

# Географическая и лингвистическая структурированность генофонда Европы – закономерности

[Олег Балановский](#)

Фрагмент из книги "Генофонд Европы"

Из всего предыдущего рассмотрения структуры европейского генофонда по Y-хромосоме очевиден высокий уровень межпопуляционных различий и четкая кластеризация популяций. Возникает естественный вопрос, каков основной принцип этой структуры, какими факторами она сформирована? Рассматривая те или иные конкретные закономерности, мы уже не раз видели кластеризацию по географическому принципу. Но видели и кластеризацию по принципу лингвистического сходства, а также – близкую по смыслу к лингвистической – кластеризацию в этнические облака. Поэтому в этом разделе описывается межпопуляционная изменчивость основных лингвистических групп Европы и анализируется роль географического и лингвистического факторов. Это вечный вопрос для генофонда любого региона мира – попробуем решить его на текущем уровне знаний применительно к генофонду Европы и к данным по Y-хромосоме.

## ВНУТРИЭТНИЧЕСКАЯ ДИФФЕРЕНЦИАЦИЯ

Этот вид анализа является дополнением графика многомерного шкалирования на субэтническом уровне, которым оканчивался предыдущий раздел этой главы. Но график основной упор делает на соотношениях разных популяций друг с другом (в том числе, становится видно, существуют или не существуют этнические облака для каждого народа). А о степени сходства разных популяций одного народа друг с другом по графику можно судить лишь по размеру этнического облака. Расчет же межпопуляционной изменчивости для популяций каждого народа, напротив, ничего не говорит о том, как генофонды разных народов соотносятся друг с другом. Но зато он дает количественный и точный ответ на вопрос, насколько сильно разные популяции одного народа отличаются друг от друга генетически, то есть оценивает параметр, который мы называем «этническая гетерогенность».

Такое исследование я тоже проводил неоднократно. В таблице 2.6 приведены результаты 2007 года для Европы (для Сибири – 2010 года), а в таблице 2.7 – результаты 2015 года по расширенному массиву данных. В таблице 2.6, кроме изменчивости популяций внутри отдельных народов (внутриэтническая дифференциация), рассчитана также изменчивость среднеэтнических частот разных народов (межэтническая дифференциация).

Для максимальной объективности, анализ 2007-2010 годов был проведен тремя методами, с использованием трех различных показателей межпопуляционной изменчивости:  $G_{ST}$ , AMOVA (оценки  $F_{ST}$ ) и средние попарные генетические расстояния. Таблица 2.6. показывает, что в пределах Европы наибольшая гетерогенность характерна для финнов, хорватов, русских и итальянцев; шведы и немцы демонстрируют умеренную гетерогенность; другие народы (греки, турки, поляки, украинцы) характеризуются сниженной гетерогенностью. Высокая этническая гетерогенность турок объясняется высоким филогенетическим разрешением, с которым изучены эти популяции (26 гаплогрупп). Если эти гаплогруппы объединить до уровня, с которым изучены другие народы (8 гаплогрупп), то средние расстояния между популяциями турок падают с 16,5 до 3,5, а  $G_{ST}$  оказывается менее чувствительным к набору гаплогрупп, снижаясь с 1,3 до 0,8.

Массив данных 2015 года включает большее число народов, и он выровнен по числу гаплогрупп, что увеличивает сопоставимость результатов по разным народам (табл. 2.7). Закономерности похожи – финны, хорваты и русские вновь в числе лидеров (средние расстояния 0,12-0,15), хотя для русских это определяется дифференциацией на северных и южных, а различия внутри каждой из этих групп не так уж велики. Этим лидерам лишь немного уступают шведы (итальянцы и немцы в таблице 2015 года отсутствуют, поскольку по их региональным популяциям нет надежных данных по панели 15 гаплогрупп). Для турок, поляков, и украинцев новыми данными подтверждена сниженная этническая гетерогенность (0,02 – 0,06). Из числа народов, проанализированных и в 2007, и в 2015 годах, различаются только результаты по грекам: им по новым данным оказывается свойственна не сниженная, а напротив, повышенная этническая гетерогенность; хотя это отчасти объясняется иным набором проанализированных популяций, а отчасти большим числом гаплогрупп, все же такие различия между старым и новым анализами заставляют быть осторожными в оценке этнической гетерогенности греков и подождать появления дополнительных данных по этому народу. Что касается народов, появившихся только в анализе 2015 года, то заслуживает упоминания неожиданно высокая этническая гетерогенность болгар (это мы видели и на графике шкалирования). А максимальная этническая гетерогенность, далеко оставляющая за собой гетерогенность иных народов Европы, оказывается свойственной башкирам. В разных местах этой книги по самым разным данным мы наталкиваемся на этот результат, который не является неожиданностью ни для антропологов (поскольку башкиры весьма гетерогенны в расовом отношении), ни для этнографов (поскольку башкиры состоят из многих родов, контрастных по своему происхождению). Остальные народы,

добавившиеся в анализ 2015 года (белорусы, бельгийцы, поляки-кашубы, литовцы, сербы, чехи) характеризуются сниженной гетерогенностью (средние расстояния 0,02-0,05) и лишь албанцы более гетерогенны (0,08).

В среднем, по результатам 2007 года, гетерогенность типичного европейского народа (его внутриэтническая дифференциация) составляет (табл. 2.6) около 0.03:  $G_{ST}=0.029$ ,  $F_{ST}=0.027$ . Различия уже не внутри одного народа, а между народами (межэтническая дифференциация) оказываются намного выше:  $G_{ST}=0.149$ ,  $F_{ST}=0.166$  (табл. 2.6). Пятикратное превышение внутриэтнической изменчивости над межэтнической является количественным выражением того, что подразделение европейских популяций по Y-хромосоме в соответствии с их этнической принадлежностью высоко информативно и объясняет львиную долю всей межпопуляционной изменчивости в генофонде Европы.

Таблица 2.6.. Закономерности изменчивости Y-хромосомы на внутриэтническом и межэтническом уровнях популяционной иерархии (данные 2007 и 2010 года).

	$N_{POP}$	N	$N_{HG}$	Средние генетические расстояния	$G_{ST}(G_{ST}*10^2)$	AMOVA ( $F_{ST} * 10^2$ )
<b>ЕВРОПА</b>						
Финны	5	107	6	0.13	8.2	8.1
Хорваты	5	100	8	0.18	8.0	7.4
Русские	10	123	8	0.14	5.2	4.5
Итальянцы	5	105	8	0.13	3.9	5.0
Шведы	4	76	12	0.13	2.7	1.4
Немцы	11	110	9	0.07	2.1	1.5
Греки	4	91	8	0.10	1.4	1.1
Турки	5	87	26(8)	0.17 (0.04)	1.3(0.8)	1.0
Поляки	8	114	7	0.02	1.1	0.3
Украинцы	4	102	10	0.02	0.9	0.4
<b>Межэтнически</b>	<b>10</b>		<b>6</b>		<b>14.9</b>	<b>16.6</b>
<b>ый уровень</b>						
<b>Внутриэтнически</b>	<b>6</b>		<b>6</b>		<b>2.9</b>	<b>2.7</b>
<b>кий уровень</b>						
<b>Внутрипопуляц</b>	<b>61</b>		<b>6</b>		<b>82.2</b>	<b>80.7</b>
<b>ионный</b>						
<b>уровень</b>						
<b>СИБИРЬ</b>						
Хакасы	4	62	30	0.85	18.2	30.3
Северные алтайцы	3	67	30	0.80	15.1	20.0
Шорцы	2	70	30	0.09	4.9	4.7
Южные алтайцы	2	104	30	0.07	3.9	3.2
Киргизы	2	123	30	0.15	2.9	5.1
<b>Межэтнически</b>	<b>5</b>		<b>30</b>		-	<b>5.5</b>
<b>ый уровень</b>						
<b>Внутриэтнически</b>	<b>3</b>		<b>30</b>		-	<b>14.7</b>
<b>кий уровень</b>						
<b>Внутрипопуляц</b>	<b>13</b>		<b>30</b>		-	<b>79.8</b>
<b>ионный</b>						
<b>уровень</b>						

Примечания:

$N_{POP}$  – число региональных популяций, изученных для данного народа;

N – средний размер выборки региональных популяций;

$N_{HG}$  – число основных гаплогрупп, включенных в анализ;

Этносы расположены в пределах региона по убыванию значений  $G_{ST}$ .

Аналогичный анализ этнической гетерогенности проведен нами для народов Южной Сибири (табл. 2.6). Суммарные межпопуляционные различия в этом регионе (0,20) оказались практически такими же, как и для популяций Европы (0,19). Но основная часть изменчивости приходится на изменчивость внутри этнотерриториальных объединений (0,15), а на различия между ними остается только 0,06. Такой результат объясняется тем, что этнотерриториальные объединения, выделяемые традиционно (хакасы, алтайцы), включают ряд малых народностей, генофонды которых значительно различаются между собой [Балаганская и др., 2011].

Таблица 2.7. Закономерности изменчивости Y-хромосомы на внутриэтническом уровне (данные 2015 года).

Этнос	Средние генетические расстояния	$N_{POP}$	N	$N_{HG}$
Башкиры	0,425	5	86	15
Греки	0,197	5	195	15
Финны	0,155	7	82	15
Хорваты	0,146	3	89	15
Болгары	0,136	9	82	15
Русские	0,129	20	104	15
<i>Русские северные</i>	<i>0,065</i>	<i>4</i>	<i>95</i>	<i>15</i>
<i>Русские центральные и южные</i>	<i>0,047</i>	<i>16</i>	<i>106</i>	<i>15</i>
Шведы	0,110	7	53	15
Албанцы	0,082	3	110	15
Турки	0,064	6	69	15
Литовцы	0,046	6	50	15
Белорусы	0,040	8	96	15
Чехи	0,037	5	51	15
Украинцы	0,032	13	92	15
Кашубы	0,030	3	68	15
Сербы	0,020	4	100	15
Бельгийцы	0,017	5	95	15
Поляки	0,016	3	234	15

*Примечания:*

Для получение более устойчивых оценок в анализ включены народы из массива «Субэтнический», представленные как минимум тремя популяциями, соответственно, не включены народы, представленные только двумя – боснийцы, коми, македонцы, мордва (мокша и эрзя).

$N_{POP}$  – число региональных популяций, изученных для данного народа;

N – средний размер выборки региональных популяций;

$N_{HG}$  – число гаплогрупп, включенных в анализ;

## ДИФФЕРЕНЦИАЦИЯ ВНУТРИ ЛИНГВИСТИЧЕСКИХ ГРУПП

Рассмотрим теперь закономерности межпопуляционной изменчивости на более высоком иерархическом уровне – не отдельных народов, а их групп. Народы можно классифицировать по-разному, но основной является лингвистическая классификация их языков. Конечно, для оценки роли не только лингвистического, но и географического фактора было бы хорошо классифицировать народы и по регионам их проживания, и затем сравнить генетическую дифференциацию при

лингвистической и географической классификациях. Такой анализ мы провели для Кавказа (глава 7), где географическая классификация достаточно очевидна. Не то для Европы – в ее пределах регионы можно выделять по-разному, и результаты слишком зависели бы от моего субъективного взгляда. Лингвистическая же классификация достаточно устойчива, почти общепризнана и уж в любом случае от меня не зависит. Поэтому в таблицах 2.8. и 2.9 приведена генетическая гетерогенность основных лингвистических групп Европы, рассчитанная как межпопуляционная изменчивость между всеми народами, относящимися к этой группе. (Отметим, что этот анализ, таким образом, основывался на среднеэтнических, а не популяционных частотах: частоты гаплогрупп во всех популяциях одного народа были усреднены, такая процедура проведена для всех народов, и сравнивались полученные среднеэтнические частоты).

При анализе по панели 17 гаплогрупп (табл. 2.8.) наиболее гетерогенными оказываются славянская, тюркская и финно-угорская группы (средние генетические расстояния между народами в пределах каждой из этих групп составляют около 0,5). Этот результат по финно-уграм и тюркам хорошо перекликается с аналогичным анализом по мтДНК, о котором речь пойдет в следующей главе 3, а причины гетерогенности славян разбираются в главе 6. Романская группа уже заметно менее гетерогенна (0,3), германская еще менее (0,1), а балтская, абхазо-адыгская и нахско-дагестанская почти гомогенны (0,01 – 0,05). Надо, впрочем, оговориться, что эти выводы зависят от набора гаплогрупп – например, доминирующие на Северном Кавказе гаплогруппы в этом наборе гаплогрупп (оптимизированном под остальную Европу) представлены только гаплогруппами **G и J** безо всякого их подразделения, тогда как с учетом их подразделения на субтипы гомогенность обеих кавказских лингвистических групп значительно бы выросла.

В наборе 40 гаплогрупп кавказские варианты опять не подразделены (это повлекло бы за собой сокращение числа анализируемых популяций по остальной Европе, не изученных по этим субвариантам), поэтому дифференциация абхазо-адыгской и нахско-дагестанской групп остается минимальной. По балтской группе по этой панели данных на момент написания книги недостаточно, гетерогенность германской группы несколько возрастает (0,2), славянской группы даже немного снижается (0,4), зато гетерогенность тюркской группы увеличивается в полтора раза (до 0,7), финно-угорской – в два (до 1,0), а романской – в три (до 0,9). Хотя очевидна большая зависимость результатов как от числа гаплогрупп, так и от степени их подразделенности, все же отметим, что при наилучшем возможном на данный момент наборе данных наиболее гетерогенной является не германская группа (хотя лучше всего подразделена доминирующая у нее гаплогруппа **R1b**), а финно-угорская группа – по всей видимости, за счет преобладающих у нее процессов дрейфа генов вследствие изоляции.

Таблица 2.8. Гетерогенность лингвистических групп Европы по панели 17 гаплогрупп.

Лингвистическая группа	Число популяций	Средние генетические расстояния	Расстояния после коррекции*	Исключенные обособленные популяции
Абхазо-адыгская	4	0,048	-	-
Балтская	2	0,009	-	-
Германская	3	0,095	-	-
Нахско-дагестанская	5	0,009	-	-
Славянская	15	0,505	0,381	Северные русские и сорбы: ген. расстояния в 1,7 и 1,9 раза больше средней
Романская	8	0,312	0,214	Румыны: ген. расстояния в 2,5 раза больше средней
Тюркская	10	0,471	-	-
Финно-угорская	8	0,470	-	-

\*В некоторых случаях одна-две генетически обособленные популяции за счет больших генетических расстояний до остальных популяций завышали уровень гетерогенности всей группы. Поэтому была проведена коррекция: повторный расчет при исключении из расчета тех популяций, среднее расстояние до которых значительно превышало аналогичный показатель для остальных популяций данной группы.

Таблица 2.9. Гетерогенность лингвистических групп Европы по панели 40 гаплогрупп.

Лингвистическая группа	Число популяций	Средние генетические расстояния	Расстояния после коррекции	Исключенные обособленные популяции
Абхазо-адыгская	3	0,058	-	-
Германская	3	0,230	-	-
Нахско-дагестанская	4	0,012	-	-
Славянская	10	0,444	0,374	Чехи: ген. расстояния в 1,6 раза больше средней
Романская	4	0,902	0,586	Румыны: ген. расстояния в 1,4 раза больше средней
Тюркская	7	0,751	0,594	Казанские татары: ген. расстояния в 1,4 раза больше средней
Финно-угорская	4	0,968	-	-

### РОЛЬ ЛИНГВИСТИЧЕСКОГО И ГЕОГРАФИЧЕСКОГО ФАКТОРОВ.

Напомню, что первое обобщающее исследование Y-хромосомного генофонда Европы [Rosser et al., 2000] называлось «Изменчивость Y-хромосомы в Европе является клинальной и задается, прежде всего, географическим, а не лингвистическим фактором». Этот вывод был основан на полученной Россером с соавторами достоверной корреляции матрицы генетических расстояний между популяциями Европы с матрицей географических расстояний между теми же популяциями (коэффициент корреляции 0,4), при недостоверной корреляции между генетическими и лингвистическими расстояниями [Rosser et al., 2000].

Поэтому наш коллектив тоже провел анализ корреляции между матрицами генетических, географических и лингвистических расстояний, включив в него и многочисленные новые данные, опубликованные после выхода в свет упомянутой статьи [Rosser et al., 2000]. Анализ проводился по точным лексико-статистическим данным и только в пределах индоевропейских народов (составляющих большинство в Европе), поскольку расстояния между разными лингвистическими семьями не поддаются точному измерению.

При использовании массива данных по 17 гаплогруппам (табл. 2.10) корреляция генетических расстояний с географическими (0,56) лишь немного больше корреляции с лингвистикой (0,45). При использовании частных корреляций величины уменьшаются, но разрыв становится рельефнее: 0,46 с географией и только 0,30 с лингвистикой.

При использовании массива данных по 40 гаплогруппам (подразделяются **R1a**, **R1b**, **N1c** – табл. 2.4) приходится исключить из анализа некоторые народы, не изученные по этой панели. Но это полностью оправдывается тем, что резко возрастают коэффициенты корреляции и с географией, и с лингвистикой. То есть подразделение гаплогрупп не только лучше выявляет их географическую приуроченность на картах, но и использование этих новых субгаплогрупп позволяет лучше увидеть характер структурированности европейского генофонда, его связь с географическим, и с лингвистическим факторами. Парная корреляция генетических расстояний (табл. 2.11) с географическими составляет 0,54, а с лингвистическими даже выше (0,74). Но анализ частных корреляций вновь расставляет все по местам: связь с лингвистикой хотя и остается существенной (0,42), но значительно уступает связи с географией: 0,69. Отметим, что этот полученный нами коэффициент частной корреляции между матрицами генетических и географических расстояний практически достигает 0,7, что более, чем в полтора раза превышает величину 0,4, полученную ранее Россером с соавторами. Это убедительно доказывает, что по мере накопления данных по изменчивости Y-хромосомы географическая структурированность генофонда Европы становится все более и более явной. В целом же, оба фактора – и география, и лингвистика – играют каждый свою роль, совместно формируя генетический ландшафт Европы.

Таблица 2.10. Роль лингвистического и географического факторов в структурировании Y-хромосомного генофонда Европы (первый вариант — 17 гаплогрупп, 26 народов).

<b>Вид корреляции</b>	<b>Параметры корреляции</b>	<b>Коэффициент корреляции</b>
Парная корреляция	<b>Генетика и лингвистика</b>	<b>0,45</b>
Парная корреляция	<b>Генетика и география</b>	<b>0,56</b>
Частная корреляция	<b>Генетика и лингвистика</b> (география остается постоянной)	<b>0,30</b>
Частная корреляция	<b>Генетика и география</b> (лингвистика остается постоянной)	<b>0,46</b>

Таблица 2.11. Роль лингвистического и географического факторов в структурировании Y-хромосомного генофонда Европы (второй вариант — 40 гаплогрупп, 18 народов).

<b>Вид корреляции</b>	<b>Параметры корреляции</b>	<b>Коэффициент корреляции</b>
Парная корреляция	<b>Генетика и лингвистика</b>	<b>0,54</b>
Парная корреляция	<b>Генетика и география</b>	<b>0,74</b>
Частная корреляция	<b>Генетика и лингвистика</b> (география остается постоянной)	<b>0,42</b>
Частная корреляция	<b>Генетика и география</b> (лингвистика остается постоянной)	<b>0,69</b>