



УДК 577.17

## МОДЕЛИРОВАНИЕ ГЕНЕТИЧЕСКОГО ЛАНДШАФТА НАСЕЛЕНИЯ ЮГА ЦЕНТРАЛЬНОЙ РОССИИ \*

**М.И. Чурносов**  
**И.И. Сорокина**

*Белгородский  
государственный  
университет*

*e-mail: sorokina@bsu.edu.ru*

В статье по данным о частотах всех фамилий (50412 фамилий) среди 849399 человек различными методами многомерной статистики (кластерный анализ, многомерное шкалирование, факторный анализ) установлено наличие на территории юга Центральной России определенной, упорядоченной системы группировки 22 районных популяций. Все районы объединяются в четыре кластера, соответствующие реальному географическому положению популяций и географическим расстояниям между ними.

Ключевые слова: генетическая структура популяции, фамилии, генофонд.

При изучении генетической структуры популяций человека используются различные подходы, позволяющие получить представление о подразделенности популяций, характере генетических взаимоотношений между ними. Среди последних важное место занимают подходы, основанные на оценке генетических расстояний между популяциями [1,2] с последующим их анализом с помощью методов многомерной статистики. На основе матриц генетических расстояний проводится кластерный анализ, обладающий большой гибкостью [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] и для оценки влияния на генетическую структуру населения таких факторов, как, например, перемещения административных границ популяций, необычные миграционные потоки в популяцию и т.д. [4,5].

К настоящему времени в терминах генетических расстояний с последующей их обработкой методами многомерной статистики описаны генетические взаимоотношения популяций Волго-Уральского региона [6], населения Костромской [3], Кировской [7] областей, республики Адыгея [5,7,8], Марий Эл [5], Чувашии [5] и др. [5,8]

В данной работе представлены результаты таксономического анализа населения юга Центральной России – всего населения Белгородской области (20 районов), Репьевского района Воронежской области, Пристенского района Курской области на уровне элементарных популяций (район) по данным антропоники.

**Материалы и методы.** Оценка генетических расстояний и таксономический анализ проведен на уровне района (элементарная популяция в 1990-х гг. [9]). Объектами исследования послужили 22 районные популяции юга Центральной России: все 20 районов Белгородской области (за исключением одного района – Белгородского, который согласно данным работы [10] не является элементарной популяцией), Пристенский район Курской области и Репьевский район Воронежской области.

Материалом для исследования послужили данные о распределении всех фамилий (50412) среди 849399 человек. На основе частот всех фамилий были рассчитаны матрицы генетических расстояний с помощью программы DJ genetic (версия 0,03 beta), разработанной Ю.А. Серегиним и Е.В. Балановской в ГУ МГНЦ РАМН. При рас-

\* Работа выполнена в рамках Федеральной целевой программы "Научные и научно-педагогические кадры инновационной России" (государственный контракт №02.740.11.0496 "Генетические факторы мультифакториальных заболеваний человека").



чете расстояний между элементарными популяциями использовали общепринятый в популяционно-генетических исследованиях метод сравнения популяций по частотам аллелей полиморфных маркеров по М. Nei [11, 12]. В данном случае аналогами аллелей являлись фамилии (квазигенетический маркер).

На основе полученной матрицы генетических расстояний в программе Statistica 6.0. проводился кластерный анализ. Для построения кластеров использовали иерархическую агломеративную процедуру. Построение дендрограммы осуществлялось двумя методами: средневзвешенной связи и методом Уорда [1]. В работе использовалась дендрограмма, соответствовавшая результатам, полученным другими методами многомерного анализа – многомерного шкалирования и факторного анализа.

В соответствии с результатами кластерного анализа (дендрограмма) была построена схема генетического ландшафта изучаемых популяций. Графическая схема данного ландшафта представляла собой проекцию дендрограммы матрицы генетических расстояний, на которой элементарные популяции, относящиеся к определенным кластерам, объединялись на карте эквидистантными фигурами. Эквидистантные линии были проведены через определенную единицу генетического расстояния, так, чтобы число уровней объединения не превышало 10, так как большее число уровней делает схему громоздкой и затрудняет анализ [8].

С использованием матриц генетических расстояний проводилось многомерное шкалирование. Для оценки качества результатов многомерного шкалирования применяли показатели стресса  $S_0$  ( $\leq 0,10$ ), коэффициент отчужденности или алиенации  $K_0$  и диаграмму Шепарда [1]. На основе рассчитанных корреляционных матриц проводили факторный анализ по методу главных факторов. При факторизации для определения числа значимых главных факторов применяли критерий Кайзера и критерий отсеивания Кеттела. Для наглядного представления данные изображались в графическом виде, где рассматриваемые популяции располагались в 2-3 мерном пространстве выявленных значимых главных факторов.

С целью изучения роли географических расстояний в формировании подразделенности генетической структуры населения нами проведен корреляционный анализ (рассчитывался ранговый коэффициент корреляции Спирмена) матриц генетических расстояний и географических расстояний. Для этого с использованием карты Белгородской области (масштаб 1:400000) были рассчитаны географические расстояния между исследованными районными популяциями.

**Результаты и обсуждение.** Был проведен расчет матрицы генетических расстояний по частотам всех фамилий, на ее основе с использованием кластерного анализа была построена дендрограмма, характеризующая генетические взаимоотношения между 22 исследуемыми районными популяциями (рис.1).

Анализ дендрограммы показал, что исследованные популяции образуют четыре группы кластеров. Первый кластер районных популяций самый многочисленный: он объединяет 10 районов (Яковлевский, Шебекинский, Прохоровский, Корочанский, Ивнянский, Чернянский, Новооскольский, Старооскольский, Губкинский районы Белгородской области и Пристенский район Курской области). Районы данной группы имеют общие территориальные границы и расположены преимущественно в центральной части области. Во второй по численности кластер вошли Волоконовский, Валуйский, Красногвардейский, Алексеевский, Ровеньской и Вейделевский районы, располагающиеся на юго-востоке области и также имеющие общие границы между собой. В третий кластер вошли районы, находящиеся на западе области (Краснояружский, Ракитянский, Грайворонский и Борисовский районы) и также характеризующиеся общими территориальными границами. Красненский район Белгородской области генетически более близок к Репьевскому району Воронежской области (они также имеют общую границу), о чем свидетельствует объединение их в самостоятельную четвертую группу.

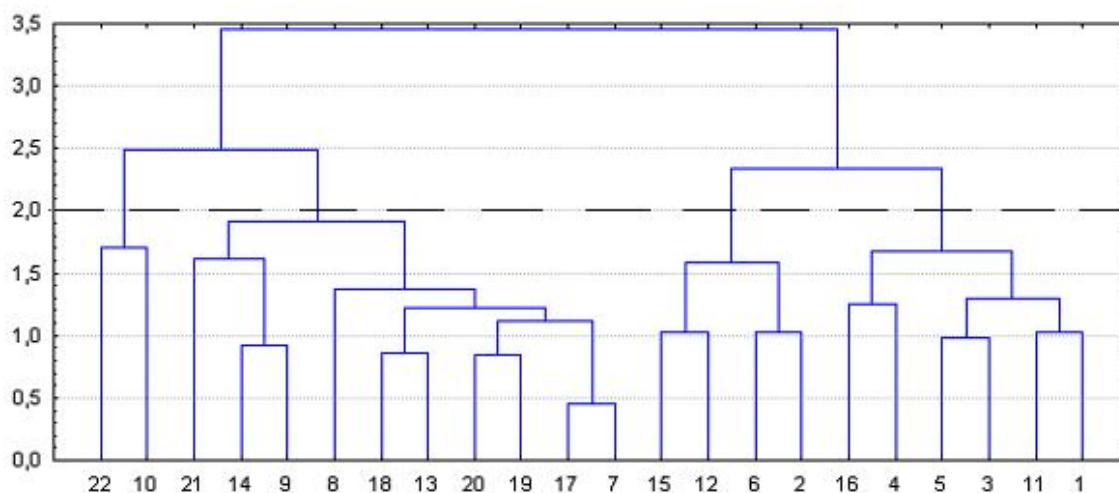


Рис. 1. Дендрограмма генетических взаимоотношений двадцати районов Белгородской области, Пристенского района Курской области и Репьевского района Воронежской области (построена методом Ward's) (пунктирной линией указан уровень выделения значимых кластеров): 1–Алексеевский, 2–Борисовский, 3–Валуйский, 4–Вейделевский, 5–Волоконовский, 6–Грайворонский, 7–Губкинский, 8–Ивнянский, 9–Корочанский, 10–Красненский, 11–Красногвардейский, 12–Краснояржужский, 13–Новооскольский, 14–Прохоровский, 15–Ракитянский, 16–Ровеньской, 17–Старооскольский, 18–Чернянский, 19–Шебекинский, 20–Яковлевский районы Белгородской области, 21–Пристенский район Курской области, 22–Репьевский район Воронежской области.

В соответствии с данными кластерного анализа была построена схема генетического ландшафта всех анализируемых популяций (рис. 2). Эквидистантные линии были проведены через 0,2 единиц генетического расстояния. Пространственное расположение результатов кластерного анализа показало наличие на территории Центрального Черноземья определенной, упорядоченной системы группировки районных популяций. Все районы объединяются в четыре относительно самостоятельных кластера. Важно подчеркнуть, что построенные по результатам кластерного анализа 22 районов эквидистантные фигуры не пересекаются и описывают реальное географическое положение изученных районных популяций. Об этом свидетельствует и значимый положительный коэффициент корреляции Спирмена между матрицами генетических и географических расстояний ( $\rho=0,60, p<0,001$ ).

Другие методы многомерной статистики – многомерное шкалирование (проведенное на основе матрицы генетических расстояний) (рис. 3) и факторный анализ по методу главных компонент (проведен на основе корреляционной матрицы) (рис. 4) показали практически одинаковые результаты, полностью соответствующие данным кластерного анализа. Все анализируемые районные популяции в трехмерном пространстве (рис. 3 и рис. 4) формируют 4 группы кластеров. На графиках, отражающих результаты многомерного шкалирования (рис. 3) и факторного анализа (рис. 4) популяции, объединяющиеся на дендрограмме в отдельные группы кластеров, обведены овалами.

Полученные нами результаты о генетических взаимоотношениях 22 районных популяций Центрального Черноземья, отличаются от результатов работы Г.И. Ельчиновой и др. [8], которые считают, что в русских сельских популяциях при использовании фамилий в качестве генетического маркера, не искаженную оценку генетического родства популяций можно получить для одного района, в крайнем случае, двух-трех сельских районов, если общая граница между ними весьма значительна. В то же время наши результаты свидетельствуют о возможности использования фамилий для корректного описания «генетического ландшафта» крупных областных популяций численностью более 1,5 млн. человек, состоящих из большого количества (более 20) районных популяций.

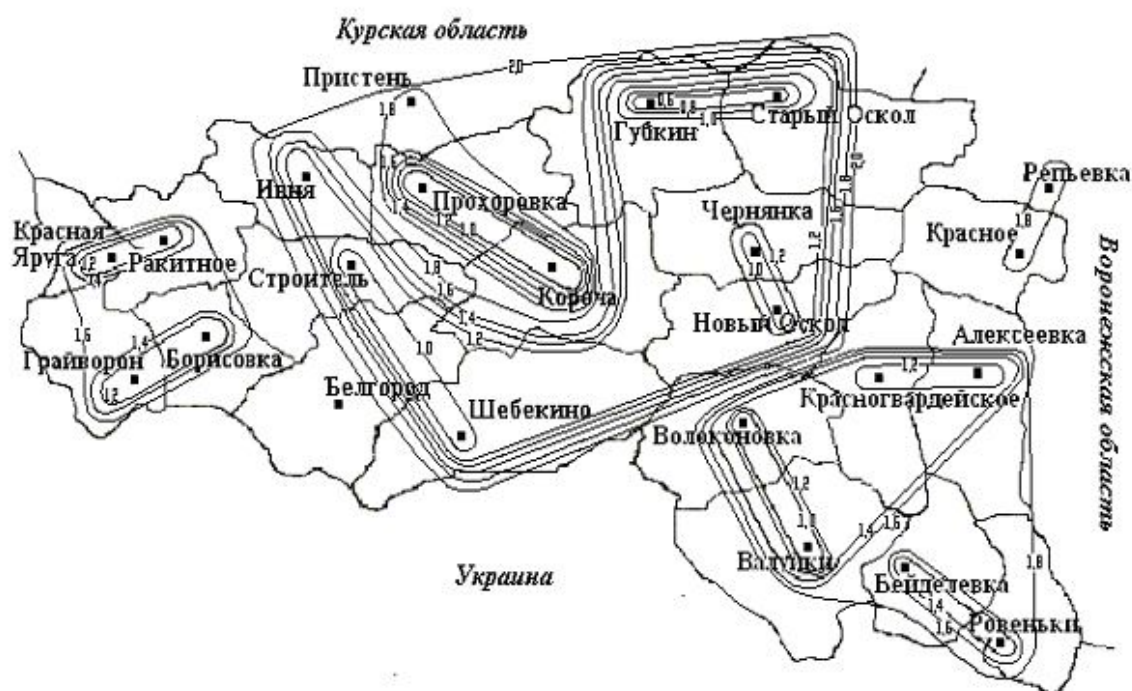


Рис. 2. Схема генетических взаимоотношений районных популяций Белгородской области (по данным о распределении всех фамилий)

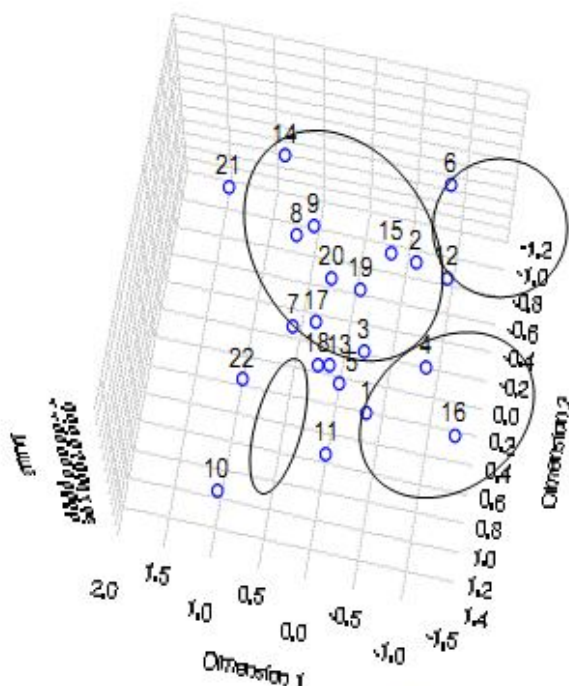


Рис. 3. График трехмерного шкалирования, отражающий расположение в пространстве 20 районных популяций Белгородской области, Пристенского района Курской области и Репьевского района Воронежской области (совершено 48 итераций, величина стресса  $S_0=0,13$ , коэффициент алиенации  $K=0,15$ ):

1-Алексеевский, 2-Борисовский, 3-Валуйский, 4-Вейделевский, 5-Волоконовский, 6-Грайворонский, 7-Губкинский, 8-Ивнянский, 9-Корочанский, 10-Красненский, 11-Красногвардейский, 12-Краснояржский, 13-Новооскольский, 14-Прохоровский, 15-Ракитянский, 16-Ровенькой, 17-Старооскольский, 18-Чернянский, 19-Шебекинский, 20-Яковлевский районы Белгородской области, 21-Пристенский район Курской области, 22-Репьевский район Воронежской области.

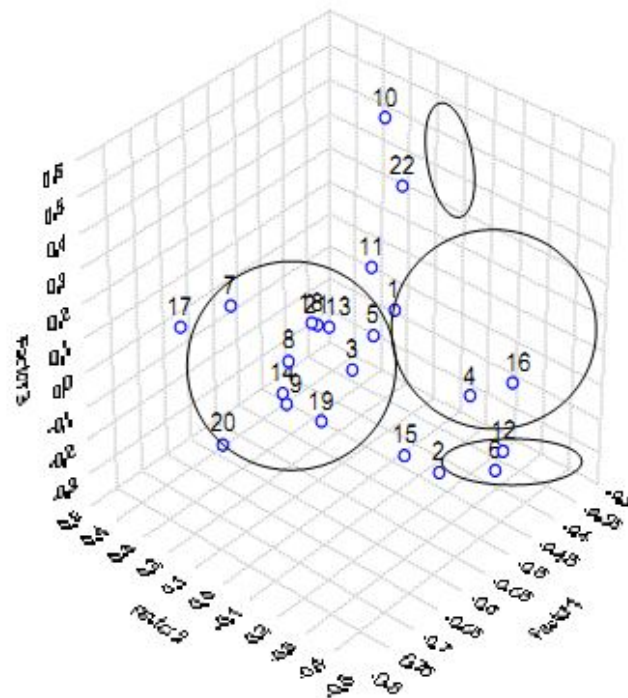


Рис. 4. Размещение в пространстве трех главных факторов изученных районных популяций Белгородской области, Пристенского района Курской области и Репьевского района Воронежской области (ось X – первый фактор, ось Y – второй фактор, ось Z – третий фактор):

1–Алексеевский, 2–Борисовский, 3–Валуйский, 4–Вейделевский, 5–Волоконовский, 6–Грайворонский, 7–Губкинский, 8–Ивнянский, 9–Корочанский, 10–Красненский, 11–Красногвардейский, 12–Краснояржский, 13–Новооскольский, 14–Прохоровский, 15–Ракитянский, 16–Ровенькой, 17–Старооскольский, 18–Чернянский, 19–Шебекинский, 20–Яковлевский районы Белгородской области, 21–Пристенский район Курской области, 22–Репьевский район Воронежской области.

Таким образом, с использованием различных методов многомерной статистики (кластерный анализ, многомерное шкалирование, факторный анализ) на основе анализа матриц генетических расстояний и корреляций, рассчитанных по частотам всех фамилий, проведено моделирование генетического ландшафта населения юга Центральной России, установлено наличие на территории Белгородской области определенной, упорядоченной системы группировки районных популяций в четыре самостоятельных кластера, соответствующей их реальному географическому положению и географическим расстояниям между ними.

**Литература**

1. Дерябин, В.Е. Многомерные биометрические методы для антропологов. / В.Е. Дерябин. – М.: ВИНТИ, 2001. – С.105-265.
2. Пасеков, В.П. Генетические расстояния. В.П. Пасеков / Итоги науки и техники. Сер. Общая генетика. Теоретическая и популяционная генетика, 1983. – Т.8. – С. 3–75.
3. Парадеева, Г.М. Медико-генетическое изучение населения Костромской области. Сообщение III. Коэффициент инбридинга и его динамика в сельских популяциях и городах районного масштаба / Г.М. Парадеева, Л.П. Большакова, А.Н. Петрин // Генетика, 1986. – Т.22, №9. – С. 2355–2360.
4. Ельчинова, Г.И. Выявление особенностей генетической структуры популяции с помощью метода описания «генетического ландшафта» / Г.И. Ельчинова, М.Ю. Кадошникова, Р.А. Мамедова // 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 / Э.К. Хуснутдинова. – Уфа: Гилем, 1999. – 238 с.

7. Мамедова, Р.А. Влияние генетического дрейфа на территориальное распределение груза наследственной патологии в Кировской области / Автореф. дисс...канд. мед. наук. – М., 1993. 23 с.

8. Ельчинова, Г.И. Опыт применения методов популяционно-генетического анализа при изучении популяций России с различной генетико-демографической структурой. / Автореф. дисс. .... д-ра биол. наук. – М., ГУ МГНЦ РАМН, 2001. – 48 с.

9. Сорокина, И.Н. Генофонд населения Белгородской области. Динамика индекса эндогамии в районных популяциях // Генетика, 2008. – Т.44, №8. – С. 1117-1125.

10. Сорокина, И.Н. Генофонд населения Белгородской области. II “Фамильные портреты” в группах районов с разным уровнем подразделенности и роль миграций в их формировании / И.Н. Сорокина, М.И. Чурносков, Е.В. Балановская // Генетика, 2007. – Т.43, №8. – С. 1120-1128.

11. Nei, M. Analysis of gene diversity in subdivided populations / M. Nei // Proc. Nat. Acad. Sci. U.S.A., 1973. – V. 70., № 12. – P. 3321–3323.

12. Nei, M. F-statistics and analysis of gene diversity in subdivided populations / M. Nei // Ann. Hum. Genet., 1977. – V.41. –P. 225-233.

## MODELLING OF GENETIC LANDSCAPE OF THE SOUTH OF CENTRAL RUSSIA REGION

**M.I. Churnosov**  
**I.N. Sorokina**

*Belgorod  
State  
University*

*e-mail: sorokinat@bsu.edu.ru*

In the review using the data of frequencies of all surnames (50412 surnames) among 849399 persons by various methods of multidimensional statistics (cluster analysis, multidimensional scaling, the factorial analysis), presence in the territory of the south of the Central Russia of defined ordered system of grouping of 22 regional populations was established. All areas can be united in four clusters in concordance with real geographical position of populations and geographical distances between them.

Key words: genetic structure of population, surname, genofund.