

## Можно ли о генетическом сходстве судить по грамматике?

[Надежда Маркина](#)

Язык указал на генетическую близость европейских популяций точнее, чем география

**Исследователи математически доказывают связь между лингвистическим и генетическим разнообразием в популяциях Европы. Эту связь они изучили у 12 народов, говорящих на индоевропейских языках, и трех из других языковых семей. По их мнению, для изученных народов язык точнее, чем география, указывает на генетическое сходство популяций. А в области лингвистики синтаксис оказался теснее связан с генетикой, чем лексика.**

**Читайте также ниже в разделе «Мнения экспертов» экспертный комментарий д.б.н. Олега Балановского «Два однотипных исследования пришли к противоположным выводам» и д.филол.н. Светланы Бурлак «Грамматика не лучше лексики».**

Люди, живущие в разных частях планеты, отличаются друг от друга и по внешности, и по языку, на котором разговаривают. Идею о том, что между биологическим разнообразием и разнообразием языков существует связь, высказал еще Дарвин. Он впервые предположил механизм этой связи. Факторы, изолирующие популяции друг от друга, способствуют накоплению как биологических, так и культурных различий, а контакты между популяциями, напротив, приводят к сходству. Поэтому и биологическое, и языковое разнообразие можно использовать для изучения истории популяций. Но во времена Дарвина и много позже эту идею невозможно было проверить на строгом количественном уровне – не было инструментов. Теперь же стало возможно математически просчитать соответствия между генами и языками.

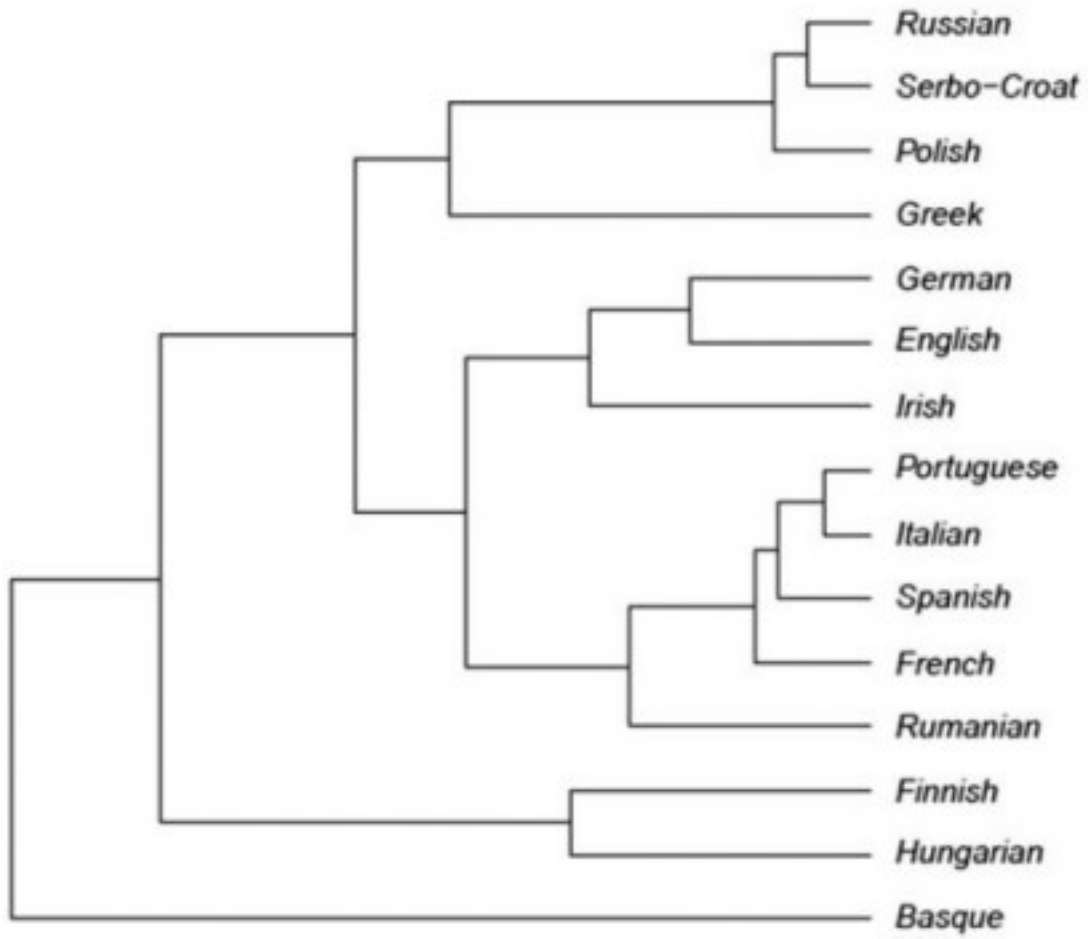
О Дарвине напоминают авторы нового исследования, которое в августе [будет опубликовано в журнале Am J Phys Anthropol](#). Команда итальянских и американских исследователей под руководством Джузеппе Лонгобарди (Giuseppe Longobardi) не в первый раз берется за сопоставление лингвистического и генетического разнообразия. В этой работе ученые использовали новые подходы для сравнения лингвистического и геномного разнообразия в Европе, причем к 12 языкам индоевропейской семьи они добавили 3 языка из других семей: финский, венгерский (финно-угорская группа) и язык басков.

Новый подход авторов к сравнительной лингвистике заключается в следующем. Во-первых, они использовали последние достижения в составлении лексических списков, применяемых для изучения родства языков. Для индоевропейских (ИЕ) языков таким последним достижением являются списки IELex (Bouckaert et al.'s (2012)). Во-вторых, для сравнения языков авторы обращаются не только к лексике, но и грамматике, разработав новый подход (систему РСМ). В этой системе грамматика любого языка представляется в виде бинарных символов, которые кодируют 56 параметров. Она позволяет сравнивать любые языки, независимо от их лексики. Исследователи подчеркивают, что грамматика меньше, чем лексика, подвержена влиянию случайных внешних факторов.

Геномные данные по 13 популяциям авторы брали из базы POPRES, включающей 5886 человек и более 500 тысяч точек замены нуклеотидов (однонуклеотидный полиморфизм, SNP), а также данные по баскам и финнам из проекта «1000 геномов». Анализ, на котором построены выводы авторов, основан на выборке 805 человек из 15 популяций, изученных по 178 тысячам аутосомных SNP маркеров (т.е. находящихся на неполовых хромосомах).

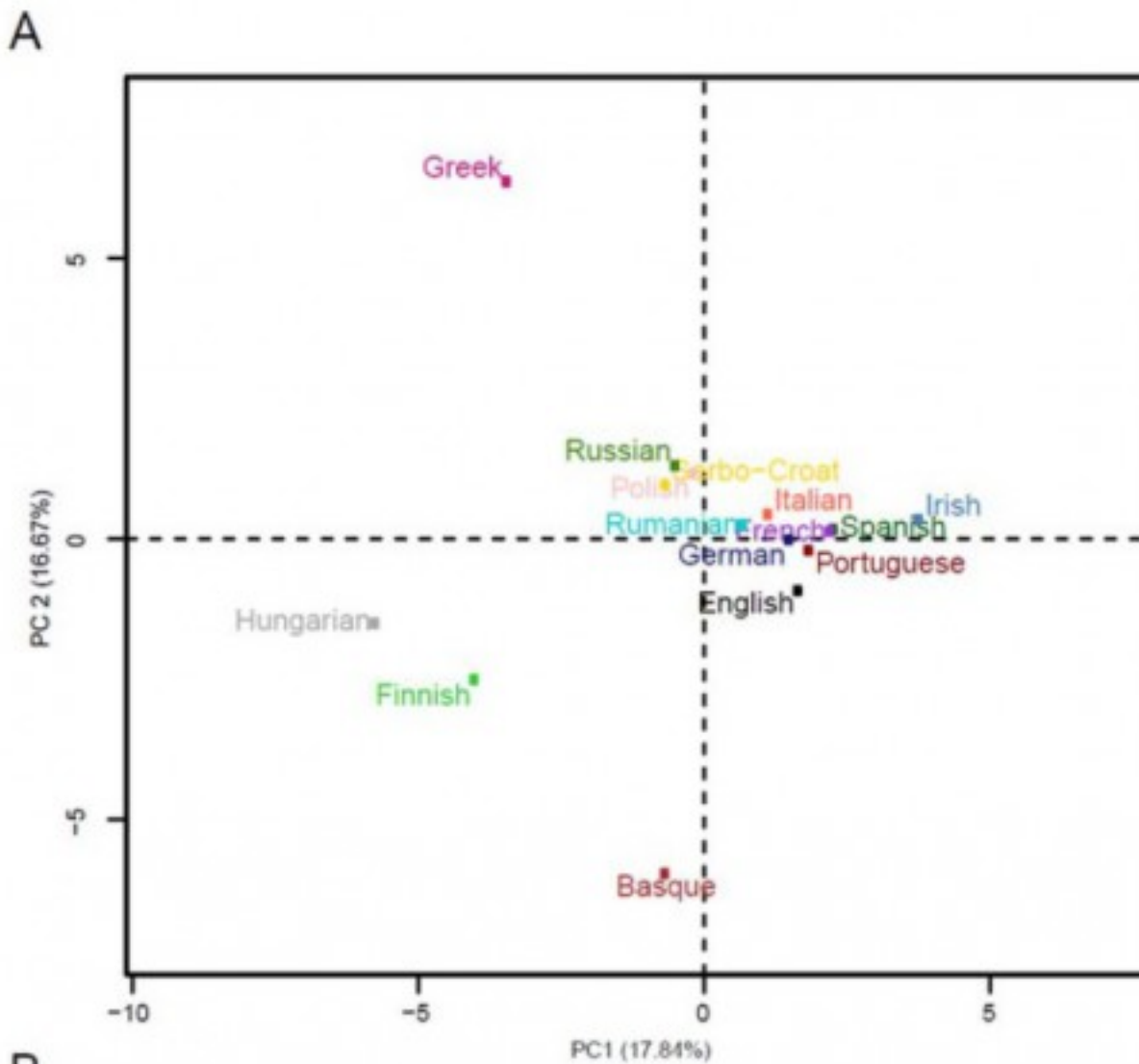
Исследователи построили четыре матрицы расстояний – как между 12 популяциями, говорящими на ИЕ языках, так и между всеми 15 европейскими популяциями. Географическая матрица строилась по географическим расстояниям между ареалами популяций. Геномная матрица – по генетическим расстояниям между популяциями, вычисленным по 178 тысячам SNP. Лингвистических матриц было две: лексическая и синтаксическая. Лексические расстояния (от 0 до 1) вычислялись по сравнению лексических списков из разных языков. Синтаксические расстояния (от 0 до 1) – по особенностям и различиям синтаксиса разных языков (56 параметров). Затем, используя тест Мантеля, рассчитали корреляции между каждой парой матриц.

C



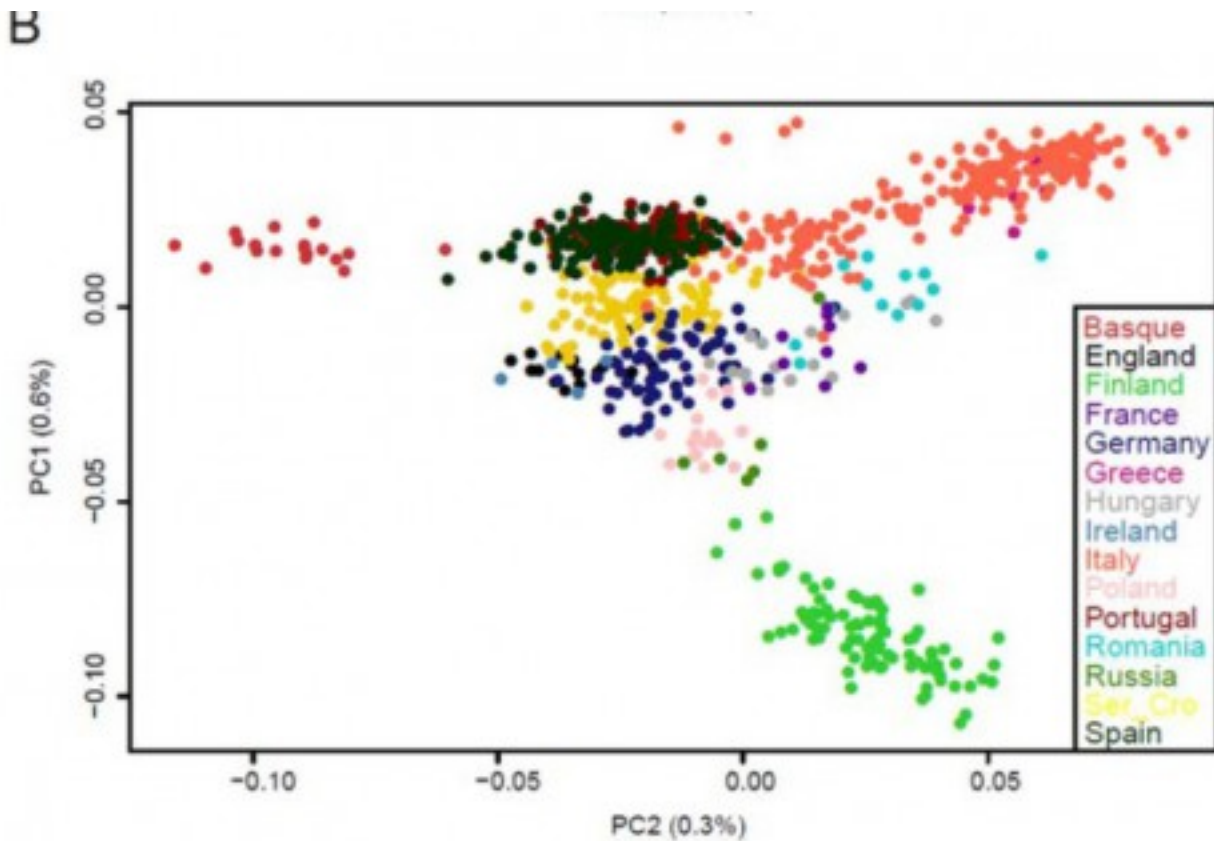
Синтаксическое дерево, построенное для 15 языков (12 индоевропейских и 3 других (басков, венгерский, финский)).





D  
 Результаты метода анализа главных компонент (PCA) по лингвистическим (синтаксическим) данным.

Авторы применили и другой традиционный метод анализа — главных компонент (PCA) к лингвистическим и генетическим данным. Как видно на рисунке PCA по синтаксическим параметрам, от ИЕ языков отделились баски, финны и венгры, а ИЕ языки сгруппировались в кластер, внутри которого выделились три субкластера: романский, германский и славянский. Выпадает из ИЕ кластера греческий язык.

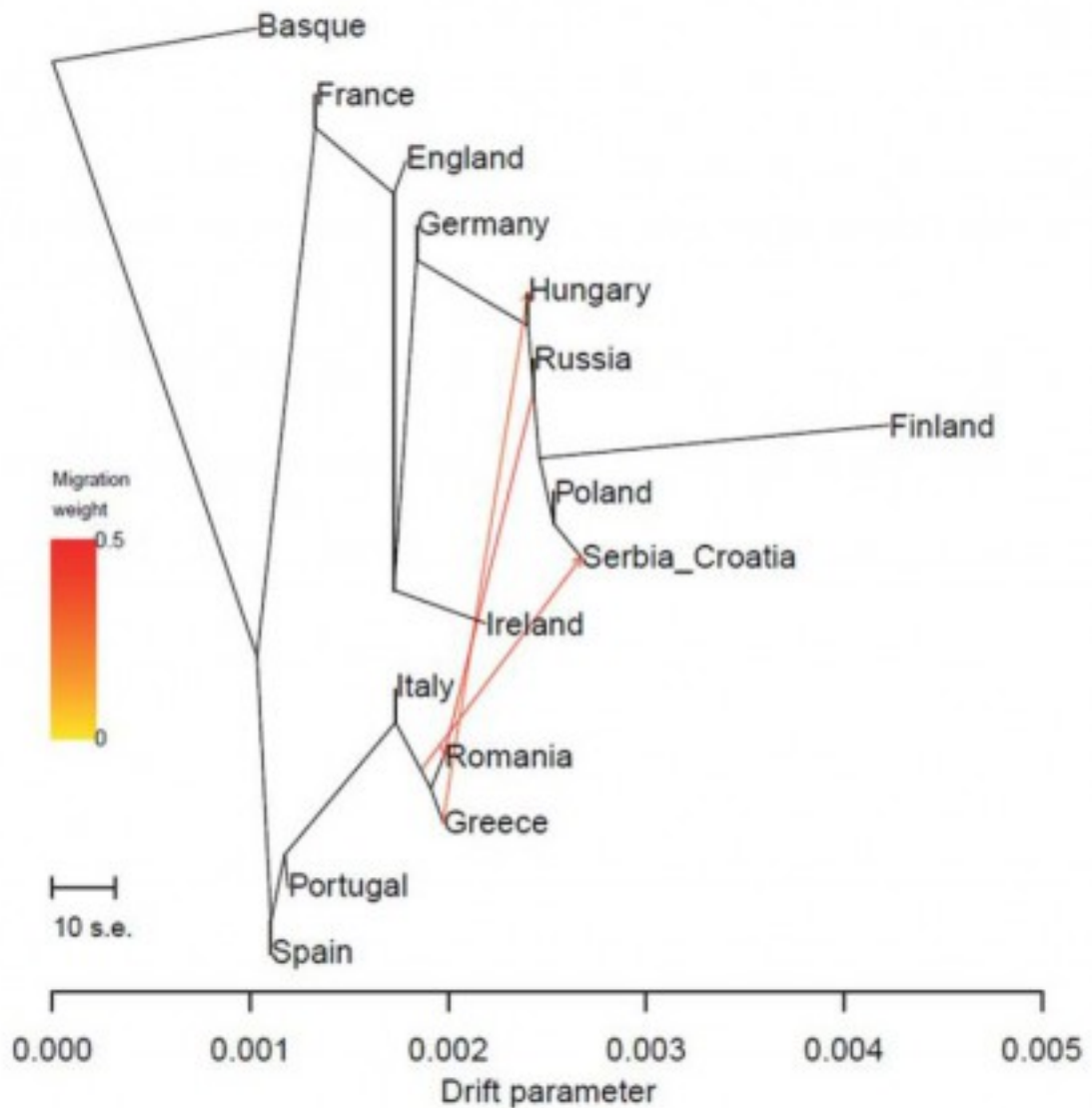


Результаты метода анализа главных компонент (PCA) по индивидуальным геномным данным. Каждая точка соответствует индивидууму из той или иной популяции (принадлежность к популяции обозначена разными цветами указана на шкале справа).

PCA, проведенный по генетическим расстояниям, показал, что финны и баски отличаются от ИЕ-популяций также и генетически. А вот венгры генетически сходны с другими популяциями Центральной Европы, у них наблюдается расхождение между генетикой и языком (по словам специалистов, это является общеизвестным фактом). Авторы указывают на некоторое генетическое сходство венгров с популяциями Северной и Центральной Азии и интерпретируют его как поток генов на запад в средние века в соответствии с известным историческим фактом миграции IX века, когда угорский язык был принесен в Венгрию. Сегодня частота азиатских маркеров у венгров невелика. Анализ древней ДНК X века показал, что в массовых захоронениях сельского населения преобладают европейские гаплотипы (в митохондриальной ДНК), а азиатские гаплотипы достигают высокой частоты у высокостатусной элиты. Это подтверждает старый вывод антропологов, что мигранты из Азии не замещали местную популяцию, а становились социальной элитой.

Исследователи использовали и еще одну программу TreeMix, которая строит деревья не только с вертикальными, но и с горизонтальными связями. Таким образом, учитываются потоки генов, которые между популяциями уже после их разделения из предковых популяций. Обычно традиционно построенные генетические и лингвистические деревья не учитывают этих вторичных миграционных потоков генов. На дереве, построенном по «сетевому» принципу, выявились, в частности, контакты между Венгрией и Грецией, Россией и Румынией и шире — между романскими языками и балканской языковой группой.

A



Дерево, построенное в программе TreeMix. Красными линиями обозначены потоки генов между географически близкими популяциями.

Авторы делают два вывода. Во-первых, для изученной ими панели этносов и SNP маркеров оказалось, что популяции, говорящие на схожих языках, генетически ближе друг к другу, чем географически близкие популяции. Также, по их данным, синтаксис определяет эту близость точнее, чем лексика. Во-вторых, они продемонстрировали потоков генов в пределах ИЕ популяций и между ИЕ и соседними популяциями других языковых семей. Эти результаты, считают авторы, можно использовать для междисциплинарного анализа популяций, относящихся к разным языковым семьям.

#### Источник:

Across Language Families: Genome Diversity Mirrors Linguistic Variation Within Europe

Giuseppe Longobardi, Silvia Ghirotto, Cristina Guardiano, Francesca Tassi, Andrea Benazzo,

Andrea Ceolin, and Guido Barbujani

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 00:00–00 (2015)

аффилиация авторов по ссылке <http://www.ncbi.nlm.nih.gov/pubmed/26059462>

статью можно сказать здесь [Longobardi 2015 Genome Diversity Mirrors Linguistic Variation AJPA](#)

