժերի անցուղի: Հայկական բարձրավանդակի դերը այդ միգրացիաներում և դրանց ազդեցությունը հայերի էթնոգենեզի վրա որոշելու ուղղությամբ վերջերս իրականացվել են մի շարք պոպուլյացիոն գենետիկական հետազոտություններ: Այդուհանդերձ, հայկական պոպուլյացիայի մայրագծային գենետիկական պատկերը դեռևս թերի է ուսումնասիրված:

Այս աշխատանքում ուսումնասիրվել է հայերի մայրագծային գենետիկական կառուցվածքը՝ օգտագործելով պատմական Հայաստանի տարբեր աշխարհագրական տարածքները ներկայացնող խմբերից ստացված միտոքենոմային մեծածավալ տվյալներ: Բացի այդ, աշխատանքում առաջին անգամ նկարագրվել և ուսումնասիրվել են Հայաստանի և Արցախի տարբեր հնավայրերից հավաքված և տարբեր պատմական ժամանակաշրջաններ ներկայացնող հնագույն միտոքենոմային նմուշներ:

Հայերի ժողովրդագրական և գենետիկական պատմությունը լուսաբանելու նպատակով աշխատանքում կիրառվել են արդի կենսաինֆորմատիկական և տվյալների վերլուծության ինչպես առկա մեթոդներ, այնպես էլ մեր կողմից ստեղծված տարբեր կենսաինֆորմատիկական գործիքներ՝ մեծածավալ տվյալների արդյունավետ վերլուծության համար: Բացի այդ, մշակվել և ստեղծվել է մարդու միտոքենոմային տվյալների նոր հանրամատչելի բազա:

Ստացված արդյունքները ցույց են տվել, որ աշխարհագրորեն տարբեր հայկական խմբերն ունեն բարդ և բազմաշերտ գենետիկական կառուցվածք՝ ներկայացված տարբեր հապլոխմբային ընտանիքներով, որոնք բաժանված են մի շարք ենթախմբերի: Միտոքոնդրիումային գենոֆոնդը մեծամասամբ ներկայացված է H, U, J, T և այլ խոշոր տոհմագծերով, որոնցից H-ը մոդալ է բոլոր հայկական խմբերի համար: Հապլոխմբերի հաճախականության վերլուծությունը ցույց է տվել նաև, որ հայկական մայրական գենոֆոնդում գրեթե բացակայում են կենտրոնական և արևելյան ասիական ծագում ունեցող A, B, C, D և G հապլոխմբերը: Այս արդյունքը վկայում է, որ չնայած Հակական լեռնաշխարհում վերջին հազարամյակում տեղի ունեցած բազմաթիվ ներխուժումների՝ հայերի գենոֆոնդը չի ենթարկվել ասիական

ազդեցությանը: Հայերի ներխուժային համեմատության արդյունքում հավաստի տարբերություն չի արձանագրվել, ինչը կարող է բացատրվել հայրատեղային ամուսնությունների ավանդությամբ:

Հայերի և այլ պոպուլյացիաների տվյալների բազմաչափ վերլուծությունը բացահայտել է, որ հայակական խմբերը գենետիկորեն մոտ են միմյանց, իսկ հայկական կլաստերը Միջին Արևելքի գենետիկական բնապատկերում միջանկյալ դիրք է գրավում Լևանտի, Գովկասի և Եվրոպայի պոպուլյացիաների միջև: Մա մատնանշում է Հայկական լեռնաշխարհում հայերի բնիկ ծագման մասին:

Հայերի ժողովրդագրական պատմությունն ուսումնասիրելու համար կառուցվել են Բայեսյան հորիզոնային պատկերներ (BSP), որոնք արտացոլում են *պոպուլյացիայի էֆեկտիվ չափի* փոփոխությունները ժամանակի ընթացքում: Ըստ ստացված արդյունքների՝ հայկական խմբերի ժողովրդագրական փոփոխությունները միանման են և ներառում են չորս հիմնական իրադարձություն՝ պոպուլյացիայի ընդլայնում շուրջ 40 հազար տարի առաջ, կայուն չափի պահպանում մինչև 20 հազար տարի առաջ, պոպուլյացիայի կտրուկ աճ մոտ 15 հազար տարի առաջ և դանդաղ անկում մինչ այժմ: Արդյունքները վկայում են նաև, որ Հայկական բարձրավանդակը վերջին սառցե դարաշրջանի ընթացքում եղել է այսպես կոչված՝ ռեֆուգիալ (ապաստանային) գոտի, ինչը, նեոլիթյան երկրագործության վերելքի հետ մեկտեղ, կարող էր նպաստել պոպուլյացիայի աճին: Հայկական խմբերի միասնական տվյալների BSP վերլուծությունը համանման արդյունքներ է արձանագրել՝ մատնանշելով Հայկական բարձրավանդակի ողջ տարածքի ժողովրդագրական պատմությունների նմանությունը:

Ժամանակակից հայկական և տարածքի հնագույն տվյալների միջև գենետիկական փոխկապակցվածությունը հետազոտվել է կոալեսցենտային սիմուլյացիոն մոդելավորման և բայեսյան հաշվարկների հիման վրա: Ստացված արդյունքները ցույց են տվել, որ հայերի ժողովրդագրական պատմության 5 այլընտրանքային մոդելներից ամենահավանականը այն է, որն արտացոլում է ժամանակակից և հնագույն տվյալների միջև գենետիկական շարունակականություն: Դա նշանակում է, որ հայկական մայրագծային գենոֆունդն առնվազն 3.5-4 հազար տարվա պատմություն ունի:

Ստեղծվել է մարդու ամբողջական միտոգենոմային տվյալների նոր բազա՝ mtMart ([www.mtmart.org](http://www.mtmart.org)), որը, ի տարբերություն համանման այլ բազաների, ավելի կատարելագործված է և զինված զանազան կենսաինֆորմատիկական գործիքներով: Այն օգտագործողին թույլ է տալիս ավելի արդյունավետ և ճշգրիտ հավաքագրել, մշակել և հետազայում վերլուծել մեծածավալ միտոգենոմային տվյալներ:

**MOLECULAR PHYLOGEOGRAPHY OF ARMENIANS  
BASED ON COMPLETE MITOGENOMES**

**SUMMARY**

**Key words:** Armenians, mitochondrial DNA, ancient DNA, demographic history, database

Armenians are the indigenous people of the Armenian Highland with strong and distinct ethnocultural characteristics. Being a crossroads linking Europe and Asia, the Armenian Highland has experienced numerous ancient and recent migrations of different tribes and ethnic groups. To describe the role of the Armenian plateau in these migrations and how the last influenced the ethnogenesis of the Armenian population, numerous population genetics studies have been recently performed on the Armenians. However, the matrilineal genetic legacy of the Armenian population is still poorly investigated.

In our work, we have studied Armenian matrilineal genetic structure by analysing large amount of high quality mitogenomic data from different geographic regions covering the whole area of Historical Armenian. Moreover, here, we have studied the first so far available data on ancient mitogenomes spanning different historical epochs, collected from the different archaeological sites of the Republic of Armenia and Artsakh.

To elucidate the genetic and demographic history of the Armenians, we have applied sophisticated bioinformatics and data analysis methods that were already available, as well as developed our own scripts for efficient data parsing and further analysis. Moreover, in frames of our project, we have also designed and developed a new publicly available database for human mitogenomic studies.

We have demonstrated that the Armenian groups from different geographic regions have a very complex haplogroup composition, represented by multiple large haplogroups families, which in turn are divided into numerous subclades. The majority of the mtDNA gene-pool is represented by several large lineages, as H, U, J, T, etc., with H haplogroup being the modal in each group. The haplogroup frequency analysis has also demonstrated that the Central and East Asian haplogroups A, B, C, D and G are found in negligible amount in Armenian mtDNA gene-pool, demonstrating that despite numerous invasions in the Armenian Highland that took place during the past millennium, the Armenian gene pool was not affected by the Asian influences. Despite the genetic complexity of the Armenian maternal gene pool, our analysis has shown that while comparing different geographic groups with each other, in most of the cases there is no marked genetic differentiation between them, which might be explained by cultural practice of patrilocality in the Armenian society.

In order to define the location of Armenians on the genetic landscape of the Middle East, in our work we have performed a variety of comparative analyses between Armenians and numerous comparative datasets. Based on the results of multivariate analysis, we have shown that Armenian groups display close genetic affinity with each other and the Armenian cluster occupies an intermediate position between the Levant, the Caucasus and European populations, thus indicating an autochthonous origin of the Armenians in the area of the Armenian Plateau.

Further, we have elucidated the demographic history of Armenians via constructing Bayesian skyline plots which reflect the changes of *effective population size* through time. We have partitioned mitogenomes into several regions and assigned different mutation rates for

each of them, thus strengthening the accuracy and likelihood of the obtained results. The results show that the patterns of demographic changes are very similar in all the Armenian groups – there are four major demographic events: population expansion ca 40 kya, constant size until 20 kya, intensive population expansion ca 15 kya, slow decline until current time. Obtained results indicate also that Armenian Highland was a refugial zone during the Last Glacial Maximum, and that population expansion might be also explained by the raise of agriculture. Notable that when performing the BSP analysis for the aggregated population we have obtained a similar result indicating that on the whole the territory of the Armenian Highland the demographic histories of populations were very similar.

To further investigate the genetic and demographic history of the Armenians and to conclusively characterise relationships between ancient and modern Armenian data, we have applied coalescence simulation modelling techniques paired with Approximate Bayesian computations. The results of simulation modelling analysis strikingly demonstrate that among 5 alternative models of Armenian demographic history the most plausible one is the model showing a genetic continuity between modern days Armenians and ancient DNA samples. This results indicates that the ancient samples represent direct biological ancestors of modern Armenian populations, evidently highlighting that Armenian maternal genetic history is at least 3,500-4,000 years old.

In our dissertation we have also developed a new database specifically for human mitochondrial DNA, which we have called mtMart ([www.mtmart.org](http://www.mtmart.org)). It has advantages of previously published and reported similar databases, however lacks their drawbacks and has many other convenient functional characteristics that enable researchers to compile, parse and analyse human large-scale mitogenomic data more efficiently and accurately.

