Spatial autocorrelation in biology
2. Some biological implications and four applications of evolutionary and ecological interest

ROBERT R. SOKAL, F.M.L.S.

Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794, U.S.A.

AND

NEAL L. ODEN

Department of Ecology and Evolutionary Biology, The University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

Spatial autocorrelation analysis tests whether the observed value of a variable at one locality is significantly dependent on values of the variable at neighbouring localities. The method was extended by us in an earlier paper to include the computation of correlograms for spatial autocorrelation. These show the autocorrelation coefficient as a function of distance between pairs of localities, and summarize the patterns of geographic variation exhibited by the response surface of any given variable. Identical variation patterns lead to identical correlograms, but different patterns may or may not yield different correlograms. Similarity in the correlograms of different variation patterns suggests similarity in the generating mechanism of the pattern.

The inferences that can be drawn from correlograms are discussed and illustrated. Examination and analysis of variation patterns of several characters or gene frequencies for one population, or of several populations in different places or at different times, permit some conclusions about the nature of the populational processes generating the observed patterns.

Autocorrelation analysis is applied to four biological situations differing in the nature of the data (interval or nominal), in the type of grid connecting the localities (regular or irregular), and the field of application (evolution or ecology). The examples comprise genotypes of individual mice, blood group frequencies in humans, gene frequency variation in a perennial herb, and the distribution of species of trees. The implications of our findings are discussed.

KEY WORDS: − spatial autocorrelation − geographic variation analysis − spatial variation patterns − *Mus musculus* − ABO blood groups − *Liatris cylindracea* − tree distribution patterns.

CONTENTS

INTRODUCTION

In an earlier paper in this journal (Sokal & Oden, 1978) we have brought the technique of spatial autocorrelation developed by statistical geographers (Cliff & Ord, 1973) to the attention of biologists. This technique tests whether the observed value of a nominal, ordinal, or interval variable at one locality is significantly dependent on values of the variable at neighbouring localities. The method was extended by us to include the computation of spatial correlograms which show changes in the autocorrelation coefficients as a function of increasing distances in the study area and summarize the patterns of geographic variation exhibited by the response surface for any given variable.

Spatial correlograms describe the underlying spatial relationships of a surface, as discussed by us in Sokal & Oden (1978). The paper cited describes the method in detail by means of two simple examples, furnishes computational formulae, expected values and standard errors, and illustrates one example (in depth)—on microgeographic variation in the snail *Helix aspersa*. Readers are referred to that paper for the technical background needed for an understanding of the biological examples and interpretations presented below.

The present paper features discussion of the biological implications of spatial autocorrelation analysis, expecially as it relates to spatial correlograms, followed by an account of four applications of the method to evolutionary and ecological problems. The latter were chosen not only because of their intrinsic interest but because they are representive of the types of problems of interest to population biologists.

BIOLOGICAL IMPLICATIONS OF AUTOCORRELATION AND CORRELOGRAMS

Population differentiation

Sokal (1978) has considered four models for the origin of patterns of population differentiation. In the first, a character is differentiated in response to an environmental gradient, producing a cline. In the second, environmental patches are heterogeneous among themselves but homogeneous within. The patches may or may not exhibit spatial ordering in the terms of the environmental variables they comprise. A third model is differentiation due to the classical isolation-by-distance model. A fourth model ascribes the differ-

entiation to historical factors. These could be of several kinds: (a) accidental migration and establishment of different types—the so-called founder effect resulting in a patchy distribution—or (b) accidental settlement by two or more types with subsequent production of a gradient between them by diffusion processes. (c) Another historical factor may be systematic migration into an area by several populations that had differentiated elsewhere.

In the first two models selection brings about population differentiation, while the last two imply stochastic differentiation. The models are unlikely to be found pure for any one character and real situations in nature will probably constitute mixtures of two or more types.

Statistical heterogeneity and patterns

Differentiation of a variable distributed over a geographic space subsumes two concepts that need to be distinguished—statistical heterogeneity and pattern. Clearly, statistical heterogeneity of values, as demonstrated, for example, for means by an analysis of variance, does not necessarily imply geographic pattern (i.e. a departure from a random spatial arrangement of these values as tested by the methods of Royaltey et al. (1975), or by the techniques expounded in this paper). Statistically heterogeneous values, such as means or gene frequencies, can be randomly positioned in a geographic space. But it is at least of theoretical interest that significant pattern can exist in the absence of statistical heterogeneity. For example, a variable measured on a population of organisms distributed in colonies may have nonsignificant among-colony variance even if colonies are spatially arranged so that colony means are autocorrelated. Such a situation could be created artificially by arranging a series of random samples from the same population in some deliberate spatial pattern, say, to have the means form a cline—the means will continue to be statistically homogeneous (by analysis of variance), but tests for patterns will reject the hypothesis of spatial independence of the means. Thus, heterogeneity and spatial patterning of means are potentially independent of each other and geographic variation patterns must in all cases be examined for both of these properties. When we do so for an interval variable we can logically obtain four possible combinations.

Significant heterogeneity of means accompanied by significant spatial patterns will be the most common result. We present four situations that could lead to such outcomes, followed by the number of the differentiation model considered above.

(1) Migration (with or without interbreeding) causes similarity between neighbouring populations, whereas distant populations differ for the variable studied. (Model IV, founder effects with gene flow among neighbouring populations tending to diminish the original differences and to produce local autocorrelation.)

(2) There are frequent local extinctions followed by the establishment of new colonies by relatively nearby founders. (Model IV, founder effects.)

(3) The selective agents are patterned in space, either as small patches (diameter less than interlocality distances) arranged as gradients or other patterns, or as larger patches that are internally homogeneous but differ from

other such large patches. (Models I or II, selection in response to environmental gradients or patterned patches.)

(4) There has been a transfer en masse of a set of phenotypically distinct populations from outside the study area. Under this model either previously existing patterns are replicated, or new patterns are formed by waves of cohesive, phenotypically distinct populations. (Model IV, systematic migration.)

Significant heterogeneity of means in the absence of pattern could occur in at least three situations.

(1) Drift determines the character. Migration and selection are weak. (Model III, isolation-by-distance with low vagility.)

(2) There are frequent local extinctions followed by the establishment of new colonies by founders. Colony size is small relative to average interlocality distance, but the founders come with near equiprobability from the entire array of colonies over the range of the species. Migration and selection are weak. This is effectively the same situation as the previous one, although the mechanism differs. (Model IV, founder effects.)

(3) The selective agents are either unpatterned or patterned such that the patterns are composed of components much smaller than average interlocality distance. An example might be populations occurring on tufts of vegetation surrounded by uninhabitable areas, with distances between tufts in excess of the tuft diameter and with the phenomes of populations on these tufts strictly controlled by the local environment through selection or phenetic plasticity. (Model II, unpatterned heterogeneous patches.)

Homogeneity of means, accompanied by significant spatial patterns are conceptually possible. Such a combination could arise from a reshuffling of means by an external directive agent—an unlikely occurrence in nature. The combination would also result if populations of organisms aligned themselves on a gradient or in some other pattern (Model I or II) as a function of their phenotype; but this would imply that samples stay together and responded as a function of their means and not of their individual phenotypes, an unlikely situation. A probable explanation for significant patterns in the absence of statistical heterogeneity is large within-sample error coupled with inadequate sample sizes for the analysis of variance or similar tests.

Homogeneity of means in the absence of patterns implies lack of population differentiation. Two plausible explanations are: (1) The vagility of the organisms is such as to comprise the entire study area; given interbreeding this leads to effective panmixia.

(2) A less likely explanation would be uniformity of environment over the area, hence uniformity of selection. The intensity of selection should be high or population sizes large, preventing drift.

Similar reasoning applies in the case of nominal data where the units of study are individuals, but we shall not pursue this matter.

Biological explanations of spatial autocorrelation

We next examine what biological phenomena will cause various types of autocorrelation. It is useful to visualize spatial autocorrelation as predictability of means a given distance apart. Positive autocorrelation predicts homogeneity or sameness; negative autocorrelation, heterogeneity or dissimilarity. Analogs

to information statistics should be obvious here, although we shall not pursue the point. Within the discussion that follows we shall first discuss autocorrelations of continuous variables. Cases in point are sample means in morphometric data or gene frequencies. We shall subsequently take up nominal cases where each sampling unit is an individual.

Low order (short distance) positive autocorrelation can be due to migration (probably though not necessarily associated with gene flow) over the distances involved. Another model would assume positive autocorrelation of the selective agent, presumably some aspect of the environment. If the variable studied is patchy, the relation between distances within and among patches and among sampling points is the key to the profile of the correlogram, as pointed out in the earlier paper (Sokal & Oden, 1978). Low order positive autocorrelation could stem from small favourable patches whose diameter is less than interlocality distances. In such a case these small patches would need to show clumping of similar environments in space, a quite plausible assumption for many ecological islands. Or the patches could be intermediate in size with diameters large enough to accommodate a few locality samples. In such a case the patches would have to be sufficiently apart so that most if not all the short-distance correlations are within patches and thus positive. Such environmental structure should be quite common in nature in the case of large ecological islands. Low order positive autocorrelation can also result from large patches that are homogeneous within patches, while differing among patches. The pattern must be such that most of the short distance comparisons are within patches. The plausibility of large, differing, but internally homogeneous patches depends on the nature of the biological material being studied and the resolution at which it is being studied. In studies involving distinct geographic races low order positive autocorrelation should be frequent within areas inhabited by any one race. Here the factor responsible cannot be directly attributed to ongoing natural selection and might more properly be considered a historical factor. Even in a fine-grained environment it is conceivable that different genotypes "choose" their patches.

Low order (short distance) negative autocorrelations can be observed when there are spatially heterogeneous small patches in the environment in populations subject to strong selective pressures or to considerable phenetic plasticity. Patch size must be less than interlocality distances. Migratory patterns resulting in low-order negative autocorrelations are conceivable but implausible.

High order (long distance) positive autocorrelation will occur with circular or at least symmetrical gradients. Thus in a circular gradient, localities a full diameter apart will presumably experience the same environmental conditions. A similar ordering of patches in terms of their mean would give such a result. Populations on mountain slopes or along shore lines of bogs and ponds come to mind readily.

High order (long distance) negative autocorrelations would be found when the most different localities are farthest apart. Coupled with low order positive autocorrelation, this should be the most common correlogram profile observable in natural populations studied over large distances, as in a cline.

Let us briefly turn to a discussion of the implications of autocorrelations for *nominal data.* These will usually be cases where individual organisms are the

sampling unit (e.g. individual genotypes of mice in a barn or individual species of trees in the study area). Positive low order correlation indicates excessive numbers of homotypic pairs. This suggests attraction of like individuals, or clumping, because nearby organisms are genetically related. Various forms of vegetative reproduction such as budding and reproduction by rhizomes will give rise to such a pattern. Natural selection on homogeneous patches encompassing several sampling units (individuals) will also give rise to such autocorrelation. Thus if a moist area favours genotype A, whereas a dry area favours genotype B, there should be positive autocorrelation among the A's and the B's if there are alternating wet and dry patches in the area and more than one individual can be contained per patch.

Low order negative correlation in the nominal case indicates repulsion of like individuals. The allelopathy model comes to mind immediately but various genetic mechanisms can also result in this pattern. Examples are mating types in protozoa and plants and various genetic mechanisms resulting in the association of unlike parent and offspring or unlike offspring.

Higher order positive autocorrelation in the nominal case seems unlikely. Such a condition implies an excess of like pairs at large distances. Very strong cases of self-incompatibility in the ecological sense might be responsible for such patterns, but such cases are probably rare in nature.

Finally, high order negative autocorrelations for nominal data would occur in large homotypic aggregations of individuals which are replaced in space at greater distances by different aggregations, as for example on a gradient.

Comparison of patterns and correlograms

Before comparing different patterns and correlograms, either for several characters in the same area or for the same characters in different areas, we need to make a subtle but important distinction between the realization of the variation pattern and its summarization by means of a correlogram. Loosely speaking, two patterns over the same geographical area are similar if their surfaces are strongly positively correlated.

It is self-evident that similar patterns should yield similar correlograms. However, similar correlograms do not ensure parallel patterns in the sense described above. The correlograms summarize the relationships among means of the space, but identical relationships may be realized from quite different patterns. It is easiest to see this on a regular square grid. Assume the means have been mapped on the vertices to indicate a clinal trend. A monotonically decreasing correlogram typical for a cline would result. If we now rotate the lattice by 90°, what may formerly have been a north-south cline has now become an east-west cline, and in terms of fixed geographic location will result in a very different pattern. Yet the correlogram for these new rotated data will be identical to the old one. Spatial correlograms thus describe the underlying spatial relationships of a surface, and for this reason they are probably closer guides to some of the processes that have generated the surface than is the surface itself. Thus, while identical variation patterns will yield identical correlograms, different variation patterns may or may not yield different correlograms. Conversely, different correlograms must be based on different patterns, but identical correlograms may result from identical or different

patterns. These relations are illustrated in Fig. 1. Observed biological response surfaces are presumably the result of both fixed, exogenous influences (environmental patches, etc.), which may or may not be autocorrelated, and endogenous influences arising from interaction among the organisms themselves (e.g., gene flow, bird dialects, allelopathy). These endogenous influences by their very nature, cause autocorrelation (migration, etc.). If two variables have different patterns but similar correlograms, this probably means that both variables are largely determined by the same endogenous process. If, however, the patterns are the same, this probably means that exogenous influences, are significant determinants of the surface.

IDENTICAL VARIATION PATTERNS   ⟶   IDENTICAL CORRELOGRAMS

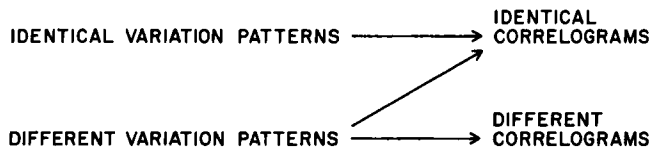DIFFERENT VARIATION PATTERNS   ⟶   DIFFERENT CORRELOGRAMS

Figure 1. Relation between patterns and correlograms based on different characters of the same populations. For explanation see text.

When different characters (or gene frequencies) measured for the same populations in a given area show the same variation pattern and therefore identical (or similar) correlograms, this suggests a common response to selective agents that differ over the area. Other explanatory hypotheses, such as several selective agents which are themselves correlated over space, are less probable. Morphometric characters could be correlated over space because of pleiotropy.

Different variation patterns yielding the same correlograms should occur in populations where migration strongly affects geographic variation of gene frequencies or morphometric characters. The several gene frequency patterns need not correspond for historical reasons such as different immigration waves or because of differential selection patterns. But since a migrant individual must carry with it all its genes, autocorrelation should be the same for all loci, and correlograms, or at least portions thereof, should resemble each other. Similar spatial autocorrelation of separate environmental variables leading to different variation patterns but similar correlograms is conceivable but unlikely.

Different geographic variation patterns associated with different correlograms should be found in populations in which differential selection among loci predominates. This should be the most frequent pairing of conditions encountered in nature. The higher order autocorrelations could also differ for historical reasons such as invasions by populations differing for one character but not the other. In the discussion above, we have ignored the possibility of gene frequency dependent migration which would further complicate our conclusions.

When comparing the same characters for different populations and areas, we cannot in any strict sense speak of similar geographic variation patterns since any geographic area is unique in space. So it is best to compare correlograms as indicators of the exogenous and endogenous processes making the geographic pattern.

Variation of the same character in the same area but at different times can be compared in patterns and correlograms and as we have seen in earlier

sections, autocorrelations can be computed for the differences as well. The general reasoning applied in the earlier instances should apply in this case also. It is important that the two successive times be far enough apart (in terms of the biology of the organisms being studied) so that the patterns can be expected to be uncorrelated. However, one further instance needs to be considered. If two variation patterns of the same character studied over the same area but examined at different times have different patterns but identical correlograms, we have the likelihood (but not proof) of stationarity. That is, the generative process has brought the phenomenon to a stationary steady state, a dynamic equilibrium in which the moments of the distribution are constant, although its realization varies through time. Numerous stochastic models which would give such results can be postulated.

APPLICATION TO FOUR EXAMPLES

We shall now apply the methodology of spatial autocorrelation analysis and the inferences developed in the last section to four examples especially chosen because of their inherent interest and because they represent types of situations and problems of interest to population biologists.

Mus musculus in a barn

These well-known data are an example of autocorrelation of nominal data. They are an analysis of population structure at its finest level—genotypes of individuals. The data were collected from populations of the house mouse trapped in a chicken barn (Selander, 1970) and have been cited extensively (e.g. Lewontin, 1974; Ricklefs, 1973). Selander sampled mice in a 15 x 57 grid (with intersections 1.6 ft apart) over an area of 24 x 96 ft in half of one barn. The results of an assay of the esterase-3 genotypes of the 378 adult and juvenile mice are shown in Fig. 2. Selander, leaving out some marginal grid squares and grouping the data into 22 quadrats of 25 basic grid squares each, found that there was significant heterogeneity of gene frequencies for this locus as well as two other loci, Esterase-2 and *Hbb*. These three loci were also analyzed in a geographic variation study of mice in Texas (Selander, Yang & Hunt, 1969). While these loci show significant heterogeneity within the barn, one, hemoglobin, did not show macrogeographic variation, although it exhibits the same degree of microgeographic variation as the other loci. The other two allozyme frequencies show only regional structure, i.e. over relatively short distances, rather than over the entire state (Royaltey, Astrachan & Sokal, 1975; and unpublished autocorrelation analysis). Selander explains the heterogeneity within the barn in terms of the social behaviour of the mice. These form tribes with an estimated modal number of ten individuals comprising one dominant male, several females, and several subordinate males. Territorial defence by the male and continued inbreeding by the male with the females, some of which are his daughters, is likely to produce the heterogeneity. Intertribal migration is reportedly rare, involving only females, while males remain within their tribe and its territory.

Our analysis of these data involved autocorrelation of individual Es-3 genotypes using the index for nominal data discussed in Sokal & Oden (1978).
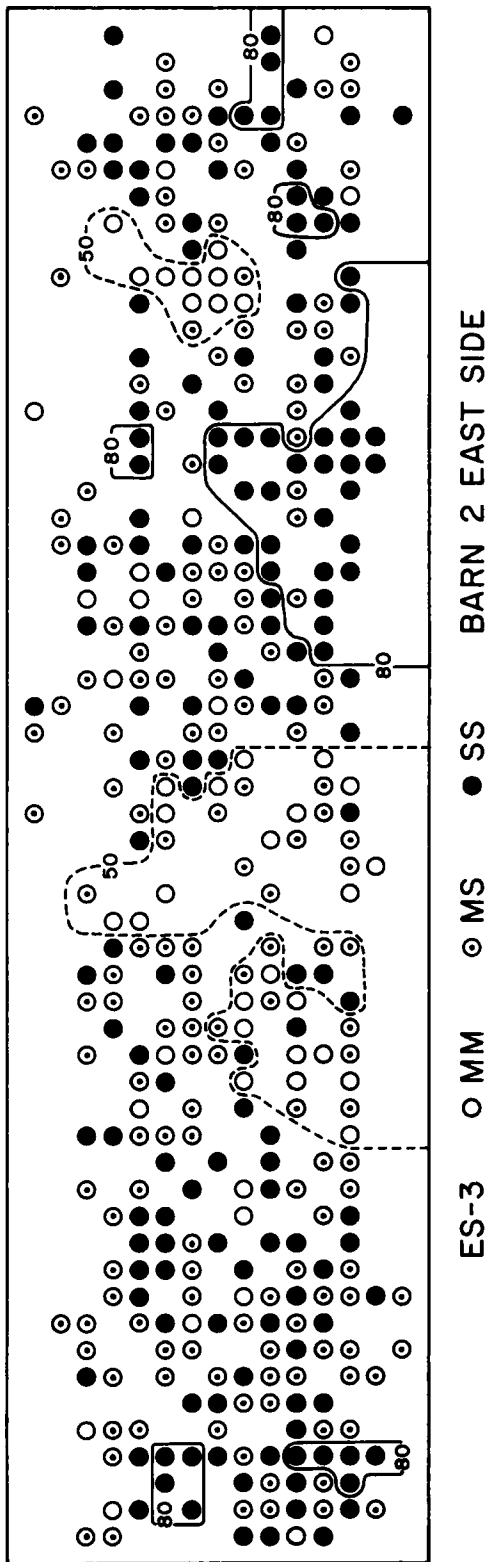
Figure 2. Spatial distribution of three genotypes of esterase-3 of house mice caught in a trap grid on the east side of barn 2 at Hildreth Farm, Austin, Texas. The contours are estimates of homogeneous frequencies entered by eye-fit by the original author (from Selander, 1970).

The points displayed in Fig. 2 were placed on a regular lattice and connected by rook's, bishop's or queen's moves. Note that the grid here lacked some values. The results are shown in Table 1. There is a significant excess of MM × MM neighbours and a significant deficiency of MM × SS neighbours under any of the three schemes for connecting grid intersections. The picture suggested by these statistics is of groups of MM individuals in homogeneous patches surrounded by heterozygotes MS rather than SS homozygotes. Various models of inbreeding could account for such patterns. There is no evidence that the heterozygotes are found more frequently near each other than random expectation would suggest. Note that the numbers of these genotypes are quite unequal and thus the significant clumping among MM homozygotes is based on a relatively small sample, while the more frequent SS homozygotes are not sufficiently above expectations to exhibit significant clumping. Yet on casual inspection of Fig. 2, the SS homozygotes (the black circles) appear clumped, suggesting that human perception of such patterns might be misled by the intensity of colour and frequency of types of points displayed in the area.

Table 1. *Observed and expected frequencies and significance tests of the deviations from expectation for pairs of neighbouring mice of three genotypes. (Original data from Selander, 1970)*

| Pairs | Moves | Frequencies | | S.E. | t_S |
		Observed	Expected		
MM × MM	Rook's	15	7.41	2.54	2.98**
	Bishop's	14	7.10	2.51	2.75**
	Queen's	29	14.53	3.62	4.00**
MS × MS	Rooks's	68	78.61	6.77	−1.56
	Bishop's	81	75.38	6.78	0.83
	Queen's	149	154.19	10.00	−0.52
SS × SS	Rook's	84	75.77	6.70	1.23
	Bishop's	83	72.66	6.70	1.54
	Queen's	167	148.62	9.91	1.86
MM × MS	Rook's	50	48.89	5.96	0.19
	Bishop's	36	46.89	5.90	−1.84
	Queen's	86	95.90	8.60	−1.15
MM × SS	Rook's	36	48.00	5.92	−2.03*
	Bishop's	29	46.00	5.86	−2.90**
	Queen's	65	94.15	8.54	−3.42***
MS × SS	Rook's	162	155.31	9.40	0.71
	Bishop's	154	148.93	9.26	0.55
	Queen's	316	304.62	13.33	0.85
Total of	Rook's	248	252.21	9.71	−0.43
different	Bishop's	219	241.85	9.53	−2.40*
pairs	Queen's	467	494.67	13.65	−2.03*

Notes. Moves refers to the type of connection between different cells in a grid laid out on the floor of a barn. Significant deviations are marked with asterisks (* = $0.01 < P < 0.05$, ** = $0.001 < P < 0.01$, *** = $P < 0.001$).

Human blood groups

The ABO blood group frequencies from the 26 counties of the Republic of Ireland reported in Mourant, Kopec & Bomoniewska-Solcjak (1958) provide an example for autocorrelation of continuous data, in this case gene frequencies.

Because the frequencies are reported for the same counties for 1947 and again for 1958 there is an opportunity to examine shifts in trends in these data and their possible implication. The graph of geographical contiguities upon which autocorrelations were computed featured an edge for each pair of counties sharing a common boundary. This net of geographical contiguities in Ireland has been employed in classical studies of spatial autocorrelations by Geary (1954), Cliff & Ord (1973), and Royaltey, Astrachan & Sokal (1975) for other variables. However, in these earlier studies the county of Dublin, included here, was excluded for demographic considerations. We chose to employ a binary connection matrix because of the difficulty of postulating reasonable weights in this case. To construct such weights, not only distance but also ease of transportation between contiguous counties would need to be considered, as well as the extent of their common boundaries.

Analysis of the gene frequencies yielded significant positive autocorrelations between contiguous counties for the i allele for 1947 and for the A and i alleles for 1958, but not for the B allele. The A and i alleles are present in high frequency and are probably nearly complementary to each other in their relationships. We therefore need to discuss only one of the two alleles, and we have chosen A allele. The B allele, quite rare in Ireland, does not provide much information. The geographic patterns of the gene frequencies of A shown in Fig. 3A and B are clearly nonrandom, being generally lower in the west and higher in the east of Ireland. The differences appear to be clinal and may reflect historical patterns of settlement and migration, specifically successive invasions by Anglo-Saxons and Anglo-Normans from Britain (higher frequency of A) pushing the older inhabitants westward (Hooper, 1947; Hackett & Dawson, 1958).

Changes during the time period under consideration are suggested by a slight decrease in the frequency of A in the east in 1958 compared with 1947. However, the differences in gene frequencies over time were not significant ($t = 1.177$ by sign test), nor were there any significant spatial autocorrelations among the differences, which are shown in Fig. 3C. Yet a comparison of the gene frequency at the two times suggest increased definition and compactness of the high frequency area in the east in 1958. This is reflected in the significant autocorrelation for 1958 ($I = 0.526$) compared with the nonsignificant auto-correlation for 1957 ($I = 0.148$) and is further substantiated by correlograms for both years (Fig. 3E). While the 1958 correlogram shows a clinal pattern (significant positive autocorrelations between adjacent counties, negative autocorrelations between counties separated by two or three intervening counties), the 1947 correlogram is not significant but shows the same profile.

In these data heterogeneous means are accompanied by a significant pattern at least for 1958. It is possible that the populations are dominated by selection, but diligent research by numerous workers has so far failed to uncover clear-cut evidence of selection at the ABO locus (Cavalli-Sforza & Bodmer, 1971). It is more likely that initial gene frequency differences between native stock and immigrants are being made into a cline by migration patterns among adjacent counties. The clinal pattern is substantiated by an analysis of Rh frequencies from the data set for 1958. Again there is significant ($P < 0.05$) positive autocorrelation between adjacent counties ($I = 0.194$) and negative autocorrelation between counties separated by two intervening counties ($I = -0.238$).

The differences in pattern over time may be artifacts of the sampling designs of the respective studies. The 1947 study (Hooper, 1947) was based on smaller samples (10,035 persons in all) and classified blood donors by county of residence. In the 1958 study (Hackett & Dawson, 1958) 55,696 persons were
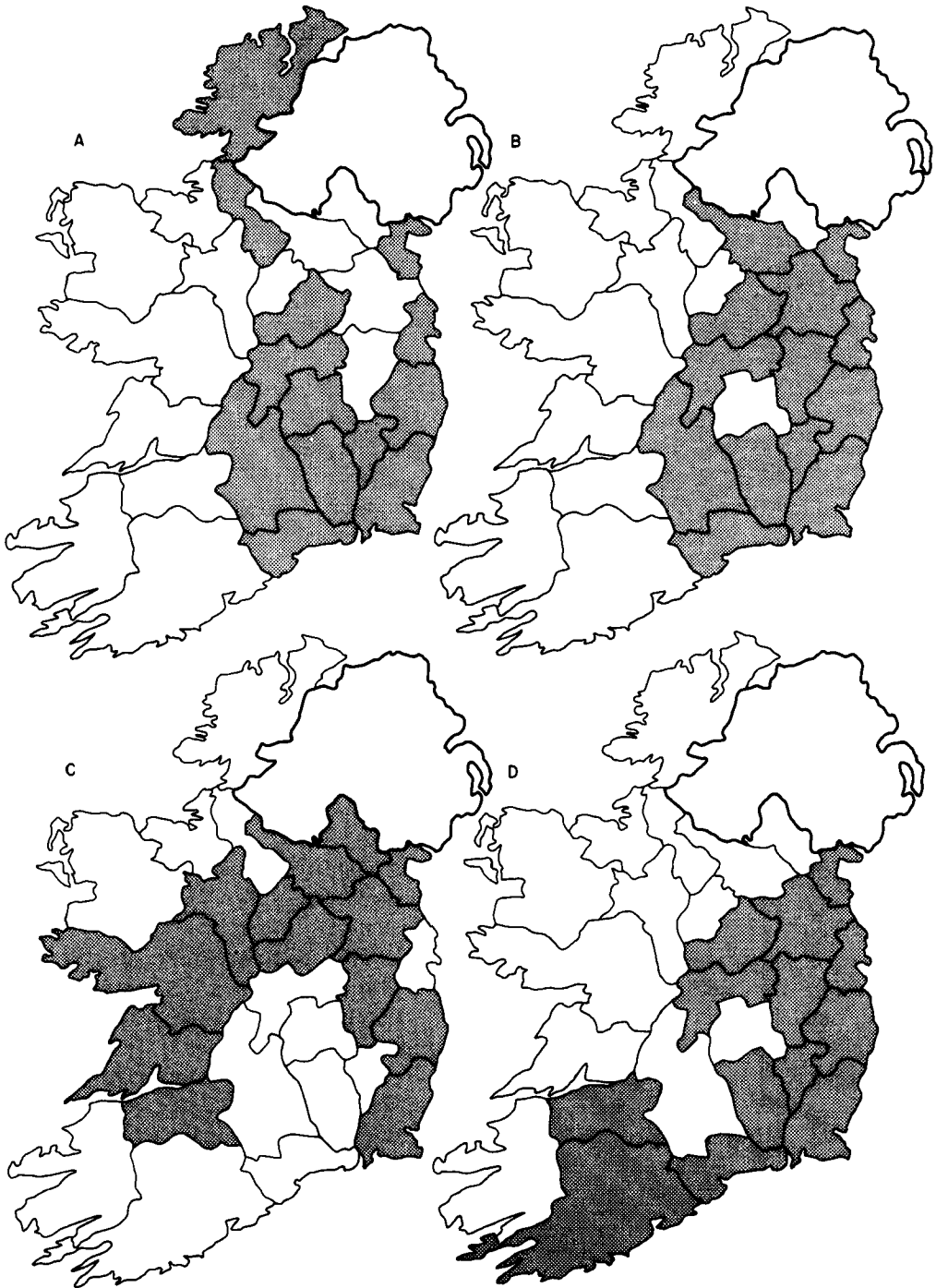
blood typed and donors were classified by county of birth. These facts alone may suffice to account for the sharper definition of the clinal patterns in 1958. If we consider the differences in pattern to be real and attempt a biological interpretation we must conclude first of all that in view of the short interval of time elapsed, this trend is most unlikely to have been caused by selection. This would leave migration as the most likely explanation. One possibility would be migration of persons from the west of Ireland toward the east in consonance with the general urbanization trend during these years in much of Europe. However, the pattern is apparently not that simple, since straightforward migration from west to east should have produced a western area with very small changes and an eastern area with large changes, as well as significant autocorrelation of the differences, none of which was found. The problem is complicated in these data by the large proportion of emigrants from the area. From the census data available to us (Anonymous, 1963) differential population change with greater losses in the west of the island is evident (see Fig. 3D). In fact, there is highly significant positive autocorrelation
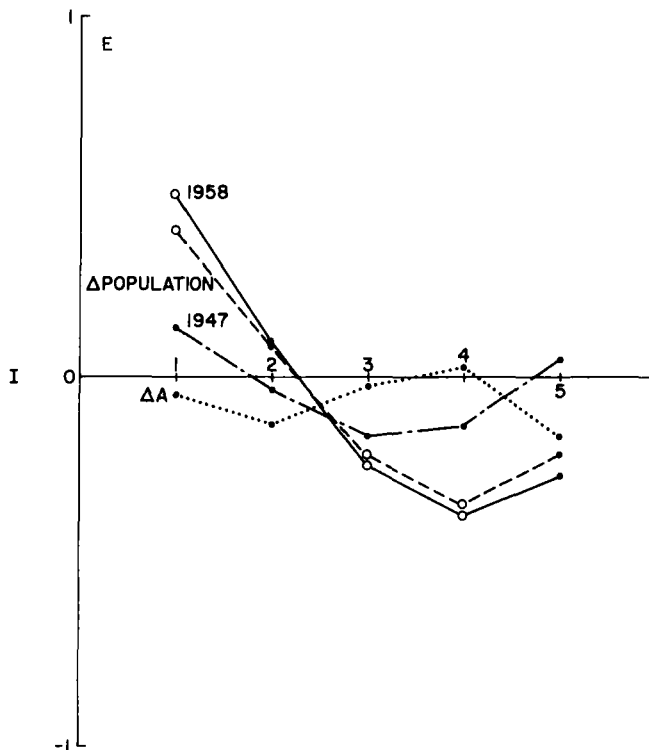

Figure 3. ABO blood group and demographic variables for the Republic of Ireland. A. Frequency of the A allele for Ireland in 1947 (based on Hooper, 1947). B. Frequency of the A allele for 1958 (based on Hackett & Dawson, 1958). C. Map of the differences between 1958 and 1947. D. Percentage change in population over the period 1946 to 1956 (data from Anonymous, 1963). All maps are recorded based on counties as a sampling unit. Counties where the score of the variable is above the median have been shaded. E. Correlograms based on the four variables mapped in Fig. 3A to D. Abscissa: distance expressed as minimum number of county boundaries that need to be traversed to move from one county to the other. Ordinate: Moran's autocorrelation coefficient I. Significant autocorrelation coefficients ($P < 0.05$) are indicated by open circles.

(I_R = 0.421; P < 0.001) among contiguous counties for per cent of population change over the period 1946-1956. But the census does not inform us how many persons leaving Connacht Province, Republican Ulster, and the western counties of Munster Province moved to the eastern tiers of counties and how many left Ireland. The possibility of differential migration rates among bearers of different genotypes cannot be ruled out. Without a more detailed analysis of the demographic history of the area studied, it is futile to try to speculate further about the causes of the observed phenomena, but this example should furnish an illustration of the type of reasoning and method of analysis which can be carried out in situations where variation patterns are compared at two time levels.

Liatris cylindracea

An example of data sampled from a regular grid or lattice is found in the work of Schaal (1974, 1975) on the perennial obligate outbreeder *Liatris cylindracea*. A sand prairie hillside was partitioned into a grid 18 m across and 33 m down the slope and marked off into 3 x 3 m quadrats, yielding 6 columns running downhill and 11 rows across the slope. The density of plants was very high with a maximum of 175 plants per m^2. From each quadrat 60 plants were chosen for allozyme analysis on 27 loci, 15 of which were polymorphic.

Schaal found considerable heterogeneity among quadrats but she found no patterns related to either the downhill gradient or the transects across the slope. The heterogeneity among the local populations was ascribed by her to reduced gene flow and the possibility of localized selection in the microhabitat.

In data arranged on a regular grid localities may be connected in ways reminiscent of chessboard moves, as discussed in our earlier paper (Sokal & Oden, 1978). Such connections provide one way to test autocorrelation in alternative directions. The values for each of the 15 allozyme frequencies recorded for each quadrat were connected by rook's moves, bishop's moves, and queen's moves in turn. The results are shown in Table 2, and four of the patterns are illustrated in Fig. 4. The values of I are relatively low, ranging from −0.30 to +0.30. Only five of the 45 autocorrelation coefficients were significant, one for rook's moves, three for bishop's moves, and one for queen's moves. There is some evidence that rook's moves and queen's moves are related since the two patterns significant for one type of move are also significant for the other. On the whole, however, there seems little correspondence between the different types of moves, and the total number of significant correlation coefficients is only twice as high as the number expected under 5% type I error.

Correlogram analysis was carried out for four of the patterns, chosen to illustrate one lacking autocorrelation by single moves (GOT; Fig. 4A), one with significant autocorrelation by bishop's moves (Pep; Fig. 4B), and two with significance by rook's and queen's moves (PGI and Het; Fig. 4C and D, the latter discussed below). The correlograms, based on all conceivable Euclidean distances between the quadrats (Fig. 4E), corroborate the findings based on chess board moves. Patterns correlated by rook's and queen's moves are positively autocorrelated at distance class 1 (length of one quadrat) and that correlated by bishop's moves is positively correlated at distance class 2 (which includes the length of the diagonal of one quadrat). Negative autocorrelations

greater distances apart reflect transitions from areas of high gene frequency to areas of low frequency and *vice versa*. The correlogram for the unpatterned GOT allele was entirely nonsignificant.

Since there are differences among patterns as well as correlograms in the
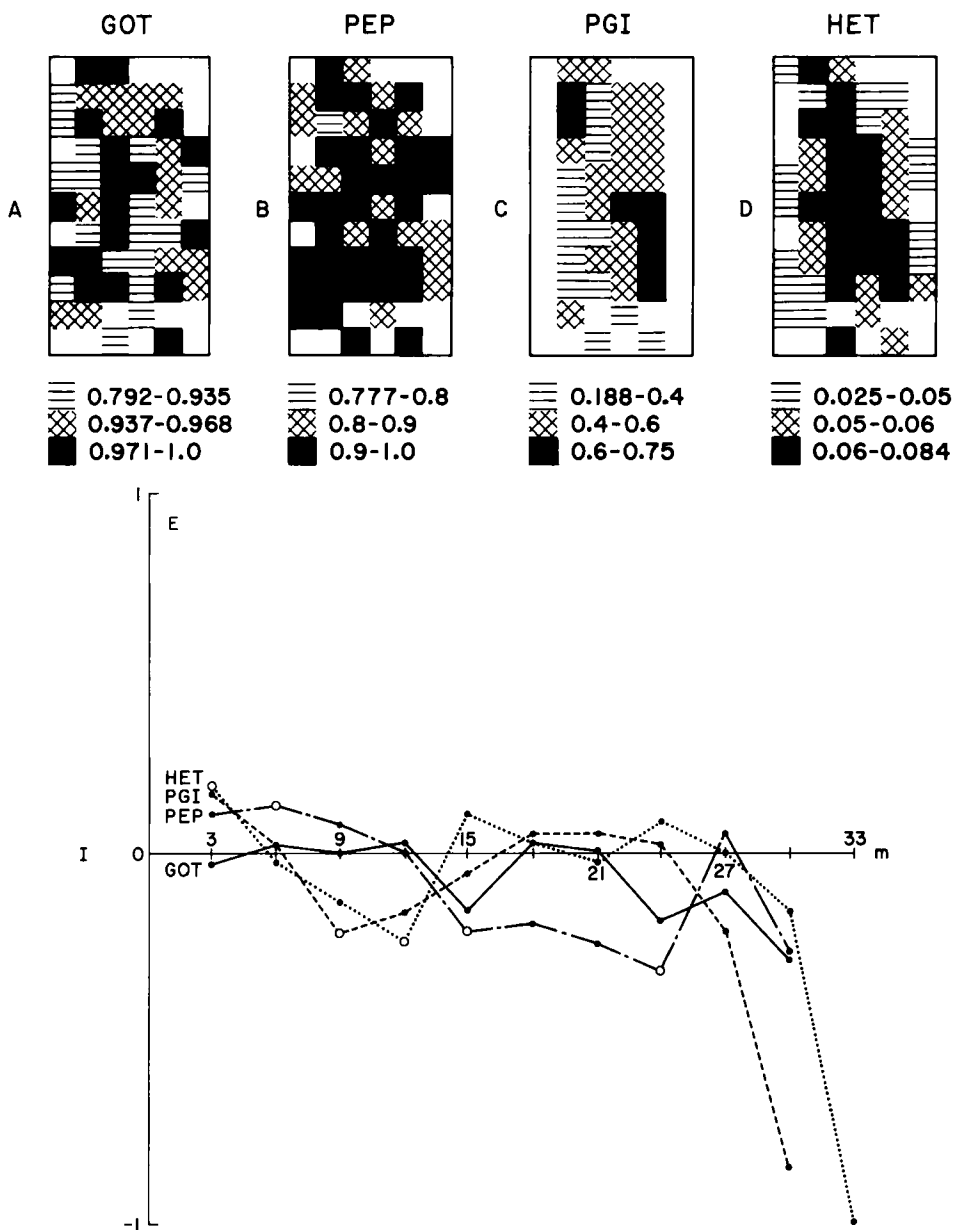

Figure 4. Allozyme frequencies and per cent heterozygosity in a hillside population of *Liatris cylindracea* laid out as a regular grid. The shading indicates magnitude of the variable as shown beneath panels. A. Glutamate-oxaloacetate transaminase. B. Peptidase. C. Phosphohexose isomerase. D. Percentage heterozygosity. E. Correlogram of the four patterns in Fig. 10A to D. Abscissa: Euclidean distance in metres. Ordinate: Moran's autocorrelation coefficient *I*. Significant autocorrelation coefficients ($P < 0.05$) are indicated by open circles. Data based on Schaal (1974).

Table 2. Moran's coefficient I for 15 gene frequencies and average heterozygosity in populations of *Liatris cylindracea* studied on a regular grid. (Data from Schaal, 1974)

Variables	Types of connections		
	Rook's move	Bishop's move	Queen's move
GOT	−0.01	−0.01	−0.01
MDH	0.09	0.05	0.07
ADH	0.13	−0.15	−0.01
AP$_1$	0.06	−0.02	0.02
AP$_2$	−0.11	−0.10	−0.10
Est$_1$	0.04	0.01	0.02
Est$_2$	0.18	−0.30	−0.05
Pep	−0.02	0.24*	0.11
G-6-PDH	0.00	−0.24*	−0.11
6-PGDH	0.17	−0.13	0.03
Per$^-$	0.18	−0.14	0.03
Per$^+$	0.19	−0.13	0.04
PGI	0.30*	0.04	0.18*
AlkP	0.20	−0.06	0.08
Est$_3$	0.01	0.26*	0.13
Het	0.30**	0.12	0.21**

Notes. Standard abbreviations are used for the 15 allozyme frequencies. Details about these can be obtained from the original publication. Het stands for average heterozygosity. Significant values of I are indicated by asterisks ($* = 0.01 < P < 0.05$, $** = P < 0.01$). Correlogram analysis of four variables based on Euclidean distances yielded no significant autocorrelations for GOT, and significant autocorrelations with the following signs and distance classes: +2, −5, −8 for Pep; +1, −3 for PGI; and +1, −4 for Het. A distance class comprises the length of one quadrat.

gene frequencies subjected to analysis and most loci showed no short-distance autocorrelation, the overall conclusion for this data set is that for most loci there is very little substructuring of the population. Our findings support Schaal's hypothesis of limited gene flow, and suggest that there is no spatial autocorrelation of the microhabitat. If these plants are being actively selected in response to micropatches in the habitat as Schaal (1975) and Schaal & Levin (1976) suggest, then the micropatches themselves are not autocorrelated. There are differences in correlograms among those few loci that show significant autocorrelations, indicating that differential selection among loci predominates in these plants.

One last test on these data is of interest. The values of the average heterozygosity per individual for each quadrat were autocorrelated. This heterozygosity is the percentage of an individual's 27 loci that are heterozygous. This variable is significantly autocorrelated at $P < 0.01$ for rook's and queen's moves (Table 2). The level of heterozygosity has been graphed in Fig. 4D. There appears to be high average heterozygosity in a fairly homogeneous patch at the centre of the grid with lower heterozygosity near the periphery. Schaal, and Schaal and Levin claim an increase of heterozygosity with age; looking at table 21 in Schaal (1974) one finds that the plants at the margins of the study plot are indeed younger, with older plants at the centre.

Tree distribution

As our last example, we feature an analysis from plant ecology. The data are nominal with 17 different classes and an irregular grid. In a recent study of the Red Ash (*Fraxinus pennsylvanica*) and the White Ash (*F. americana*) Taylor (1972) mapped all trees over 10 cm in diameter at a study site measuring 210 x 310 m. Since the amount of information in her map was excessive for our purposes, we elected to perform separate analyses for two regions roughly equal in population size but different in area. The low density region containing 236 trees was roughly one and a half times the area of the high density region containing 269 trees. For the connecting graph all trees closer than 1.5 m were linked. The 17 species of trees recorded in the study are identified in the notes below Table 3. The data were analyzed as nominal (join counts) with 17 classes, under the hypothesis of sampling with replacement. The results are displayed in Table 3. In both regions the total number of joins between unlike species was highly significantly below expectations indicating clumping of like species in the area. There were appreciable differences in autocorrelation between the two density areas (see Table 3). The following neighbouring pairs of species occur significantly more frequently than expected in both study regions: 2-2, 3-3, 4-4, 5-5, and 2-12; while the following pairs of neighbouring trees occur significantly less often than expected in both regions: 2-1, 2-3, 2-4, 2-5, 6-3, and 6-4. All of the latter pairs (negative autocorrelation) involve trees that differed in their tolerance to conditions of high moisture. Four of the positively autocorrelated pairs involve the same species; the single heterotypic linkage is between two moisture-loving species, the Red Ash and the Swamp Oak. Since moisture conditions varied in a patchy way on the study site, the autocorrelations confirm the differential responses of the species of trees to these conditions. It is interesting that no homotypic neighbours occur significantly below expectations, suggesting that allelopathy-like mechanisms are not effective over the inter-tree distances in these data.

Two further points can be made. In such a large data set with many individual trees and numerous species, relationships between individuals of the less numerous species might be swamped. By removing the strongly auto-correlated numerous species from the data set, which in the connection matrix chosen for these data will not change the linkages between the residual individuals (but would do so were the data Gabriel-connected), we can recompute the expected number of joins for the remaining trees. In most instances autocorrelation or the absence thereof remained as it was in the full study, but in three cases autocorrelation emerged or disappeared in the residual data set.

An underlying assumption of the autocorrelation statistic is that the process underlying the observed distribution is stationary; that is, that its mean and the autocorrelation coefficients for each distance class are constant for all data points. We expected that this assumption would be violated in these ecological data and that the autocorrelations in the high density area would differ from those in the low density area. This point can be examined by comparing the two matrices of Table 3. Altogether 38 significant departures from expected join counts are shown for the two regions when columns and rows not present in both matrices are removed. Of these 38 instances of

R. R. SOKAL AND N. L. ODEN

Table 3. Significant departures from expectations in join counts for species of trees at two densities (Data from Taylor, 1972)

Species	Low density												
	1	2	3	4	5	6	7	8	9	10	11	12	15
1													
2	−	+											
3		−	+										
4		−		+									
5		−			+								
6		−		−									
7							+						
8													
9			−	−	−				+				
10		−								+			
11								+					
12	−	+	−	−			+					+	
15	−												

Species	High density										
	1	2	3	4	5	6	7	9	12	14	17
1											
2	−	+									
3	−	−	+								
4		−		+							
5		−		+	+						
6	−		−	−		+					
7											
9											
12		+									
14											
17											

Notes. Plus and minus signs indicate significant (at $P < 0.05$) positive or negative departures from expectations. The code numbers refer to the following species:

1	*Fraxinus americana*	White ash
2	*F. pennsylvanica*	Red ash
3	*Carya ovata*	Shagbark hickory
4	*C. ovalis*	Sweet pignut hickory
5	*C. glabra*	Pignut hickory
6	*Salix nigra*	Black willow
7	*S. fragilis*	Crack cherry
8	*Prunus serotina*	Black cherry
9	*Ulmus americana*	American elm
10	*Acer rubrum*	Red maple
11	*Quercus alba*	White oak
12	*Q. bicolor*	Swamp oak
13	*Q. rubra*	Red oak
14	*Populus deltoides*	Common cottonwood
15	*Crataegus sp.*	Hawthorns
16	*Pyrus malus*	Domestic apple
17	*Ostrya virginiana*	Hornbeam

autocorrelations, 22 (11 in each matrix) are common to both matrices and 16 are present in only one. This indicates that the interrelationships found at the low density are appreciably different from that found at the high density. Most of the differences affect species 7, 9, and 12, which apparently occur in patches in the low density area but not in the high density area.

SUMMARY

The technique of spatial autocorrelation analysis summarized by Cliff & Ord (1973) tests whether the observed value of a variable at one locality is significantly dependent on values of the variable at neighbouring localities. In an earlier paper (Sokal & Oden, 1978) we extended the method to include the computation of spatial correlograms which show the autocorrelation coefficient as a function of distance between pairs of localities. Correlograms summarize the patterns of geographic variation exhibited by the response surface of any given variable. While identical variation patterns will lead to identical correlograms, different patterns may or may not yield different correlograms. Similarity in the correlograms of different variation patterns suggests similarity in the generating mechanism of the pattern. The study examines some biological implications of various combinations of significant or nonsignificant heterogeneity of character means with significant or nonsignificant variation patterns for these means. By examining and analyzing variation patterns of several characters or gene frequencies for one population, or of several populations in different places or at different times, some conclusions about the nature of the populational processes generating the observed patterns can be reached.

These methods of autocorrelation analysis are applied to four biological situations differing in the nature of the data (interval or nominal), in the type of grid connecting the localities (regular or irregular), and the field of application (evolution or ecology). The examples comprise genotypes of individual mice in a barn, blood group frequencies in humans, allozyme frequencies in the perennial *Liatris cylindracea,* and the distribution of species of trees in a lowland area. We conclude that the mice are nonrandomly distributed because of the genetic consequences of social structure in the mouse colony; that the human blood groups show clinal structure probably as the result of successive invasions of Ireland from the east; that there is little substructuring and limited gene flow in the population of *Liatris*; and that the distribution of trees in the lowland area is governed by environmental heterogeneity within the area.

9

REFERENCES

ANONYMOUS, 1963. *Census of Population of Ireland 1961:* 152 pp. Dublin: Stationery Office.
CAVALLI-SFORZA, L. L. & BODMER, W. F., 1971. *The Genetics of Human Populations:* 965 pp. San Francisco:W. H. Freeman.
CLIFF, A. D. & ORD, J. K., 1973. *Spatial Autocorrelation:* 175 pp. London: Pion.
GEARY, R. C., 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician, 5:* 115-145.
HACKETT, W. E. R. & DAWSON, G. W. P., 1958. The distribution of the ABO and simple "Rhesus" (D) blood groups in the Republic of Ireland from a sample of 1 in 37 of the adult population. *Irish Journal of Medical Science, 6, Ser. No. 387:* 99-109.
HOOPER, I., 1947. Blood group distributions in Ireland. *Irish Journal of Medical Science, 6, Ser. No. 259:* 471-479.
LEWONTIN, R. C., 1974. *The Genetic Basis of Evolutionary Change:* 346 pp. New York: Columbia University Press.
MOURANT, A. E., KOPEC, A. C. & BOMONIEWSKA-SOLCZAK, K., 1958. *The ABO Blood Group:* 276 pp. Springfield, Illinois: C. E. Thomas.
RICKLEFS, R. E., 1973. *Ecology:* 861 pp. Newton, Massachusetts: Chiron Press.
ROYALTEY, H. H., ASTRACHAN, E. & SOKAL, R. R., 1975. Tests for patterns in geographic variation. *Geographical Analysis, 7:* 369-395.
SCHAAL, B. A., 1974. Population structure and balancing selection in *Liatris cylindracea.* Yale University Ph.D. dissertation.
SCHAAL, B. A., 1975. Population structure and local differentiation in *Liatris cylindracea. American Naturalist, 109:* 491-510.
SCHAAL, B. A. & LEVIN, D. A., 1976. The demographic genetics of *Liatris cylindracea* Michx (Compositae). *American Naturalist, 110:* 191-206.
SELANDER, R. K., 1970. Behavior and genetic variation in natural populations. *American Zoologist, 10:* 53-66.
SELANDER, R. K., YANG, S. Y. & HUNT, W. G., 1969. Polymorphism in esterases and hemoglobin in wild populations of the house mouse (*Mus musculus*). *Studies 5: In Genetics. University of Texas Publication No. 6918:* 271-338.

SOKAL, R. R., 1978. Population differentiation: Something new or more of the same? In P. Brussard & O. Solbrig (Eds), *Genetics and Ecology*. New York: Academic Press.

SOKAL, R. R. & ODEN, N. L., 1978. Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society, 10:* 199-228.

TAYLOR, S. M. O., 1972. Ecological and genetic isolation of *Fraxinus americana* and *Fraxinus pennsylvanica*. University of Michigan Ph.D. Dissertation.