Структура генофонда Европы в зеркале полногеномных маркеров-2

Олег Балановский

Фрагмент из книги "Генофонд Европы"

Продолжаем публиковать фрагмент из книги О.П.Балановского «Генофонд Европы», посвященный анализу полногеномных маркеров ДНК — самых современных и наиболее информативных для анализа генофонда. В этой части описан метод анализа предковых компонентов и его отображение на геногеографических картах народов Европы.

4.3. Анализ предковых компонентов (ADMIXTURE)

Я не сам по себе – я лишь тень отдаленного предка.
Постоянно мне кажется – с ним я один человек.
Я на дереве общем всего лишь случайная ветка,
От которой появится новый зеленый побег.

Миллионы за мною и передо мной миллионы.
Жар свирепой пустыни, зола от костров на снегу.
Как измученный зек, из невидимой этой колонны
Я и рад убежать бы, да только, увы, не могу.

А. Городницкий

ПРИНЦИП МЕТОДА

Чтобы сравнивать генофонды по широкогеномным данным, часто используется анализ «предковых компонентов» (не путать с «главными компонентами»!). Этот метод реализован в программе STRUCTURE и в ее усовершенствованном варианте — программе ADMIXTURE. Метод требует только двух вещей: широкогеномных данных по большому числу популяций и задания числа гипотетических предковых популяций, из которых сформировались все эти современные популяции. Тогда программа вычисляет, каковы должны быть частоты каждого аллеля в каждой из предковых популяций, и рассчитывает для каждой современной популяции доли этих предковых популяций в ее генофонде. Эти разные предковые компоненты принято отображать разными цветами, и каждая популяция характеризуется своим спектром этих цветов. Понятно, что такая модель условна — в реальности вряд ли современные генофонды сформировались в результате смешения фиксированного числа предковых популяций. Но выявляемые предковые компоненты, как мы сейчас увидим, имеют реальный смысл.

ВЫБОР ЧИСЛА КОМПОНЕНТОВ (k)

Основной проблемой является выбор числа предковых компонентов, который должен задать пользователь. Это число принято обозначать **k**. Хотя и существуют приемы определения оптимального **k**, они не всегда срабатывают, и выбор обычно остается на усмотрение исследователя. Ситуацию облегчает то, что выделяемые компоненты весьма устойчивы — при анализе самых разных массивов данных часто выявляются одни и те же предковые компоненты. Так, на низких **k** выделяются континент-специфичные генофонды, а при их увеличении выделяются давно известные из антропологических и геногеографических работ ареалы сходных популяций внутри континентов.

АНАЛИЗ В МИРОВОМ МАСШТАБЕ: ГЕНЕТИЧЕСКАЯ РЕАЛЬНОСТЬ РАС

Наиболее репрезентативно исследование [Pagani et al., submitted], основанное на секвенировании полных геномов, и поэтому набор использованных SNP-маркеров не смещен в сторону европейских популяций, как это свойственно большинству широкогеномных панелей. В этой работе к тому же проанализировано большое число популяций изо всех основных регионов мира.

В результате при **k** =2 один компонент включил все популяции Африки и Западной Евразии, а другой – все популяции

Восточной Евразии и Америки.

При увеличении до **k** =3 западно-евразийские и африканские популяции разделяются на отдельные компоненты.

При **k** =4 выделяется отдельный компонент, доминирующий в популяциях американских индейцев, но также составляющий около половины генофонда и в Сибири (остальная часть сибирского генофонда приходится на восточно-евразийский и в меньшей мере на западно-евразийский компоненты).

Но уже при **k**=5 население Сибири приобретает свой собственный компонент, вбирающий основную часть сибирского генофонда.

При **k**=6 население Сахула (Австралия + Папуа Новая Гвинея, правда, представленное в этом исследовании лишь папуасскими популяциями), приобретает свой собственный компонент. До этого оно было частью восточноевразийского компонента, а после отделения сибирского – частью оставшегося восточноазиатского компонента.

Эта картина удивительно схожа (чтобы не сказать совпадает) с традиционной расовой классификацией. Ведь при k=3 мы, по сути, видим привычное со школы деление на три большие расы — негроидов, европеоидов и монголоидов. А если физическим антропологам сказать, что требуется разделить все антропологическое многообразие на 4 группы, тогда большинство из них выделит американоидную расу из состава монголоидной. Небольшое отличие от антропологических классификаций состоит в том, что австралоидная раса приобретает свой собственный компонент лишь начиная с k=6, тогда как в традиционных расовых классификациях австралоидная раса имеет чуть более высокий иерархический статус. Поразительно и печально, что при таком полном подтверждении традиционных расовых классификаций новейшими генетическими данными, до сих пор широко распространено мнение, что генетика «доказала» отсутствие рас. Впрочем, об этом уже давно писалось [Раса: миф или реальность, 1998].

Все эти расы, выделившиеся при **k**=6 (негроидная, европеоидная, американдская, северомонголоидная, южномонголоидная, австралоидная), продолжают существовать и при более высоких k, лишь при каждом его повышении какая-то из них дробится на два компонента.

Так, при **k**=7 из европеоидного (западноевразийского) компонента выделяется южноазиатстко-ближневосточный, который при **k**=8 разделяется на ближневосточный и южноазиатский (индийский), при **k**=9 выделяется компонент северо-восточной Сибири, при **k**=10 начинает дробиться негроидная раса и так далее.

АНАЛИЗ В ЕВРОПЕЙСКОМ МАСШТАБЕ

Но вернемся к генофонду Европы. В ходе анализа генофонда балто-славянских народов (глава 6) был проведен анализ ADMIXTURE, включивший не только балтов и славян, но и остальные изученные по широкогеномным маркерам популяции Европы, и, в несколько меньшей мере, популяции других регионов мира. Это позволяет нам воспользоваться этими результатами для анализа в европейском масштабе.

Приведем карты распространения предковых компонентов, полученных при контрастных величинах **k**, равных 6 и 18. Первый вариант близок к разумному минимуму, поскольку делит все генофонды Старого Света на шесть основных доменов: африканский домен, два домена в северной половине Евразии (европейский и сибирский), и три домена в южной половине Евразии (средиземноморский, южноазиатский и восточноазиатский). Это те регионы мира, которые выделяются и географически, и известны из предшествовавших геногеографических работ, что подтверждает осмысленность результатов, получаемых методом ADMIXTURE. Второй вариант (**k**=18) близок к максимуму, поскольку при таких высоких **k** многие популяции приобретают «собственные» компоненты, и картина становится чрезвычайно дробной.

Чтобы различать предковые компоненты, выделенные при **k**=6 и **k**=18, будем далее обозначать их в первом случае (**k**=6) как 6.1, 6.2, … 6.5, 6.6; а во втором случае (**k**=18) как 18.1, 18.2, …. 18.17, 18.18.

РАСПРОСТРАНЕНИЕ ПРЕДКОВЫХ КОМПОНЕНТОВ ПРИ K=6

На рисунке 4.4 картографировано распространение по Европе и смежным регионам всех шести предковых компонентов, выделяемых при **k**=6. Поскольку они выделены в масштабе мира, максимумы только двух из них приходятся на изучаемый регион – это североевропейский и средиземноморский компоненты. Геногеографические карты всех шести основных компонентов, в разной мере созидающих генофонд Европы, позволяют объективно оценить и степень влияния, и регионы их

влияния (рис. 4.4. А-Е).

Североевропейский компонент (6.1, рис. 4.4.А) охватывает всю Европу, но максимума достигает на территориях вокруг Балтийского моря, где составляет более 90% генофонда. Паттерн его распространения поразительно сходен – почти совпадает – с распространением светлой пигментации волос и глаз. Это ведущий антропологический признак выделения североевропеоидной малой расы, и этот генетический компонент, по всей видимости, совпадает с антропологическим компонентом, давно уже выделенным в физической антропологии.

Средиземноморский компонент (6.2, рис. 4.4.Б) составляет более трети генофонда на европейском побережье и более половины генофонда – на азиатском и африканском побережьях Средиземного моря.

В Европе доля североевропейского компонента снижается к югу, а доля средиземноморского навстречу ему – к северу. Но в целом североевропейский компонент преобладает по всей Европе: даже на южных полуостровах Европы он составляет большую долю генофонда, чем средиземноморский предковый компонент. Центр тяжести средиземноморского компонента приходится на Юго-Западную Азию. Его можно сопоставить с южноевропеоидной (средиземноморской) малой расой. Основное отличие от нее кроется за пределами картографированного ареала – ареал средиземноморской малой расы включает и Индостан, тогда как генетический предковый компонент в Индостане свой, равноправный средиземноморскому.

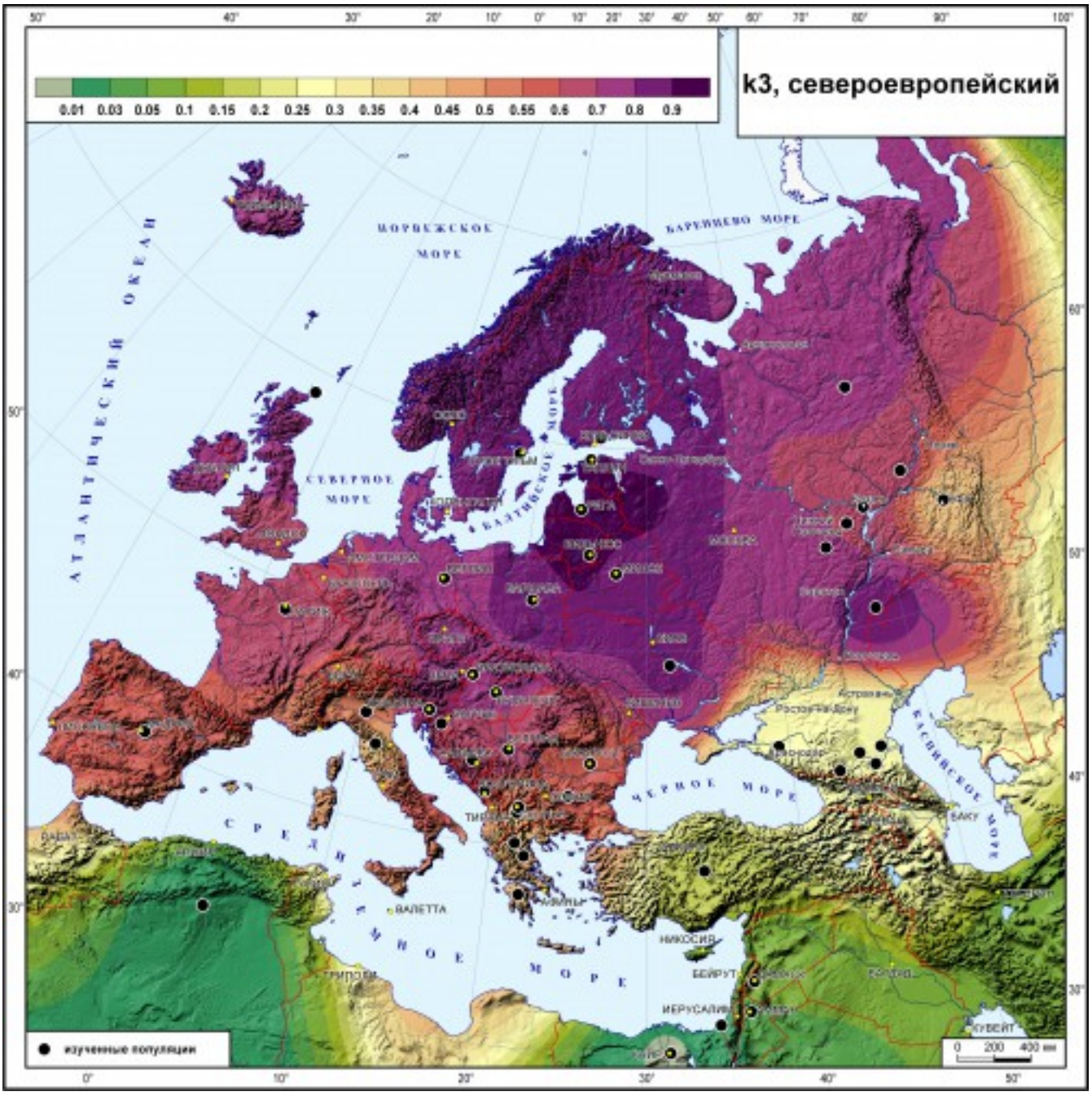


Рис. 4.4 А. Распространение предковых компонентов ADMIXTURE при k=6. Североевропейский компонент (6.1).

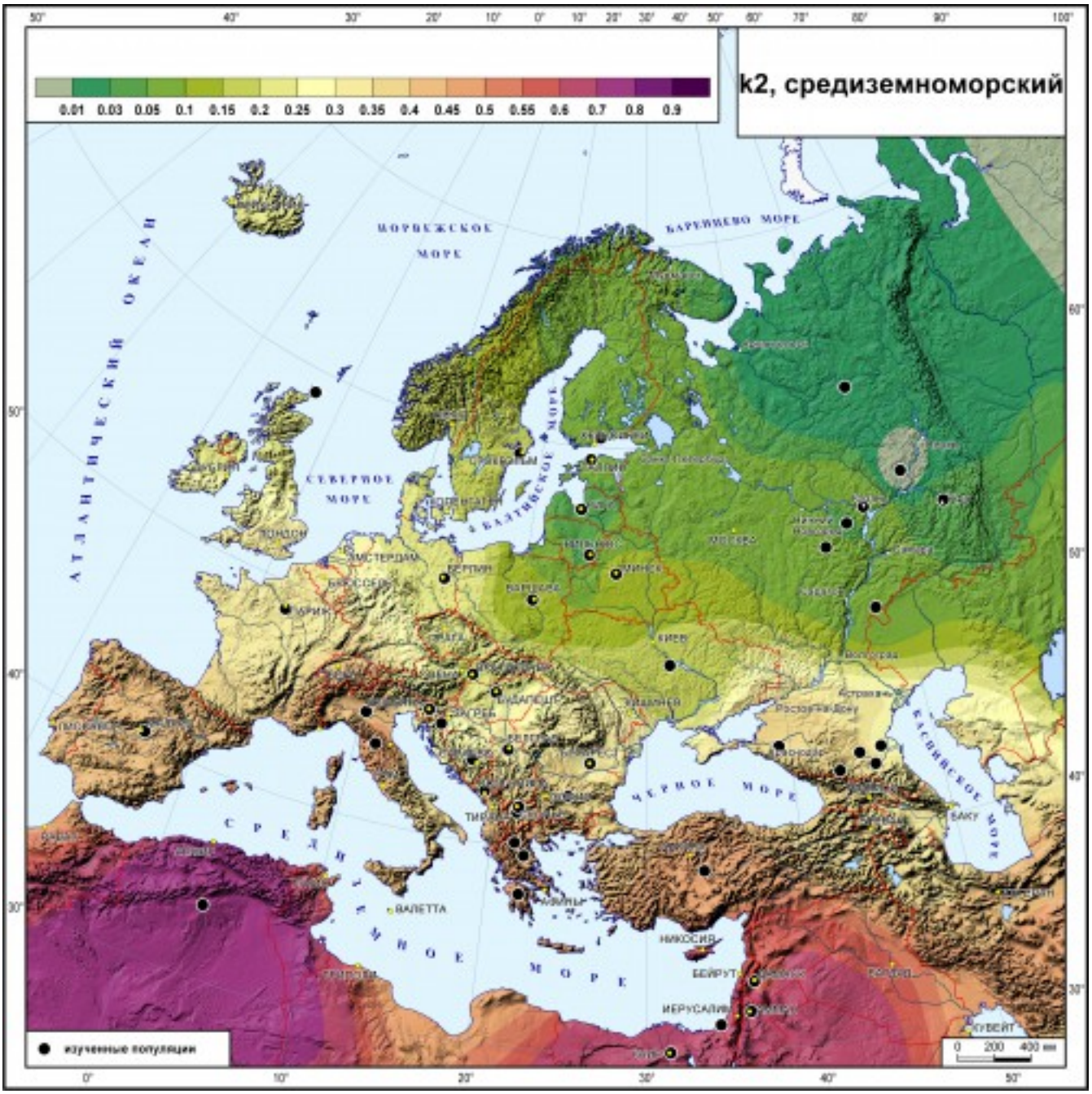


Рис. 4.4 Б. Распространение предковых компонентов ADMIXTURE при k=6. Средиземноморский компонент (6.2).

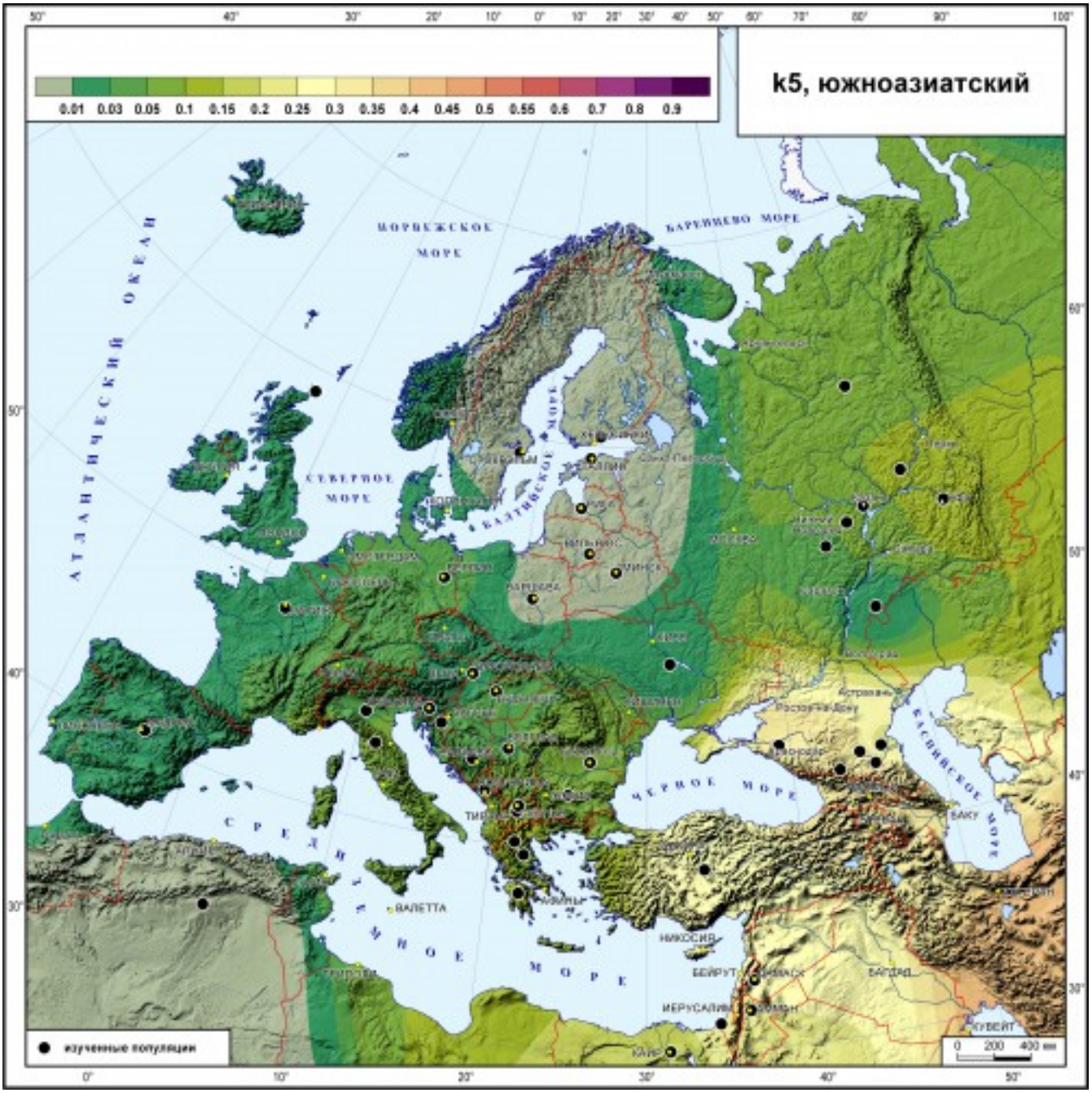


Рис. 4.4 В. Распространение предковых компонентов ADMIXTURE при k=6. В. Южноазиатский компонент (6.3).

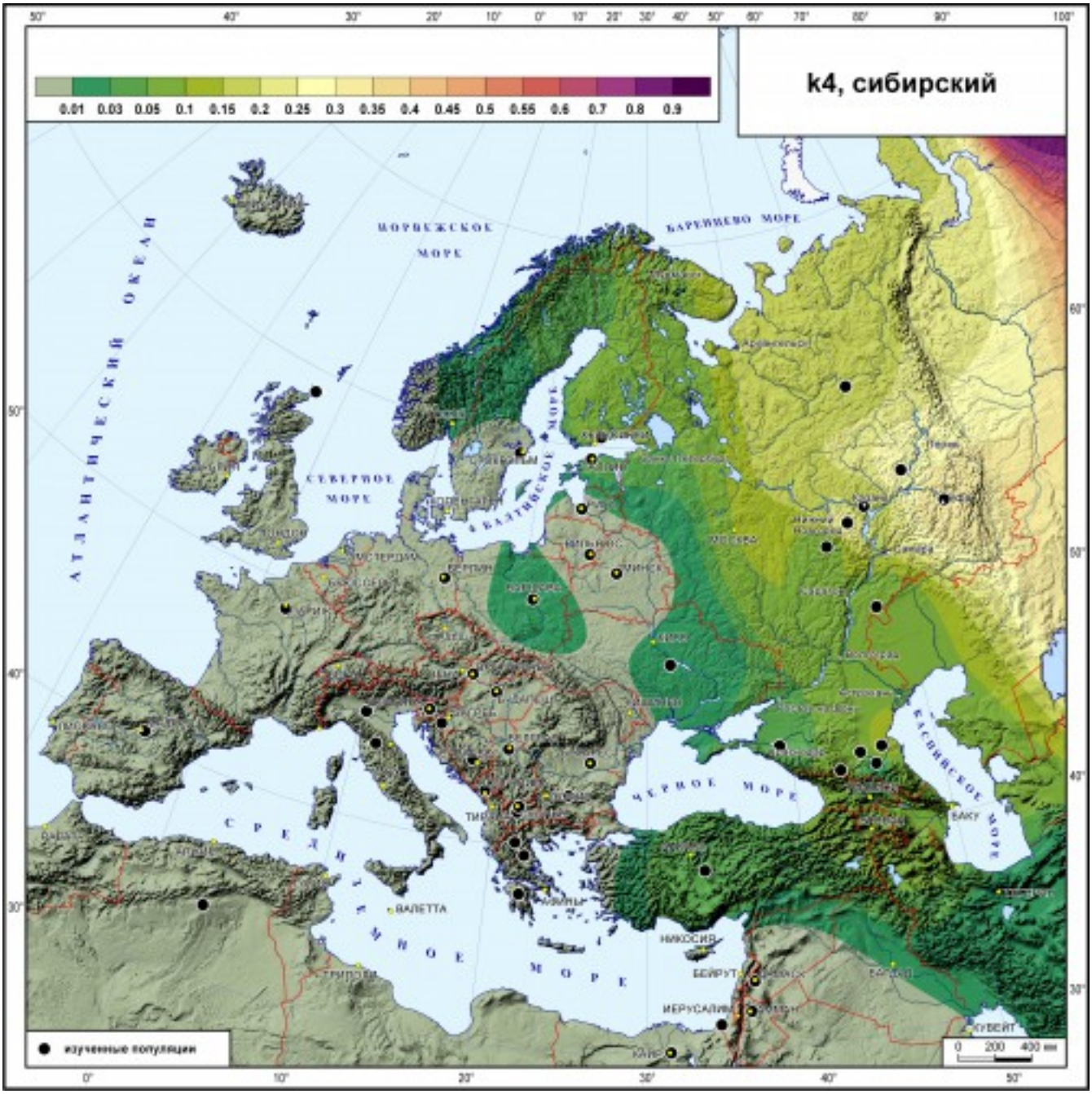


Рис. 4.4 Г. Распространение предковых компонентов ADMIXTURE при k=6. Сибирский компонент (6.4).

Впрочем, и средиземноморский компонент заходит в Индостан, и индийский (южноазиатский) компонент заходит в Средиземноморье.

Южноазиатский компонент (6.3, рис. 4.4.В) в пределах Европы наиболее выражен на Северном Кавказе, где составляет около трети генофонда. Примерно так же он выражен и по всей Юго-Западной Азии. Заметную долю генофонда (10-15%) он составляет на Балканском и Аппенинском полуостровах и в южных частях Центральной Европы, но далее к северу не распространяется, и вокруг Балтийского моря почти полностью отсутствует.

Сибирский компонент (6.4, рис. 4.4.Г). Примерно с той же степенью влияния, как и южноазиатский компонент, в Европе представлен и сибирский. На отдельных территориях (на южном Урале) он составляет более четверти генофонда, но чуть западнее у народов Поволжья частота его уже снижается, далее к западу на основной части Восточно-Европейской равнины он составляет совсем небольшую часть генофонда, а еще западнее – в Центральной и Западной Европе – практически отсутствует.

Что касается двух оставшихся компонентов мирового генофонда, то их доля в европейском генофонде ничтожна.

Восточноазиатский компонент (6.5, рис. 4.4.Е) составляет от 1 до 5% процентов у отдельных народов Урала, Поволжья и Восточного Кавказа, но отсутствует на всей остальной территории Европы.

Африканский компонент (6.6, рис. 4.4.Д) сходным образом составляет пару процентов генофонда Пиренейского полуострова, и отсутствует на остальной территории Европы.

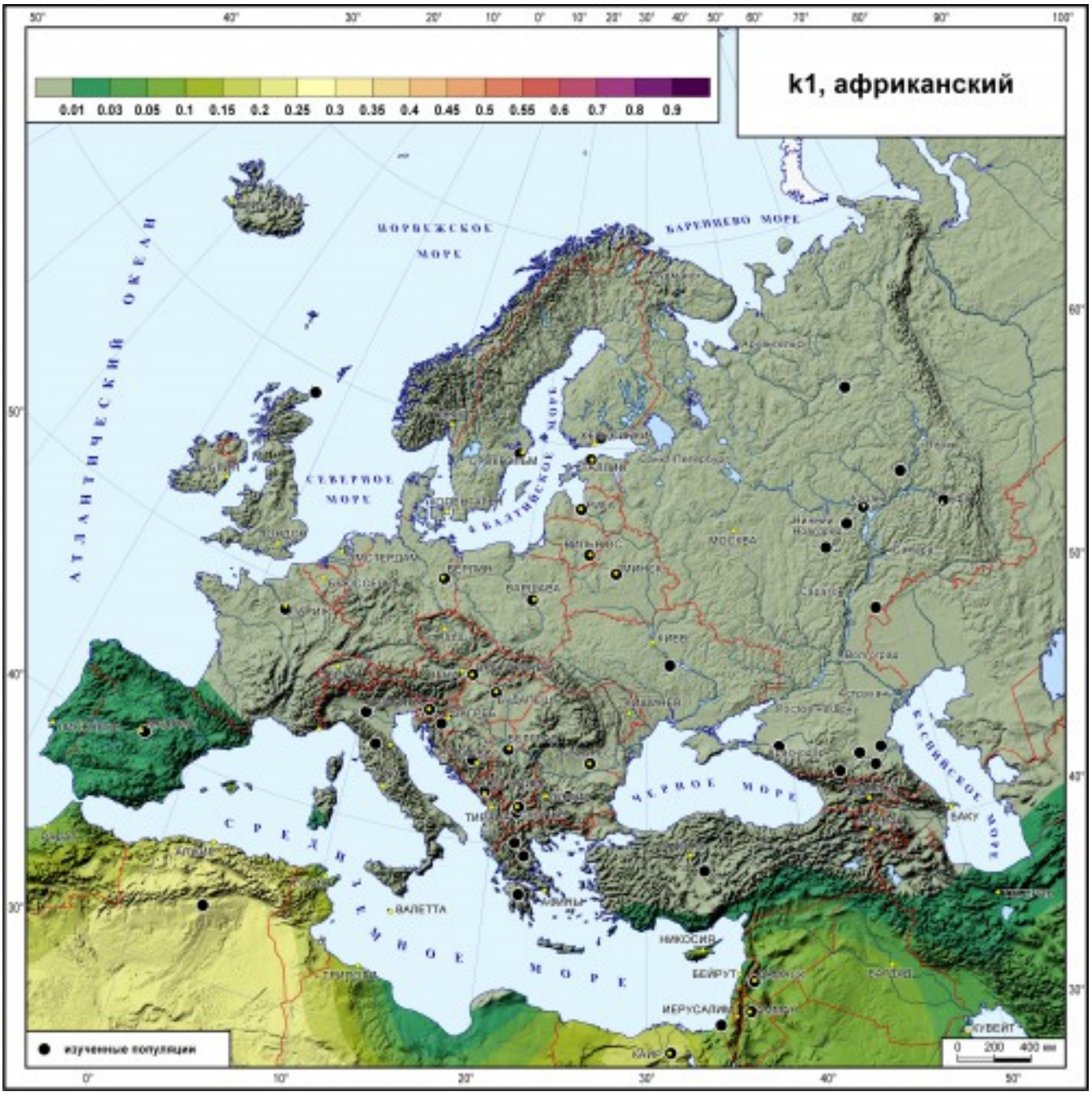


Рис. 4.4. Д. Распространение предковых компонентов ADMIXTURE при k=6. Африканский компонент (6.6).

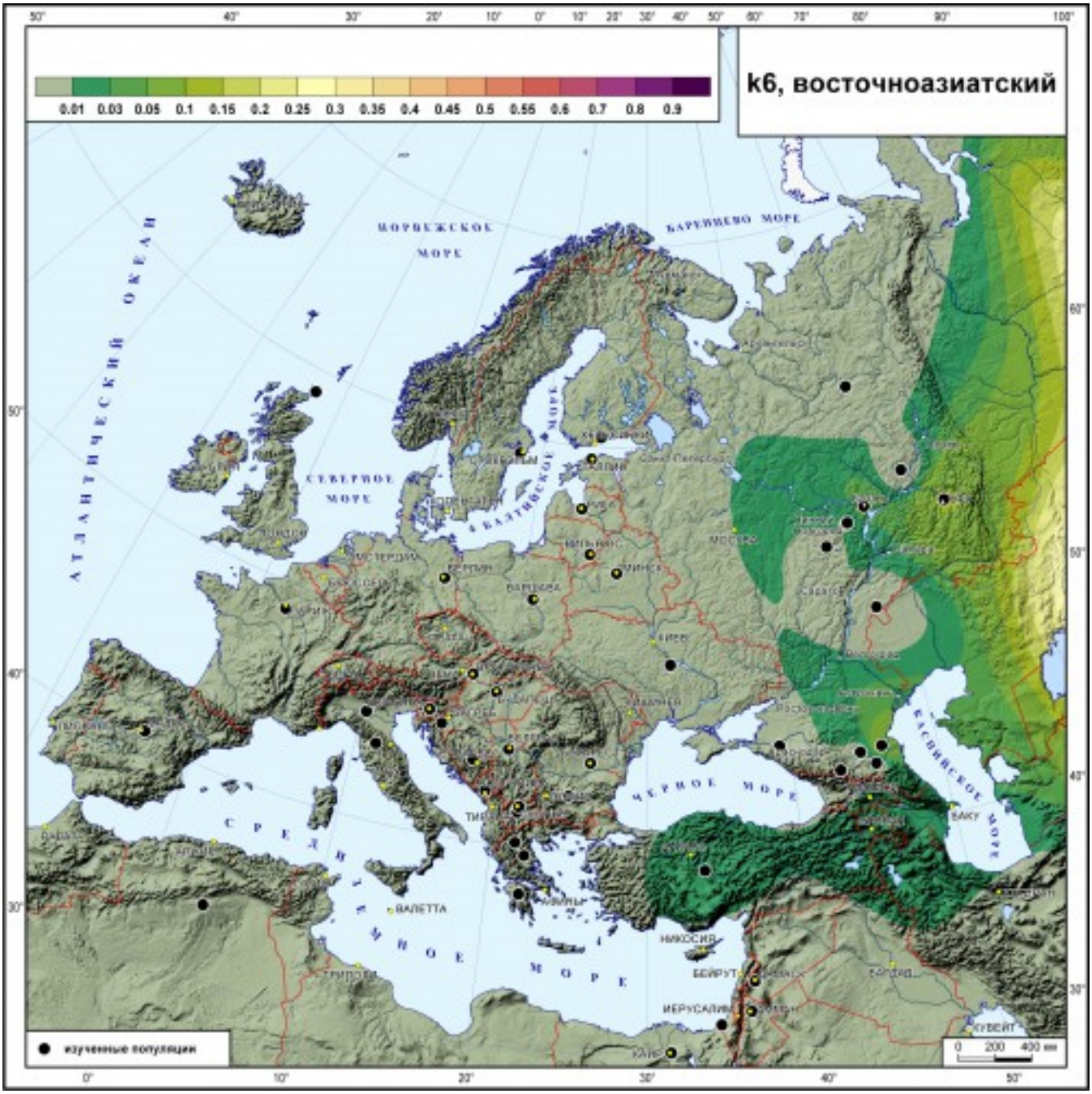


Рис. 4.4 Е. Распространение предковых компонентов ADMIXTURE при k=6. Д. Восточноазиатский компонент (6.5).

РАСПРОСТРАНЕНИЕ ПРЕДКОВЫХ КОМПОНЕНТОВ ПРИ K=18

Хотя в нашем анализе генофондов мира методом ADMIXTURE Европа и представлена куда большим числом популяций, чем другие части света, однако из 18 предковых компонентов 10 являются африканскими или азиатскими. На территории Европы они либо вообще не встречаются, либо ограничиваются ее самыми восточными или самыми южными популяциями, но даже в них частоты этих компонентов лишь немного превышают 1%. Из остальных 8 компонентов основными в Европе являются западноевропейский и восточноевропейский компоненты; чуть меньшее значение имеют средиземноморский и кавказский; поволжский и южноуральский выражены локально, хотя и очень ярко; а южноиндийский и сибирский компоненты, напротив, покрывают почти всю Восточную Европу, но составляют лишь немногим более одного процента ее генофонда. Рассмотрим эти восемь карт подробнее.

Четыре основных предковых компонента Европы (**18.1–18.4**), выделяемые при значениях **k=18**, представлены на рис. 4.5. А-Г.

Западноевропейский компонент (18.1, рис. 4.5.А) составляет более половины генофонда на Иберийской полуострове, Франции, Британских островах и юге Скандинавии. Перепрыгивая моря, он объединяет популяции всех побережий западной

трети Европы.

Восточноевропейский компонент (18.2, рис. 4.5.Б) доминирует у популяций восточной трети Европы. На всем пространстве огромной Восточноевропейской равнины – от Черного моря на юге до Балтийского и Белого на севере, от Карпат на западе до Волги на востоке – он составляет даже не половину, а две трети генофонда, и чуть менее выражен и далее на восток – до Урала.

Средняя треть Европы – Центральная Европа и Балканы – примерно в равной мере характеризуется западноевропейским и восточноевропейским компонентами, причем на Балканах они оба уступают лидерство средиземноморскому компоненту.

Средиземноморский компонент (18.3, рис. 4.5.В)**.** Максимум этого предкового компонента находится за пределами формальных границ Европы – в Леванте и Египте. Но это и показывает всю условность границ между выделяемыми для удобства частями света. Ведь население всех побережий Средиземноморья, и особенно восточной его части, издавна обменивалось миграциями. Скорее мир Средиземноморья можно было противопоставить миру более северных областей Европы, чем противопоставлять друг другу европейскую и африканско-азиатскую части Средиземноморья. Об этом и напоминает нам география средиземноморского компонента, составляющего более трети генофонда населения Балканского и Аппенинского полуостровов (особенно Греции и Италии) и чуть менее трети – генофонда остальной части южной Европы, включая и Карпаты, и Альпы, и южную половину Франции.

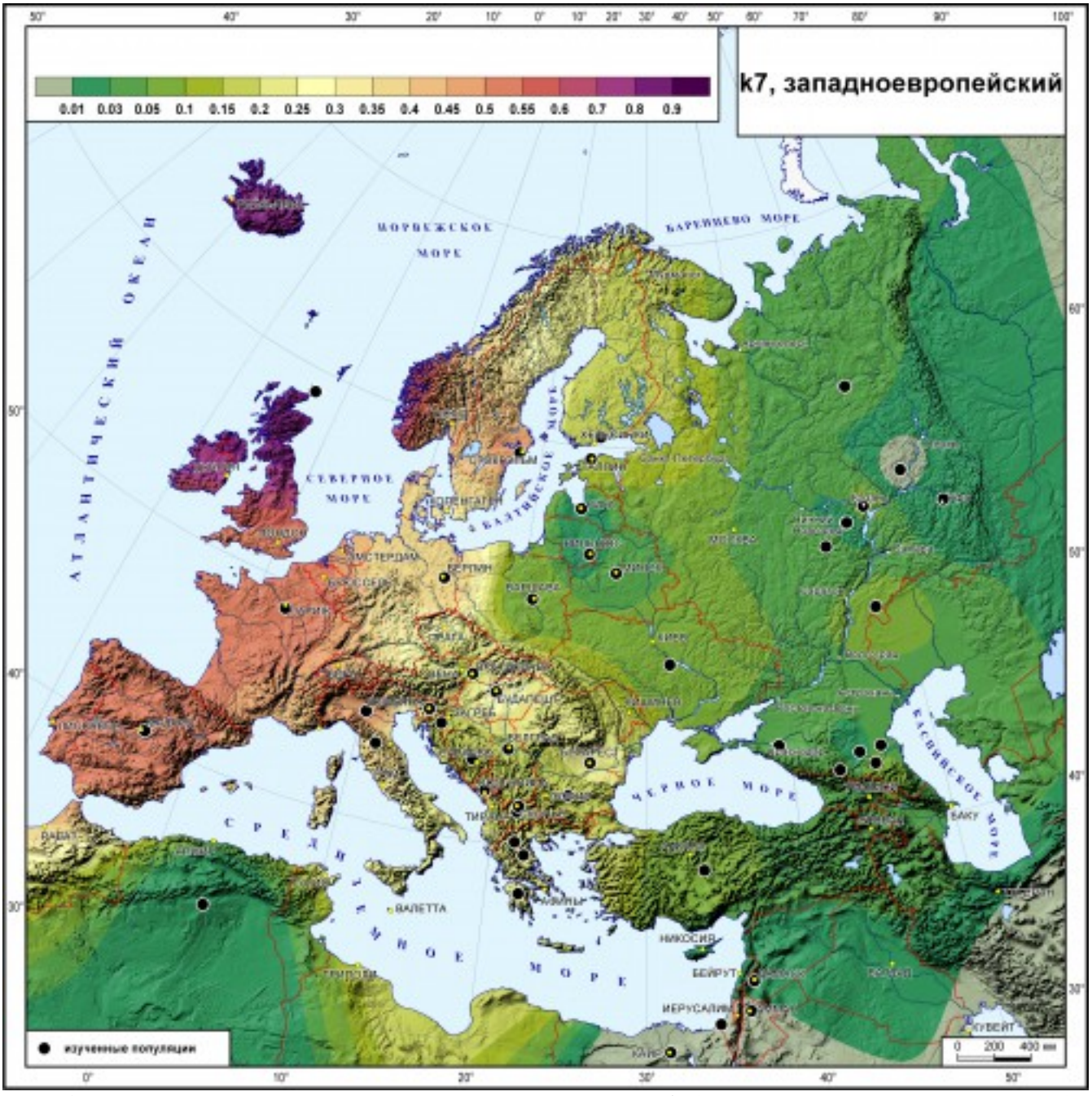


Рис. 4.5 А. Распространение предковых компонентов ADMIXTURE при k=18. Западноевропейский компонент (18.1).

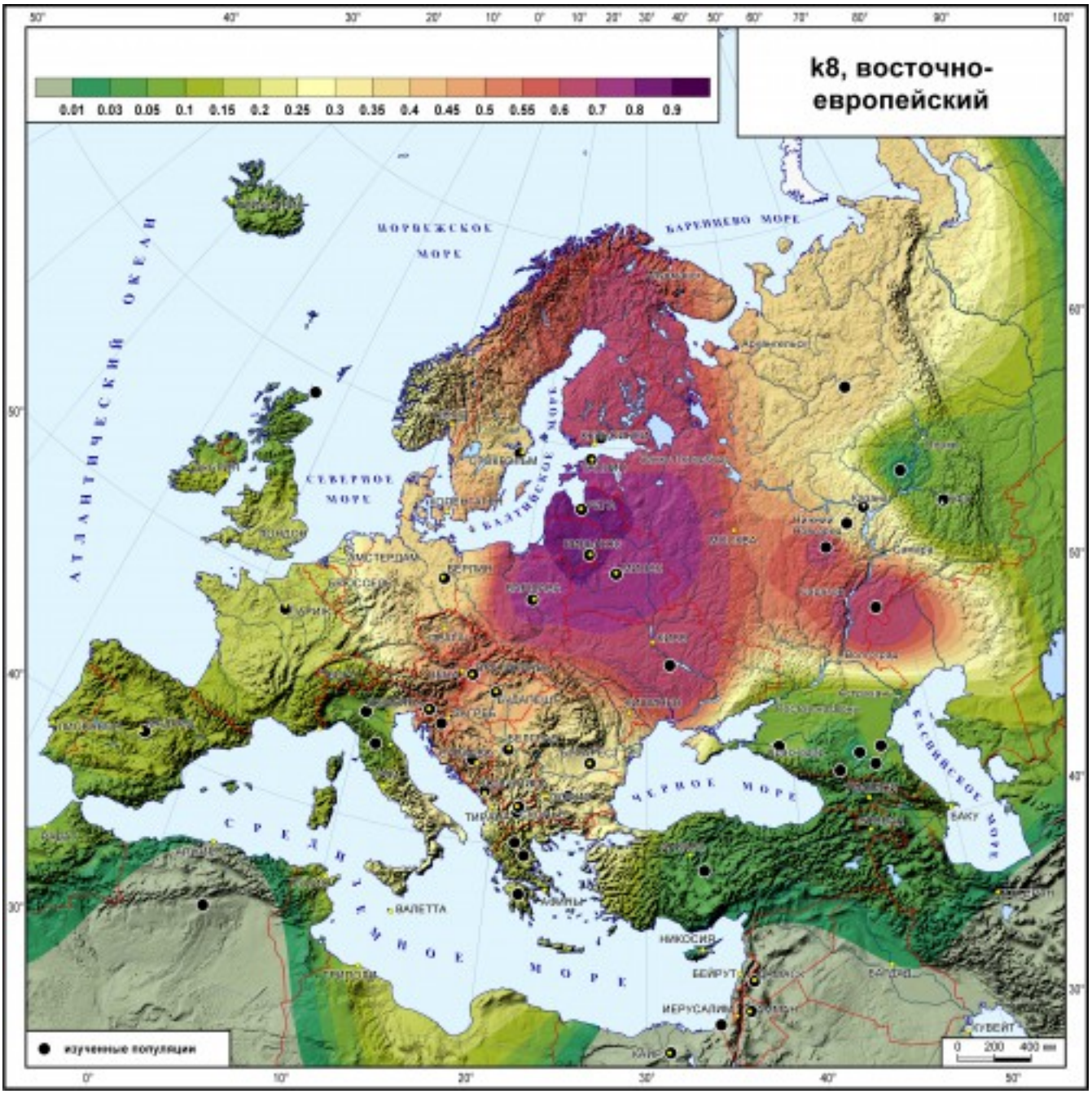


Рис. 4.5 Б. Распространение предковых компонентов ADMIXTURE при k=18. Восточноевропейский компонент (18.2).

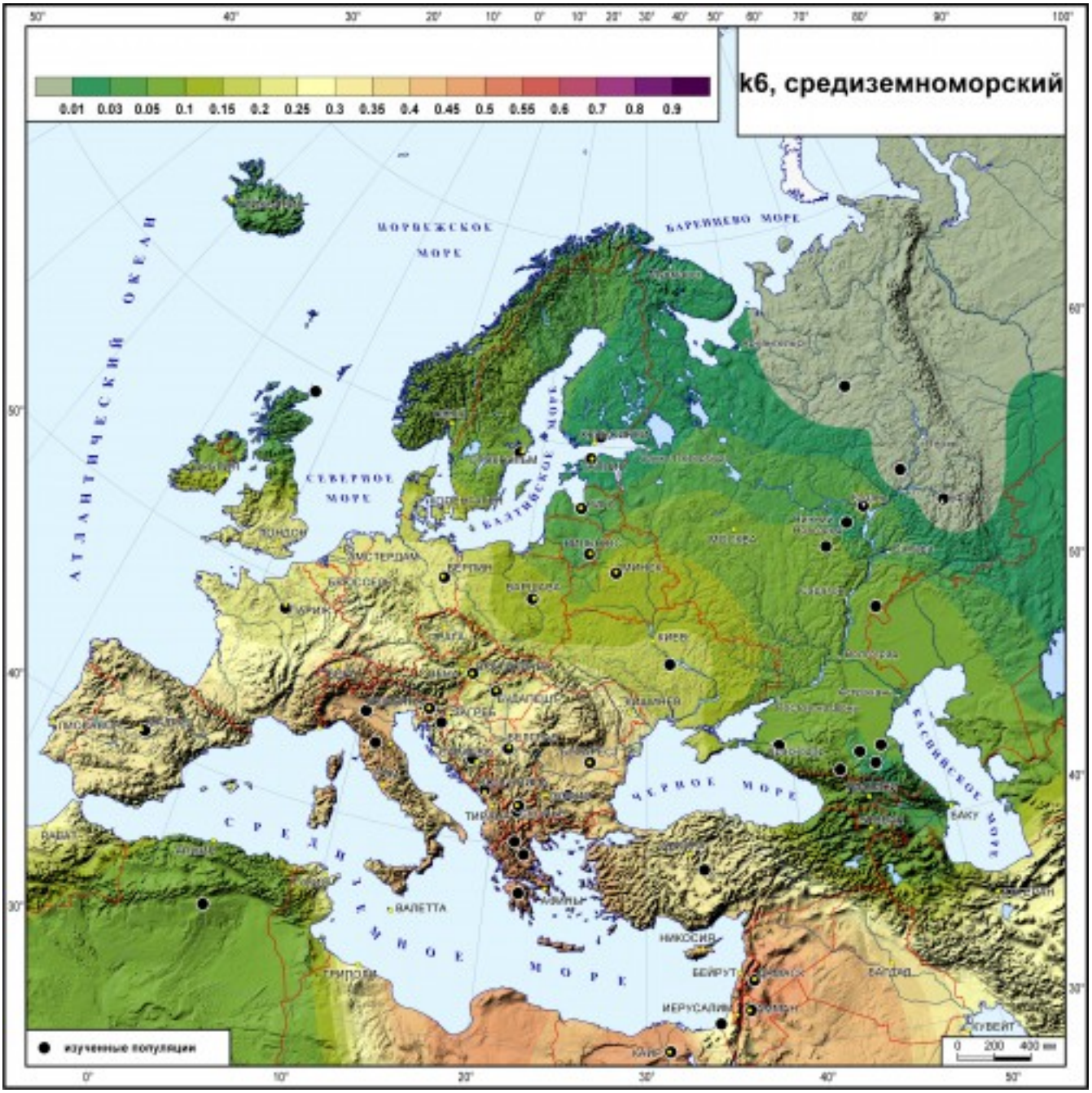


Рис. 4.5 В. Распространение предковых компонентов ADMIXTURE при k=18. А. Средиземноморский компонент (18.3).

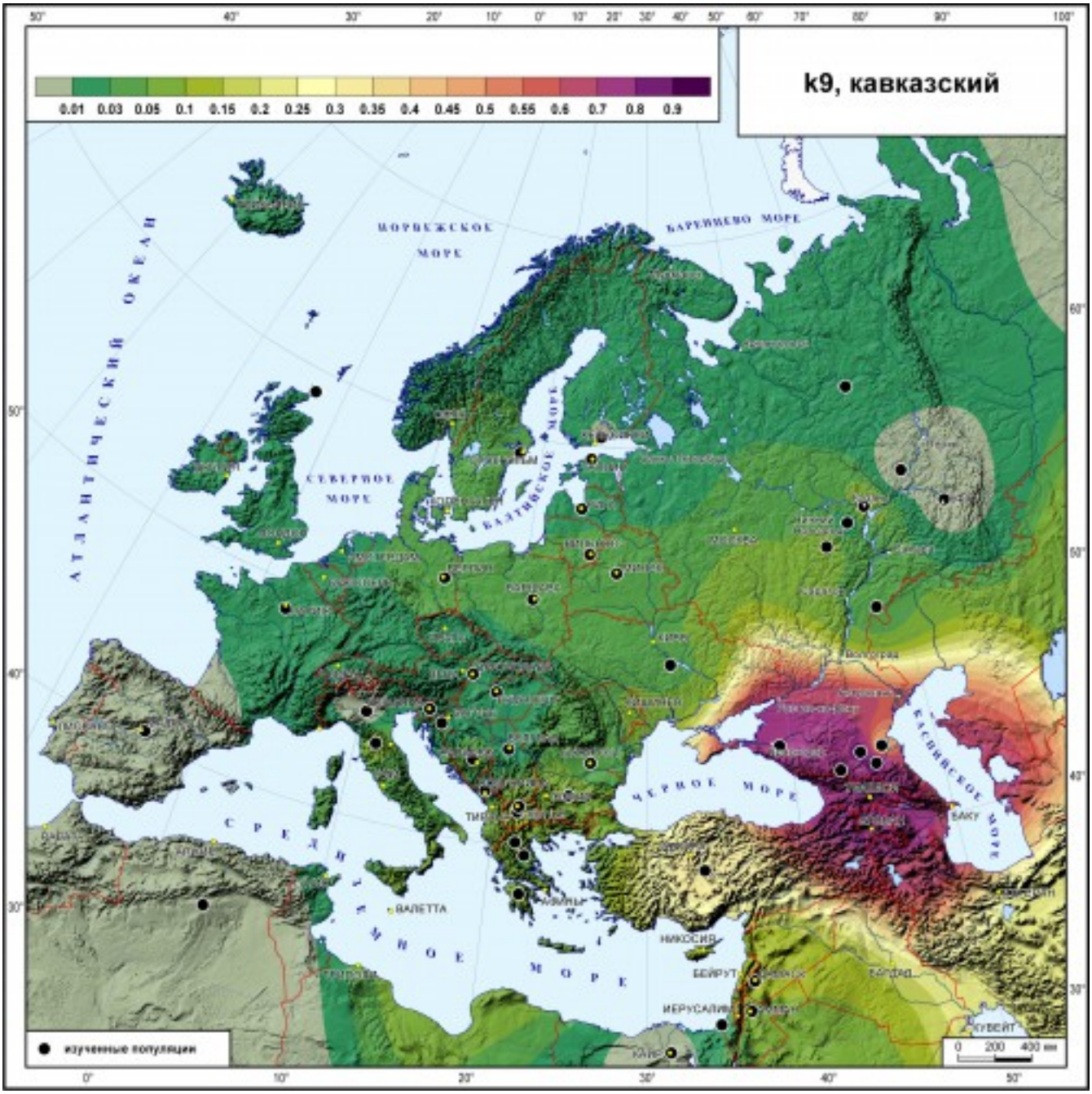


Рис. 4.5 Г. Распространение предковых компонентов ADMIXTURE при k=18. Кавказский компонент (18.4).

Кавказский компонент (18.4, рис. 4.5.Г)**.** Четвертый и последний из основных предковых компонентов Европы, напротив, распространен только в одной части Европы – на Северном Кавказе. Но зато он составляет более трех четвертей генофонда северокавказских популяций. Впрочем, генетически Кавказ является частью Передней Азии, а не Европы (как мы увидим в главе 7). Поэтому и предковый кавказский компонент на Переднюю Азию еще распространяется, а в Европе выражен очень слабо: его доля варьирует от 1% до 5% генофонда, причем популяции Восточной Европы ближе к верхней границе этого интервала, а Западной Европы – к нижней.

Этими четырьмя компонентами описывается почти весь генофонд Европы – за исключением Волго-Уральского региона. В этой восточной окраине Европы обнаруживается большое генетическое разнообразие, проявившееся сразу в двух предковых компонентах.

Приуральский компонент (18.5, рис. 4.5.Д). Один из них, названный нами приуральским, превышает 90% у удмуртов и составляет около трети генофонда коми.

Южноуральский компонент (18.6, рис. 4.5.Е). Другой – южноуральский предковый компонент – доминирует у башкир и составляет около трети генофондов татар и чувашей.

Но оба этих компонента распространены, хотя и с очень разными частотами, у всех . народов Волго-Уральского региона. На остальной территории Восточной Европы встречаются с крайне низкими частотами, а в Западной Европе отсутствуют. Наличие этих двух специфических уральских компонентов указывает на большое генетическое разнообразие и своеобразие, скрытое в самых восточных областях Европы – вывод, совпадающий с результатами анализа митохондриальной ДНК (рис. 3.5) и с результатами анализа карт генетических расстояний (глава 5).

Для полноты картины упомянем еще два компонента, которые, хотя и с низкими частотами, все же обнаружены в генофонде Европы.

Сибирский компонент (18.7., рис. 4.5.Ж) встречается, как и можно ожидать, у народов на самом востоке Европы (с частотой около 2%).

Южноиндийский компонент (18.8., рис. 4.5.З), распространенный также по всей Передней Азии, заходит, пусть и с самыми низкими частотами, на Восточно-Европейскую равнину, но отсутствует в Западной Европе.

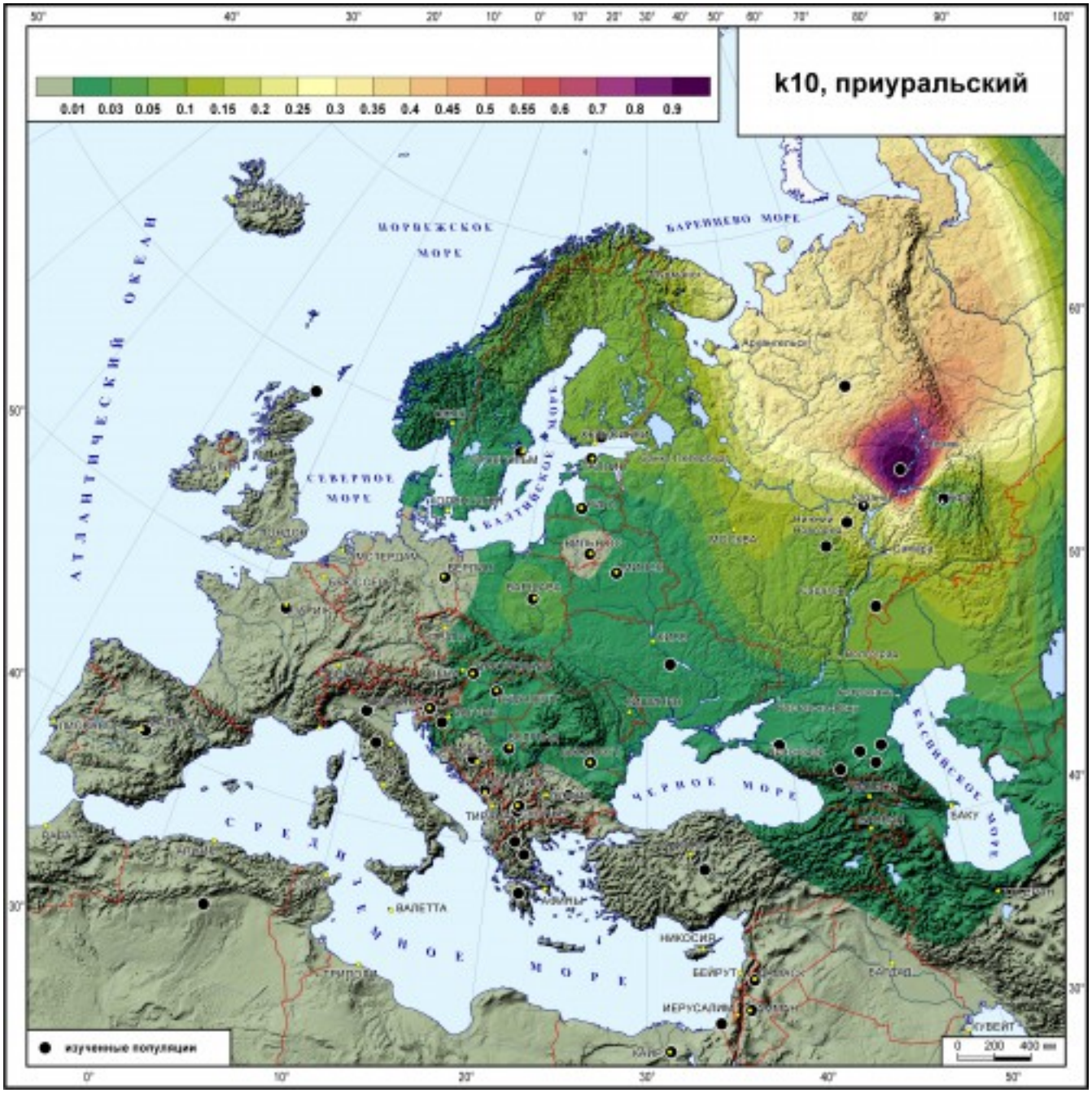


Рис. 4.5 Д. Распространение предковых компонентов ADMIXTURE при k=18. Приуральский компонент (18.5).

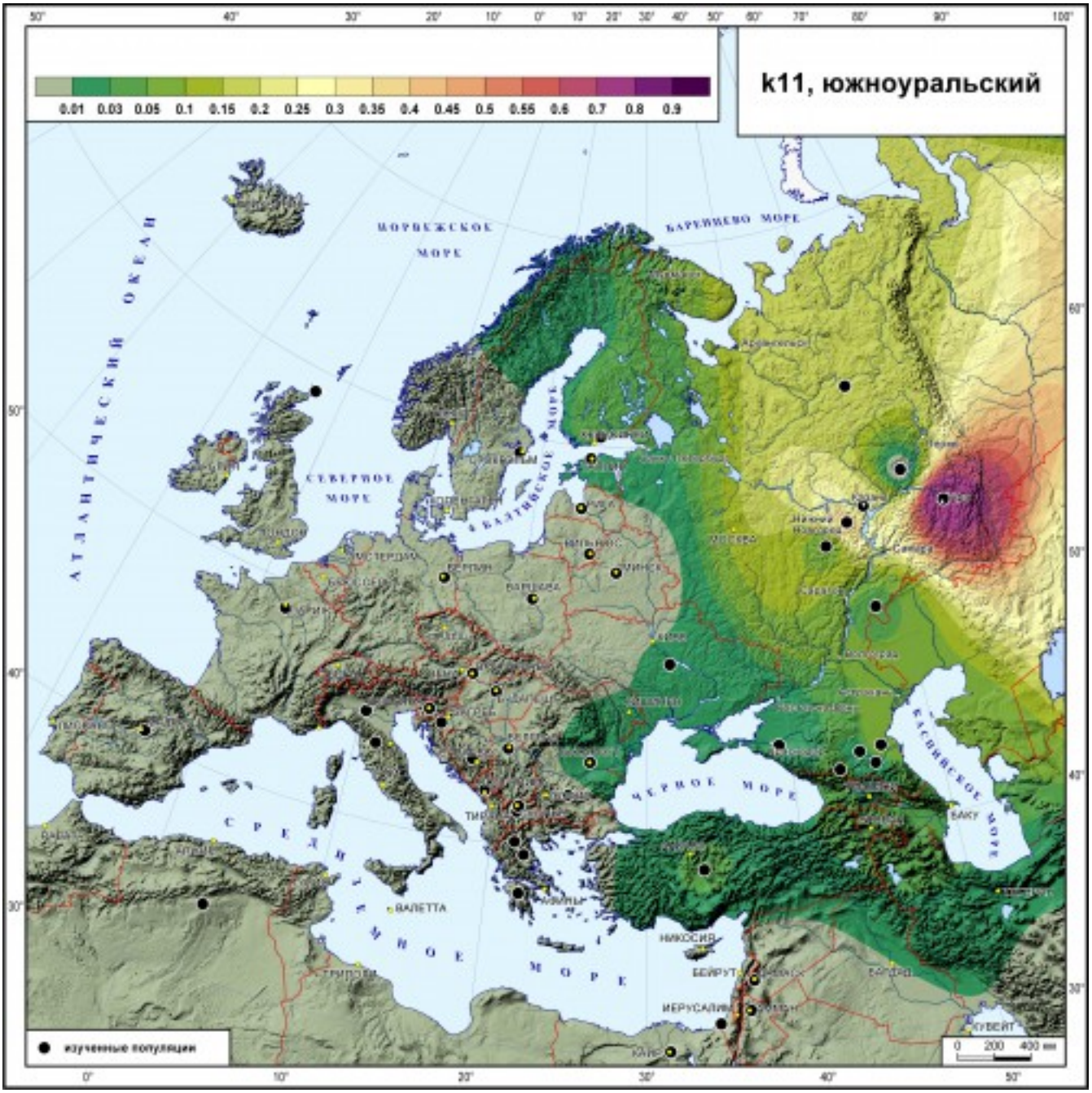


Рис. 4.5 Е. Распространение предковых компонентов ADMIXTURE при k=18. Южноуральский компонент (18.6).

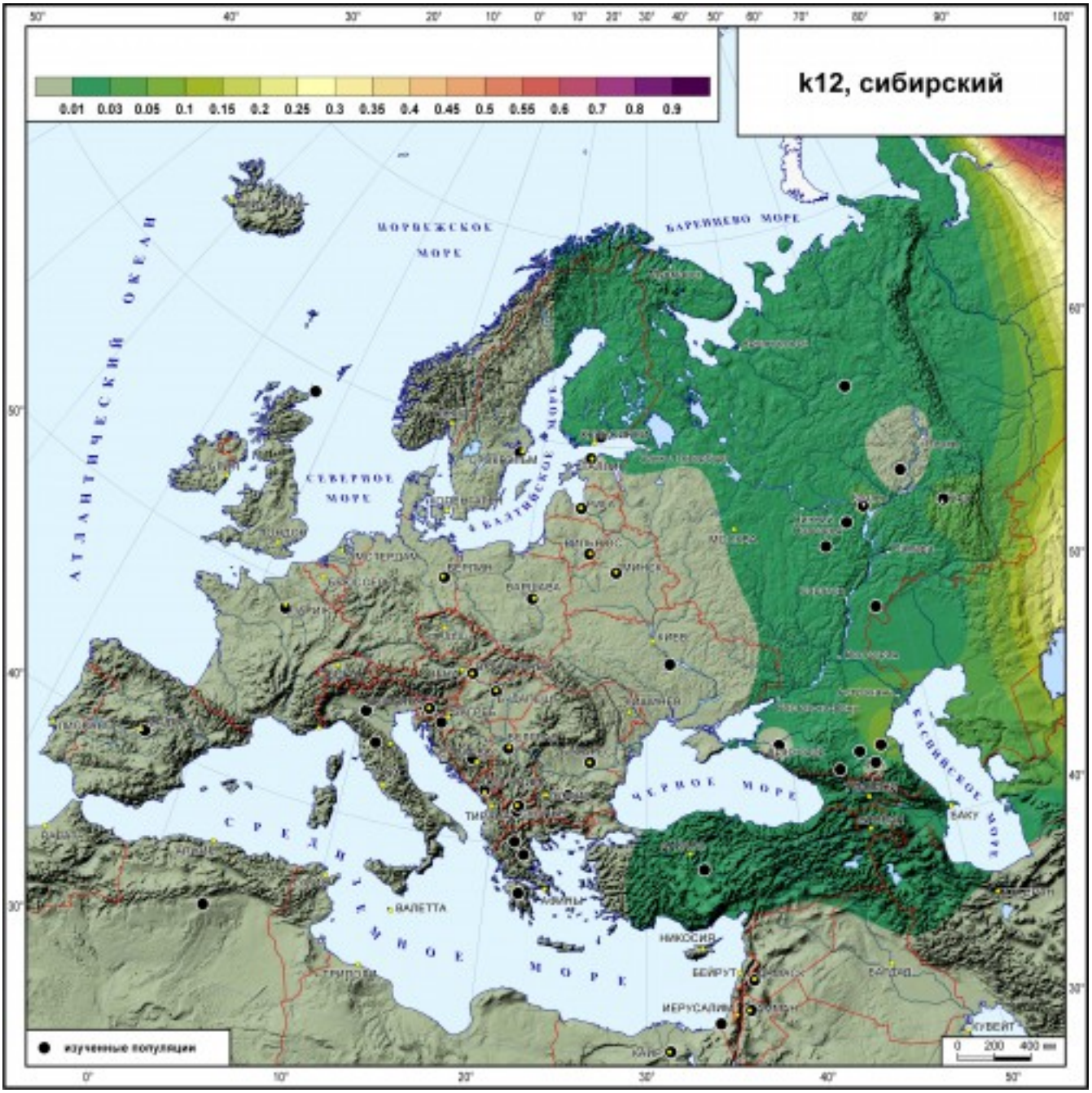


Рис. 4.5 Ж. Распространение предковых компонентов ADMIXTURE при k=18. Сибирский компонент (18.7).

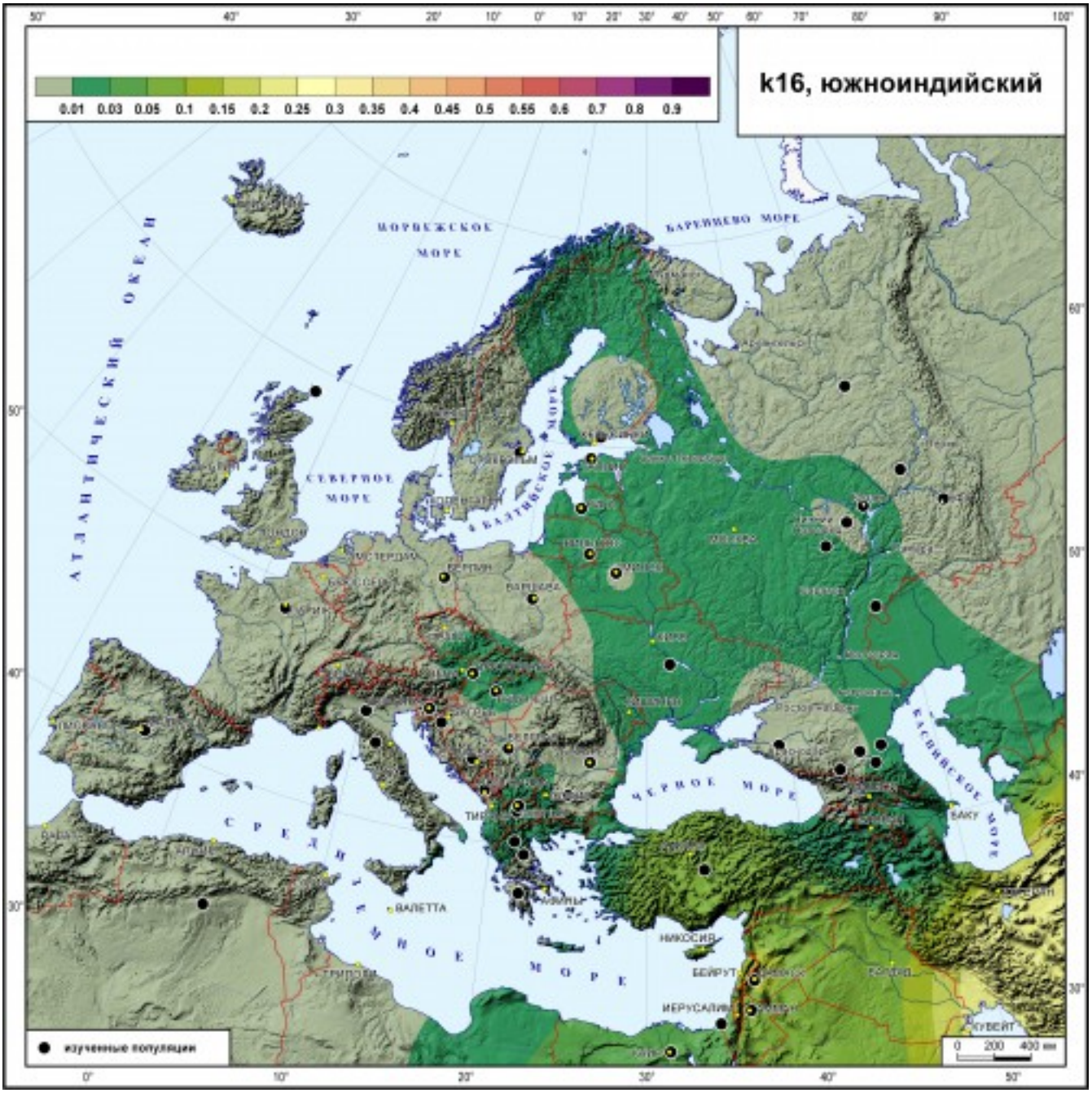


Рис. 4.5 З. Распространение предковых компонентов ADMIXTURE при k=18. Южноиндийский компонент (18.8).