Autocorrelation of Gene Frequencies Under Isolation by Distance

Guido Barbujani[1]

Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794

Manuscript received April 2, 1987
Revised copy accepted August 13, 1987

ABSTRACT

Spatial autocorrelation statistics are used for description of geographic variation of gene frequencies, but the relationship of these indices with the parameters describing the genetic structure of populations is not established. A simple relation is derived here between kinship coefficient and a measure of spatial autocorrelation, Moran's *I*. The autocorrelation coefficient of gene frequencies at a given distance is a direct function of the kinship at that distance, and an inverse function of the standardized gene frequency variance, F_{st}. Under isolation by distance, the expected values of Moran's *I* for any allele may be calculated by means of Malécot-Morton function, which predicts an exponential decline of genetic similarity in space. This allows comparison of observed gene frequency patterns with the patterns that should be caused by interaction of short range migration and random genetic drift.

ONE of the goals of population genetics is to analyze geographical distributions of gene frequencies, in order to make inferences concerning the evolutionary processes that have generated them. Several methodologies have been proposed for the study of human gene frequencies, but basically two types of approach have been employed in the last decade. MENOZZI, PIAZZA and CAVALLI-SFORZA (1978) and PIAZZA, MENOZZI and CAVALLI-SFORZA (1982) evaluated some indices describing synthetically the geographical variation shown by several polymorphisms jointly through a *multivariate analysis*. The second approach is the analysis of numerous single gene frequencies separately by spatial autocorrelation techniques (SOKAL and MENOZZI 1982; SOKAL, SMOUSE and NEEL 1986; SOKAL and WINKLER 1987; BARBUJANI 1987), using what has been called a *multivariable approach* (SOKAL 1979). The patterns described are consistent with the effects of both selection and nonselective processes on the systems considered; moreover, hypotheses on the evolutionary forces accounting for such patterns were not always explicitly tested in the cited studies. The virtual lack of specifically designed statistical techniques (FELSENSTEIN 1982) may have been an important reason for that.

It seems, therefore, that further studies of gene frequency distributions aimed at understanding their origin should include: (1) computation of indices describing the patterns with minimum loss of information; and (2) comparison of such indices with those expected under different evolutionary hypotheses. CAVALLI-SFORZA (1966) suggested that under isolation by distance all genes should display an equal fraction of their maximum possible variability, *i.e.*, show equal values of their standardized gene frequency variance, F_{st}. Heterogeneous F_{st} values should therefore be regarded as evidence that systematic pressures have affected some of the genetic systems of interest. Two tests for the heterogeneity of F_{st} values across loci have been proposed (LEWONTIN and KRAKAUER 1973; BARBUJANI 1985; for the debate on the former, see references in FELSENSTEIN 1982), and the latter has been applied to human data (BARBUJANI and MILANI 1986). The study of F_{st} values may, however, provide only a general description of gene frequency diversity. Tests of hypotheses on the evolutionary processes that account for the observed gene frequency distributions require more complex approaches; spatial autocorrelation analysis may be suitable for this purpose.

Spatial autocorrelation is defined as the association of the values of one variable with the values of the same variable at all other localities (SOKAL and ODEN 1978a). It gives a detailed description of gene frequency variation in space, and is independent of preliminary assumptions about the underlying population structure. Different loci are taken into account separately (hence the expression "multivariable" approach, rather than "multivariate") and this bypasses problems due to incomplete data matrices, which may affect other approaches. The allele frequencies for all sampled localities are compared to allele frequencies at localities within suitably chosen distances, autocorrelation coefficients are computed for each distance class, and the resulting correlogram summarizes the spatial relationships between populations for that allele. Inferences are then possible, based on the shape of the single correlograms, and on the comparison of correlograms computed at various loci (SOKAL and ODEN 1978a, b; SOKAL 1979; SOKAL and WARTEN-

[1] Permanent address, where reprints should be requested: Istituto di Zoologia, Universitá di Ferrara, via L. Borsari 46, I-44100 Ferrara, Italy.

BERG 1981). Multivariable analysis may therefore be regarded as a repeated application of univariate methods.

SLATKIN (1985) noted that the independence from prior assumptions may turn out to be a drawback of this method, if consistency of the observed gene frequency patterns with a genetical model is to be tested. Indeed, the relation of spatial autocorrelation measures with the various indices of genetic differentiation has not yet been studied (SOKAL and WARTENBERG 1983), and only recently have some general ideas been outlined (*e.g.*, in SLATKIN 1985). Investigation of the relationship between spatial statistics and the parameters describing the genetic structure of populations seems thus an important research priority (CLEGG and EPPERSON 1985).

In this paper a simple relation is described among three quantities: the kinship coefficient, WRIGHT's F_{st}, and a widely employed measure of spatial autocorrelation, MORAN's I. This allows prediction of the expected set of autocorrelation coefficients under a model of isolation by distance, assuming exponential decline of genetic similarity in space (MALÉCOT 1973; MORTON 1973b). As a consequence, consistency of the observed gene frequency patterns with the pattern that should be brought about by interaction of gene flow and short-range migration may be tested by comparing observed and expected correlograms. This appears therefore a first step toward application of spatial autocorrelation methods within the framework of current population-genetic theories.

MORAN'S I

Spatial autocorrelation methods have been developed since the early fifties (MORAN 1950) [for recent advances, see RIPLEY (1981) and CLIFF and ORD (1981)], but their use in biology was pioneered by SOKAL and his associates, whose symbolism will be used here with minor changes. Let $\mathbf{p}$ $(p_1, p_2, \cdots, p_n)$ be the array of frequencies of one allele in n different populations. The geographical sites where the populations live are initially plotted on a map, which lies either on a plane or on a spherical surface. The points on the map may be joined to other population points by a particular network (GABRIEL and SOKAL 1969); alternatively, all possible connections between pairs of localities may be taken into account. Such air distances (or great circle distances) are considered as reasonably representative of spatial distances for population genetics studies (CRUMPACKER *et al.* 1976). There is no particular reason to choose them (*e.g.*, as in BARBUJANI 1987) rather than distances along a connection network (*e.g.*, as in SOKAL and MENOZZI 1982), except that under MORTON's models, which will be outlined more accurately later, kinship values are currently regressed on air distances (MORTON 1973b). In choosing to connect the populations as the crow flies, we of course neglect environmental barriers that affect selection gradients as well as gene flow directions.

Each autocorrelation coefficient refers to a distance class. The choice of the class intervals depends on the distribution of the sites of interest, and on the scale the investigator wishes to emphasize, but there is no established standard criterion. A connectivity matrix $\mathbf{W}$, whose rows and columns represent the localities, is then constructed for each allele and each distance class. Although in some cases $\mathbf{W}$ may become complex, it is usually a square binary symmetrical matrix. Its elements, w_{ij}, assume the value of unity when the ith and jth localities are separated by a distance falling in the class considered, otherwise they are equal to 0.

Given $\mathbf{p}$ and $\mathbf{W}$, an autocorrelation coefficient, Moran's I, is calculated as

$$I = n \sum_{ij,\, i \neq j} w_{ij}(p_i - \bar{p})(p_j - \bar{p}) \Big/ \left[W \sum_i (p_i - \bar{p})^2 \right] \quad (1)$$

where n and w_{ij} have been defined above, p_i and p_j are the allele frequencies in the ith and jth population, respectively, and $\bar{p}$ is the mean over all populations; W is the sum of all w_{ij} for that distance class, *i.e.*, twice the number of edges connecting localities at that distance for a binary $\mathbf{W}$. Under a randomization hypothesis the expected value is $E(I) = -(n-1)^{-1}$, and the formulas for its standard error are in SOKAL and ODEN (1978a).

Spatial autocorrelation coefficients indicate whether the values of a variable influence each other, and measure the strength of their association. Thus, a significant positive autocorrelation indicates that at the distance considered allele frequencies are similar, *i.e.*, both deviate from the mean in the same direction; a significant negative coefficient stands for dissimilarity; a nonsignificant value means that there is no consistent relationship between pairs of allele frequencies at that distance. In general, for large samples I ranges from -1 to $+1$, the greater the absolute value, the stronger the relationship. However, it may occasionally exceed these limits when particular sets of localities are considered (DE JONG, SPRENGER and VAN VEEN 1984), hence it is not strictly a correlation coefficient (CLIFF and ORD 1981). The plot of I values versus distance is referred to as spatial correlogram of that allele.

THE KINSHIP COEFFICIENT

CAVALLI-SFORZA and BODMER (1971) define the coefficient of kinship between two individuals A and B as the probability that a gene taken at random from A, at a given locus, may be identical by descent to a gene taken at random from B at the same locus. When populations and not individuals are taken into ac-

count, genes are compared for identity of the allele they carry rather than identity by descent (MORTON 1973d). If only short-range migration and random genetic drift affect a population, that is to say, under isolation by distance, variation of kinship may be expressed through a formula due to the work of MALÉCOT (see MALÉCOT 1973) and MORTON (MORTON, MIKI and YEE 1968; MORTON 1973a, b):

$$\varphi(d) = (1 - L)ae^{-bd} + L. \qquad (2)$$

Here $\varphi(d)$ is the kinship at distance d, e is the base of the natural logarithms, and L is an estimate of kinship at infinite distance (in practice, it is the lowest value in the matrix of observed kinship). For a detailed discussion of the meaning and calculation of these quantities the reader may refer to MORTON (1973b) and WIJSMAN and CAVALLI-SFORZA (1984). Three methods allow estimation of the parameters a and b: two of them are predictive (those based on genealogies and migration), and one is inductive (bioassay of kinship from gene frequencies, metric traits, or surnames) (MORTON 1975). As far as bioassay is concerned, a convenient measure of kinship between two populations i and j is

$$f_{ij} = \frac{(p_i - \bar{p})(p_j - \bar{p})}{\bar{p}(1 - \bar{p})} \qquad (3)$$

where all the quantities involved are the same as defined above for Moran's I.

The values of f_{ij} may be averaged over several loci, leading to a square symmetrical matrix of pairwise kinship, which, according to MORTON (1973a), contains all the relevant information about population structure in any generation. The parameters a and b of Equation 2 are eventually estimated by regression of logarithm of kinship on the geographic distance between populations.

RELATING THE KINSHIP COEFFICIENT TO MORAN'S I

Let us consider the kinship coefficient between populations, estimated through bioassay at one locus. Let us assume also that k pairs of localities fall in a distance class whose lower and upper limits are, say, d_L and d_U, respectively, and whose midrange is d. We also recall that k is $W/2$ in (1). The average kinship between populations within that distance class is

$$f = \frac{\sum_{ij} (p_i - \bar{p})(p_j - \bar{p})}{k\bar{p}(1 - \bar{p})} \qquad (i > j) \qquad (4)$$

where the summation is over the k pairs of populations in the distance class of interest. Moran's I, for the same distance class and with a binary connectivity matrix, is

$$I = \frac{n \sum_{ij} w_{ij}(p_i - \bar{p})(p_j - \bar{p})}{2k \sum_{i} (p_i - \bar{p})^2} \qquad (i \neq j) \qquad (5)$$

which may be rewritten as

$$I = \frac{n\sum_{ij} (p_i - \bar{p})(p_j - \bar{p})}{2k \sum_{i} (p_i - \bar{p})^2} \qquad (i \neq j) \qquad (6)$$

(the term w_{ij} is neglected, as all pairs of populations in that distance class, and only they, have $w_{ij} = 1$ as a multiplier). Since $(p_i - \bar{p})(p_j - \bar{p})$ is equal to $(p_j - \bar{p})(p_i - \bar{p})$,

$$\sum_{ij \neq j} (p_i - \bar{p})(p_j - \bar{p}) = 2 \sum_{ij > j} (p_i - \bar{p})(p_j - \bar{p}). \qquad (7)$$

Now we may call $D = \Sigma_{ij>j} (p_i - \bar{p})(p_j - \bar{p})$, and S the sum of squares of p, *i.e.*, $S = \Sigma_i (p_i - \bar{p})^2$; substituting in (4) and (6) yields

$$f = D/[\bar{p}(1 - \bar{p})k] \qquad (8)$$

and

$$I = nD/kS. \qquad (9)$$

Then

$$I/f = \bar{p}(1 - \bar{p})n/S. \qquad (10)$$

If the sample is large, $n \simeq n - 1$, and $S/n \simeq \sigma_p^2$ *i.e.*, the variance of p. Accordingly

$$I = f\bar{p}(1 - \bar{p})/\sigma_p^2 \qquad (11)$$

and, since $\sigma^2{}_p/[\bar{p}(1 - \bar{p})] = F_{st}$, then

$$I = f/F_{st}. \qquad (12)$$

MORAN'S I, for any allele and at any distance, is therefore the ratio of the kinship as estimated for that allele and that distance to F_{st}. In populations whose genetic structure is accounted for by isolation by distance, the kinship coefficient varies in space according to Equation 2. As a consequence, in such populations the values of I as a function of distance, for any allele A, are expected to be

$$E(I(d))_A = \varphi(d)/F_{st,A}, \qquad (13)$$

where $F_{st,A}$ is the standardized variance of allele A frequencies, and $\varphi(d)$ depends only on distance, and not on the particular allele studied. This is an exact relation if d is the average distance between pairs of populations; when it is the midrange of that distance class, the relation is approximate.

DISCUSSION

In incompletely isolated populations genetic drift tends to cause local genetic differentiation, and gene flow between neighboring populations tends to contrast this process. The distribution of gene frequencies in an area where neither differential selection nor long range gene flow occur results from the balance between these two factors. They affect all loci to the same degree (SLATKIN 1985, 1987), whereas there is no reason to assume that the impact of differential selection, if any, should be equal at each locus involved. Isolation by distance models account for the pattern of genetic similarity between populations which is expected when genetic drift and short range migration determine gene frequency variation (KIMURA and WEISS 1964; MALÉCOT 1973; MORTON 1973a, b). In this case, a unique function, Equation 2, describes the decline of kinship with distance. In several regions the values of φ have actually been shown to decrease exponentially as the distance between sampled populations increases (MORTON 1982).

Under isolation by distance the expected autocorrelation of gene frequencies is therefore a function of two factors: the mode of decline of genetic relatedness, which is expressed by the parameters a, b and L of MALÉCOT-MORTON equation, and does not depend on the marker considered; and F_{st}, which assumes a particular, measurable value for each allele. This applies to any selectively neutral gene. The values on the main diagonal of the matrix of pairwise kinship should not contribute to estimation of the parameters a and b in MALÉCOT-MORTON function, since they do not contribute to calculation of MORAN's I.

At neutral loci, thus, MORAN's I should decline monotonically with distance, paralleling the decrease of kinship and tending asymptotically to $\varphi(\infty)/F_{st} \simeq L/F_{st}$. When, at distance close to 0, $\varphi(d) = F_{st}$ (WRIGHT 1969, pp. 294–295; CAVALLI-SFORZA and BODMER 1971, p. 399), then $I(d)$ tends to 1. An exponentially decreasing profile of the correlograms has actually been observed in the analysis of gene frequency patterns simulated by computer under isolation by distance (SOKAL and WARTENBERG 1983).

If the frequencies of one allele evolved under isolation by distance, then the correlogram of that allele is predictable by using Equation 13, where the theoretical values of kinship, $\varphi(d)$, have been estimated on the basis of either genealogies, or migration data, or surnames, or a set of selectively neutral genes. Deviations from this expected set of values would support non-neutrality of the allele. As an example, in Table 1 the kinship parameters estimated in two human populations are employed for prediction of autocorrelation at various distances for two erythrocyte markers.

As yet, however, the possible applications of the expression (13) are limited by at least two unsolved statistical problems. First, only methods for testing single correlograms versus their null expectations exist (ODEN 1984), whereas no method has been established for comparison of two correlograms. While the relation (13) may thus be used for comparison of pairs of expected and observed I values, and may allow detection of sharp deviations from the spatial pattern predicted by MALÉCOT and MORTON's model, it cannot at this time be applied for rigorous testing of the hypothesis that the frequencies of one allele have evolved under isolation by distance.

Second, kinship coefficients are generally calculated on the same gene frequencies on which spatial autocorrelation analysis is carried out. In this way, the same data would be used both for generating a hypothesis and for testing it, which is surely incorrect and would affect, to an unpredictable but wide extent, the reliability of subsequent inferences.

A third class of statistical problems is more easy to cope with. As previously remarked, autocorrelation studies are often based on a connection network rather than interpopulation air distances. However, this does not affect the relationships evaluated above, if the same network is taken into account in constructing also the matrix of geographic distances employed in estimation of a and b. Besides, the above relations hold true for autocorrelation of gene frequencies

TABLE 1

Expected values of MORAN's I for two markers in two human populations

Distance (km)	Kinship	Expected I (*Ada*)	Expected I (*Ak*)
POPULATION: Brazil BIOASSAY OF KINSHIP: from isonymy (IMAIZUMI and MORTON 1969) $a = 0.0231$ $b = 0.0073$ $L = -0.0012$			
100	0.0099	0.5438	0.4379
200	0.0041	0.2253	0.1815
300	0.0014	0.0769	0.0620
400	0.0001	0.0026	0.0021
500	−0.0006	−0.0325	−0.0262
Infinite	−0.0012	−0.0659	−0.0531
POPULATION: Po delta, Italy BIOASSAY OF KINSHIP: From gene frequencies (BARRAI *et al.* 1983) $a = 0.0142$ $b = 0.0380$ $L = -0.0053$			
20	0.0014	0.0756	0.0619
40	−0.0022	−0.1197	−0.0973
60	−0.0038	−0.2110	−0.1681
80	−0.0046	−0.2536	−0.2035
100	−0.0050	−0.2737	−0.2212
Infinite	−0.0053	−0.2912	−0.2345

The F_{st} estimates, based on European and Asian populations, are 0.0182 for *Adenosine deaminase* (*Ada*) and 0.0226 for *Adenylate kinase* (*Ak*) (BARBUJANI and MILANI 1986). a, b and L are the parameters of MALÉCOT-MORTON function.

involving a binary symmetrical matrix **W**. It is possible, but not effortless, to accommodate more complex weight matrices.

When kinship is bioassayed from gene frequencies, the slope of the MALÉCOT-MORTON function may be affected by mutation and selection (see *e.g.*, BARRAI *et al.* 1984; BARBUJANI and CANELLA 1987). Selective pressures, in particular, have a substantial influence on gene frequency diversity, and this may lead to incorrect assessment of the overall genetic similarity between populations (CAVALLI-SFORZA 1984; WIJSMAN and CAVALLI-SFORZA 1984). Such an error will in turn cause improper evaluation of the expected correlograms when Equation 13 is applied. In other words, if the expected correlogram includes the effect of selection, comparing it with the correlogram observed for a given allele no longer tests the consistency of that allele's frequencies with those expected under isolation by distance. Thus, observed autocorrelation coefficients should be compared with expected values computed from a kinship matrix including only neutral genes. Cross comparisons between loci whose neutrality is controversial or unknown are a possible working solution. Alternatively, the use of surnames (YASUDA *et al.* 1974) or migration matrices (MORTON 1973c) for bioassay or prediction of kinship, respectively, can be suggested.

Marital migration data, in particular, allow construction of a kinship matrix which is surely unaffected by selective processes. As a consequence, systematic pressures over some of the loci of interest should have an impact only on the observed correlograms, and would not affect the expected ones. This would also solve the second statistical problem cited above, since kinship and autocorrelation coefficients would be evaluated on the basis of two independent sets of data. Accordingly, the use of MALÉCOT-MORTON function for prediction of spatial autocorrelation seems possible especially in studies dealing with fairly small geographical areas, in agreement with the view put forward in the first works on isolation by distance (YASUDA and MORTON 1969).

Contribution 639 in Ecology and Evolution from the State University of New York at Stony Brook. This work was supported by a grant of the Italian Consiglio Nazionale delle Ricerche, Comitato per le ricerche tecnologiche, and by Ministero delle Pubblice Instruzione 40% research fund "Genetica Ecologica." I wish to thank the following, whose support, criticisms and suggestions were important in the development of this paper: ROBERT R. SOKAL, GIAN ANTONIO DANIELI, PAMELA A. DERISH, ROSALIND M. HARDING, GUIDO MODIANO, NEAL L. ODEN, SILVANO PRESCIUTTINI and ANTONELLA RUSSO.

LITERATURE CITED

BARBUJANI, G., 1985 A two-step test for the heterogeneity of F_{st} values at different loci. Hum. Hered. **35:** 292–295.

BARBUJANI, G., 1987 Diversity of some gene frequencies in European and Asian populations. III. Spatial correlogram analysis. Ann. Hum. Genet. In press.

BARBUJANI, G., and R. CANELLA, 1987 Diversity of some gene frequencies in European and Asian populations. II. Fit of an isolation by distance model. Hum. Hered. **37:** 265–272.

BARBUJANI, G., and F. MILANI, 1986 Diversity of some gene frequencies in European and Asian populations. Effects of longitude. J. Hum. Evol. **15:** 61–69.

BARRAI, I., G. BARBUJANI, M. BERETTA and C. VULLO, 1983 Heterozygosity and geographic distances in a limited area. J. Hum. Evol. **12:** 403–408.

BARRAI, I., A. ROSITO, G. CAPPELLOZZA, G. CRISTOFORI, C. VULLO, C. SCAPOLI and G. BARBUJANI, 1984 Beta-thalassemia in the Po delta: selection, geography and population structure. Am. J. Hum. Genet. **36:** 1121–1134.

CAVALLI-SFORZA, L. L., 1966 Population structure and human evolution. Proc. Royal Soc. London B **164:** 362–379.

CAVALLI-SFORZA, L. L., 1984 Isolation by distance. pp. 229–248. In: *Human Population Genetics. The Pittsburgh Symposium*, Edited by A. CHAKRAVARTI. Van Nostrand Reinhold, New York.

CAVALLI-SFORZA, L. L., and W. F. BODMER, 1971 *The Genetics of Human Populations.* W. H. Freeman, San Francisco.

CLEGG, M. T., and B. K. EPPERSON, 1985 Recent developments in population genetics. Adv. Genet. **23:** 235–269.

CLIFF, A. D., and J. K. ORD, 1981 *Spatial Processes: Models and Applications.* Pion, London.

CRUMPACKER, D. W., G. ZEI, A. MORONI and L. L. CAVALLI-SFORZA, 1976 Air distance versus road distance as a geographical measure for studies on human population structure. Geogr. Anal. **8:** 215–223.

DE JONG, P., C. SPRENGER and F. VAN VEEN, 1984 On extreme values of Moran's I and Geary's c. Geogr. Anal. **16:** 17–24.

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

GABRIEL, K. R., and R. R. SOKAL, 1969 A new statistical approach to geographic variation analysis. Syst. Zool. **18:** 259–270.

IMAIZUMI, Y., and N. E. MORTON, 1969 Isolation by distance in Japan and Sweden compared with other countries. Hum. Hered. **19:** 433–443.

KIMURA, M., and G. H. WEISS, 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49:** 561–576.

LEWONTIN, R. C., and J. KRAKAUER, 1973 Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. Genetics **74:** 175–195.

MALÉCOT, G., 1973 Isolation by distance. pp. 72–75. In: *Genetic Structure of Populations*, Edited by N. E. MORTON. University of Hawaii Press, Honolulu.

MENOZZI, P., A. PIAZZA and L. L. CAVALLI-SFORZA, 1978 Synthetic maps of human gene frequencies in Europeans. Science **201:** 786–792.

MORAN, P. A. P., 1950 Notes on continuous stochastic phenomena. Biometrika **37:** 17–23.

MORTON, N. E., 1973a Kinship and population structure. pp. 66–69. In: *Genetic Structure of Populations*, Edited by N. E. MORTON. University of Hawaii Press, Honolulu.

MORTON, N. E., 1973b Isolation by distance. pp. 76–77. In: *Genetic Structure of Populations*, Edited by N. E. MORTON. University of Hawaii Press, Honolulu.

MORTON, N. E. 1973c Prediction of kinship from a migration matrix. pp. 119–123. In: *Genetic Structure of Populations*, Edited by N. E. MORTON. University of Hawaii Press, Honolulu.

MORTON, N. E., 1973d Kinship bioassay. pp. 158–163. In: *Genetic Structure of Populations*, Edited by N. E. MORTON. University of Hawaii Press, Honolulu.

MORTON, N. E., 1975 Kinship, fitness and evolution. pp. 133–154. In: *The Role of Natural Selection in Human Evolution*, Edited by F. M. SALZANO. North-Holland, Amsterdam.

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

MORTON, N. E., C. MIKI and S. YEE, 1968 Bioassay of population structure under isolation by distance. Amer. J. Hum. Genet. **20:** 411–419.

ODEN, N. L., 1984 Assessing the significance of a spatial correlogram. Geogr. Anal. **16:** 1–16.

PIAZZA, A., P. MENOZZI and L. L. CAVALLI-SFORZA, 1982 Synthetic gene frequency maps of man and selective effects of climate. Proc. Natl. Acad. Sci. USA **28:** 2638–2642.

RIPLEY, B. D., 1981 *Spatial Statistics.* John Wiley & Sons, New York.

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

SLATKIN, M., 1987 Gene flow and the geographic structure of natural populations. Science **236:** 787–792.

SOKAL, R. R., 1979 Ecological parameters inferred from spatial correlograms. pp. 167–196. In: *Contemporary Quantitative Ecology and Related Ecometrics,* Edited by G. P. PATIL and M. ROSENZWEIG. International Co-operative Publishing House, Fairland, MD.

SOKAL, R. R., and P. MENOZZI, 1982 Spatial autocorrelations of HLA frequencies in Europe support demic diffusion of early farmers. Am. Nat. **119:** 1–17.

SOKAL, R. R., and N. L. ODEN, 1978a Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. **10:** 199–228.

SOKAL, R. R., and N. L. ODEN. 1978b Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biol. J. Linn. Soc. **10:** 229–249.

SOKAL, R. R., and D. E. WARTENBERG, 1981 Space and population structure. pp. 186–213. In: *Dynamic Spatial Models,* Edited by D. GRIFFITH and R. MCKINNON. Sijthoff and Noordhoff, Alphen aan der Rijn, The Netherlands.

SOKAL, R. R., and D. E. WARTENBERG, 1983 A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics **105:** 219–237.

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:** 259–287.

WIJSMAN, E. M., and L. L. CAVALLI-SFORZA, 1984 Migration and genetic population structure with special reference to humans. Annu. Rev. Ecol. Syst. **15:** 279–301.

WRIGHT, S., 1969 *Evolution and the Genetics of Populations. Vol. II: The Theory of Gene Frequencies.* University of Chicago Press, Chicago.

YASUDA, N., and N. E. MORTON, 1969 Studies on human population structure. pp. 249–265. In: *Proceedings of the 3rd International Congress of Human Genetics,* Edited by J. F. CROW and J. V. NEEL. Johns Hopkins Press, Baltimore.

YASUDA, N., L. L. CAVALLI-SFORZA, M. SKOLNICK and A. MORONI, 1974 The evolution of surnames: an analysis of their distribution and extinction. *Theor. Popul. Biol.* **5:** 123–142.

Communicating editor: M. T. CLEGG