Genetics and Language in European Populations

Robert R. Sokal; Neal L. Oden; Pierre Legendre; Marie-Josee Fortin; Junhyong Kim; Barbara A. Thomson; Alain Vaudor; Rosalind M. Harding; Guido Barbujani

American Naturalist, Vol. 135, No. 2 (Feb., 1990), 157-175.

Stable URL:
http://links.jstor.org/sici?sici=0003-0147%28199002%29135%3A2%3C157%3AGALIEP%3E2.0.CO%3B2-5

American Naturalist is currently published by The University of Chicago Press.

Vol. 135, No. 2 The American Naturalist February 1990

GENETICS AND LANGUAGE IN EUROPEAN POPULATIONS

ROBERT R. SOKAL,* NEAL L. ODEN,† PIERRE LEGENDRE,‡ MARIE-JOSÉE FORTIN,* JUNHYONG KIM,* BARBARA A. THOMSON,* ALAIN VAUDOR,‡ ROSALIND M. HARDING,§ AND GUIDO BARBUJANI‖

*Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245; †Department of Preventive Medicine, Division of Epidemiology, Health Sciences Center, State University of New York, Stony Brook, New York 11794-8036; ‡Département de Sciences biologiques, Université de Montréal, C.P. 6128, Succursale A, Montreal H3C 3J7, Canada; §M.R.C. Molecular Haematology Unit, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, England; ‖Dipartamento di Biologia, Università di Padova, via Trieste 75, I-35123, Italy

Submitted July 25, 1988; Revised February 6, 1989; Accepted March 14, 1989

In this paper we use the relationships between genetics and language in European populations to infer processes that led to their current population structure. The genetic structure of populations in geographic space can be examined at different scales ranging from one fine enough to record individuals to one encompassing continents. Processes that affect genetic structure include genetic drift (i.e., gene pools of limited size; Wright 1969, p. 345; Nei 1987, p. 352) and the limited mobility of individuals within the area of study. Both processes can be safely assumed to affect all natural populations, as is apparent from the various studies in which the spatial distributions of gene frequencies have been shown to follow the patterns predicted by Morton's (Morton et al. 1971; Morton 1982) and Malécot's (1973) models of isolation by distance.

In addition, spatially patterned selection and directed migration (distinct from the random dispersal of individuals underlying the isolation-by-distance model) may also cause spatial differentiation. The last two forces disturb what might otherwise be a simple relationship between genetic and geographic distance. For humans, we have ample evidence from historical sources that directed migrations have indeed taken place. Some are well known; others are inferred from archaeological and prehistoric information. The study of past migrations in animal and plant populations, which lack historical records, is much more difficult. It is of interest, therefore, to examine genetic differences in humans, whenever estimates of the history of the populations exist. Methods successful in interpreting structure in these human populations as measured against their documented histories can then be applied to populations without historical records. Studies in which historical hypotheses are directly tested against available genetic information are currently under way in our laboratory. In this article, primarily concerned with

Am. Nat. 1990. Vol. 135, pp. 157–175.

relationships between language and genetics, we make a first attempt at testing the relationships between independently obtained descriptors of historical processes affecting language boundaries and empirically observed measures of genetic change at these boundaries.

The processes listed above serve only as benchmarks of expected population structure. Neither random differentiation nor directed migration is likely to operate singly in populations at any scale. Various combinations of these factors surely act at different strengths over time and at different spatial scales. Despite this complexity, we try below to reach some overall generalizations for human populations at the continental scale.

Inferences from gene-frequency patterns to the processes that produced them must customarily be made from patterns assayed at a single moment in time (for an exception, see Sokal and Uytterschaut 1987). This greatly limits the power of potential inferences by comparison with diachronic studies. Nevertheless, various techniques for making such inferences have been suggested (Sokal and Oden 1978; Sokal and Wartenberg 1981; Felsenstein 1982; Sokal 1986*a*,*b*; Slatkin 1987; for recent applications, see Sokal et al. 1986, 1987*a*, 1989*a*).

Yet another approach is pursued here. We find a concomitant variable—language—to aid in decisions between alternative interpretations of current gene-frequency patterns. Both genetic and language patterns result from the biological and social interactions of individuals and groups in the populations concerned. For this reason, genetics and language manifest considerable similarity. Yet, biological and cultural variables of populations are also mediated by different structures and processes, and their patterns partly reflect this difference. Reflecting the similarity, language shares some of the properties of the biological variables employed in studies of population structure. Language changes over space, can be decomposed into numerous characteristics, exhibits spatial differentiation as a result of the limited mobility of its speakers, is transmitted vertically from parent to offspring, and possesses a phylogenetic history. However, language differs from genetics in that it lacks the Mendelian mechanism for segregation and recombination and is capable of horizontal and oblique (in addition to vertical) transmission (Cavalli-Sforza and Feldman 1981, p. 54).

Because a common language frequently signifies a common origin for two populations and a related language indicates a common origin further back in time (Ruhlen 1987, p. 4), linguistic relationships should be reflected by genetic relationships. Much of the comparison between genetic and linguistic evidence has therefore been with respect to phylogeny. H. M. Hoenigswald (MS) has examined the formal similarities between biological and linguistic phylogenies, and Cavalli-Sforza et al. (1988) have presented empirical evidence for such correspondence on a worldwide basis. At the population level, where phylogenetic models are less suitable because of gene flow, some quantitative comparisons have been made and are reviewed in the Discussion. In all, the evidence appears to justify the expectation that language and genetics may jointly throw light on the processes that have led to the spatial differentiation under study.

If language is used to study population structure, how is it to be measured? Phonemic, syntactic, and grammatical properties can be quantified for each lan-

guage; similarities between pairs of languages can be described as a percentage of shared cognates. Which of these measures should be chosen? (An analogous problem is encountered in the study of biological variation: which of various classes of biological information—molecular, genetic, morphological, physiological, etc.—should be chosen to represent the biological diversity?) We bypass this problem in the studies reported below by accepting a coarse classification of languages into language families, believing that it is adequate at the large scale investigated here.

An earlier paper (Sokal et al. 1988) discussed the factors affecting the correspondence between genetics and language. The correspondence is diminished by the well-documented, repeated genetic and linguistic assimilation of disparate ethnic elements into a single ethnic group with a single language. In particular, the correlation between genetics and language in our region of study, Europe, is lowered because migrant populations rarely settled in unoccupied areas. They frequently absorbed the native populations of the areas they settled in, the resulting population adopting the language of either the natives or the immigrants. A factor increasing genetic-linguistic correspondence is that language differences themselves impede free gene flow and, therefore, enhance genetic differentiation.

Here we summarize the results of seven approaches to investigating the relationships between gene frequencies and language families on a continental scale in Europe. Details of the new methods and their results have been published elsewhere (Harding and Sokal 1988; Sokal 1988; Sokal et al. 1988, 1989*b*; Barbujani and Sokal 1990; Legendre et al. 1990). We compare and reconcile the results of the various approaches employed in these studies. Next we discuss the implications of these results for the structure and origin of the European populations subsumed under the language families. We use available historical information to make predictions about the genetic differences between language families and test these predictions against our observations.

THE DATA

Our findings are based on 93 gene frequencies (erroneously stated as 97 in Sokal 1988) and 10 cranial measurements at 3466 locations in Europe. The 103 variables are grouped below into 27 systems, most corresponding to a genetic locus. The number preceding each system is that assigned to it in Mourant et al. (1976) or in our laboratory (numbers $\geq$ 100). Each conventional system abbreviation is followed, in parentheses, by the numbers of allele frequencies and samples employed, separated by a comma. Sources of the data for systems 1.1 through 65 are Mourant et al. (1976), Tills et al. (1983), and the results of an extensive computer search of the recent literature. Systems 100 and 101-102 were obtained through the courtesy of P. Menozzi, A. Piazza, and L. L. Cavalli-Sforza. Systems 200 and 201 are from Steinberg and Cook (1981), and system 901-910 is from Schwidetzky and Rösing (1984). The systems are 1.1 *ABO* (3, 870), 1.2 *ABO* with anti-*A*, -A_1, and -*B* (4, 157), 2.5 *MN* (2, 194), 2.7 *MN* with anti-*M*, -*N*, and -*S* (4, 68), 3.1 *P* (2, 102), 4.1 *Rh* (2, 568), 4.13 *Rh* with anti-*C*, -*D*, -*E*, and -*c* (8, 82), 4.19 *Rh* with anti-*C*, -*D*, -*E*, -*c*, and -*e* (8, 76), 5.1 *Lu* (2, 33), 6.1 *K* (2, 116), 6.3 *K* with anti-*K* and -*k* (2, 39), 7.1

Se (2, 53), 8.1 *Fy* (2, 108), 36.1 *Hp* (2, 175), 37.1 *Tf* (3, 38), 38.1 *Gc* (2, 112), 50.1.1 *ACP1* (3, 72), 52 *PGD* (3, 42), 53 *PGM1* (3, 70), 56 *Ak* (3, 64), 63 *ADA* (2, 53), 65 *T* (2, 62), 100 *HLA-A* (7, 66), 101-102 *HLA-B* (14, 66), 200 *Gm 1,2,5* (4, 45), 201 *Km* (*Inv*) (2, 38), 901-910 *cranial variables* (the 10 variables are listed in Schwidetzky and Rösing 1984, pp. 10, 97).

The numbers of localities sampled for each separate system range from 870 for the *ABO* system to 33 for the Lutheran system. We employed slightly different sample sizes, both in number of alleles and number of localities, for the different methods discussed below (for details, see Harding and Sokal 1988; Sokal 1988; Sokal et al. 1988, 1989*b*; Barbujani and Sokal 1990). The gene frequencies at each locality are based on sample sizes ranging from 50 to many thousands of persons and were all sampled after World War II. Although there was considerable population displacement during and after the war, this should not affect our results especially. Relocated populations (e.g., Poles and Germans) were recorded at their new locations; smaller groups of immigrants identified as such in the source publications were omitted from the data base.

The cranial measurements are means based on at least 25 skulls from populations dated between A.D. 1500 and the present (Schwidetzky and Rösing 1984). This distinguishes the cranial measurements from the gene frequencies, which all represent recent populations. Some changes in variation patterns of these cranial variables have occurred since the early Middle Ages (Sokal and Uytterschaut 1987). Further change over the 500-yr period spanned by the present samples is therefore possible. This could lead us to confound differences due to temporal change with differences due to language change. The results reported below show that cranial variables respond much like gene frequencies and thus do not lead to specific conclusions resting on the cranial data alone. We have therefore retained the cranial data in our report.

The languages spoken by the sampled European populations are grouped into 5 language phyla and 12 language families (Ruhlen 1987). The families are listed below, preceded by their phyla: INDO-EUROPEAN, Albanian, Baltic, Celtic, Germanic, Greek, Romance, Slavic; FINNO-UGRIC, Finnic, Ugric (Hungarian); ALTAIC, Turkic; AFRO-ASIATIC, Semitic (Maltese); LANGUAGE ISOLATES, Basque. We obtained language-family boundaries by consulting a number of sources (Meillet and Cohen 1952*a*,*b*; Mather et al. 1975; Cowgill 1976; Harms 1976; Ivanov 1976; Moulton et al. 1976; Posner 1976; von Czoernig 1984). We carefully investigated sample localities close to language boundaries to ascertain the language actually spoken. Where boundaries are imprecise, we assigned each locality to the language family of the majority of speakers. There are, however, few samples in the data for which we have any doubt about the language spoken. The language-family boundaries of Europe are shown in figure 1.

ANALYSES AND RESULTS

We tested whether speakers of diverse language families differ in their mean gene frequencies (Sokal et al. 1989*b*). This is a classic analysis-of-variance (ANOVA) problem. However, we could not use ANOVA because the spatial autocor-

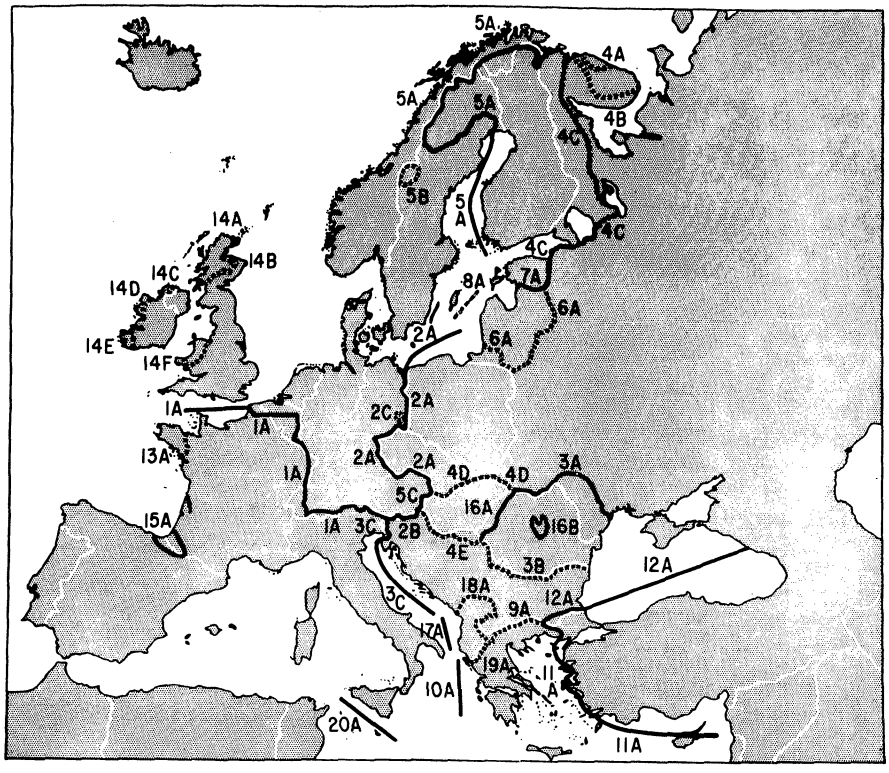

FIG. 1.—Major language-family boundaries in Europe. The boundaries for some non-Slavic populations in the Soviet Union have been omitted, since we lack the data to test them. The language-family boundaries are coded by number. All those tested in this study can be identified by looking up the appropriate number in table 1. Boundaries labeled B through E represent disjunct segments of their particular language-family boundary. *Solid lines*, Contiguous pairs that differ significantly by at least one of the methods (based on information in the table); *broken lines*, all other boundaries.

relation in these data (demonstrated in Harding et al. 1987; Sokal et al. 1989*a*) violates the independence assumption of the analysis (Cliff and Ord 1981, p. 189). This problem was overcome by employing a nonparametric permutational approach (contiguity-constrained permutational ANOVA; Legendre et al. 1990). For each allele frequency, we tested whether the partition of the sample points into groups corresponding to the observed language families yields a pooled within-group sum of squares less than that obtained when the data points are randomly partitioned into geographically compact groups. We grouped the resulting probabilities by their genetic systems, calculated a probability for each system by the Bonferroni method (Sokal and Rohlf 1987, p. 178), and found 12 systems significant at $P \leq 0.05$. This constitutes 57% of 21 systems tested, far in excess of expected type-I error. Combining the probabilities obtained for each system by Fisher's method (Sokal and Rohlf 1981, p. 779) results in a highly significant

TABLE 1

COMPARISON OF DIFFERENT APPROACHES FOR TESTING THE RELATIONSHIP BETWEEN GENETICS AND LANGUAGE IN EUROPE: RESULTS FOR CONTIGUOUS PAIRS OF LANGUAGE FAMILIES

Boundary No.	Pairs	Coded Results	Average Standardized Deviations
1A	Germanic-Romance	C Q R f W	1.676
2A	Germanic-Slavic	r	−0.258
2B	Germanic-Slavic	Q R	0.582
3A	Romance-Slavic	R	0.822
3B	Romance-Slavic		−0.675
3C	Romance-Slavic	R W	0.093
4C	Slavic-Finnic	W	—
4D	Slavic-Ugric		−1.695
4E	Slavic-Ugric		−0.800
5A	Germanic-Finnic	r f W	−0.528
5C	Germanic-Ugric	r w	−0.906
6A	Slavic-Baltic		−2.687
7A	Finnic-Baltic	F	0.386
8A	Germanic-Baltic		0.244
9A	Slavic-Greek		0.560
10A	Romance-Greek	f	0.228
11A	Greek-Turkic	R	−0.459
12A	Slavic-Turkic	R	0.429
13A	Romance-Celtic		−1.499
14A	Germanic-Celtic	R W	−0.597
15A	Romance-Basque	Q R W	0.086
16A	Romance-Ugric	R	0.702
16B	Romance-Ugric	W	0.084
17A	Romance-Albanian	F W	0.517
18A	Slavic-Albanian		−0.812
19A	Greek-Albanian		−0.112
20A	Romance-Semitic	R F w	0.652

NOTE.—C, R, and F indicate Bonferroni significance at $P \leq 0.05$ for contiguity-constrained permutational ANOVA, rate-of-change method, and difference method, respectively. Q, Language-family boundaries with appreciably higher variances in quadrats crossed by the boundaries than in uncrossed quadrats; W, a recognized boundary in the average derivatives of the surfaces by the Womble method. r and f, A significant combination of a single system and language-family pair for the rate-of-change and difference methods, respectively. w, A boundary recognized in an appreciable portion of the individual surfaces by the Womble method. The average standardized deviation for each language-family pair is a measure of average departure from expectation over the first four methods. Negative values indicate greater genetic similarity than expected for members of a given language-family pair; positive values, greater dissimilarity.

probability ($P < 0.00005$). We conclude that, for numerous genetic systems, population samples differ more among language families than within families. However, tests for the difference of means between all available pairs of language families, computed in an especially conservative manner to allow for spatial autocorrelation, show only a single difference, that between Germanic and Romance, significant at $P \leq 0.05$ (C, table 1).

An earlier study (Sokal et al. 1988) used three separate approaches to test for increased genetic change at language-family boundaries. Variances of gene-frequency or cranial-variable samples were compared in 5° × 5° map quadrats

crossed by language-family boundaries with variances in quadrats not so crossed (*quadrat-variance method*). Increased genetic change at language-family boundaries should yield greater variances in quadrats crossed by boundaries than in quadrats comprising a single language family. Because the quadrat variances for all but 10 of the variables lack spatial autocorrelation, Wilcoxon two-sample tests (Sokal and Rohlf 1981, p. 432) could be applied to the ranks of the variances of both crossed and uncrossed quadrats for each variable, to test the null hypothesis that crossed quadrats and uncrossed quadrats have the same variances. Bonferroni probabilities were calculated for the 27 different systems, based on the probabilities obtained by the Wilcoxon test for each variable within the system. Of the 27 systems, 6 (22%) show significantly increased variances ($P \leq 0.05$) for crossed quadrats. Combining the Bonferroni probabilities by Fisher's method yielded $0.025 < P < 0.05$. Overall, quadrats crossed by language-family boundaries have higher variances than uncrossed quadrats. Only three combinations of adjoining language families (Q in the table) consistently have appreciably higher variances.

Our second approach (Sokal et al. 1988), the *rate-of-change method*, tests whether gene-frequency surfaces interpolated from the available samples yield higher directional derivatives perpendicular to actual language-family boundaries than to identically configured boundaries randomly placed on the map of Europe. Derivations and computational details are given in the original reference. Bonferroni tests of each variable over the available language-family boundaries yield significant ($P \leq 0.05$) rates of change at language boundaries for gene-frequency surfaces representing six systems (22%). Some language-family boundaries are significant for a system, even though the system considered collectively over all boundaries is not. Another five systems are significant by this less stringent criterion. Fisher's method of combining probabilities was applied to the Bonferroni probabilities for each combination of a system and a language boundary, separately for each language-family boundary tested, over all systems (Sokal et al. 1988). The results (R in the table) show 10 significant ($P \leq 0.05$) combinations of pairs of language families. An overall Bonferroni test of the separate probabilities for each of the language boundaries yielded $P = 0.00145$. We may safely conclude, therefore, that there are higher gene-frequency gradients on the interpolated maps at some of the language boundaries than at randomly placed boundaries in Europe.

A third approach, the *difference method* (Sokal et al. 1988), employs original data values rather than interpolated ones. The method tests whether gene frequencies in Europe within 500 km of a language boundary differ more across actual language-family boundaries than do gene frequencies situated similarly across randomly placed boundaries. This turns out to be a less powerful approach than the other two. Only three allele frequencies from three systems are significant by a Bonferroni test over all language-family boundaries. Less stringent criteria permitted an additional five systems to be identified as possessing significant combinations within given language-family boundaries. Combining probabilities for individual tests of the 29 separate language-family boundaries by Fisher's method yielded three boundaries significant at $P \leq 0.05$ (F in the table). Significant

individual combinations of systems and boundaries were detected for an additional three boundaries (coded f). Bonferroni tests of the probabilities over all language-family boundaries yielded nonsignificant values. On the basis of this method, we were, therefore, unable to conclude that there are significant overall differences in the gene frequencies at language-family boundaries.

Another study tested whether *genetic distances* between European populations are related to their *linguistic distances* if *geographic distances* are kept constant (Sokal 1988). This approach tests the overall significance for all populations without singling out specific language-family boundaries. Separate genetic distances were computed for each system since each differs in the number of localities available for study. (For technical details, consult Sokal 1988.) Partial correlations of genetic and linguistic distances, with geographic distance kept constant, were computed and tested following the method of Smouse et al. (1986). Eleven partial correlations are significant at $P \leq 0.05$. An overall probability for rejection of the null hypothesis of no significant correlation was obtained by combining the probabilities associated with the correlation coefficients for each system by Fisher's method. After suitable adjustments for the replicated genetic systems in the data, the partial correlations of genetics and language, geography kept constant, yield a highly significant overall probability of $P \ll 0.001$. Speakers of different language families in Europe clearly differ genetically, even when one allows for geographic differentiation.

In another study, the same data were used to compute genetic distances among speakers of the European language families (Harding and Sokal 1988). The distances obtained for each system were pooled, and the overall distances subjected to numerical taxonomic procedures (Sneath and Sokal 1973), to obtain a grouping of the language families of Europe by genetic distance rather than by linguistic relationship. This classification is shown at the left side of figure 2. A Germanic-Celtic cluster is joined by a Slavic-Ugric one before both clusters join Romance to form a large cluster. Greek affiliates with Albanian; they are then joined by a Turkic-Baltic cluster. The two large clusters fuse and are joined by Semitic, Finnic, and Basque as outliers. Note that Albanian, Baltic, and Semitic are based on only two, three, and seven systems, respectively, and that their genetic affiliations are therefore not as reliable as those for the other language families. In a geographic classification based on a clustering of the great-circle distances of approximate centroids of the language-family areas (middle of fig. 2), we note a Germanic-Celtic cluster joining a Romance-Basque cluster and a Slavic-Ugric cluster joining another cluster with a tight Greek-Albanian nucleus joined by Semitic speakers. Turkic next joins these five language families. Finnic and Baltic make up an outlying cluster to the large cluster formed by the fusion of all previously mentioned languages. If the language families are grouped by their established linguistic relationships (right side of fig. 2; based on Ruhlen 1987), a large Indo-European cluster emerges, which is unresolved except for the closer affiliation of Baltic with Slavic. There is also a Finnic-Ugric cluster. The three remaining language families, Semitic, Basque, and Turkic, remain as single representatives of their phyla. On comparing the three groupings in figure 2, one notes that the genetic arrangement is closer to the geographic one than to the linguistic

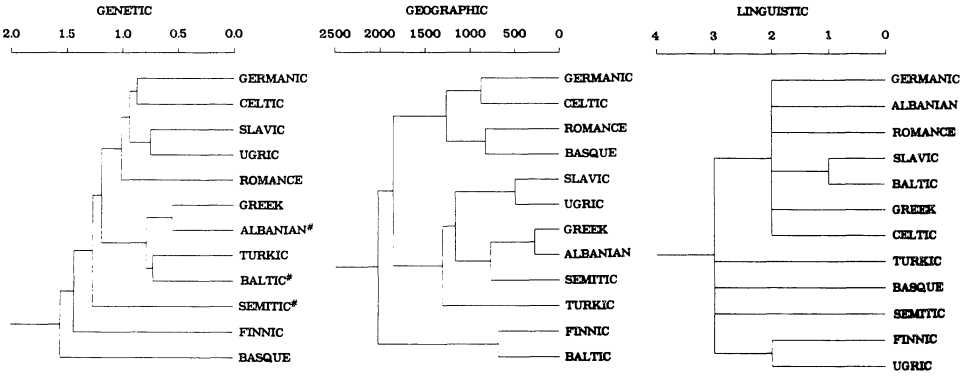

FIG. 2.—Grouping of the European language families by cluster analysis of genetic distances (*left*), geographic distances (*center*), and linguistic relationships (*right*). The dendrograms were obtained by UPGMA clustering (Sneath and Sokal 1973) and are based on the data of Harding and Sokal (1988). The three language families on the left followed by # are based on few genetic systems, and their positions in the genetic dendrogram are therefore unreliable.

one. Thus, the genetic classification largely reflects geographic propinquity. However, the distant genetic affiliations of Finnic, Basque, and Semitic indicate that some correlation persists between linguistic relationships and genetic distance. The modern European gene pools still reflect the remote origins of some ethnic units representing these major linguistic groups.

Each of the previous six approaches involved a test of the established language groupings or boundaries against the observed genetic variation of the populations concerned. A seventh approach (Barbujani and Sokal 1990) set out to discover the *zones of rapid genetic change* in Europe, irrespective of their linguistic circumstances. A newly proposed technique (Barbujani et al. 1989) based on earlier work by Womble (1951) was employed. This method averages the absolute values of the derivatives of the gene-frequency surfaces over the entire map and highlights strings of connected areas of rapid genetic change that can be recognized as genetic boundaries. This method was applied to the average derivatives of 60 gene-frequency surfaces as well as to the 32 individual surfaces based on 66 or more samples (Barbujani and Sokal 1990). A conservative recognition criterion, representing the top 5% of change in the average derivatives or in an appreciable number of the surfaces, was applied to obtain 33 gene-frequency boundaries. These boundaries are shown on a map of Europe in figure 3. Of these 33 boundaries, 15 represent all or part of 10 modern language-family boundaries (cf. fig. 1). These boundaries are shown as W or w in the table. Another 11 boundaries occur between different languages within a language family. In 7 instances, the recognized gene-frequency boundaries do not seem to mark a change in language. When these are examined more closely, however, 5 (Corsica vs. Italy, SW vs. NE Finland, C vs. S Germany, N vs. S Italy, NW vs. SE Yugoslavia) represent zones of marked linguistic (dialectal) change. One of the 2 remaining boundaries is a modern relict of ancient ethnic boundaries that have no effect on current speech.

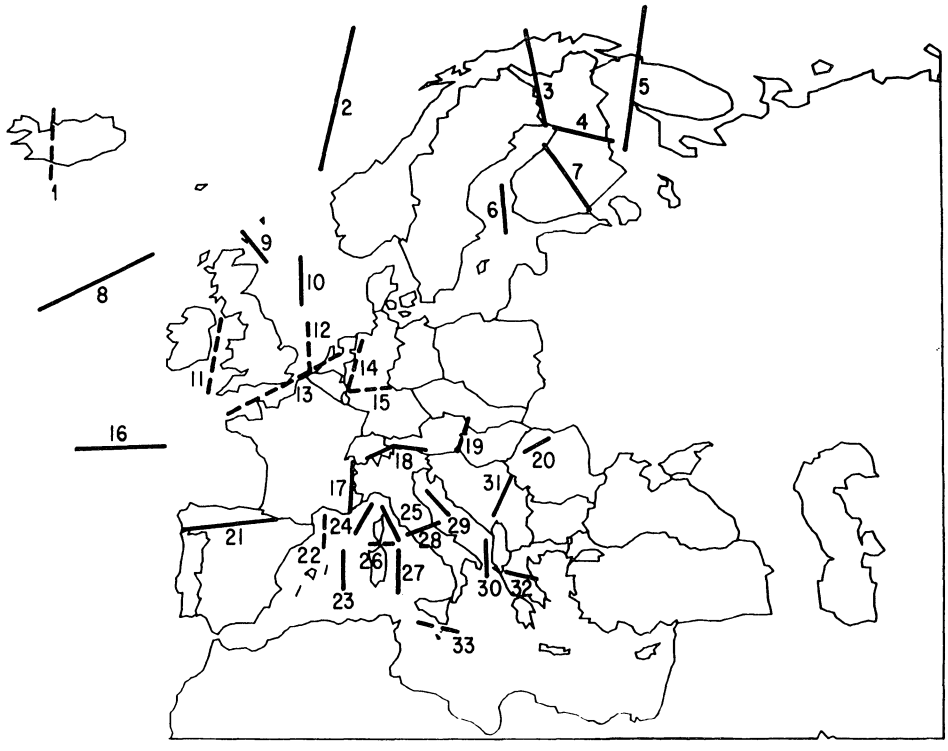

FIG. 3.—Zones of sharp genetic change in Europe recognized by the method of Womble (1951) as modified by Barbujani et al. (1989). *Solid lines*, Boundaries recognized on the basis of average derivatives of 60 gene frequencies; *dashed lines*, boundaries resulting from an analysis of individual surfaces.

It separates western from eastern Iceland and is due to the ethnic origins of the population (the west had a far higher proportion of settlers who came from Scandinavia via Ireland and brought Irish wives and servants). The final boundary, through northern Greece, does not readily correspond to a linguistic or ethnic line. However, it may mark off the region into which Greek speakers from Anatolia were resettled after World War I. The results—31 of 33 genetic boundaries are also linguistic boundaries—evidence a close association of genetic and linguistic variation.

DISCUSSION

The Different Approaches Compared

How comparable are the results of these various tests? Of 25 systems remaining after combining the results from systems 1.1 and 1.2 and those from 6.1 and 6.3, 5 (4.19 *Rh* haplotypes, 5.1 *Lu*, 7.1 *Se*, 52 *PGD*, and 56 *Ak*) are not significant by any criterion. Note that these tend to be systems based on a low number of locality samples. The remaining 20 systems are significant by one or more of the methods

applied. No common biological characteristics differentiate the significant systems from those that are not. The outcomes for the various methods differ because they test different aspects of the relationships between genetics and language. In the order in which they were presented above, the methods test for differences in gene frequency among speakers of the various language families; for increased change in gene frequencies at language-family boundaries (three methods); for a linear relationship between genetic and linguistic distances; for similarity between genetic and linguistic classifications of language families; and for zones of rapid genetic change, which were subsequently found to be largely coincident with zones of language change. The first five methods test specific null hypotheses; the last two are exploratory in nature.

The Patterns of Genetically Significant Language-Family Boundaries in Europe

When we indicate the significant language-family boundaries (table 1; including the less stringently significant results identified by lowercase letters) on the map of Europe in figure 1, an interesting pattern emerges. The language families are separated by genetic boundaries over most of Europe. The exceptions to this generalization are points of interest in the discussion that follows. Most Celtic-speaking populations (those in Ireland, Wales, and Brittany) are not differentiated genetically from the surrounding speakers of the majority language. This may reflect gradual diffusion gradients rather than sharp boundaries in these populations. West Finnic and Slavic speakers cannot be differentiated with our data base since we lack a sufficient number of Slavic samples at reasonable distances from the language-family border with the West Finnic speakers. We cannot demonstrate differences between Baltic and Slavic speakers. Extensive diffusions between these populations may account for the lack of difference: first, when the extensive area of Baltic speakers was reduced by the Slavic expansion in prehistoric times; and second, during the expansion of the Lithuanian Empire in the thirteenth and fourteenth centuries and again during the subsequent advance of Slavic speakers resulting in the modern location of the boundary. Note also that Baltic and Slavic are the only two language families that are considered closer to each other than they are to the other Indo-European language families analyzed (Ruhlen 1987, p. 37).

A striking feature of figure 1 is the absence of significant gene-frequency differences for the various language families in the Balkans and Hungary. The boundary between Romance speakers in Romania and South Slavic speakers in Yugoslavia and Bulgaria is not reflected in gene-frequency differences, nor are the boundaries between Albanian speakers and their neighbors speaking South Slavic and Greek. The boundary between South Slavic speakers in Yugoslavia and Bulgaria and Greek speakers in Greece is similarly not significant. The extensive migration and admixture of these populations throughout their history may well be responsible for the lack of present differences. Greek-speaking populations extended well up the Balkan peninsula in the past. Illyrian- and Thracian-speaking populations occupied large areas noncoincident with modern boundaries, and genetic differences resulting from their boundaries may confound the results for present populations. Slavic speakers reached well into Greece in the sixth cen-

tury, and Romance speakers (Vlachs, Arumanians) as well as Albanians and Greeks migrated extensively. Turkish speakers were found in all the Balkan countries during the Turkish occupation. We report no differences between Ugric speakers in Hungary and Slavic speakers to their north and south. This may be due to the assimilation of substantial numbers of Slavic speakers who lived in modern Hungary before the Magyar land-taking and also due to the subsequent diffusions of Magyar speakers into modern Slovakia and Croatia.

In view of the above, it seems surprising that the Turkic speakers show up as different from Greek and Slavic speakers by the rate-of-change method. However, the genetic significance of the Greek-Turkic language-family boundary is questionable (Sokal et al. 1988). The Slavic-Turkic boundary comprises a relatively short segment separating South Slavic speakers in Bulgaria from Turkish speakers in that country and in European Turkey, and a long segment between the East Slavic speakers along the north shore of the Black Sea and the Turkish speakers of Anatolia. The outcomes of tests for this boundary are dominated by the long segment, which is far better supported by sampling points than the short segment, and we cannot assert firmly that Turkic speakers differ from South Slavic speakers in the Balkans, whereas they are certainly distinct from Slavic speakers overall.

We now turn to the significant language-family boundaries. The following boundaries are substantiated by the largest numbers of genetic systems (four, three, two, and two, respectively): the Semitic-Romance boundary between Malta and Sicily; the boundary between Basque speakers and their Romance-speaking neighbors; the Romance-Germanic boundary; and the boundary between Romance and South Slavic speakers in Italy and Yugoslavia. The specific allele frequencies differentiating the language-family boundaries vary considerably. No one allele frequency characteristically differentiates even the majority of European language families. Since differences are not observed at the same loci for the various pairs of populations, random genetic drift, rather than adaptation, may account for the observed genetic divergence. The alternative processes that may account for these patterns (i.e., random genetic drift occurring where the populations are currently located, or secondary contact between populations that diverged elsewhere) can be assessed on the basis of the available historical information. Attempts to demonstrate that the pattern of significantly differentiated boundaries in figure 1 represents a statistically significant departure from randomly allocating significance to the same number of boundaries were unsuccessful.

Inferences from the Patterns

What inferences about geographic patterns of gene frequencies can be drawn from the differences among test results? It is possible to intuit the outcomes when applying five of the methods (all but the two methods employing distances) to simple models of gene-frequency surfaces: random surfaces, inclined planes, step clines, character-displacement surfaces (marked divergence on both sides of a boundary, but similarity at some distance from the boundary), or patches of

different gene-frequency values. Each of these is examined with language boundaries parallel or orthogonal to the direction of the gradients of the surfaces. The methods yield differentially diagnostic results. Based on these models, our test results indicate that most surfaces (54 of 69 independent variables) exhibit patterns consistent with either a random surface or one in which the language boundaries are not positioned at right angles to the gradients of the gene-frequency patterns. Three patterns are consistent with character displacement and four indicate inclined planes, but none of the former and only one of the latter surfaces has been independently identified as a cline (Sokal et al. 1989*a*). These negative findings may result from the pattern of language-family boundaries in Europe (see fig. 1), which is sufficiently complex in form and compass direction that even simple clinal trends would not be reflected by tests applied to the boundaries. The language-family boundaries show no preferred compass direction. We conclude, therefore, that in view of the complexity of both the observed gene-frequency surfaces and the language-family boundaries (some of which show significant differences for a given gene frequency, whereas others do not), the differences in the outcomes of the tests cannot be ascribed to certain combinations of simple geographic gene-frequency patterns with suitably positioned language-family boundaries.

What is the relationship of geographic contiguity between language families to genetic difference between them? All the language-family combinations in the table are contiguous pairs. We are not in a good position to answer this question, since, of the test methods employed, only the one testing for differences of means (employed in conjunction with the contiguity-constrained permutational ANOVA) can potentially characterize noncontiguous language-family pairs as significantly different. Yet by that method, the only significant pair is Germanic-Romance, which is contiguous. However, when we tested the relationship between differences of means and the contiguities of all pairs of language families by means of a Mantel test (Mantel 1967; Sokal 1979), differences were significantly higher ($P = 0.02$) for noncontiguous pairs. Although no single pair of noncontiguous language families is so different as to exceed the conservatively chosen threshold for statistical significance (in fact, only one pair of contiguous language families is significant by this criterion), in aggregate there is a significant trend for higher differences between language families that do not share a common boundary. Of the contiguous pairs of language families that it was possible to test, 6 of 21 are not significant by any of the approaches: Romance-Celtic, Slavic-Ugric, Slavic-Greek, Slavic-Baltic, Slavic-Albanian, and Greek-Albanian. We have already discussed possible reasons for this finding.

Note that nearly all non-Indo-European language families are significantly separated from their Indo-European neighbors by one or more of the tests. The proportion of significant boundaries among those separating Indo-European from non-Indo-European language families is 72.7%, whereas 53.3% of boundaries between Indo-European language families are significant. However, this difference is not statistically significant.

The conclusions of the seven studies reported above agree that the genetic

structure of these populations cannot entirely be the result of (random) geographic differentiation after their arrival at their present areas of settlement. It has been shown that such a simple geographic-differentiation model yields only a partial explanation for the observed geographic differences in the populations (Sokal et al. 1989*b*). We have seen that the observed patterns cannot stem from clinal trends that undoubtedly exist in Europe. This is especially true since all the significant differences found occur between populations representing contiguous language families. These would be the least differentiated if they were based on large-scale clinal patterns. The demonstration of a linguistic component in addition to geographic differentiation (Harding and Sokal 1988; Sokal 1988) and the near ubiquity of linguistic boundaries along the zones of rapid genetic change in Europe (Barbujani and Sokal 1990) suggest historical as well as geographic components responsible for the genetic differentiation of the European language families. Some of the observed genetic differences can clearly be associated with historical migration patterns of populations representing speakers of various languages, these speakers differing aboriginally in gene frequencies.

We attempted to find explanations for the observed differences between contiguous language-family areas. Explanations for given results and specific language boundaries, such as were furnished earlier in this paper, are suspect; there is the danger of selecting explanatory facts to fit the evidence. Accordingly, we followed the preferable course of assembling all potentially explanatory historical facts for all boundaries without knowing the outcomes for specific methods and testing these globally against the test results obtained by our methods. This avoids the risk of employing post hoc explanations. We developed two prediction vectors based on eight historical scenarios (described in Sokal et al. 1988). Examples of such scenarios are displacement of population A by population B and repeated advances of one population into the territory of the other. The prediction vectors describe the number of events that have tended to sharpen or damp the genetic differences across a given boundary. Next, we quantified the information summarized in the first four coded columns of table 1 by calculating a standardized deviation from expectation for each combination of approach and language-family boundary. The fifth column could not be incorporated in the computations since it is not based on a proper significance test and did not yield a test statistic. The averages of these values over the first four approaches are given in the last column of the table. Negative values indicate greater-than-expected genetic similarity by members of a given language-family pair; positive values, greater dissimilarity. The Kendall rank-correlation coefficient of the vector of averages in the table with the sharpening vector is 0.495 ($P = 0.0014$); with the damping vector, it is -0.347 ($P = 0.0232$). These two vectors together determine 33% of the variance of the averages. We conclude that the observed genetic differences can be partially predicted in terms of the history of the populations involved.

The genetic dissimilarities of these population stocks may have come about by chance (founder effects) in the small population isolates that originally gave rise to the languages ancestral to the language families of Europe. These populations settled in various areas of Europe and expanded. The original genetic differences

would be maintained during the subsequent (and still ongoing) genetic differentiation of the populations, which is in part geographically based. In addition, there was genetic and linguistic amalgamation with native populations that presumably were already genetically differentiated at the time of contact with the immigrants. The combination of aboriginal genetic differentiation of immigrants plus differentiation of the residual native populations may have led to the differences found among present-day speakers. Genetic differences at language boundaries are maintained despite the amalgamation because (1) the settlers are sufficiently different genetically that genetic differences with speakers of other languages are retained despite admixture with the native populations or (2) there are natural geographic barriers that define distinct regions, which in turn are filled by speakers of one language. (In the latter case, if both immigrants and the natives in the region were genetically differentiated from their neighbors, the two sets of differentiae for newcomers and natives may reinforce one another.) The genetic differences may be retained by language differences, which inhibit intermarriage. One needs, formally at least, to consider the alternative of differential selection along language-family lines, but we believe this to be quite implausible.

The above model is compatible with the established view of the origin of the Indo-European peoples in the Pontic steppes and their arrival in Europe in the fifth and fourth millennia B.C. as the bearers of the Kurgan culture (Gimbutas 1986). This view has been challenged by Renfrew (1987), who pushed back the times for the development of the Indo-European languages by several thousand years, had them originate in Anatolia, and tied the spread of the Indo-European populations and languages to the spread of agriculture as explicated by Ammerman and Cavalli-Sforza's (1984) demic-diffusion model. The difficulty with corroborating Renfrew's model by using modern genetic data is that, since his model is tied to the hypothesis of the origin of agriculture by demic diffusion, it is impossible to refute one without the other. (Evidence for the latter hypothesis, based on modern gene frequencies, has already been presented in Menozzi et al. 1978 and Sokal and Menozzi 1982.) A second difficulty with Renfrew's hypothesis is that it does not feature a detailed scenario for the origins of the separate Indo-European language families. Under the gradual differentiation visualized by Renfrew, no agent would keep the relatively simple organized social units cohesive enough to form and retain the relatively few Indo-European language families. For each language family, a centripetal force (political organization, internal migration, mating patterns, etc.) would be needed to maintain cohesion.

To make the chronology and distribution pattern of the Renfrew model compatible with our findings of genetic differences among language families above and beyond those caused by geographic differentiation would require a stochastic-branching model based on small population sizes and founder effects. Such models have been demonstrated in populations on a smaller scale (e.g., in the Yanomama tribe of Amerindians; Neel 1981; Sokal et al. 1986).

Studies by others showing biological correlates of language include anthropometric and dermatoglyphic, as well as genetic, variables. Various authors (Parsons and White 1973; Dow and Cheverud 1985; Sokal et al. 1986, 1987*b*; Dow et

al. 1987; Sokal and Winkler 1987) have investigated the relationships between differences in these variables and linguistic differences in populations. Other references were cited by Jorde (1980). These studies range from small spatial scales (9 km; Smouse and Wood 1987) to the continental (7000 km for sub-Saharan Africa; Vecchi and Passarello 1977–1979; Rösing 1984–1985). In all studies with adequate sample sizes, some relationship between language and these biological variables can be established. However, the processes by which these relationships are established differ with the spatial scale of the populations investigated. Neutral models of population structure, that is, those based on the concept of isolation by distance, are currently contrasted with models that include directed migration and/or selection. Restriction of gene flow because of physical or cultural barriers is not usually recognized as a factor causing departures from the gene-frequency patterns expected under isolation by distance, but it should be, as the studies reviewed in this paper demonstrate. Indeed, whereas gene flow acts as a homogenizing force (Slatkin 1985, 1987), the effect of a cultural barrier such as a language barrier is the opposite. Language boundaries maintain sharp genetic differences that otherwise would be blurred by population admixture. Therefore, the factors that impair gene flow should be taken into account as major determinants of the genetic structure of populations. Language differences are one such factor in humans.

SUMMARY

Migration, selection, and spatial differentiation determine the patterns of geographic variation in the gene frequencies of human populations. Inferences about past processes must be made from current patterns. The use of language differences as a variable concomitant to gene frequencies allows such inferences despite the complex relationship between language and genetics in populations. Seven methods that test varying aspects of this relationship show genetic differences among speakers of different language families in Europe, in addition to differences among the populations due to geographic differentiation. A model, based on the known history of each language-family boundary, was constructed to predict the likelihood of genetic differences at the boundaries. The model is in good agreement with the observed results. The genetic-linguistic patterns observed in Europe are consistent with the combined operation of spatial differentiation and aboriginal genetic differences among speakers of different languages before they moved to their present locations on the continent.

ACKNOWLEDGMENTS

D. DiGiovanni provided dedicated assistance with the computations and other technical matters, J. Schirmer prepared the maps, and C. Daly did the word processing of the various drafts of the manuscript. This work was supported by grant GM 28262 from the National Institutes of Health and grant BSR 8614384 from the National Science Foundation (both to R.R.S.). Contribution 705 in Ecology and Evolution from the State University of New York at Stony Brook.

LITERATURE CITED

Ammerman, A. J., and L. L. Cavalli-Sforza. 1984. The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, N.J.

Barbujani, G., and R. R. Sokal. 1990. The zones of sharp genetic change in Europe are also linguistic boundaries. Proc. Natl. Acad. Sci. USA (in press).

Barbujani, G., N. L. Oden, and R. R. Sokal. 1989. Detecting regions of abrupt change in maps of biological variables. Syst. Zool. (in press).

Cavalli-Sforza, L. L., and M. W. Feldman. 1981. Cultural transmission and evolution: a quantitative approach. Princeton University Press, Princeton, N.J.

Cavalli-Sforza, L. L., A. Piazza, P. Menozzi, and J. Mountain. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc. Natl. Acad. Sci. USA 85:6002–6006.

Cliff, A. D., and J. K. Ord. 1981. Spatial processes: models and applications. Pion, London.

Cowgill, W. 1976. Indo-European languages. Pages 431–438 *in* The new Encyclopaedia Britannica. Vol. 9. Macropaedia. Encyclopaedia Britannica, Chicago.

Dow, M. M., and J. M. Cheverud. 1985. Comparison of distance matrices in studies of population structure and genetic microdifferentiation: quadratic assignment. Am. J. Phys. Anthropol. 68:367–373.

Dow, M. M., J. M. Cheverud, and J. S. Friedlaender. 1987. Partial correlation of distance matrices in studies of population structure. Am. J. Phys. Anthropol. 72:343–353.

Felsenstein, J. 1982. How can we infer history and geography from gene frequencies? J. Theor. Biol. 96:9–20.

Gimbutas, M. 1986. Remarks on the ethnogenesis of the Indo-Europeans in Europe. Pages 5–19 *in* W. Bernhard and A. Kandler-Pálsson, eds. Ethnogenese europäischer Völker. G. Fischer, Stuttgart, W. Germany.

Harding, R. M., and R. R. Sokal. 1988. Classification of the European language families by genetic distance. Proc. Natl. Acad. Sci. USA 85:9370–9372.

Harding, R. M., R. R. Sokal, and N. L. Oden. 1987. An examination of spatial patterns of gene frequencies in Europe. [Abstract.] Am. J. Phys. Anthropol. 72:207–208.

Harms, R. T. 1976. Uralic languages. Pages 1022–1032 *in* The new Encyclopaedia Britannica. Vol. 18. Macropaedia. Encyclopaedia Britannica, Chicago.

Ivanov, V. V. 1976. Slavic languages. Pages 866–874 *in* The new Encyclopaedia Britannica. Vol. 16. Macropaedia. Encyclopaedia Britannica, Chicago.

Jorde, L. B. 1980. The genetic structure of subdivided human populations: a review. Pages 135–208 *in* J. H. Mielke and M. H. Crawford, eds. Current developments in anthropological genetics. Vol. 1. Plenum, New York.

Legendre, P., N. L. Oden, R. R. Sokal, A. Vaudor, and J. Kim. 1990. Approximate analysis of variance of spatially autocorrelated regional data. J. Classif. (in press).

Malécot, G. 1973. Isolation by distance. Pages 72–75 *in* N. E. Morton, ed. Genetic structure of populations. University of Hawaii Press, Honolulu.

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

Mather, J. Y., H. H. Speitel, and G. W. Leslie, eds. 1975. The linguistic atlas of Scotland. Archon Books, Hamden, Conn.

Meillet, A., and M. Cohen, eds. 1952*a*. Langues Indo-Européennes d'Europe (Carte). Carte I *in* Les langues du monde. Centre National de la Récherche Scientifique, Paris.

———. 1952*b*. Langues Ouraliennes (Carte). Carte VII *in* Les langues du monde. Centre National de la Récherche Scientifique, Paris.

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

Morton, N. E. 1982. Estimation of demographic parameters from isolation by distance. Hum. Hered. 32:37–41.

Morton, N. E., S. Yee, D. E. Harris, and R. Lew. 1971. Bioassay of kinship. Theor. Popul. Biol. 2:507–524.

Moulton, W. G., E. Haugen, and M. I. Herzog. 1976. Germanic languages. Pages 19–31 *in* The new Encyclopaedia Britannica. Vol. 8. Macropaedia. Encyclopaedia Britannica, Chicago.

Mourant, A. E., A. C. Kopéc, and K. Domaniewska-Sobczak. 1976. The distributions of the human blood groups, and other polymorphisms. 2d ed. Oxford University Press, London.

Neel, J. V. 1981. On being headman. Perspect. Biol. Med. 24:277–294.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Parsons, P. A., and N. G. White. 1973. Genetic differentiation among Australian Aborigines with special reference to dermatoglyphics and other anthropometric traits. Pages 81–94 *in* R. L. Kirk, ed. The human biology of Aborigines in Cape York. Hum. Biol. Ser. 5. Australian Institute of Aboriginal Studies, Canberra, Australian Capital Territory.

Posner, R. 1976. Romance languages. Pages 1025–1045 *in* The new Encyclopaedia Britannica. Vol. 15. Macropaedia. Encyclopaedia Britannica, Chicago.

Renfrew, C. 1987. Archaeology and language. Jonathan Cape, London.

Rösing, F. W. 1984–1985. Relations between morphometric distances and linguistic classification in the populations of sub-Saharan Africa: a comment on Vecchi and Passarello's analysis. Riv. Antropol. 63:259–262.

Ruhlen, M. 1987. A guide to the world's languages. Vol. 1. Classification. Stanford University Press, Stanford, Calif.

Schwidetzky, I., and F. W. Rösing. 1984. Vergleichend-statistische Untersuchungen zur Anthropologie der Neuzeit (nach 1500). Homo 35:1–49.

Slatkin, M. 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16:393–430.

———. 1987. Gene flow and the geographic structure of natural populations. Science (Wash., D.C.) 236:787–792.

Smouse, P. E., and J. W. Wood. 1987. The genetic demography of the Gainj of Papua New Guinea: functional models of migration and their genetic implications. Pages 211–224 *in* B. D. Chepko-Sade and Z. Halpin, eds. Mammalian dispersal patterns: the effects of social structure on population genetics. University of Chicago Press, Chicago.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. 35:627–632.

Sneath, P. H. A., and R. R. Sokal. 1973. Numerical taxonomy. Freeman, San Francisco.

Sokal, R. R. 1979. Testing statistical significance of geographic variation patterns. Syst. Zool. 28: 227–231.

———. 1986*a*. Die räumliche Analyse der menschlichen Populationsstruktur. Homo 37:50–71.

———. 1986*b*. Spatial data analysis and historical processes. Pages 29–43 *in* E. Diday, Y. Escoufier, L. Lebart, J. Pages, Y. Schektman, and R. Tomassone, eds. Data analysis and informatics, IV. Proceedings of the Fourth International Symposium on Data Analysis and Informatics, Versailles, France, October 9–11, 1985. North-Holland, Amsterdam.

———. 1988. Genetic, geographic, and linguistic distances in Europe. Proc. Natl. Acad. Sci. USA 85:1722–1726.

Sokal, R. R., and P. Menozzi. 1982. Spatial autocorrelation of HLA frequencies in Europe support demic diffusion of early farmers. Am. Nat. 119:1–17.

Sokal, R. R., and N. L. Oden. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. 10:199–228.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. 2d ed. Freeman, San Francisco.

———. 1987. Introduction to biostatistics. 2d ed. Freeman, New York.

Sokal, R. R., and H. Uytterschaut. 1987. Cranial variation in European populations: a spatial autocorrelation study at three time periods. Am. J. Phys. Anthropol. 74:21–38.

Sokal, R. R., and D. E. Wartenberg. 1981. Space and population structure. Pages 186–213 *in* D. Griffith and R. McKinnon, eds. Dynamic spatial models. NATO Adv. Study Inst. Ser.: Ser. D, Behav. Soc. Sci. 7. Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands.

Sokal, R. R., and E.-M. Winkler. 1987. Spatial variation among Kenyan tribes and subtribes. Hum. Biol. 59:147–164.

Sokal, R. R., P. E. Smouse, and J. V. Neel. 1986. The genetic structure of a tribal population, the Yanomama Indians. XV. Patterns inferred by autocorrelation analysis. Genetics 114: 259–281.

Sokal, R. R., N. L. Oden, and J. S. F. Barker. 1987*a*. Spatial structure in *Drosophila buzzatii* populations: simple and two-dimensional spatial autocorrelation. Am. Nat. 129:122–142.

Sokal, R. R., H. Uytterschaut, F. W. Rösing, and I. Schwidetzky. 1987*b*. A classification of European skulls from three time periods. Am. J. Phys. Anthropol. 74:1–20.

Sokal, R. R., N. L. Oden, and B. A. Thomson. 1988. Genetic changes across language boundaries in Europe. Am. J. Phys. Anthropol. 76:337–361.

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

Sokal, R. R., N. L. Oden, P. Legendre, M. J. Fortin, J. Kim, and A. Vaudor. 1989*b*. Genetic differences among language families in Europe. Am. J. Phys. Anthropol. 79:489–502.

Steinberg, A. G., and C. D. Cook. 1981. The distribution of the human immunoglobulin allotypes. Oxford Monogr. Med. Genet. Oxford University Press, Oxford.

Tills, D., A. C. Kopéc, and R. E. Tills. 1983. The distribution of the human blood groups, and other polymorphisms. Suppl. 1. Oxford Monogr. Med. Genet. Oxford University Press, Oxford.

Vecchi, F., and P. Passarello. 1977–1979. Relazioni tra distanza morfologica e caratteristiche linguistiche nelle popolazioni dell'Africa Sub-sahariana. Riv. Antropol. 60:187–194.

von Czoernig, C. 1984. Ethnographische Karte der oesterreichischen Monarchie. Pages 240ff *in* H. Kühnel, E. Vavra, and G. Stanger, eds. Von der Revolution zur Gründerzeit. Katalog des niederösterreichischen Landesmuseums, N.F. 147. Amt der niederösterreichischen Landesregierung, Vienna.

Womble, W. H. 1951. Differential systematics. Science (Wash., D.C.) 114:315–322.

Wright, S. 1969. Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies. University of Chicago Press, Chicago.