**Анализ аутосомного генофонда беларусов**

Наряду с ставшими уже традиционными методами анализа структуры генофонда популяций по Y-хромосоме и митохондриальном ДНК (о которых мы упоминали выше), в последнее время все большее распространение получают новые и более точные методы анализа генофонда популяций, в основу которых положены данные о частотах аллелей большого числе генетических локусов на неполовых хромосомах. В силу того, что в передаче генетического материала на неполовых (аутосомальных) хромосомах участвуют как мужчины, так женщины, подробный и детальный анализ распространения аллельных вариант способен дать более точную характеристику процессов исторических и эволюционных процессов в отдельно взятой популяции. Кроме того, большая часть человеческого генома сосредоточена в неполовых хромосомах - количество генов (точнее определенных генетических локусов) локализованных на неполовых хромосомах, несоизмеримо выше числа генов в митоДНК и половых хромосмах. Поэтому анализ аутосомных хромосом обладает более высокой степенью информационного разрешения. Выбор типа локусов в нашем исследовании также обусловлен сугубо прагматическими соображениями.
Поскольку существует разные типы генетических маркеров, мы ограничились снипами - точечными мутациями, представляющими собой замену одиночного нуклеотида. Как было показано в других исследованиях, математико-статистические методы исследования распределения аллельных вариантов (генотипов) отдельных снипов характеризуются наивысшой разрешающой способности в плане изучения воздействия всех известных генетических эффектов на историческую демографию популяции.

Перед тем, как перейти к весьма сжатому и краткому описанию результов анализа популяции беларусов по определенному набору типированных снипов, следует вкратце описать предысторию изучения аутосомных маркеров беларуской популяции. Как уже отмечалось выше, новые методы исследования генофонда популяция с помощью изучения статистических характеристик одиночных нуклеотидных вариантов (SNV/SNP), были введены в популяционной генетике и биоинформатике относительно недавно. В более ранней парадигме, анализ (в том числе, и сравнительный) аутосомной составляющей генофонда состоял в определении набор-панели высокополиморфных и слабо сцепленных между собой аутосомных ДНК маркеров. В то же самое время, насколько нам известно, аналогичный анализ с использованием другого типа маркеров –снипов- до совсем недавнего времени не проводился. Например, в исследованиях аутосомных ДНК-маркеров часто использовался стандартный набор CODIS, который первоначально разрабатывался для нужд ДНК-профилирования в криминалистике. В данную систему CODIS входят 16 аутосомных STR локусов и амелогениновый тест на определение половой принадлежности Однако относительно простота амплификации сателлитной ДНК в формате мультиплексной PCR, и доступность методов статистической обработки с использованием современных статистических подходов, применяемых в популяционной генетике и судебной медицине, привела к тому что эта система часто использовалась в ранних исследованиях генофонда популяций Восточной Европы.

Cамой масштабной работой такого рода ( а именно, c использованием классических аутосомных маркеров STR), является совместная статья Балановского и Тегако. В свете нашего исследования представляется нужным отметить наиболее интересующий нас аспект вышеупомянутой работы, а именно сравнительный анализ генофонда беларусов с генофондами других популяций Западной Евразии. Как отмечают авторы, cравнительный анализ генофондов Западной Евразии показал, что «по [аутосомным ДНК](http://ru.wikipedia.org/w/index.php?title=%D0%90%D1%83%D1%82%D0%BE%D1%81%D0%BE%D0%BC%D0%BD%D0%B0%D1%8F_%D0%94%D0%9D%D0%9A&action=edit&redlink=1) маркерам все восточные славяне входят в один кластер Восточной Европы. Однако **русские и украинские популяции расположены ближе друг к другу, чем к белорусам**… Русские и украинцы оказываются ближе к популяциям Западной и Южной Европы, чем белорусы. К сожалению, по этим маркерам не изучены ближайшие западные и северные соседи белорусов, и в частности, балтские народы. Можно предполагать, что некоторое своеобразие генофонда белорусов связано **с вкладом балтского субстрата**»[[1]](#footnote-1)

Использованные в работе Балановского и Тегако генетические данные по генофонду популяций беларусов были взяты из баз данных Института генетики НАН Беларуси, где хранятся базы данных по 35 аутосомным генам, изучаемых, главным образом, в целях медицинской генетики. Однако, как отмечается в статье О.Давыденко и Е. Кушнеревич «Гаплогруппы Y-хромосомы и происхождение национального генофонда», для эволюционных исследований эти маркеры не совсем удобны, так как на частоту и характер распространения аллелей зачастую оказывает влияние *жесткий естественный отбор*, и поэтому по одним только результатам анализа распределения частот невозможно судить о генетическом родстве разных популяций.[[2]](#footnote-2)

Несмотря на то, что основное внимание в статье Давыденко и Кушнеревич уделяется анализу Y-хромосомных гаплогрупп беларусов, в конце статьи авторы мимоходом упоминают о результатх исследования аутосомного генофонда, проведенного междунродным коллективом под руководством Дорона Бехара[[3]](#footnote-3):

«В работе израильского ученого Дорона Бехара с соавторами изучались различные популяции евреев и анализировалась генетическая взаимосвязь между ними и их «хозяйскими» популяциями. В большинстве случаев эта связь была незначительной. Белорусская «хозяйская» популяция также исследовалась с использованием именно нашего ДНК-банка, то есть эти данные весьма сопоставимы, поскольку сделаны на одной той же представительной выборке. Ученые сравнивали 100 тыс. однонуклеотидных замен (то есть изучали 100 тыс. точек, разбросанных по всему геному человека). Результаты работы демонстрируют достаточно близкий спектр изменений у белорусов, литовцев и русских и свидетельствуют о присутствии **трех основных «древних» компонентов в сопоставимых пропорциях. Это также косвенно указывает на возможность общего происхождения славян и балтов**, а также на то, что основы генофонда различных европейских народов заложены еще в доисторическое время. Если сопоставить данные математической лингвистики с данными по геномному анализу, нельзя не увидеть определенного параллелизма не только в дивергентном характере, но и во временных оценках эволюции – как языка, так и генетики. Эти данные также могут свидетельствовать в пользу предположения общности происхождения славян и балтов из некогда единой популяции.»

В приведенной выше цитате мы выделили самую важную часть, в которой авторы высказывают предположение о том, что **присутствие трех основных древних компонентов указывает на возможность общего происхождения славян и балтов**. Очевидно, что предложенная авторами спорная (на наш взгляд) интерпретация противоречит выводам Балановского и Тегако о том, что своебразие аутосомного генофонда беларусов может быть связано **с вкладом балтского субстрата.**  К сожалению, в обеих статьях авторы не сочли нужным подробно аргументировать свои выводы по столь важному вопросу. Поэтому одной из задач нашего небольшого исследования мы ставим критический разбор обеих утверждений на основане полученных нами данных о структуре аутосомного генофонда беларусов.

# Анализ структуры аутосомного генофонда популяции беларусов: методы, технические параметры и предварительные замечания.

В целях сопоставимости выводов данных анализа с приведенными выше выводами профессиональных популяционных генетиков, мы использовали в своем исследовании референтную выборку беларусов из ДНК-банка Института Генетики Беларуси в том же объеме, в котором она была задействована в исследовании группы ученных под руководством Бехара (модифицированная выборка Генбанка с кодом доступа: GSE21478)[[4]](#footnote-4). Наряду с референтной группой беларусов (обозначена как Belarusian), мы использовали данные лиц беларуского происхождения из нашего собственного проекта этно-популяционного анализа лиц, предки которых проживали на территории Беларуси минимум 100-150 лет (обозначено как Belarusian\_V).

Для проведения сравнительного анализа генофонда популяций нам понадобился референтный набор популяций. Референтный набор популяций в этом калькуляторе был собран в программе [PLINK](http://www.google.com/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&cad=rja&sqi=2&ved=0CCoQFjAA&url=http%3A%2F%2Fpngu.mgh.harvard.edu%2F~purcell%2Fplink%2F&ei=BZdxUa-RK4X74QT-joGQBg&usg=AFQjCNHI9oFtR2YOxmlkKaGIMkTnNxTbSA&sig2=Jr-W1O-HAyNVp7Wsousv5A&bvm=bv.45373924,d.bGE)  методом “intersection&thinning” ( дословно “пересечением и истончением”) образцов из различных источников данных: [HapMap 3](http://www.sanger.ac.uk/resources/downloads/human/hapmap3.html) (отфильтрованный набор данных КЕС, YRI, JPT, CHB), [1000genomes](http://www.1000genomes.org/), [Rasmussen et al. (2010),](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22494)  [HGDP](http://hagsc.org/hgdp/files.html) (кураторская база данных Стэнфордского университета), [Metspalu et al. (2011),](http://dienekes.blogspot.com/2011/12/population-structure-in-south-asia.html) Yunusbayev et al (2011), [Chaubey et al. (2010)](http://mbe.oxfordjournals.org/content/early/2010/10/26/molbev.msq288.short) и т.д. Кроме того, мы отобрали произвольным образом по 10 сэмплов (или максимальное количество доступных сэмплов в тех случаях, когда общее число сэмплов в популяции было меньше 10) от каждой европейской страны, представленной в панеле базе данных [POPRES.](http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000145.v4.p2) Наконец, для того чтобы оценить степень корреляции между современным и древним генетическим разнообразием населения Европы, мы также включили в выборку образцы древней ДНК [Эци](http://ru.wikipedia.org/wiki/%D0%AD%D1%86%D0%B8) ([Keller et al. (2012))](http://www.nature.com/ncomms/journal/v3/n2/full/ncomms1701.html) , образцы жителей шведского неолита Gök4, Ajv52, Ajv70, Ire8, STE7 ( [Skoglund et al. (2012))](http://dienekes.blogspot.com/2012/04/ancient-dna-from-neolithic-sweden.html) и 2 образца La Braña – останков мезолитических жителей Пиренейского полуострова ([Sánchez-Quinto et al.(2012](http://www.blogger.com/))).

Затем мы добавили 90 образцов – анонимизированных данных - участников моего проекта. После слияния вышеупомянутых наборов данных и истончения набора SNP с помощью особой команды PLINK, мы исключили SNP-ы с менее чем 0.5% минорных аллелей. После чего мы отфильтровал дубликаты, лиц с высоким уровнем общих по происхождению идентичных сегментов (IBD). В качестве критерия фильтрации были использованы расчеты IBD в Plink, где IBD представлена как средняя доля аллелей общих между двумя людьми по всем анализируемым локусам. Затем мы удалили из выборки лиц с высоким коэффициентом предпологаемого родства, коэффициенты родства были вычислены в программном обеспечении [King](http://bioinformatics.oxfordjournals.org/content/26/22/2867.full)).

Для получения более стабильных результатов, мы также отфильтровали сэмплы с более чем 3 стандартными отклонениями от средних данных по популяции. Поскольку коэффициент родства может быть надежно определен с помощью оценки HWE (ожидания, вытекающего из закона [Харди-Вайнберга](http://www.google.com/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&cad=rja&ved=0CC0QFjAA&url=http%3A%2F%2Fmga.bionet.nsc.ru%2F~yurii%2Fcourses%2F2012.04_Kazan%2Fslides%2FHWE_RU.pdf&ei=qp1xUb3FDOvU4QTYqIGYDA&usg=AFQjCNHWJZ-PAuhtc02erLPcu3Sh7R_45w&sig2=SGVUME4xUQZUVD08qUQQeQ&bvm=bv.45373924,d.bGE)) между SNP-ами с той же базовой частотой аллелей, то SNP-ы с существенным отклонением (*p* < 5.5 x10−8) от ожидания Харди-Вайнберга были удалены из объединенного набора данных. После этого мы выделили те SNP-ы, которые присутствовали в чипах Illumina / Affymetrix, и затем произвели фильтрацию снипов на основе расчетов степени неравновесного сцепления (в этом эксперименте мы использовалт хромосомное ‘окно’ размером в 50 базовых пар, с шагом 5 базовых пар и пороговым значением уровня сцепления R ^ 2, равным 0,3).

По окончанию этой сложной последовательности операций, мы получил окончательноый набора данных, который включал в себя 80 751 снипов, 2516 человек и 225 референсных популяций.

## Анализ этно-популяционного адмикс

 В ходе следующеего этапа, окончательный набор данных по референсным популяциям (которые я храню в linkage-формате PLINK) был обработан в программе [Admixture](http://www.genetics.ucla.edu/software/admixture/). Во время выбора подходящей модели проведения теста на этно-популяционный адмикс, мы столкнулись с крайне трудной задачей: как было показано в профильных научных исследованиях ([Patterson et al.2006](http://www.plosgenetics.org/article/info%3Adoi/10.1371/journal.pgen.0020190)) количество маркеров, необходимых для надежной стратификации популяций в анализе обратно пропорциональна генетическому расстоянию (фСТ) между популяциями. Согласно рекомендациям пользователей программы Admixture, считается что примерно 10 000 генетических SNP-маркеров достаточно для выполнения интер-континентальной GWAS -коррекции обособленных популяций (например, уровень дивергенции между африканскими, азиатскими и европейскими популяциями FST > 0.05), в то время как для аналогичной коррекции между внутриконтинентальными популяциями требуется более чем 100000 маркеров (в Европе, например, ФСТ < 0.01). Для повышения точности результатов Admixture мы решили использовать метод, предложенный Dienekes. Этот метод позволяет [преобразовать частот аллелей в “синтетические” индивид](http://dodecad.blogspot.com/2011/05/how-to-create-zombies-from-admixture.html)ы (см. также [пример Зака Аджмала](http://www.harappadna.org/2011/07/admixture-supervised-zombies-vs-unsupervised/) из проекта HarappaDNA). Идея метода довольно проста: сначала необходимо запустить unsupervised анализ Admixture с целью вычисления частот аллелей в так называемых предковых компонентов, а затем на основании аллельных частот сгенерировать “фиктивные популяции”. Именно эти фиктивные популяции и индивиды будут использоваться в ходе чистых референсов в ходе последующего анализа этно-популяционного анализа. Впрочем, как и любые другие исследователи, работающий над четким решением проблемы этно-популяционного адмикса, мы были вынуждены считаться с ограничениями этого подхода. Хотя мы отдаем себе отчет в существовании явных методологических подвохов в использовании смоделированных искусственных индивидов для определения адмикса в реальной популяции, мы полагаем что полученные в ходе аллельно-частотного моделирования “фиктивных индивидов” представляют самую лучшую аппроксимацию древних генетических компонентов предпологаемых древних компонентов. В ходе применения простого моделирующего метода, нами были получены значимые результаты в ходе создания нового калькулятора. Сначала мы произвели unsupervised Admixture (при значении К = 22, т.е 22 кластера частот аллель или предковых компонентов). По выполнению анализа нами были получены оценки коэффициентов адмикса в каждой из этих 22 аллельных кластеров, а также частоты аллелей для всех SNP-ов в каждой из 22 родовых популяций.

Затем мы использовали мнемонические обозначения для каждого компонента (имена для каждого из компонентов выведены в порядке их появления). Нужно помнить, что обозначения этих компонентов носят скорее мнемонический условный характер:

**Pygmy**

**West-Asian
North-European-Mesolithic
Tibetan
Mesomerican
Arctic-Amerind
South-America\_Amerind
Indian
North-Siberean
Atlantic\_Mediterranean\_Neolithic
Samoedic
Proto-Indo-Iranian
East-Siberean
North-East-European
South-African
North-Amerind
Sub-Saharian
East-South-Asian
Near\_East
Melanesian
Paleo-Siberean
Austronesian**

Вышеупомянутые частоты аллель, вычисленные в ходе unsupervised (безнадзорного) анализа (Admixture K = 22) объединенного набора данных, были затем использованы для симуляции синтетических индивидов, по 10 индивидов на каждую из 22 предковых компонент. Это симуляционное моделирование проводилось с помощью PLINK команды -simulate Расстояние между между симулированными «искусствеными» индивидами было визуаилизировано с использованием многомерного масштабирования.



На следущем этапе, я включил группу смоделированных индивидов (220 индивидов) в новую эталонную популяцию. После чего я запустил новый анализ А, на этот раз в полном “поднадзорном” режиме для K = 22, причем полученные в ходе симуляционного моделирования фиктивные популяции фиктивных индивидов использовались в качестве новых референсных эталонных групп. На конвергенцию 22 априорно заданых предковых компонентов было затрачено 31 итераций (3 7773,1 сек) с окончательным loglikelihood: -188032005,430318 (ниже, на следущей странице, приведена таблица значений Fst между расчетными ‘предковыми’ популяциями):



Рисунок 1. FST-дистанции между компонентами

Приведенная выше матрица Fst дистанций была использована для определения наиболее вероятной топологии NJ-дерева всех 22 предковых компонентов ( примечание: в качестве outgroup-таксона использовался South-African component).



Рисунок 2.Наиболее вероятная топология NJ-дерева всех 22 предковых компонентов

# Анализ структуры аутосомного генофонда популяции беларусов: результаты анализа этнического адмикса.

После проведения анализа этно-популяционного адмикса мы получили следущие результаты, обсуждению которых будет посвящена следущая часть нашего исследования. Результаты представляют собой разбивку аллельных частот на 22 кластера, каждый из которых представляет собой гипотетическую предковую популяцию. Поскольку в цели данного небольшого исследования не входит подробный анализ всех популяций, мы ограничимся сравнительном анализом структуры (компонентов) беларусов c географически близкими популяциями, а также с теми популяциями, которые могли входить в исторические контакты с предками современных беларусов:



Рисунок 3. Результаты анализа ADMIXTURE K=22

У рассматриваемых здесь европейских популяций наиболее часто представлены следующие компоненты:

**North-East-European,Atlantic\_Mediterranean\_Neolithic,North-European-Mesolithic, West-Asian, Samoedic, Near\_East.**

Разберем вкратце каждый из них. **В ракурсе нашего исследования самым важным компонентом представляется – северо-восточно-европейский компонент North-East-European**, он присутствует почти у всех европейцов, и в самой значительной степени - у балтов и славян: литовцы (81,9), латыши (79,5), **беларусы (76,4),** эстонцы (75,2), поляки (70,2), русские (67- 70,4), украинцы (62,1- 67,1), сорбы (65,9), карелы (60,2), вепсы (62,5), чехи (57,4), северные немцы (54,6), южные- 42,6, у британцев от 46 до 49, норвежцы- 48,1, шведы- (53,7).

Второй по значимости компонент - **Atlantic\_Mediterranean\_Neolithic (юго-западно-европейский или просто западно-европейский неолитический компонент)**.[[5]](#footnote-5)У восточноевропейцев он выражен в умеренной степени- чехи (27,8), поляки (18,4), украинцы ( от 17 до 21%), **беларусы (13%)**, русские (от 11 у северных до 17,3 у южных), у коми (8,9 %), манси (8,8 %).

Третьй компонент – **северо-европейский мезолитический компонент -North-European-Mesolithic[[6]](#footnote-6)**: cаамы (76,4 %), финны (от 30,1 до 37,3 %), вепсы (24,1), карелы (23,2), ижорцы (22, 7). Заметен этот компонент и у северных русских (10,5 %), норвежцев (9,8 %), шведов (7,8 %), эстонцев (7,1 %). У **беларусов он практически отсутствует (1.1%).**

Четвертый компонент – **западно-азиатский (кавказский) West Asian[[7]](#footnote-7)**. На интересуемой нас территории этот компонент чаще встречается у казанских татар (9,9 %), южных немцев (8,4), украинцев (от 6,6 до 7,7 %), южных русских (6,2%). На западе высок процент у итальянцев (21,5 % у центральных итальянцев), французов (6,7 %), **у беларусов (2.2%).**

Пятый компонент - **уральский** Samoedic. Значительно присутствует у селькупов (68,1%), хантов (64,6), ненцы (37,1), манси (30,9 %-), удмурты (29,6), марийцы (27, 8), шорцы (22,0 %), башкиры (21,7%), чуваши и хакассы по 17,6 %, коми- 16,4 %, казанских татар (11,9 %). У западноевропейцев этот компонент практически не встречается, у русских (от 1,0% у центральных до 4,7 % у северных), у карел (1,6%), словаков (1,4%), западных украинцев (1,7 %), **беларусы (0.5%).**

Шестой компонент – ближневосточный Near\_East[[8]](#footnote-8) У южных немцев (3,5), украинцы (от 2,3 у восточных до 3,8 % у западных), чехи (3,0), **беларусы (3,4),** словаки (3,2), у русских от 1,0 до 1,5%, у литовцев- 1,4%, у поляков- 1,3 %.

## Анализ разделяемых аутосомных сегментов между популяциями Северо-Восточной Европы.

С целью верификации результатов анализа главных компонентов генетического разнообразия я подготовил новую выборку популяций, которая включает в себя ряд референсных евразийских популяций и анализируемую группу участников моего проекта MDLP. В совокупности, выборка включала в себя 900 индивидов, каждый из которых был типирован по 350 000 снипам.В ходе нового экспериментального теста в ходе статистической обработки общих по генетическому происхождению сегментов хромосом в составе выборки было выделенно 15 групп кластеров генетически близких популяций Как нам представляется, ключевым моментом для понимания принципов этого анализа, а также результатов, является понятие эффективной популяции или эффективный размер (Ne) популяции, т.е размера той популяции которая участвовала в репродукции или обмене генами в некоем отдаленном временном промежутке. Собственно говоря, эффективная популяция – это даже не число уникальных предков, а математическая абстракция разброса гамет, размер которого оценивается исходя из разброса числа гамет относительного к гамет, передаваемых родителям репродуктивного возраста следующему поколению. Он отличается от репродуктивоного объема Nr в той мере, в какой существует неравный вклад лиц родительского поколения в генофонд следующего поколения. Это создает разброс значений числа гамет к, того родителя относительно числа гамет к, передаваемых родителям следующему поколению (Wright, 1931, Li Ch. Ch., 1955). Новая программа Chromopainter позволяет оценить этот размер, исходя из числа наблюдаемых рекомбинаторных гаплотипов и значений LD. Когда я производил оценку этого размера, то для каждой из 22 неполовых хромосом он получился разный, однако среднеарифметическое значение составило 22 000. Это близко к значениям Neрекомендованным к использованию профессионалами (например, авторами программы IMPUTE V2). Как видно из приведенных ниже результатов, даже 22 000 для совокупности эффективного размера элементарных популяций – это более, чем достаточно. 

Рисунок 4. Расположение популяций в пространстве 1 и 3 главных генетических компонентов

Изложим ниже некоторые закономерности размещения популяци

1. Финны оказались ближе к русским и поволжским финно-угорам (эрзя и мокша)

2. Все литовцы (участники проекта + референсы из вышеупомянутой статьи Бехара) и часть референсных белорусов образовали отдельный кластер, тесно примыкающий к кластеру белорусов, поляков, украинцев

3. Следущим кластером является центрально-европейский кластер, представленный главным образом венграми, хорватами, а также частью немцев.

4. Ниже находится балканский кластер (румыны, болгары и часть венгров).

5. К этому кластеру примыкают турки и часть армян

6. В центре плота находятся западноевропейцы из моего проекта (французы, немцы, бельгийцы и жители британских островов).

7. Выше находятся два оркнейских кластера, в которых находится и часть скандинавских сэмплов.

8. Еще левее находится кластер образованный референсными северо-итальянцами и тосканцами.

9. Ниже находятся армяне и слево итало-иберийский кластер (часть итальянцев и испанцы).

10. Левее этой группы популяций находится кластер ашкеназов.

11. Наконец, самый крайний слева кластер представлен изолированной популяцией сардинцев.

12. Ниже итало-иберийского и армянского кластеров расположен целый ряд кавказский кластеров. Это прежде всего адыгейцы и абхазцы, затем северные осетины.

13. Вышеназванные кластеры частично перекрывают кластер ногайцев (что свидетельствует о наличии генетического обмена между северокавказскими популяциями и ногайцами)

14. Кластер ногайцев плавно переходит в кластер узбеков, который в свою очередь примыкает к изолированному кластеру чувашей

15. Наконец самым изолированным кластером является кластер французских басков (в нижнем левом углу плота).[[9]](#footnote-9)

## Обсуждение результатов и выводы

Как отмечалось в введении к нашей статье, главной задачей нашего исследования являлась проверка двух рабочих гипотез, озвученных в предыдущих исследованиях профессиональных попгенетиков. Во-первых, это гипотеза о **присутствии трех основных древних компонентов , которая указывает на возможность общего происхождения славян и балтов**. Во-вторых, это утверждение о том, что своеобразие аутосомного генофонда беларусов может быть связанно **с вкладом балтского субстрата.**

**После внимательного изучения результатов нашего исследования,** можно сказать, что оба из приведенных выше заключений представляют собой крайне упрощеные варианты сложного процесса формирования аутосомного генофонда беларусов. Хотя мы и не можем предоставить окончательных аргументов в пользу или опровержение каждой из этих версий, мы может предоставить более полное и подробное обозрение структуры аутосомного генофонда. В отличие от трех основых компонентов, упомянутых выше, в нашем исследовании мы выделили шесть основных компонентов, типичных для европейцев в целом. Основу генофонда составляет компонент, который мы обозначили как северо-восточно-европейский компонент. Именно этот компонент выделяет беларусов среди других восточных славян, приближая их к современным балтийским популяциям (у литовцев процент компонента составляет 81,9, у латышей - 79,5%, у **беларусов -76,4%,** у эстонцев - 75,2%). Примечательно, по мере удаления от территории Беларуси на север в с торону Латвии и Эстонии, увеличивается процент северо-европейского генетического компонента (как мы полагаем, этот компонент доминировал в генофонде доисторических жителей Скандинавии в эпоху до распространения финно-угоров и индо-европейцев). С другой стороны, беларусов и других восточных славян отдаляет от балтов и сближает друг к другу более высокий процент так называемого западно-азиатского или кавказского компонента (любопытно, что в этом случае эта закономерность может свидетельствовать в пользу западно-азиатской теории происхождения индо-европейцев).

Далее, как показывает анализ в программе fineStructure, генофонд беларусов характеризируется высокой степенью генетических контактов как с балтами, так и остальными славянами, а также с рядом финно-угорских популяций (например, c эрзя и мокша). О симметричном характере межпопуляционного обмена свидетельствует симметричное расположение популяции беларусов относительно этих трех групп.

Исходя из вышенаписанного, представляется логичным сделать вывод о том, что основной критический этап становления аутосомного генофонда пришелся на период смешивания носителей северо-восточно-европейского генетического компонента с носителями западно-азиатского (кавказского) генетического компонента, после чего предковый аутосомный генофонд беларусов приобрел относительную стабильность. Разумеется, данная модель не исключает позднейшие эпизоды смешивания популяций, но они оставили меньший след в структуре аутосомного генофонда беларусов. В этой связи возникает очевидный вопрос – в какой именно исторический период произошло смешение носителей северо-восточно-европейского генетического компонента с носителями западно-азиатского (кавказского) генетического компонента, и кто были их носителями?

В начале сентября 2012 года известная американская лаборатория популяционной генетики доктора Райха опубликовала альфа-версию программного продукта [ADMIXTOOOLS 1.0](http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&ved=0CDEQFjAA&url=http%3A%2F%2Fgenetics.med.harvard.edu%2Freich%2FReich_Lab%2FSoftware_files%2FREADME_ADMIXTOOLS.pdf&ei=dUrSUKScKIvCswar84DwAQ&usg=AFQjCNGw8GpC5Bck-wjNXJ_2y9r8y8xS9g&bvm=bv.1355534169,d.Yms). Альфа-версия была разработана для внутреннего использования, поэтому modus operandi этого продукта вряд ли является кристально понятным для стороннего пользователя. Положительным аспектом на мой взгляд является то, что ADMIXTOOLS пакет обеспечивает полную совместимость с форматом другой очень популярной программы[EIGENSOFT](http://genepath.med.harvard.edu/~reich/Software.htm), которая была разработана в той же лаборатории. Это немаловажное обстоятельство намного упрощает процесс обучения в ADMIXTOOLS .

Вышеупомянутый пакет включает в себя 6 приложений, среди которых я считаю наиболее полезной **qp3Pop** и утилиты для вычисления частотной характеристики аллелей. Впрочем, я не собираюсь обсуждать **qp3pop** во всех деталях и в контексте данной заметки достаточно отметить, что эта программа реализует тест three\_pop (F\_3), подробно описанный в известной статье  [Рейха и соавт. 2009.](http://www.nature.com/nature/journal/v461/n7263/abs/nature08365.html)

Однако другой имплементированный в пакете метод, – метод rolloff – нуждается в более пристальном внимании. Этот метод позволяет производить математическую оценку как времени, так как и уровня адмикса. Оценка производится на основании анализа неравновесия по сцеплению между SNP-ами. Тут необходимо вспомнить стандартное определение неравновесия по сцеплению.Неравновесием по сцеплению (часто используется английская аббревиатура **LD**) называется неслучайная связь между двумя аллелями, в силу которой определенные комбинации аллелей встречаются наиболее часть. В теории, чем дальше друг от друга находятся SNP-ы ,тем меньше будет уровень LD. Темп угасания снижения LD в адмиксе напрямую связана с числом поколений, прошедших с момента адмикса, так как c возрастанием числа поколений увлечивается число рекомбинаций произошедших между двумя отдельными SNP-ами. Проще говоря: Rolloff соответствует экспоненциальной кривой угасания уровня LD от расстояния, и эта скорость экспоненциального снижения как раз и используется для оценки числа поколений, так и уровня адмикса в анализируемой популяии. Учитывая, что одно поколение примерно равно 29 лет, можно преобразовать число поколений в года.

Этот метод открывает интересные перспективы. Для целей этого анализа, я создал специальный набор SNP-данных, который включает в себя около 750 000 cнипов, частично или полностью в 250 различных популяциях человека. Далее, я разбил популяции 3 \* 62 000 трио в следующем виде (X, Y, Z), где X и Y – пара рефренсных групп, а Z – белорусы из коллекции [Behar et al.2010](http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21478). После этого я провел q3Pop анализ этих трио.

Результаты изложены в нижеприведенной таблице

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Indian | Polish | Belarusian | -0.000736 | 0.000251 | -2.935 |
| Polish | Indian | Belarusian | -0.000736 | 0.000251 | -2.935 |
| Karitiana | Sardinian | Belarusian | -0.001278 | 0.000517 | -2.471 |
| Sardinian | Karitiana | Belarusian | -0.001278 | 0.000517 | -2.471 |
| Otzi | North\_Amerind | Belarusian | -0.002556 | 0.001126 | -2.271 |
| Cirkassian | Polish | Belarusian | -0.000488 | 0.000231 | -2.113 |
| Polish | Cirkassian | Belarusian | -0.000488 | 0.000231 | -2.113 |
| Pima | Otzi | Belarusian | -0.002727 | 0.00137 | -1.99 |
| Pima | Sardinian | Belarusian | -0.000794 | 0.000431 | -1.843 |
| Sardinian | Pima | Belarusian | -0.000794 | 0.000431 | -1.843 |
| Otzi | Surui | Belarusian | -0.002938 | 0.001931 | -1.522 |
| Surui | Otzi | Belarusian | -0.002938 | 0.001931 | -1.522 |

На первый взгляд, результаты нашего эксперимента с 3qPop, кажется, неплохо согласуются с выводами, содержащимися в работе [Паттерсон и др. 2012](http://www.genetics.org/content/early/2012/09/06/genetics.112.145037.abstract): “Самый поразительный вывод состоит в обнаружени четкого сигнала адмикса в северной Европе, один из элементов которого связан с предками населения наиболее близкого по своей генетике к баскам и жителям Сардинии, а другой – с предками современного населения северо-восточной Азии и Америки. Этот явный сигнал, вероятно, отражает историю смешивания неолитических мигрантов с коренным населением Европы, что подтверждается недавним генетическим анализом древних костей Швеция и секвенированием полного генома Отци Тирольца”. Что касается собственно белорусов, то источники сигнала смешивания с посторонними популяцими менее ясны и расплывчаты. Как было показано ранее, с точки зрения формального анализа примесей (f3 статистики), белорусы могут быть представлены в виде популяционного микса поляков и индусов / черкессов. Первый компонент смеси может быть связан с носителями культуры шнуровой керамики/боевых топоров и культуры колоковидных кубков; второй, в соответствии с результатами, должен быть общим для индусов и черкесов.

**Белорусы = ((неолитические культуры Европы) + “носители культуры колоковидных кубков”) + (мезолитическое население Европы) + компонент носителей культуры шнуровой керамики)) + скифо-сарматский тип**

Для оценки дата события базового адмикса в белорусской популяции, мы использовали в качестве референсных популяций поляков и индусов (Примечание: мы снизили порог генетических дистанции в параметрах Rolloff для снижения уровня шума от более поздних адмиксов).



Как вы можете видеть, сигнал присутствия адмикса обнаруживается гораздо хуже, и в силу этого, погрешности в оценке временного промежутка высоки:

**154,158 + -87,024 поколений назад (или, 4470 + -2523 года до настоящего времени / 2510 – +2523 лет до н.э.).**

Исходя из этого, мы решили модифицировать Rolloff-анализ генофонда белорусов, используя на этот раз в качестве референсов **литовцев и пуштунов**. Следуя этому совету, я решил предпринять вторую попытку формального анализа адмикса в двух имеющихся у нас выборках беларусов ( выборка беларусов из статьи Behar et al. 2011), и выборка беларусов, собранная в нашем проекте.Ниже приведены результаты эксперимента с двумя этими группам (в отличие результатов нашей предыдущей попытки, результаты данного эксперимента менее “зашумленные”):



Интервал числа поколений, прошедших со времен анализируемого адмикса (105.086+-52.59) или 3069 +- 1525 лет до настоящего времени, что соответствует временном интервалу 2 тыс. до нашей эры – 6 век нашей эры. Принимая во внимание эти выводы, мы можем предположить, что основной аутосомный эпизод смешивания предковых популяций беларусов произошел в течении довольно таки продолжительного времени, охватывающего несколько тысяч лет. В этой связи, вопрос о том, кто именно был носителями северо-восточно-европейского генетического компонента с носителями западно-азиатского (кавказского) генетического компонента, остается открытым.

1. Балановский О. П., Тегако О. В*.* Генофонд белорусов по данным о трех типах генетических маркеров — аутосомных, митохондриальных,Y-хромосомы. [↑](#footnote-ref-1)
2. Давыденко Олег, Кушнеревич Елена.2011. Гаплогруппы У-хромосом и происхождение национального генофонда. *Наука и Инновации.Тайна происхождения белорусов*. 9(103),12-15. [↑](#footnote-ref-2)
3. Behar, Doron.M.2010.The genome-wide structure of the Jewish people.Nature 466, 238–242. [↑](#footnote-ref-3)
4. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21478> [↑](#footnote-ref-4)
5. Больше всего у сардинцев (68,1 %), басков (59,2 %), иберийцев (48,8), корсиканцев (47,8), португальцев (46,6), северных итальянцев (44,3), французов (43,5 %). Данный компонент достаточно выражен у всех западноевропейцев- более 30 % [↑](#footnote-ref-5)
6. Название связано с тем, что этот компонент достигает значительных частот в древней ДНК жителей мезолитической Иберии, неолитических жителей Швеции и современном ДНК жителей Фенноскандии [↑](#footnote-ref-6)
7. Наибольший процент на Западном Кавказе- абхазы (64, 9%), имеретинцы (63,7), лазы (56,6), аварцы (56,8), лезгины (55,4). [↑](#footnote-ref-7)
8. Евреи Йемена (60,9 %), Сауд. аравия (59,5), бедуины (56,7), евреи Эфиопии (52,5), египтяне (43,8).В Европе oтносительно много у итальянцев (центр- 17,4), португальцев (11,9). [↑](#footnote-ref-8)
9. Такое поведение на плоте объясняется только изолированным положением популяции и небольшим числом эффективной популяции.То есть все эти баски являются многократными родственниками между собой т.е., положение басков на графике есть следствие *к*лассического генного дрейфа, который можно наблюдать на карте.На самом деле положение басков на данном плоте не может ни подтвердить, ни опровергнуть гипотезу о континуитете баскской популяции , т.к PCA-координаты (eigenvalues и eigenvectors) вычислялись в Chromopainter исходя из количества shared DNA chunks между **популяциями-донорами и популяциями-рецепиентами.**То есть баски изоляты в том смысле, что уровень обмена ДНК между ними и другими популяцими ничтожен.

Исходя из этого можно сделать вывод о том что баски эта экстремально-эндогенная популяция изолянтов, при этом генетическое разнообразие басков низко, т.к. размер эффективной популяции басков низок. [↑](#footnote-ref-9)