УДК 577.17

МНОГОФАКТОРНЫЙ АНАЛИЗ ГЕНЕТИЧЕСКИХ СООТНОШЕНИЙ ПОПУЛЯЦИЙ ЦЕНТРАЛЬНОЙ РОССИИ

© И.Н. Сорокина, И.Н. Лепендина, И.В. Батлуцкая, Е.Н. Крикун, М.И. Чурносов

Ключевые слова: фамилии; митохондриальная ДНК; генетические соотношения.
Проведен анализ генетических соотношений тринадцати районов семи областей Центральной России. Установлена «широтная» изменчивость генофонда населения Центральной России по данным о фамилиях. Генетический ландшафт популяций Центральной России, полученный по частотам фамилий, полностью соответствует таковому по частотам 29 аллелей 11 иммуно-биохимических локусов и отличается от данных, установленных на основе митохондриальной ДНК.

ВВЕДЕНИЕ

Важное место при изучении генетической структуры популяций человека занимают подходы, основанные на оценке генетических расстояний между популяциями [1, 2] с последующим их анализом с помощью методов многомерной статистики [3]. На основе матриц генетических расстояний проводится кластерный анализ [1] и строится дендрограмма. Для получения еще более наглядной картины взаимосвязей между популяциями по соответствующим дендрограммам составляется «генетический ландшафт» местности [4]. Он представляет картографические схемы, на которых элементарные популяции, относящиеся к определенным кластерам, объединяются на карте линиями, образуя эквидистантные фигуры. Способ описания генетической структуры популяций с помощью эквидистантных фигур, последовательно объединяющих популяции в соответствии с их генетическими расстояниями друг от друга и создающий таким образом «генетический ландшафт» исследуемого населения, может эффективно использоваться как для определения границ и размеров элементарных популяций [4], так и для оценки влияния на генетическую структуру населения различных факторов, как, например, перемещения административных границ популяций, миграционные потоки в популяцию и т. д. [4–7].

В настоящей работе в терминах генетических расстояний с последующей их обработкой методами многомерной статистики описаны генетические соотношения ряда популяций Центральной России.

МАТЕРИАЛЫ И МЕТОДЫ

В качестве объектов исследования были выбраны тринадцать районов семи областей Центральной России: Михайловский и Спасский районы Рязанской области, Боровский и Барятинский районы Калужской области, Болховский и Ливенский районы Орловской области, Петровский район Тамбовской области, Пристенский и Черемисиновский районы Курской области, Репьевский район Воронежской области, Прохоровский, Яковлевский и Красненский районы Белгородской области. Источником данных о фамилиях послужили списки избирателей, которые охватили «тотально» все население старше 18 лет. В анализ включены данные об изменчивости 21556 фамилий среди 242672 человек Центральной России. Вычисление популяционно-генетических характеристик осуществлялось на районном уровне (элементарная популяция [8]). Частота фамилий рассчитывалась через отношение числа носителей данной фамилии в изучаемой популяции к общему числу жителей старше 18 лет в данной популяции.

С целью оценки генетических соотношений районных популяций Центральной России на основе матрицы генетических расстояний проводился кластерный анализ, по результатам которого строилась схема генетического ландшафта, проводилось многомерное шкалирование, по корреляционной матрице – факторный анализ в программе Статистика (версия 6).

Полученные по данным квазигенетических маркеров результаты о соотношениях районных популяций Центральной России сравнивались с аналогичными результатами на основе распределения иммунобиохимических и молекулярно-генетических маркеров.

Для выявления роли географических расстояний в формировании генетической структуры популяций проводился корреляционный анализ (ранговый коэффициент Спирмена) матриц генетических и географических расстояний.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

По данным о распределении всех фамилий была проведена оценка генетических соотношений между 13 районами 7 областей Центральной России. На основе матрицы генетических расстояний проведен кластерный анализ и построена дендрограмма методом Уорда (рис. 1), анализ которой показал дифференцировку изучаемых тринадцати районов Центральной России на три кластера. Первый кластер сформирован Боровским и Барятинским районами Калужской области, Михайловским и Спасским районами Рязанской области,

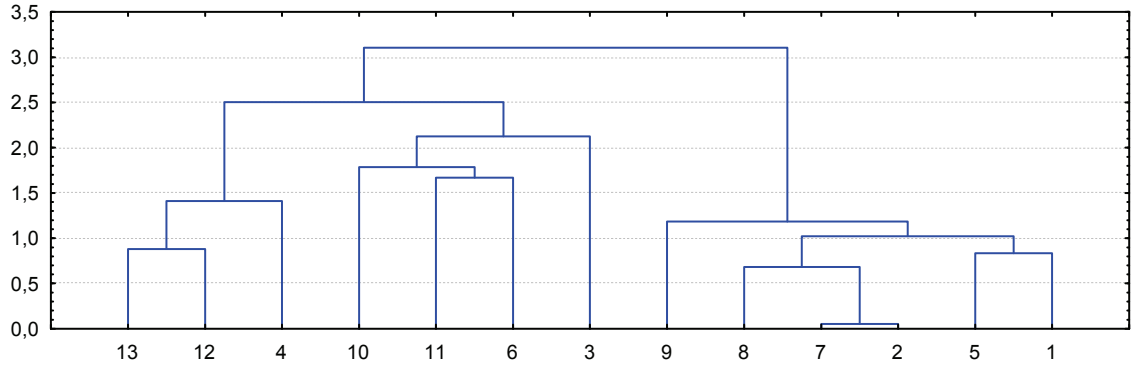

Рис. 1. Дендрограмма генетических соотношений тринадцати районных популяций Центральной России (построена методом Уорда по частотам всех фамилий): 1 – Барятинский район Калужская область; 2 – Боровский район Калужская область; 3 – Черемисиновский район Курская область; 4 – Пристенский район Курская область; 5 – Болховский район Орловская область; 6 – Ливенский район Орловская область; 7 – Михайловский район Рязанская область; 8 – Спасский район Рязанская область; 9 – Петровский район Тамбовская область; 10 – Репьевский район Воронежская область; 11 – Красненский район Белгородская область; 12 – Прохоровский район Белгородская область; 13 – Яковлевский район Белгородская область

Рис. 2. Схема генетических соотношений и географического расположения тринадцати районных популяций Центральной России (получена на основе распределения всех фамилий) (овалами обведены популяции, которые формируют кластеры при кластерном анализе)

Болховским районом Орловской области и Петровским районом Тамбовской области. Второй – включает Яковлевский и Прохоровский районы Белгородской области, Пристенский район Курской области. Третий – Ливенский район Орловской области, Красненский район Белгородской области, Репьевский район Воронежской области и Черемисиновский район Курской области.

На основе результатов кластерного анализа была построена схема генетических соотношений 13 изучаемых районов. Эквидистантные линии проведены без пересечения через 0,5 единиц генетического расстояния. Все районные популяции Центральной России дифференцируются на три «широтные» группы (рис. 2). Шесть популяций Центральной России (Михайловский и Спасский районы Рязанской области, Боровский и Барятинский районы Калужской области, Болховский район Орловской области и Петровский район Тамбовской области) образуют единый «северный» кластер. Второй – «северо-восточный» – включает Красненский район Белгородской области, Репьевский район Воронежской области, Черемисиновский район Курской области и Ливенский район Орловской области. Третий – «юго-западный» кластер объединил районы Белгородской области (Яковлевский и Прохоровский), а также Пристенский район Курской области.

Далее был проведен сравнительный анализ результатов, полученных на основе частот фамилий с данными по этим же популяциям Центральной России по иммуно-биохимическим маркерам.

Следует отметить, что результаты изучения генетической дифференциации 13 районных популяций Центральной России, полученные по частотам всех фамилий, согласуются с данными, полученными по частотам 29 аллелей 11 иммуно-биохимических локусов (рис. 3). При этом важно подчеркнуть, что генетическая дифференциация популяций Центральной России на три кластера полностью согласуется с их реальным географическим расположением (рис. 2). Коэффициент корреляции Спирмена между матрицами генетических расстояний, построенных по частотам фамилий и по частотам генов, составил $\rho = 0{,}34$ ($p < 0{,}05$).

В результате проведения многомерного шкалирования получен график, представленный на рис. 4. При построении графика совершено 70 итераций. Коэффициент стресса данного графика равен $S_0 = 0{,}075$, коэффициент алиенации $K = 0{,}10$; кривая Шепарда удовлетворительная, что позволяет считать результаты многомерного шкалирования приемлемыми.

В трехмерном пространстве, как и на дендрограмме, сохраняется дифференциация тринадцати районов Центральной России на три кластера: первый образуют Красненский район Белгородской области, Репьевский район Воронежской области, Ливенский район Орловской области и Черемисиновский район Курской области; второй – Прохоровский и Яковлевский районы Белгородской области и Пристенский район Курской области; третий – Михайловский и Спасский районы Рязанской области, Боровский и Барятинский районы Калужской области, Болховский район Орловской области и Петровский район Тамбовской области.

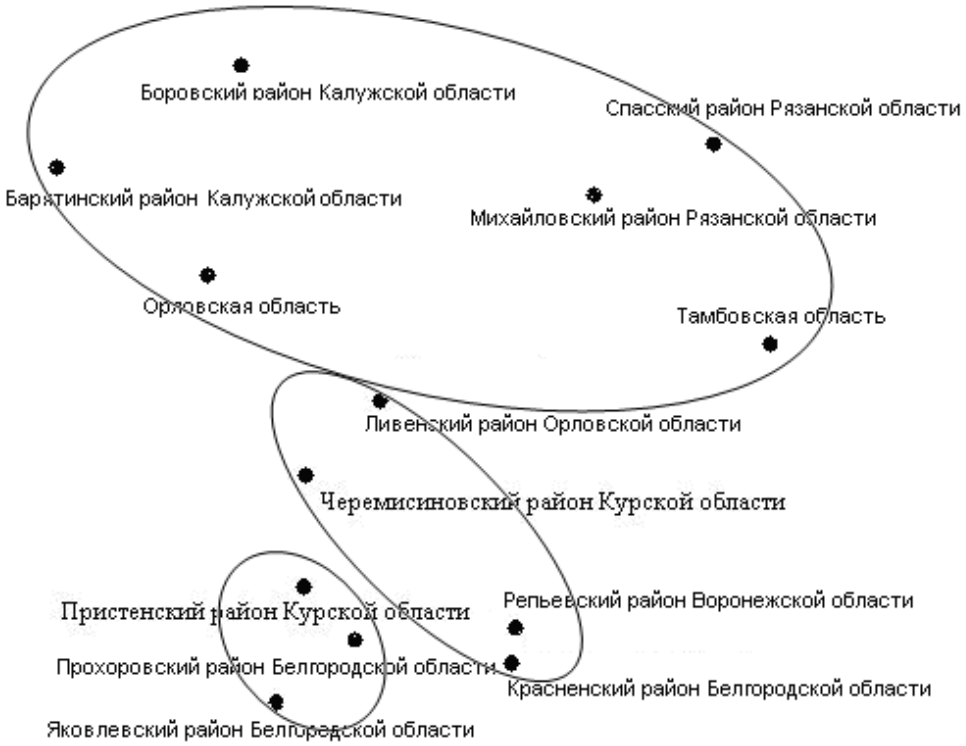

Рис. 3. Схема генетических соотношений тринадцати популяций Центральной России по частотам 29 аллелей 11 иммуно-биохимических локусов (овалами обведены популяции, которые формируют кластеры при кластерном анализе и многомерном шкалировании)

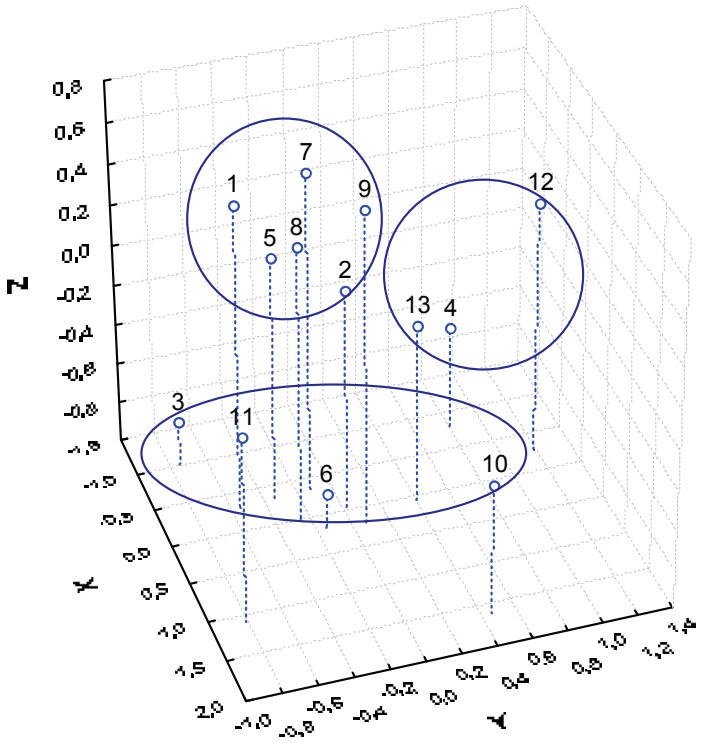

Рис. 4. График трехмерного шкалирования, отражающий расположение в пространстве тринадцати популяций Центральной России (построен по частотам всех фамилий): 1 – Барятинский район Калужская область; 2 – Боровский район Калужская область; 3 – Черемисиновский район Курская область; 4 – Пристенский район Курская область; 5 – Болховский район Орловская область; 6 – Ливенский район Орловская область; 7 – Михайловский район Рязанская область; 8 – Спасский район Рязанская область; 9 – Петровский район Тамбовская область; 10 – Репьевский район Воронежская область; 11 – Красненский район Белгородская область; 12 – Прохоровский район Белгородская область; 13 – Яковлевский район Белгородская область

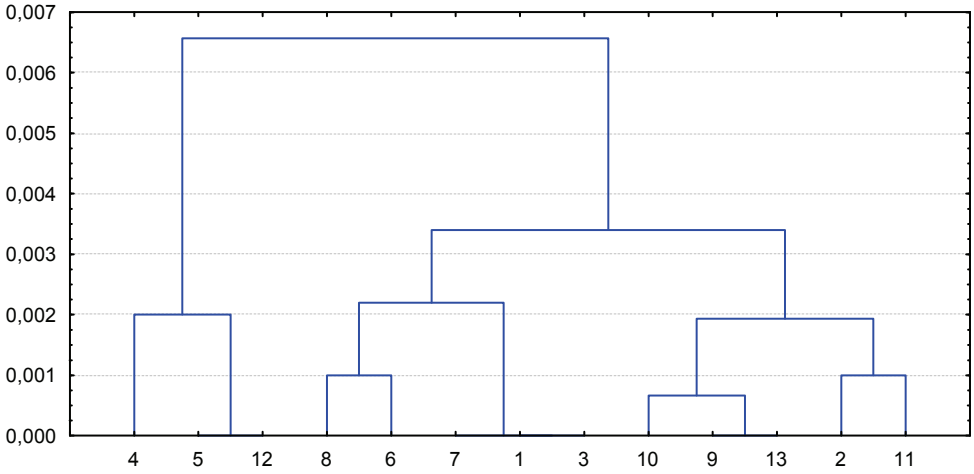

Рис. 5. Дендрограмма генетических соотношений тринадцати районных популяций Центральной России (построена методом Уорда по частотам гаплогрупп мтДНК): 1 – Барятинский район Калужская область; 2 – Боровский район Калужская область; 3 – Черемисиновский район Курская область; 4 – Пристенский район Курская область; 5 – Болховский район Орловская область; 6 – Ливенский район Орловская область; 7 – Михайловский район Рязанская область; 8 – Спасский район Рязанская область; 9 – Петровский район Тамбовская область; 10 – Репьевский район Воронежская область; 11 – Красненский район Белгородская область; 12 – Прохоровский район Белгородская область; 13 – Яковлевский район Белгородская область

Таким образом, изучение генетических соотношений 13 популяций Центральной России, проведенное по данным о частотах фамилий, выявило, во-первых, их дифференциацию на три кластера. Во-вторых, установлена «широтная» изменчивость генофонда населения Центральной России по данным о фамилиях. В-третьих, генетический ландшафт 13 популяций, полученный по частотам фамилий, полностью соответствует таковому по частотам генов.

На следующем этапе исследования мы провели сравнение результатов, полученных с использованием нейтральных квазигенетических маркеров с данными, полученными по однородительским генным маркерам – по гаплогруппам митохондриальной ДНК. Митохондриальная ДНК является особым типом генетических маркеров, т. к. наследуется только по материнской линии, что позволяет дифференциально прослеживать популяционные процессы (миграции, динамики численности и т. д.), в различной степени затрагивающие представителей разных полов. Проведенный кластерный анализ (рис. 5) матрицы генетических расстояний, построенной по частотам гаплогрупп мтДНК, показал объединение исследуемых популяций в два основных кластера. Первый кластер объединяет десять районов – Спасский и Михайловский районы Рязанской области, Ливенский район Орловской области, Барятинский и Боровский районы Калужской области, Черемисиновский район Курской области, Петровский район Тамбовской области, Репьевский район Воронежской области, Красненский район Белгородской области, Яковлевский район Белгородской области. Второй кластер включает Прохоровский район Белгородской области, Болховский район Орловской области и Пристенский район Курской области.

При проведении многомерного шкалирования подтверждается выявленная при кластерном анализе закономерность: генетическая дифференциация населения Центральной России по частотам гаплогрупп митохондиральной ДНК на две группы.

Установленные различия в генетической дифференциации исследуемых тринадцати районных популяций Центральной России по разным типам маркеров (на три группы по квазигенетическим маркерам и на две группы по гаплогруппам митохондиральной ДНК) могут быть связаны с различным характером передачи данных маркеров из поколения в поколение: квазигенетические маркеры отражают популяционно-генетические процессы, происходящие среди населения обоих полов, а гаплогруппы мтДНК – только лишь в женской части населения. При этом использование нескольких типов маркеров, дополняющих друг друга, позволяет более точно и адекватно прослеживать популяционно-генетические процессы (миграции и др.), протекающие в популяциях.

Таким образом, полученные результаты многофакторного анализа генетических соотношений популяций Центральной России по различным типам маркеров (квазигенетические, иммунно-биохимические, мтДНК маркеры) позволяют заключить, что для корректного проведения популяционно-генетических исследований следует одновременно использовать несколько типов маркеров.

ЛИТЕРАТУРА

1. *Дерябин В.Е.* Многомерные биометрические методы для антропологов. М.: ВИНИТИ, 2001. С. 105-265.
2. *Пасеков В.П.* Генетические расстояния // Итоги науки и техники. Сер. Общая генетика. Теоретическая и популяционная генетика. 1983. Т. 8. С. 3-75.
3. *Парадеева Г.М., Большакова Л.П., Петрин А.Н. и др.* Медико-генетическое изучение населения Костромской области. Сообщение III. Коэффициент инбридинга и его динамика в сельских популяциях и городах районного масштаба // Генетика. 1986. Т. 22. № 9. С. 2355-2360.
4. *Ельчинова Г.И., Кадошникова М.Ю., Мамедова Р.А.* Выявление особенностей генетической структуры популяции с помощью метода описания «генетического ландшафта» // Генетика. 1991. Т. 27. № 11. С. 1994-2001.
5. Наследственные болезни в популяциях человека / под ред. Е.К. Гинтера. М.: Медицина, 2002. 304 с.
6. *Хуснутдинова Э.К.* Молекулярная этногенетика народов Волго-Уральского региона. Уфа, 1999. 238 с.
7. *Ельчинова Г.И.* Опыт применения методов популяционно-генетического анализа при изучении популяций России с различной генетико-демографической структурой: автореф. дис. … д-ра биол. наук / ГУ МГНЦ РАМН. М., 2001. 48 с.
8. *Чурносов М.И., Сорокина И.Н., Балановская Е.В.* Генофонд населения Белгородской области. Динамика индекса эндогамии в районных популяциях // Генетика. 2008. Т. 44. № 8. С. 1117-1125.

Поступила в редакцию 12 сентября 2009 г.

Sorokina I.N., Lependina I.N., Batlutskaya I.V., Krikun E.N., Churnosov M.I. The multifactorial analysis of genetic parities of populations of the Central Russia. The analysis of genetic parities of thirteen areas of seven areas of the Central Russia is carried out. The «width» variability of a genofund of the population of the Central Russia by the surnames data is established. The genetic landscape of populations of the Central Russia, received on frequencies of surnames completely corresponds to that on frequencies of 29 alleles 11 immuno-biochemical loci and differs from the data established on a basis of mt DNA.

Key words: surnames; mt DNA; genetic parities.