Языки и гены – глобальный анализ

Используя единую базу генетических и лингвистических данных для 397 популяций, говорящих на 295 языках, специалисты провели глобальный анализ генетически-языковых соответствий и несоответствий в населении мира. В большинстве случаев наблюдается закономерность – если популяции используют родственные языки, они близки генетически. Но в каждом пятом случае эта закономерность нарушается, что происходит во всех регионах и во всех языковых семьях. Таким образом, взаимоотношения генов и языков оказываются очень динамичными, как в пространстве, так и во времени.

История человечества записана в генах и в языках. Так начинают свою статью в журнале PNAS специалисты из Цюрихского университета и Института эволюционной антропологии Общества Макса Планка, в которой они сопоставляют пути генов и языков в человеческих популяциях. До какой степени генетическая и лигвистическая истории совпадают и в каких случаях они расходятся? Для исследования этого вопроса авторы создали базу данных GeLaTo (Genes and Languages Together), в которой отражено разнообразие языков и глобальное генетическое разнообразие.

О том, что эволюция генов и языков имеет концептуальное сходство, писал еще Чарлз Дарвин. Его идея получила подтверждение в трудах Кавалли-Сфорца в 1980-х годах. Глобальное филогенетическое дерево и глобальное дерево языков во многом совпадают, в отдельных частях света, например, в Европе, процессы генетической и лингвистической дифференциации зеркально отражают друг друга. Эти совпадения являются результатом локальных процессов совместной диффузии и усиливаются при масштабных популяционных экспансиях. Есть гипотеза, что миграции, вызванные переходом к земледелию и скотоводству в неолите, дали начало основным языковым семьям. Но из генетически-языковых совпадений есть и исключения. Несовпадения возникают, когда популяции перенимают язык у своих соседей без (или с мимимальным) генетическим смешением, или же когда популяций генетически ассимилирует соседскую популяцию без изменения языка.

База GeLaTo включает информацию о более чем 4000 человек, представляющих 397 популяций, говорящих на 295 языках. Участники были генотипированы на панели Human Origins, включающей около 500 тысяч SNP.

Сначала авторы оценили, в какой степени географические соседи говорят на языках одной семьи. Оказалось, что в радиусе 500 км больше половины популяций из базы GeLaTo имели соседей, относящихся к другим языковым семьям. Когда радиус увеличили до 1000 км, их доля возросла до 84%. При сопоставлении генетических и лингвистических данных выяснилось, что в большинстве случаев между ними наблюдается соответствие: популяции, принадлежащие к одной языковой семье, генетически также более близки между собой. Но приблизительно в одной пятой случаев (79 из 397, 18% популяций) отмечены генетически-лингвистические несоответствия.

Авторы вводят понятия анклавов – это популяции, лингвистически и/или генетически отличающиеся от своих географических соседей и в то же время сходные с другими, географически далекими популяциями. Если популяция оказывается в изоляции как по языковому, так и по генетическому критериям, она представляет собой парный генетически-лингвистический анклав.

Непарный генетический анклав возникает, когда популяция принимает язык географических соседей, генетически далеких от нее. Причины могут быть разными, часто перенимают язык культурно или политически доминирующего народа. Например, кочинские евреи, живущие в Индии, говорят на языке дравидийской семьи; грузинские евреи говорят на языке картвельской семьи.

Лингвистические анклавы гораздо более редки. Наиболее известный пример – это венгры, которые генетически близки к окружающим балканским популяциям, но лингвистически от них далеки. Их язык относится к уральской языковой семье, его принесли мадьяры, которые вторглись в Карпатскую котловину в XIX веке. Как показывает анализ древней ДНК, вскоре после миграции переселенцы были генетически ассимилированы окружающими индоевропейскими народами Карпатской котловины. Тот факт, что при этом они сохранили свой язык, это уникальный случай. Более сложный вариант представляют собой жители Мальты. Генетически они близки к сицилийцам, а говорят на языке афроазиатской семьи, на который оказывают влияние тюркские и индоевропейские языки.

Полученные данные продемонстрировали, что распадение соответствия «гены-языки» происходит среди разных языковых семей и в разных регионах. Авторы задаются вопросом, является ли это отклонением от нормы или вариантом нормы? Чтобы более детально разобраться в этом, они исследовали восемь языковых семей, которые были наиболее полно представлены в базе GeLaTo.

Сравнили генетические расстояния Fst между популяциями одной языковой семьи и разных языковых семей и проследили

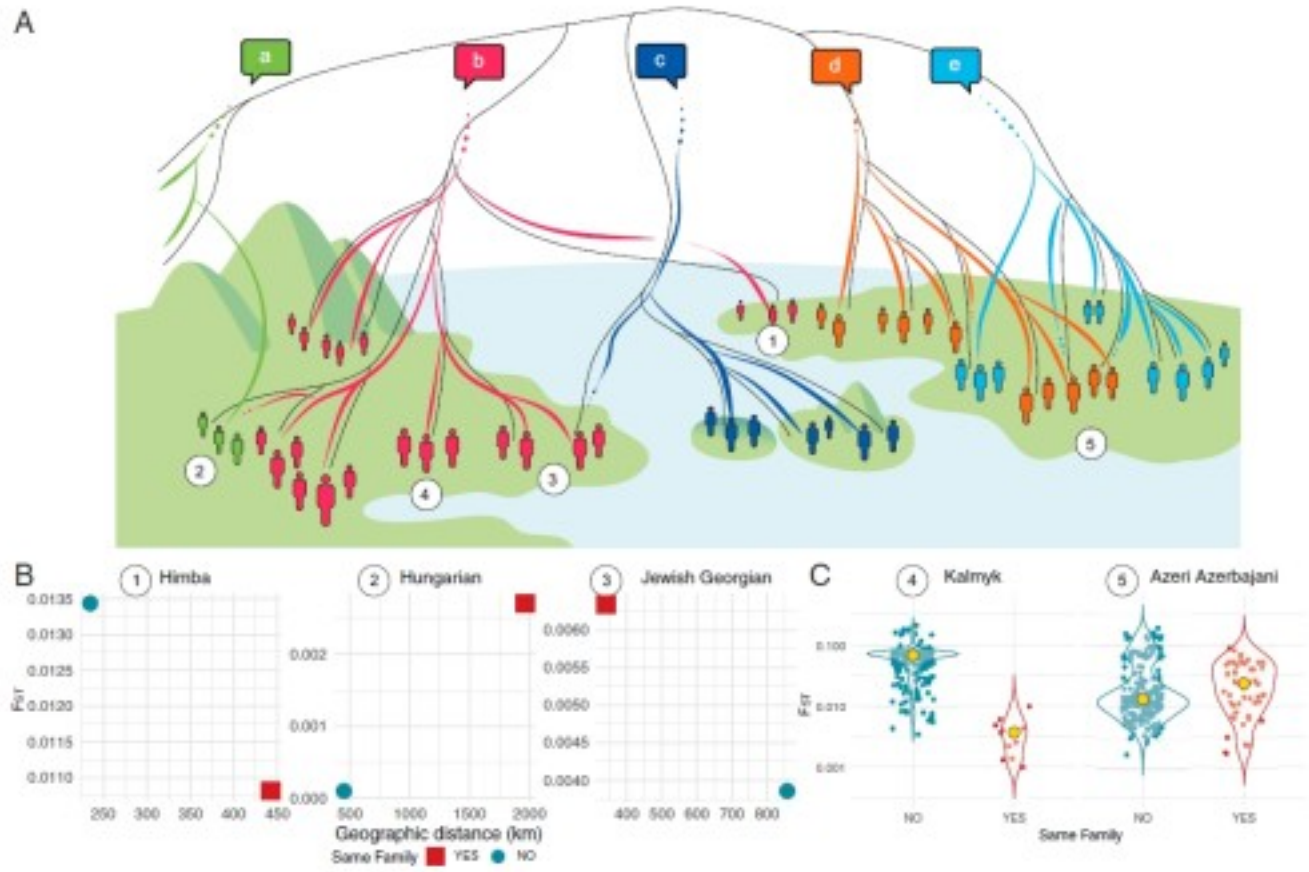
их корреляцию с географическими расстояниями. Для индоевропейской, атлантик-конго, монгольской и сино-тибетской языковых семей генетические расстояния внутри семьи (при сравнении их с расстояниями между популяциями разных семей) были меньше чем соответствующие географические расстояния. Для популяций афроазиатской и тюркской языковых семей сравнительные показатели генетических и географических расстояний частично перекрывались. Для популяций уральской и австранезийской языковых семей генетические расстояния в пределах одной семьи превышали соответствующие географические расстояния. Таким образом, генетически-лингвистическая согласованность в разных языковых семьях градуально изменяется.

На согласованность генов и языков могут влиять демографические и культурные события, происходящие в популяциях в разные времена. Для оценки временного масштаба авторы вычислили для пар популяций из базы GeLaTo время жизни общего генетического и языкового предка, таким образом, сравнили лингвистическую и генетическую временную глубину.

В некоторых языковых семьях (афроазиатской, дагестанской и уральской) время жизни генетического общего предка было таким же или меньшим, чем время жизни лингвистического общего предка. Во всех других семьях генетические предки оказались намного древнее, чем лингвистические. Более детально проанализировали три языковых семьи: индоевропейскую, австроазиатскую и тюркскую. В каждой семье провели сравнение генетического и лингвистического деревьев, наибольшее сходство найдено для индоевропейской семьи (коэффициент корреляции 0,68), для афроазиатской он составил 0,65, для тюркской — 0,57.

Таким образом, авторы показали, что генетически-лингвистические несовпадения – это не отдельные исключения, а регулярно встречающиеся ситуации, которые есть на любом континенте и во всех языковых семьях. Для отдельных популяций время жизни генетических и лингвистических общих предков может различаться, это показывает, что отношения генов и языков во временном разрезе очень динамичны. Примечательно, что наиболее близкое соотношение между генами и языками обнаружено в популяциях индоевропейской семьи.

На схеме представлены варианты лингвистически-генетических соотношений. Генетическая (демографическая) история показана черными линиями, лингвистическая история – цветными линиями; a-e – языковые семьи. Иногда эти линии идут параллельно, иногда нет. Цифрами обозначены разные варианты. 1. Лингвистически-генетический анклав (пример – популяция химба, атлантик-конго семья). 2. Линвистический анклав (пример – венгры, уральская семья). 3. Генетический анклав (пример – грузинские евреи, картвельская семья). 4. Вариант, где генетические расстояния между популяциями внутри одной семьи меньше, чем между популяциями разных семей (пример — калмыки (могнольская семья). 5. Вариант, где такая закономерность не соблюдается (пример – азербайджанцы, тюркская семья).



Варианты лингвистически-генетических соотношений. Генетическая (демографическая) история показана черными линиями, лингвистическая история – цветными линиями; a-e – языковые семьи. (Barbieria et al. , 2022)

текст Надежды Маркиной

Источник:

Chiara Barbieria et al. A global analysis of matches and mismatches between human genetic and linguistic histories // PNAS. 2022 Vol. 119 No. 47

https://doi.org/10.1073/pnas.2122084119