Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm

FRANZ MANNI,[1] ETIENNE GUÉRARD,[1] AND EVELYNE HEYER[1]

Abstract When sampling locations are known, the association between genetic and geographic distances can be tested by spatial autocorrelation or regression methods. These tests give some clues to the possible shape of the genetic landscape. Nevertheless, correlation analyses fail when attempting to identify where genetic barriers exist, namely, the areas where a given variable shows an abrupt rate of change. To this end, a computational geometry approach is more suitable because it provides the locations and the directions of barriers and because it can show where geographic patterns of two or more variables are similar. In this frame we have implemented Monmonier's (1973) maximum difference algorithm in a new software package to identify genetic barriers. To provide a more realistic representation of the barriers in a genetic landscape, we implemented in the software a significance test by means of bootstrap matrices analysis. As a result, the noise associated with genetic markers can be visualized on a geographic map and the areas where genetic barriers are more robust can be identified. Moreover, this multiple matrices approach can visualize the patterns of variation associated with different markers in the same overall picture. This improved Monmonier's method is highly reliable and can be applied to nongenetic data whenever sampling locations and a distance matrix between corresponding data are available.

The classical way to portray genetic variability is to visualize DNA sequences or populations in dendrograms, with multidimensional scaling (Seber 1984; Torgerson 1958), or in principal-components analysis (Gabriel 1968) plots. These methods display similarities and dissimilarities in a virtual space that corresponds to the plot itself. This approach is particularly suitable for identifying clusters or outliers that are informative about the kind of differentiation underlying the genetic variability. On such bases, different genes, species, populations, etc. can be putatively identified. Nevertheless, when the analysis is put forward to recognize

[1]Département Hommes, Natures, Sociétés, Human Population Genetics Group, CNRS UMR 5145, Musée de l'Homme, 17 Place du Trocadéro, Paris, France.

Human Biology, April 2004, v. 76, no. 2, pp. 173–190.

KEY WORDS: GENETIC BARRIERS, MONMONIER'S ALGORITHM, DELAUNAY TRIANGULATION, BOOTSTRAP ANALYSIS, GENE GEOGRAPHY, GENETIC STRUCTURES

genetic spatial patterns related to geography, these methods of analysis can be less appropriate.

One of the first attempts to study genetic differences in light of geographic distances was published by Malécot (1948). The association was formalized in the so-called model of isolation by distance (IBD), meaning that a double logarithmic regression can be expected between a matrix of genetic distances and a matrix of corresponding geographic distances. The IBD method has been proven effective in a number of studies and is often assumed as a null hypothesis in population-genetic studies. When a significant regression cannot be computed, the association between the two matrices can still be assessed by using the Mantel test (Mantel 1967; Manly 1997), which computes the significance with a permutation approach.

One of the general methods of geographic analysis is spatial autocorrelation, which is used to assess whether or not the genetic and geographic distances are interrelated and whether or not a correlation has a spatial pattern. Spatial autocorrelation measures the level of interdependence between the variables and the nature and strength of the interdependence. Spatial autocorrelation can be either positive or negative; in positive spatial autocorrelation all similar values appear together, whereas in negative spatial autocorrelation dissimilar values appear in close spatial association. In geographic applications there is usually positive spatial autocorrelation. This statistic is generally computed with Moran's I by dividing the spatial covariation by the total variation (several other methods exist). The use of spatial autocorrelation analysis of trend residuals has often been advocated to distinguish IBD from regional or long-distance dispersal processes. Although the correlation methods can assess whether there is an association between genetic and geographic distances, they convey no information about the specific patterns of variation of a given variable in two-dimensional space. The correlation can be high in a certain area and much less significant in another area, and these differences can be hard to see in a regression chart that never says *where* the genetic discontinuities are.

To provide a real graphical representation of genetic differences in geographic space, investigators have suggested alternative approaches. One method consists in performing a principal-components analysis and then plotting separately the first three principal components on the z axis of maps, where the x and y axes are used to plot spatial coordinates (Menozzi et al. 1978). This method is still popular in population genetics even though it does not provide a statistical analysis of the pattern of change of genetic frequencies; rather, it portrays an interpolated genetic landscape in the geographic space.

As a historical note we would like to point out that genogeography, as a field of interdisciplinary investigation, was introduced in 1928 by the Russian geneticist A.S. Serebrovsky. Today, this field of investigation is better known as gene geography. This second method of visualizing geographic space differs from Menozzi's method in that the two key points of the representation are the principle of fusion–fission of genes in the homogeneous geographic space

(equally free-for-all human genes) and the principle of local–linear (but not of the high orders) interpolation of gene frequencies onto a spherical surface of geographic space (Rychkov et al. 1990).

A third method of geographic visualization of patterns was originally published by Womble (1951) and rediscovered by Barbujani et al. (1989). This method focuses on a detailed visualization of the geographic areas associated with a considerable genetic change, what we are going to call boundaries or barriers. The method allows different variables within the same landscape to be considered together. First, individual surfaces are differentiated such that steep slopes become peaks and flat plains fall to zero. Second, the magnitudes of the derivatives of surfaces from different variables can be added to get a composite picture of barriers derived from all variables.

Womble's method has been further developed (Barbujani et al. 1989; Oden et al. 1993; Bocquet-Appel and Bacro 1994), and within a continuous landscape a consideration of significance was introduced to determine how high the cutoff is between being a barrier and not being one. By introducing a percentile consideration of significance (i.e., by considering values in the top $X\%$ to represent barriers), values within the landscape can be compared, thus controlling for the effect of IBD, if it is applicable. A limitation of the Womble procedure, as well as of the principal-components method (synthetic spatial maps of gene frequencies), is that it implies an interpolation of the landscape, which leads to potential artifactual continuities or discontinuities (Sokal et al. 1989, 1999a, 1999b).

Because genetic structures can show correspondence with geography, general methods of geographic analysis can be successfully applied to population genetics. This implies the computation of neighboring problems (computational geometry). Here, we discuss a fourth method, Monmonier's maximum difference algorithm (Monmonier 1973), which was designed for the visualization on a geographic map of the trend data contained in matrices. The algorithm finds the edges associated with the highest rate of change in a given distance measure, which can be genetic, morphologic, or something else. The algorithm is applied to a geometric network that connects all the populations (sampled locations) using Delaunay triangulation (Brassel and Reif 1973).

Even though we have advocated using Monmonier's method in genetic studies (Manni and Barrai 2000, 2001; Manni and Pagnoni 2001; Manni et al. 2002; Palmé et al. 2003) after the early publication of Barbujani et al.'s (1996) paper, we never provided a detailed discussion of Monmonier's method. To this end, we have studied the geometric constraints of the triangulation on Monmonier's algorithm and have developed a method to test the significance of boundaries using an analysis of bootstrap matrices, because no suitable statistics have been proposed to date. This improved version of the method has been implemented in a software package, available upon request, called Barriers. The identification of the most significant barriers can be generalized to all cases where a distance matrix between items is available and where the sample locations are known.

Materials and Methods

The Triangulation. Delaunay triangulation (Brassel and Reif 1979) is the fastest triangulation method for connecting a set of points (localities) on a plane (map) with a set of triangles (Figure 1A). It is the most direct way to connect (triangulate) adjacent points on a map. Delaunay triangulation is the dual structure of the Voronoi diagram (Voronoi 1908), and one can be derived from the other (Figure 1). Given a set of populations whose geographic locations are known, however, only Delaunay triangulation is possible. Voronoi diagrams, as defined by the author, imply that all possible points inside a polygon are closest to its centroid (the location of the sampled population) than to any other polygon (Figure 1B). It means that we divide the geographic space S into m subspaces S_i, satisfying the following properties:

$$\cup_i S_i = S, \tag{1}$$

$$S_i \cap S_j = \varnothing, \qquad \forall i \neq j, \tag{2}$$

$$\mathrm{Dist}(x_k, w_i) < \mathrm{Dist}(x_k, w_j), \qquad \forall i \neq j,\ x_k \in S_i, \tag{3}$$

where w_i is the centroid of S_i.

Once a network connecting all the localities has been obtained, each edge of the network is associated with its distance value from a matrix, as shown in Figure 2.

Monmonier's Algorithm. Monmonier's (1973) maximum difference algorithm is used to identify boundaries, namely, the areas where differences between pairs of populations are largest. The first boundary is traced perpendicular to the edges of the network (Figure 2). Starting from the edge for which the distance value is maximum and proceeding across adjacent edges, the procedure is continued until the forming boundary has reached either the limits of the triangulation (map) or closes on itself by forming a loop around a population. For multiple barriers that are constructed one after another in a hierarchical order according to user settings, the procedure can stop at a previously computed boundary (Figure 2). Note that when two edges have the same value, the one followed by a triangle with higher values is included in the boundary.

Sample Coordinates. When populations are sampled on a flat surface, they can easily be connected by a Delaunay triangulation. On the contrary, some problems can arise when samples lie on a curved surface, such as the earth's surface, because it is not possible to project the position of these samples on a plane without some kind of error, whatever the kind of projection. As an example, because the geographic representation of geographic maps is more exact in its central part, the measuring of x and y coordinates of samples located near the

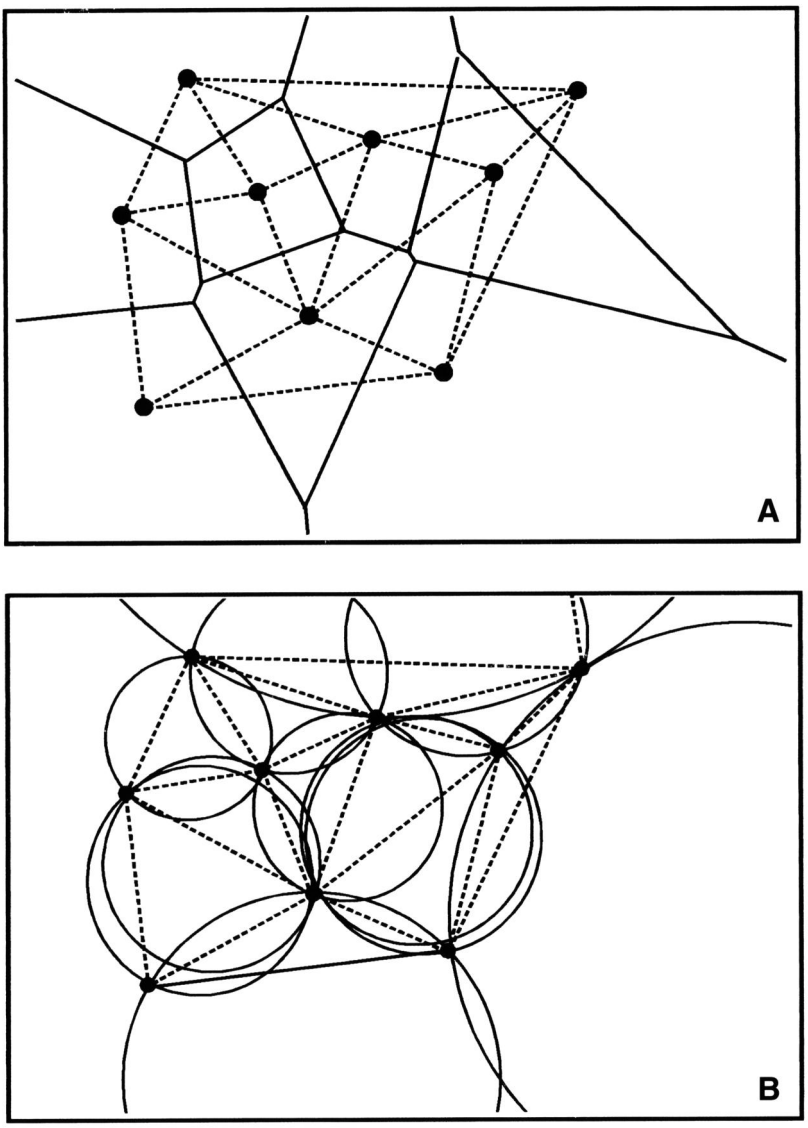

Figure 1. (A) When populations (solid dots) are located on a surface, a Voronoi tessellation (solid lines) and the corresponding Delaunay triangulation (dotted lines) can be computed. Because of their geometric properties, one can be obtained from the other. Monmonier's algorithm computes barriers that lie on the Voronoi tessellation, because their edges are equidistant from pairs of populations. (B) The geometric definition of Voronoi diagrams implies that all points inside a polygon are closer to its centroid (solid dots) than to any other polygon. This implies that any triangle in the Delaunay triangulation (dotted lines) contains no populations and therefore can be inscribed in an empty circle. Populations (solid dots) lie on the circumference of the circle (circumcircle property).

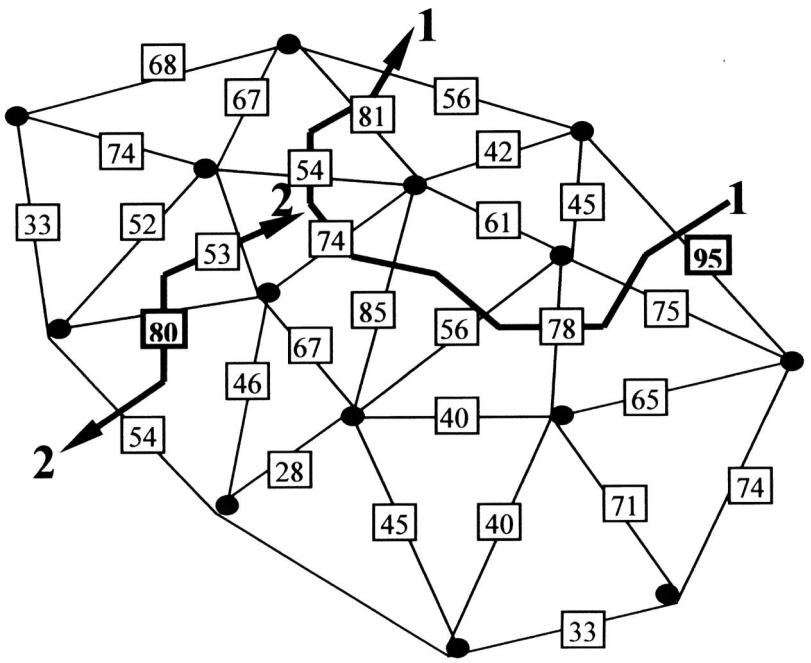

Figure 2. Example of computation of a barrier using Monmonier's algorithm. Once a triangulation between populations (solid dots) is obtained (solid lines), the edges are associated with pairwise distance measures according to the distance matrix used (genetic, morphologic, etc.). Then the highest distance measure associated with the triangulation (95 in the example) is taken as the starting edge of the first computing barrier. The barrier is always extended across the edge associated with the highest distance, 75 instead of 45, then 78 instead of 65, and after that 56 instead of 40, etc. The procedure is continued until the forming boundary has reached either the limits of the triangulation or has closed on itself by forming a loop around a population. For multiple barriers, which are constructed one after another in a hierarchical order according to user settings, the process stops when a preexisting barrier is reached (e.g., barrier 2 in the figure). The starting edge of a barrier can be either at a border of the triangulation (barrier 1) or inside it (barrier 2). In the first case the extension takes place in only one direction; in the second case the extension takes place in two different directions, as the arrows show.

borders will be affected by a considerable error. These distortions, related to projection of a curved surface on a plane, can result in a triangulation different from the one obtainable in a curved space; thus the geometry of barriers may be affected. A possible way to overcome this limitation is to compute a matrix of the real geographic distances (on the curved surface) between the points, project the position of samples on a multidimensional scaling plot, and use the new coordinates to compute the Delaunay triangulation. In this way the topological errors in the definition of sample locations will be averaged, giving, as a result, a more accurate triangulation.

Testing the Significance of Barriers Using a Multiple Matrix Approach. The definition of Monmonier's algorithm recalls the dichotomous process of arborescence of phylogenetic trees because, once a barrier passes across the edge of a triangle, it can be extended only across one of the two remaining edges, in what we define as a right or left decision. To assess the statistical significance of computed barriers, we have implemented a test that is based on the analysis of resampled bootstrap matrices (e.g., from molecular sequences). As with bootstrap phylogenetic trees, a score is associated with all the different edges that constitute barriers and indicates how many times each edge is included in one of the N boundaries computed from the N matrices (typically $N \geqslant 100$). The scores are visualized by the Barriers software by representing the thickness of each edge proportionally to its bootstrap score (Figure 3).

The Barriers Software. The software we have developed computes barriers on a Delaunay triangulation using Monmonier's algorithm and runs under Microsoft Windows (Windows 2000 or higher recommended). Minimal system requirements are 128 MB of RAM and a Pentium II processor. A good video card is recommended to get a fast visualization of ongoing analyses. The program has a graphic clicky interface.

Input files (data coordinates and the distance matrix) can be imported as text files by using an interactive window that provides a preview of the resulting file in order to correctly set the line and column from which the import has to be started. Computed triangulation and barriers are saved as specific files with the extension .dvb (as a contraction of the words Delaunay, Voronoi, and barriers). Results can be exported in Windows BMP format (*.bmp) and in Postscript vectorial format (*.ps). For each run of barrier analysis a report file (in text format) is generated. The report gives all the details of the different steps of the algorithm in constructing the barrier, edge by edge. Options such as color, thickness of Voronoi tessellation, Delaunay triangulation, barriers, and sample points can also be set.

The most common bug is related to the presence of samples listed with identical x and y coordinates. If this happens, the triangulation between these sample points cannot be computed. Another bug is related to matrices with identical distance measures, because in this case the right or left decision cannot be made. The way out, suggested by Barbujani et al. (1996), is to include in the barrier the edge associated with the shortest geographic distance. This solution may not be appropriate because it implicitly assumes the IBD model, even when it was not tested. We undertook a more conservative approach by including, in forming the barrier, the edge that drives the boundary toward the triangle (of the two possible ones) associated with the higher distance measure.

Testing Barriers: Two Experimental Examples. To illustrate the wide range of applications of Monmonier's algorithm, we discuss the application of

the method to genetic and surname data. Real experimental examples were chosen to avoid the oversimplification of simulated samples.

Genetic Data. We present the results of the typing of unique event polymorphisms (UEPs) on the nonrecombinant region of the human Y chromosome (Manni et al. 2002). Seventeen populations were sampled around the Mediterranean basin, resulting in data for 650 individuals (Figure 3). An F_{st} distance matrix between these population was then computed. These markers enable the definition of haplotypes whose frequencies can be helpful in identifying genetic differences between populations, differences that are related to the populations' past demographic history.

Surnames. Surnames can be considered one locus on the nonrecombining portion of the Y chromosome, and their analysis allows us to infer genetic structures of populations. We studied the surnames of 2.4 million Dutch telephone subscribers (Manni 2001) and computed Lasker's distance matrix. The significance of barriers was further tested by resampling original surnames and then recomputing barriers from 100 bootstrap matrices (Figure 4). This example was chosen because of the huge sample and because the geometry of the sampling grid provided an excellent case study.

Results

Y-Chromosome Variation. The first barrier, computed on the F_{st} distance matrix, divides western European populations from the surrounding populations (top of Figure 3). In this example we have plotted the thickness of the barrier proportionally to the inverse ratio between the higher F_{st} distance (0.603) and F_{st} values associated with crossed edges (0.603/0.603, 0.389/0.603, 0.376/0.603, etc.). This method provides a basic significance test that can be applied to barrier analysis when bootstrap matrices are not available. The results show that the thickness of the barrier decreases from the Strait of Gibraltar to southern Italy, thus suggesting that the genetic barrier is strong only between northwestern Africa and the Iberian Peninsula. These results are in agreement with several previously published papers, which we will not cite here because we are discussing only the method.

We can get a deeper insight into Monmonier's method by comparing the barrier analysis and a multidimensional scaling plot built from the same F_{st} distances. A multidimensional scaling analysis (bottom of Figure 3) indicates that there is no perfect association between genetic and geographic distances, because populations are unequally spaced on the plot, thus suggesting that genetic barriers could exist. Because in the multidimensional scaling plot there is only one well-defined cluster, formed by the two Moroccan samples, a genetic barrier separating them from all the other populations could be expected, in contrast to the

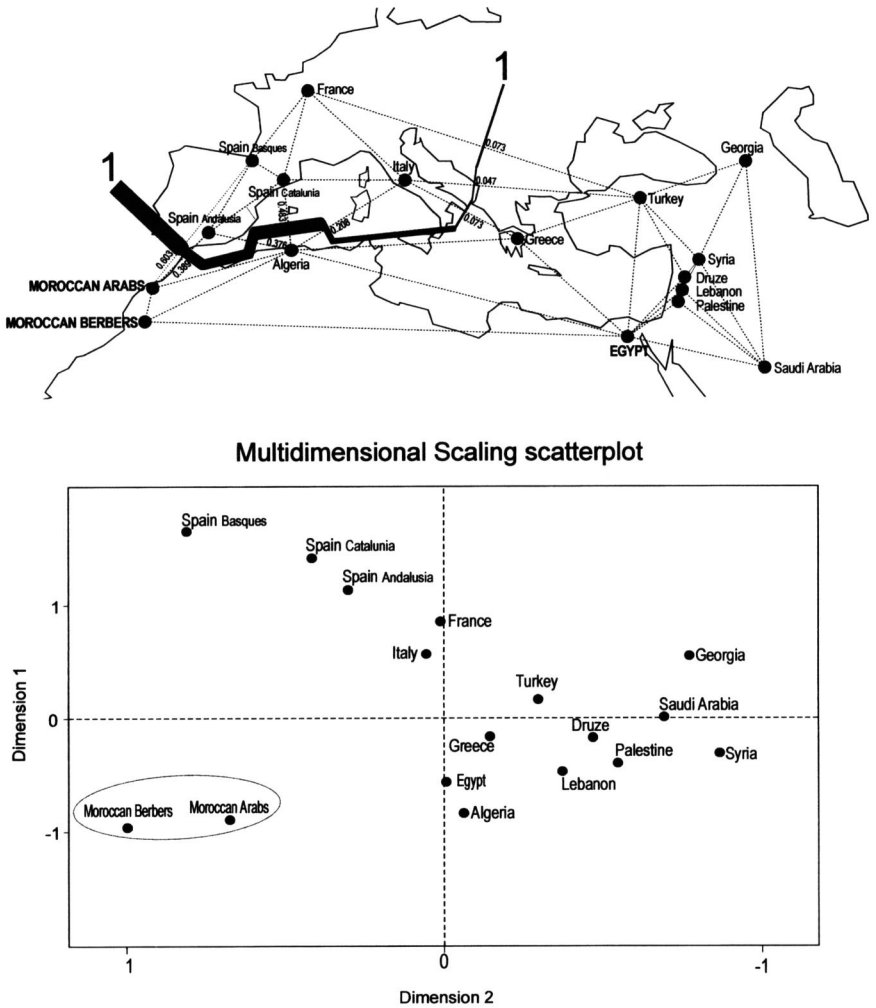

Figure 3. Human Y-chromosome differences around the Mediterranean basin. (Top) A Delaunay triangulation (dotted lines) and the first genetic barrier (solid line) computed on a F_{st} distance matrix between populations. [Redrawn from Manni et al. (2002).] (Bottom) The multidimensional scaling analysis representing the same F_{st} distance matrix as in the top figure. Multidimensional scaling takes a set of dissimilarities (as in a distance matrix) and returns a set of points such that distances between the points in the plot are approximately equal to the dissimilarities. [Redrawn from Manni et al. (2002).]

obtained results (top of Figure 3). This apparent discrepancy clearly points to the differences between a classic multivariate analysis and a geographic one, which is illustrated in Figure 5.

The computed barrier does not enclose the Moroccan populations because the genetic distances between the Algerian and the Moroccan samples are lower

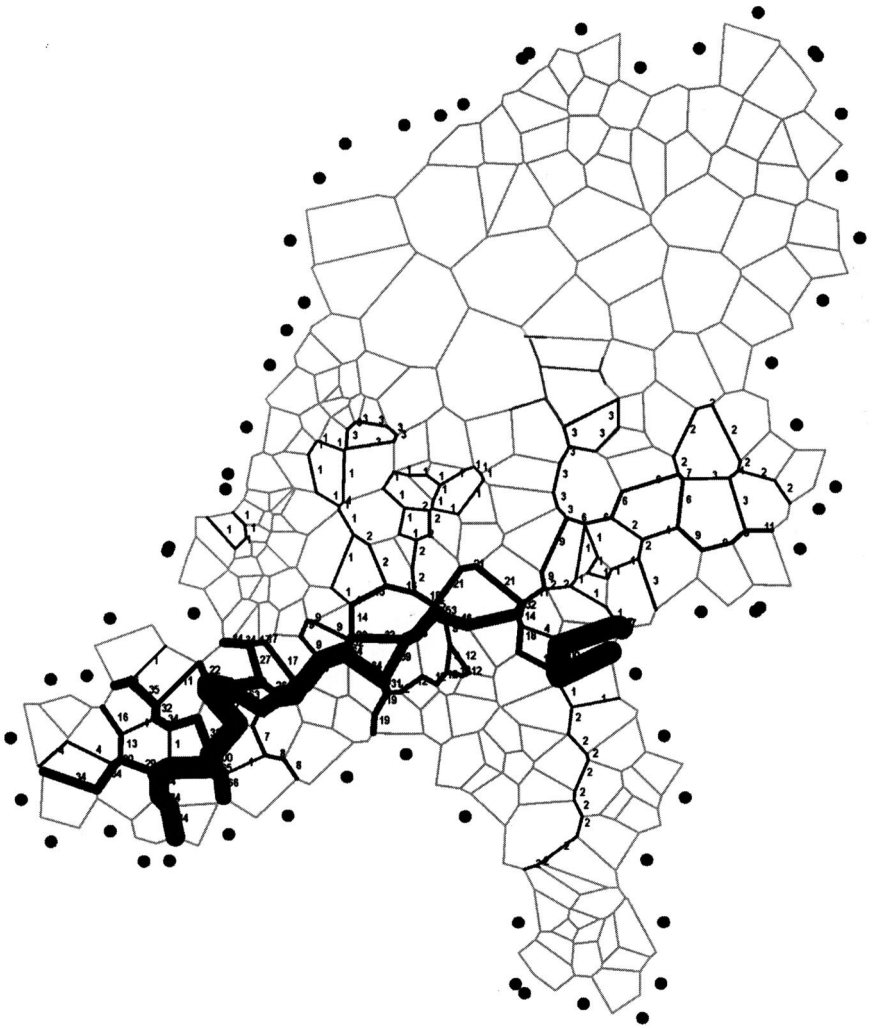

Figure 4. We analyzed the first five barriers (in black) on 100 bootstrap matrices obtained by randomly resampling original surnames. The thickness of each edge of a barrier is proportional to the number of times it was included in one of the 500 computed barriers (small numbers). In gray is the Voronoi tessellation (the Delaunay triangulation is not shown). Small black dots outside the triangulation are the virtual points used to close the Voronoi tessellation (see text and Figure 6A). The analysis refers to the surname distribution of 226 localities. Differences were also summarized in a single overall distance matrix used to compute the first five barriers with Monmonier's method (Figure 6A).

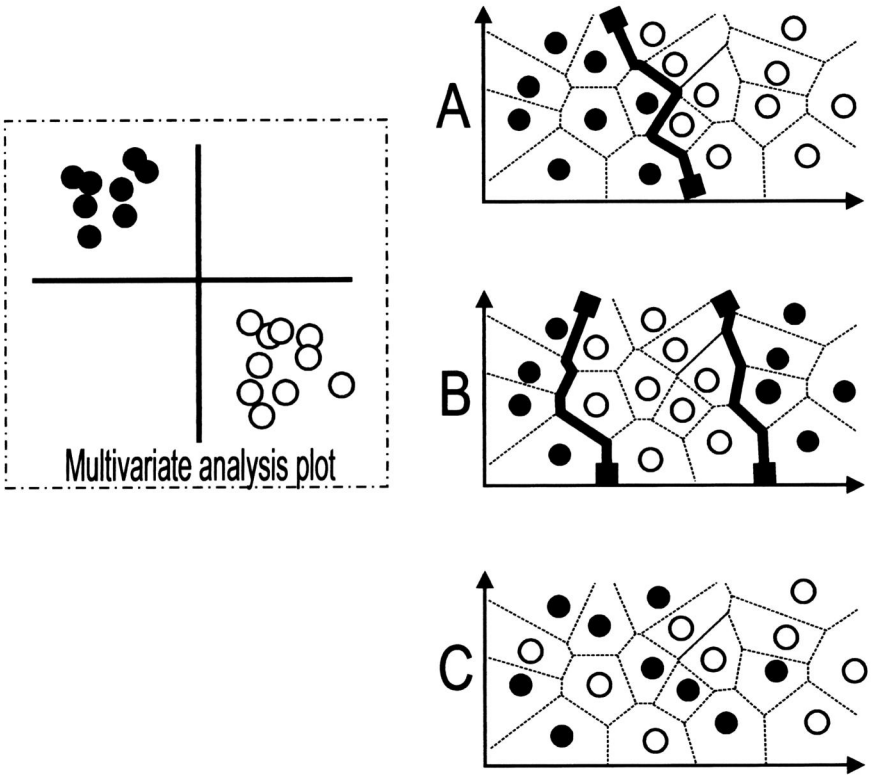

Figure 5. Three different scenarios corresponding to the same multivariate analysis plot (principal-components analysis, multidimensional scaling, etc.). Eighteen populations belong to two different and nonoverlapping groups (black and white dots). (A) All the populations belonging to the two clusters are spatially contiguous; therefore one main genetic barrier is expected to be present between them. (B) The two clusters can still be identified, but only the white populations are geographically contiguous; therefore two main genetic barriers are expected. (C) The population is intermixed, so there is no correspondence between the geographic location of a population and its genetic (morphologic) differences. Thus no main barriers are expected to occur.

than the distances between the Algerian and the Spanish populations. As a consequence, the geometry of the boundary mirrors the differences between the Iberian Peninsula and northwestern Africa. Besides this differentiation, the remaining European and Middle Eastern populations are closer on the multidimensional scaling plot, as suggested by the decreasing thickness of the barrier in its final part (top of Figure 3).

Genetic Structures of the Netherlands Inferred from Surnames. We applied Monmonier's method to identify surname barriers (the zones where the differences in the distribution of surnames are maximized) from a matrix of Lasker's distances among 226 sampled localities. We computed the first five barriers

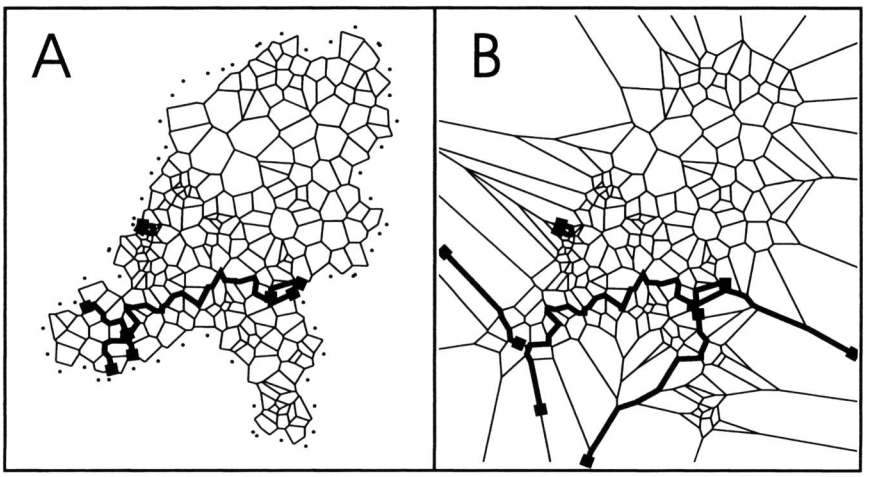

Figure 6. By definition, the more external part of a Voronoi tessellation (in gray) tends to infinity (part B). It often happens that one of these borders is coupled with the highest genetic distance value of the matrix; therefore the origin of the barrier is outside the triangulation itself. We compare the first five barriers after (part A) and before (part B) adding the virtual points (small dots in part A) that close the Voronoi tessellation. Virtual points can be considered virtual populations that locally modify the neighborhood of the triangulation (not shown), thus being interpreted as the borders. In part B barriers originate outside the triangulation, in contrast to the definition of the method. The example refers to surname differences in The Netherlands (see text and Figure 4).

that were contiguous to each other (Figure 6A). The geometry of these barriers suggests a main differentiation zone between the north and the south of the Netherlands. The significance of the boundaries was tested by a multiple analysis of 100 bootstrap matrices. The results confirm that the main differentiation in the distributions of Dutch surnames occurs along a southwestern-northeastern direction (see Figure 4).

It should be noted that bootstrap analysis conveys additional information because it shows that fragmentation patterns are more complex than suggested by the previous analysis of a single overall matrix, particularly in the southwestern area (Figure 4). To satisfy the curiosity of the reader, we will say that obtained barriers almost perfectly overlap the boundary between Roman Catholics and Protestants (van Heek 1954). Few mixed marriages between the two religious groups took place in the past centuries, an occurrence that is mirrored by the surname distribution.

Editing the Triangulation. It is possible to edit the triangulation by adding virtual points (see Figure 6). Virtual points locally modify the neighborhood being interpreted as the borders of the triangulation. This addition is of great importance because the external links of a Voronoi tessellation, by definition, tend to approach infinity (Figure 6). It often happens that one of these borders is

coupled with the highest genetic distance value of the matrix; therefore the origin of the barrier will appear outside the triangulation itself. In Figure 6 we compare the first five barriers after (Figure 6A) and before (Figure 6B) adding the virtual points that close the Voronoi diagrams. Any other triangulation program that does not correct this geometric property of Voronoi diagrams is likely to create fake barriers when Monmonier's algorithm is applied to it.

In summary, the editing of triangulation enables the user to adapt the network to specific features of the geographic space as, for example, in the case of deserts or internal lakes. Moreover, this tool can be useful to delete some long links between distant populations; this is often the case with external samples when the general shape of the triangulation is not convex, because the samples will be considered adjacent in a Voronoi tessellation. An extreme case of removing external links is provided in Figure 6A, where all the triangulation lies inside the administrative borders of the Netherlands and only links between close neighbors are preserved. A further example is shown in Figure 3; here, the long link between the French and the Georgian samples was removed (in this case its presence or absence does not change the geometry of the barrier; data not shown).

Redundancy of Data. The possibility of analyzing multiple matrices enables the separate analysis of single markers, thus visualizing the degree of redundancy in the data. Two scenarios can arise: (1) A large proportion of the markers can exhibit the same geographic pattern of variation, or (2) almost each different marker (sequence) can show a different geographic pattern. This multimatrix analysis gives a more realistic view of the noise associated with each marker and enables the user to get estimates about the markers' informative power. The separate plot of barriers obtained from different matrices can be helpful in deciding on the number of different markers to be analyzed to get accurate results. If all patterns are different, that could mean that each marker adds some information to an overall distance matrix and that new markers might lead to more detailed results. If recurrent patterns are observed, it means that the number of studied markers is sufficient and that no additional markers are needed to improve barrier detection; that is, there is redundancy in the data. This case applies to the example on Dutch surnames (Figure 4), because the patterns of resampled barriers are geographically quite stable, meaning that different surnames (randomly resampled in bootstrap matrices) drive to the same boundary shapes. This conclusion was expected because 2.4 million surnames were studied. At any rate, even when dealing with large samples, some differences among the computed patterns can still be observed. They give a more realistic vision of geographic patterns of variability when compared with the analysis of a single matrix (Figure 6A).

Discussion

The Ideal Case. When attempting to identify boundaries with Monmonier's algorithm, the best results will be obtained with regularly spaced populations,

where the area under investigation approximates a convex polygon. Irregularly spaced populations can lead to ambiguous results because barriers tend to fall between the most widely spaced populations, which under an IBD model are logically expected to be significantly different from one another. In such a case it is impossible to discriminate whether there is a real change in genetic features among populations separated by the boundary or whether there is a regular genetic pattern of change (related to gene flow) that is not detected because intermediate populations were not sampled. This issue points to the crucial definition of the sampling grid, which should be considered before the sampling is performed to ensure more accurate analyses. As a general rule, large unsampled areas should be avoided when possible. One of the merits of a triangulation-based method is that the irregular distribution of samples is not masked by interpolation, as it is in geographic principal-components analysis (Menozzi et al. 1978), Womble's method (Womble 1951), or the genogeographic approach (Rychkov et al. 1990).

Some topological configurations of sampled points are not suitable for Monmonier's barrier analyses. A good example of inappropriate data is represented by transects, because sampled points are monodimensionally distributed. When the algorithm is applied to such cases, the barriers are forced to cut the triangulation vertically, where the highest pairwise distances among contiguous populations are. Moreover, caution should be taken in concave-shaped triangulations in which some links connect distant populations. These long links can result in factitious barriers because they are probably associated with the highest distances in the matrix, as in an IBD model, as discussed earlier. In such cases it may be useful to test alternative scenarios by editing the triangulation with virtual points to remove these links. The possibility of close Voronoi tessellation, implemented in the software, seems necessary because it is the only way, besides computing barriers by hand, to correctly apply Monmonier's method without obtaining barriers originating outside the triangulation (Figure 6B).

Population Genetics and Isolation by Distance. Barriers represent zones of abrupt change in the pattern of genetic variation. This means that when populations fit the IBD model, the chances of finding discontinuities in the pattern of change dwindle. On the other hand, the presence of isolating factors (cultural, geographic, or morphologic) is likely to weaken the gene flow by increasing the chances of finding significant barriers. In general, there are many barriers when the associations between genetic and geographic distances are low, and vice versa. This does not mean that the algorithm cannot be applied to those cases where genetic distances are demonstrated to increase according to geographic distances. When this dependence can be statistically tested as significant, a good strategy is to compute a matrix of expected genetic distances (according to the kind of regression computed) and to subtract it from the original matrix, thus obtaining a matrix of residuals The resulting new matrix will enable us to get a further evaluation of the pattern of barriers because it will portray the variation not related to migration phenomena.

Range of Application of the Method. Because only spatial coordinates and a distance matrix (whatever the kind of measure) are needed to run Monmonier's algorithm, the areas of application can be as wide as the computational sciences. We note that when a distance matrix cannot be obtained, the method is not applicable. In this respect, Monmonier's (1973) method differs from Womble's (1951) approach because Womble's method does not handle distance matrices and enables the analysis only of frequency vectors (alleles, haplotypes, linguistic features, morphologic traits). We have applied Monmonier's algorithm to linguistics (Manni 2001), to anthropological data (Demeter et al. 2003), to surnames (Manni 2001; Manni and Barrai 2000, 2001), and to genetics (Manni 2001; Palmé et al. 2003; Manni et al. 2002). This wide range of possible applications increases the chances to undertake multidisciplinary and comparative studies that would be impossible with other approaches. Moreover, the application of Monmonier's method can be useful in a number of biological and medical studies, for example, identification of areas where disease rates change rapidly or identification of different expression rates of genes.

A further, indisputable merit of Monmonier's algorithm is its simplicity, which enables the user to supervise the details of ongoing barrier analysis. This simplicity goes together with reliable results, as recently shown by Dupanloup et al. (2002). Dupanloup compared Monmonier's algorithm with a new method (SAMOVA, for spatial analysis of molecular variance) to identify highly differentiated populations on a geographic landscape. The SAMOVA, an addendum to the Arlequin package (Schneider et al. 2000), is intended to identify maximally differentiated groups of populations without the a priori definition needed before (e.g., according to genetic, morphological, or linguistic classifications). In this respect, SAMOVA, because it does not deal with a truly geometric approach, is able to identify maximally differentiated populations, whereas Monmonier's approach is better at finding genetic barriers between sets of populations (Dupanloup et al. 2002). This latter property of Monmonier's algorithm seems particularly interesting because maximally differentiated samples may already be visible in multivariate analysis plots (principal-components analysis or multidimensional scaling), whereas barriers are not (see Figure 5).

Another advantage of our improved version of Monmonier's algorithm is that the SAMOVA, in its currently available version, is applicable only to genetic data (sequences, alleles), whereas Monmonier's approach can be applied to any kind of data if differences are computed as a distance matrix. Moreover, the geometric properties of Monmonier's method are of particular interest because they make possible an estimate of the robustness of the different edges of the computed barriers. In this sense we advocate in resampling techniques (bootstrap, jackknife, etc.) a more appropriate way to obtain a realistic picture of the significance of barriers, because results can illustrate the differential robustness of the different edges formed. According to their differences, the two methods were proved to behave differently in population-genetic studies. In a simulation-based approach Dupanloup et al. (2002) showed that Monmonier's algorithm works

better than SAMOVA in finding highly differentiated population groups, especially when the amount of gene flow within groups is small. In addition, the performance of both methods decreases when there is a high level of gene flow between groups and when gene flow within groups is low compared to the gene flow between groups (Dupanloup et al. 2002). These findings can be considered a different way to state the already cited antagonism between the presence of genetic barriers and an IBD scenario, because it is trivial that no significant barriers can exist in a perfect cline of frequencies. In this perspective SAMOVA, more than a method to trace boundaries, can be considered a test of significance for population clusters. Thus a Monmonier boundary plot should always be compared with the corresponding plots from multidimensional scaling or principal-components analysis because the two approaches are complementary (the difference between them being the additional information conveyed by Monmonier's analysis), and they can reach cogent conclusions only when undertaken and discussed together (see Figures 3 and 5).

Conclusions

Monmonier's maximum difference algorithm enables a better interpretation of microevolutionary processes, such as gene flow, genetic drift, and selection. It also helps to identify hidden boundaries resulting from secondary gene flow among previously isolated populations. The application of Monmonier's algorithm can lead to an understanding of the processes that caused the patterns. The application to population genetics using geographic methods and techniques of analysis seems necessary for a better understanding of the environmental constraints on human demography. Similarly, in ecological studies phylogeographic approaches progressively shift toward a landscape genetics approach (Manel et al. 2003). There is an increasing interest in those statistical investigations focused on the detection of selection processes across geographic space (Nielsen 2001), and the popularization of Monmonier's algorithm is likely to put forward the understanding of such phenomena. Future developments will probably be directed toward comparing different boundary maps, thus enabling the geometric assessment of the similarities and dissimilarities between the geographic patterns of variation of different variables (Jacquez 1995). Monmonier's algorithm, because it can be applied to any kind of distance matrix and is independent of the metrics adopted, is a promising method to directly compare the patterns of genetic, cultural, or ethnological differentiation.

Acknowledgments We wish to thank M. Schroor and G. Barbujani for their past help. We are also indebted to S. Blanc-Manel and G. Luikart for reading the manuscript.

Received 2 September 2003; revision received 3 November 2003.

Literature Cited

Barbujani, G., G.M. Jacquez, and L. Ligi. 1990. Diversity of some gene frequencies in European and Asian populations. V. Steep multilocus clines. *Am. J. Hum. Genet.* 47:867–875.

Barbujani, G., N.L. Oden, and R.R. Sokal. 1989. Detecting areas of abrupt change in map of biological variables. *Syst. Zool.* 38:376–389.

Barbujani, G., M. Stenico, L. Excoffier et al. 1996. Mitochondrial DNA sequence variation across linguistic and geographic boundaries in Italy. *Hum. Biol.* 68:201–205.

Bocquet-Appel, J.P., and J.N. Bacro. 1994. Generalized wombling. *Syst. Zool.* 43:442–448.

Brassel, K.E., and D. Reif. 1979. A procedure to generate Thiessen polygons. *Geogr. Anal.* 325:31–36.

Demeter, F., F. Manni, and Y. Coppens. 2003. Late Upper Pleistocene human peopling of the Far East: Multivariate analysis and geographic patterns of variation. *C. R. Palevol.* 2:625–638.

Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. *Molec. Ecol.* 11:2,571–2,581.

Epperson, B., and T. Li. 1996. Measurement of genetic structure within populations using Moran's spatial autocorrelation statistics. *Proc. Natl. Acad. Sci. USA* 93:10,528–10,532.

Gabriel, K.R. 1968. The biplot graphical display of matrices with application to principal component analysis. *Biometrika* 58:453–467.

Jacquez, G.A. 1995. The map comparison problem: Tests for the overlap of geographic boundaries. *Stat. Med.* 14:2,343–2,361.

Malécot, G. 1948. *Les mathematiques de l'hérédité.* Paris: Masson.

Manel, S., M.K. Schwartz, G. Luikart et al. 2003. Landscape genetics: Combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18:189–197.

Manly, B.F.J. 1997. *Randomization, Bootstrap, and Monte Carlo Methods in Biology,* 2nd ed. London: Chapman and Hall.

Manni, F. 2001. *Strutture genetiche e differenze linguistiche: Un approccio comparato a livello micro e macro regionale.* Doctoral thesis. Ferrara, Italy: University of Ferrara.

Manni, F., and I. Barrai. 2000. Patterns of genetic and linguistic variation in Italy: A case study. In *Archaeogenetics: DNA and the Population Prehistory of Europe,* C. Renfrew and K. Boyle, eds. Cambridge: McDonald Institute for Archeological Research Monographs, 333–338.

Manni, F., and I. Barrai. 2001. Genetic structures and linguistic boundaries in Italy: A microregional approach. *Hum. Biol.* 73:335–347.

Manni, F., P. Leonardi, A. Barakat et al. 2002. Y-chromosome analysis in Egypt suggests a genetic regional continuity in northeastern Africa. *Hum. Biol.* 74:645–658.

Manni, F., and A. Pagnoni. 2001. Lineamenti genetici e linguistici del territorio ferrarese. In *Storia di Ferrara*, v. 1, *Territorio e Preistoria.* Ferrara, Italy: Corbo Editore, 409–426.

Mantel, N.A. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.

Menozzi, P., A. Piazza, and L.L. Cavalli-Sforza. 1978. Synthetic maps of human gene frequencies in Europeans. *Science* 201:786–792.

Monmonier, M. 1973. Maximum-difference barriers: An alternative numerical regionalization method. *Geogr. Anal.* 3:245–261.

Nielsen, R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* 86:641–647.

Oden, N.L., R.R. Sokal, M.-J. Fortin et al. 1993. Categorical wombling: Detecting regions of significant change in spatially located categorical variables. *Geogr. Anal.* 25:315–336.

Palmé, A.E., Q. Su, A. Rautenberg et al. 2003. Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula. Molec. Ecol.* 12:201–212.

Rychkov, Iu.G., A.V. Rychkov, E.V. Balanovskaia et al. 1990. Genogeography of human populations: Computer mapping of population genetics data. *Genetika* 26:332–340 (in Russian).

Schneider, S., D. Roessli, and L. Excoffier. 2000. *Arlequin, ver. 2.0: A Software for Population*

Genetics Data Analysis. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Seber, G.A.F. 1984. *Multivariate Analysis.* New York: Wiley.

Sokal, R.R., R.M. Harding, and N.L. Oden. 1989. Spatial patterns of human gene frequencies in Europe. *Am. J. Phys. Anthropol.* 80:267–294.

Sokal, R.R., N.L. Oden, B.A. Thompson et al. 1993. Testing for regional differences in means: Distinguishing inherent from spurious spatial autocorrelation by restricted randomizations. *Geogr. Anal.* 25:199–210.

Sokal, R.R., N.L. Oden, and B.A. Thomson. 1999a. A problem with synthetic maps. *Hum. Biol.* 71:1–13.

Sokal, R.R., N.L. Oden, and B.A. Thomson. 1999b. Problems with synthetic maps remain: Reply to Rendine et al. *Hum. Biol.* 71:447–453.

Torgerson, W.S. 1958. *Theory and Methods of Scaling.* New York: Wiley.

van Heek, F. 1954. Het geboorteniveau der Nederlandse Rooms-Katholieken. In *Demografisch-sociologische studie van een geëmancipeerde minderheidsgroep.* Leiden.

Voronoi, M.G. 1908. Nouvelles application des paramètres continus à la théorie des formes quadratiques, deuxième mémoire, recherche sur le parallélоedres primitifs. *J. Reine Angew. Math.* 134:198–207.

Womble, W.H. 1951. Differential systematics. *Science* 114:315–322.