Spatial Structure of Two-Locus Genotypes Under Isolation by Distance

Bryan K. Epperson

Department of Forestry, Michigan State University, East Lansing, Michigan 48824

Manuscript received April 13, 1994
Accepted for publication January 19, 1995

ABSTRACT

Extensive Monte Carlo simulations are conducted of spatial distributions of two-locus genotypes in large, continuous populations under isolation by distance models. The results show that substantial patches of double homozygotes are present in the spatial structures, even when loci are unlinked. The stochastic spread of identical two-locus genotypes largely outpowers the tendency for recombination to decouple patterns for separate loci. A spatial patch is a large area containing mostly one double homozygous genotype in a highly contiguous constellation. This patch structure is reflected in high positive spatial autocorrelations and large excesses of pairs, or joins, of identical double homozygotes at short-to-intermediate distances of spatial separation. Although spatial patches of double homozygotes are the dominant spatial feature, and the major contributors to overall high levels of autocorrelations among two-locus genotypes, other substantial features include areas of concentrations of identical genotypes heterozygous at only one locus. One implication of the patch structure is the presence of high levels of linkage disequilibrium, caused by isolation by distance even for unlinked loci, at some spatial scales; yet the disequilibrium in the large total populations is near 0. Thus linkage disequilibrium produced by isolation by distance is highly dependent on spatial scale. Another implication is that high degrees of spatial structuring and autocorrelations are produced for genetic variation controlling quantitative traits, at least when the number of loci is relatively small, under a wide range of situations, even if the trait is selectively neutral. The significance of the results to field studies is also examined.

IN classical theoretical models of genetic isolation by distance within large continuous populations, the effects of limited dispersal and mating by proximity are expressed in terms of inbreeding and consanguinity coefficients (WRIGHT 1943; MALÉCOT 1948). More recently, simulation studies have been conducted on complex models that generate explicit spatial distributions of single-locus *genotypes*, and these have incorporated important stochastic sampling events that occur as individual genotypes disperse (*i.e.*, "stochastic migration") within a population. These studies have shown that large spatial patches (see Figure 1), or areas of concentrations of one homozygous genotype, are the dominant spatial features of the distributions of genotypes for single loci (RHOLF and SCHNELL 1971; TURNER *et al.* 1982; SOKAL and WARTENBERG 1983; SOKAL *et al.* 1989; EPPERSON 1990a). However, the spatial distributions of multilocus genotypes have not been characterized.

Views on the spatial structure of multilocus spatial distributions range widely. On the one hand, it has been suggested that the spatial patterns for two unlinked loci might be essentially completely decoupled by recombination (SOKAL and WARTENBERG 1983; EPPERSON 1990b). Large spatial patches of single-locus homozygotes would exist, but not large patches of multilocus homozygotes. One apparent implication of this view is that generally there should be little spatial autocorrelation for quantitative traits that are controlled by more than a few loci (EPPERSON 1993). In contrast, other results for the very different models on the spatial correlations among discrete populations support the view that the correlations should be similar regardless of the number of loci (ROGERS and HARPENDING 1983; LANDE 1991; NAGYLAKI 1994).

The spatial distribution of multilocus genotypes is of immediate theoretical and experimental interest, in addition to its implications for quantitative traits. For example, linkage disequilibrium can arise from overlapping spatial structures for two loci (PROUT 1973). Such spatial cross-correlations are involved in the formation of novel combinations of genes and thus may affect the evolution of species (WRIGHT 1943). Moreover, theoretical and experimental results on spatial structure provide hypothesis testing frameworks, that are particularly powerful when applied to multiple locus data (LEWONTIN and KRAKAUER 1973; SOKAL 1979, 1988; SOKAL and JACQUEZ 1991; SOKAL and ODEN 1991; EPPERSON 1990b, 1993).

Most simulation studies have used Moran's I-statistics as summary measures of autocorrelation among the allele frequencies in local groups of individuals. Recently, extensive analyses of join-count statistics have been conducted (B. K. EPPERSON, unpublished data). Both sets of spatial autocorrelation statistics measure several essential features of patchy spatial structure, in both theoretical models (SOKAL and WARTENBERG 1983; SOKAL

Author e-mail: bryan@alpha.for.msu.edu

and JACQUEZ 1991) and in real populations (SOKAL and ODEN 1978a; EPPERSON and CLEGG 1986; SOKAL 1988).

Isolation by distance with stochastic migration results in autocorrelations in terms of pairs of diploid genotypes, which includes some information that is used in join-count statistics but not in Moran's I-statistics. In the stochastic migration case, the join-count statistics use additional information that is in a sense similar to that used in GILLOIS' set of identity measures in contrast to a single coefficient of consanguinity (GILLOIS 1966). This information is intimately connected to other important parameters, such as diploid mating-type frequencies and levels of biparental inbreeding. Join-count statistics have also provided important experimental tools and, in a multilocus context, they may meet the repeated calls for the development of efficient measures for analyzing spatial multilocus data in real populations. One goal is to describe the overall degree of multilocus genotypic isolation by distance. One promising summary join-count statistic is based on the total number of joins between nonidentical multilocus genotypes (EPPERSON and ALLARD 1989; EPPERSON 1993).

The aim of the present study is to quantify the spatial structure of multilocus genotypes and quantitative traits, using join-count and I-statistics, in theoretical models of limited dispersal and mating by proximity. In addition, we address several aspects of how the values of these statistics are changed by dispersal, recombination, and allele frequency parameters.

Join-count statistics are test statistics that compare the numbers of each type of pair of genotypes, separated by a given distance, with the expected number under the null hypothesis that the distribution is random (SOKAL and ODEN 1978a; CLIFF and ORD 1981). The computations involve counting all joins or *pairs* of genotypes, which for a population or sample of size n are on the order of n^2. This presents computational difficulties when n is very large, and in the present study of large populations (n = 10,000), this difficulty is faced through the use of supercomputers.

Join-count statistics have been used to efficiently characterize spatial structure and isolation by distance in samples of much smaller sizes (*e.g.*, n = 50–400) in populations of several species (SOKAL and ODEN 1978b; EPPERSON and CLEGG 1986; EPPERSON and ALLARD 1989; SCHOEN and LATTA 1989; WAGNER *et al.* 1991; reviewed in EPPERSON 1993), and to detect selection among patterns for multiple loci (EPPERSON and CLEGG 1986). They have been applied to multilocus data of populations of lodgepole pine (EPPERSON and ALLARD 1989), and Moran's I statistics were applied to quantitative traits in *Impatiens capensis* (ARGYRES and SCHMITT 1991).

MATERIALS AND METHODS

Simulations: Each simulated population consisted of 10,000 individuals with diploid two-locus genotypes, located on a 100 × 100 square lattice. Initially, each simulation consisted of individuals with randomly assigned two-locus genotypes (for two diallelic loci, A/a and B/b), with probabilities according to Hardy-Weinberg equilibrium, 0 linkage disequilibrium, and allele frequencies of the a and b alleles, q_a and q_b, respectively, set variously at 0.5 or 0.2. Generations of mating and other features of life cycles were simulated using standard Monte Carlo methods, and details of the simulation Fortran source code program, including how matings near the boundaries were "reflected," were described previously (EPPERSON 1990a). The sets of simulations included three different levels of dispersal, and either members of only one sex dispersed (*e.g.*, male or pollen dispersal, no female or seed migration) or both sexes dispersed. Each dispersing parent was stochastically chosen (with replacement) from either the nearest 8, 24, or 48 locations with uniform probabilities, $^1/_9$th, $^1/_{25}$th, or $^1/_{49}$th, respectively (N_m or N_f = 9, 25, or 49 respectively), and there was additionally chances of no dispersal with probabilities $^1/_9$th, $^1/_{25}$th, or $^1/_{49}$th, respectively, which also equals the rate of selfing (EPPERSON 1990a). Next, gametes were stochastically sampled from the two chosen parents according to equal segregation rules and the recombination rate, r. Then the two gametes were combined to form the offspring which remains at the same location for the next generation. Sets of five replicate simulations were conducted for each combination of r, q_a,q_b, and N_m,N_f (Table 1). For each simulation, discrete (nonoverlapping) generations of the life cycle were repeated for 200 generations (because previous results show that quasistationarity is established within this time period; *e.g.*, SOKAL and WARTENBERG 1983), and the spatial distribution of the 10 different two-locus genotypes was output at every 10 generation increments, using the codes in Table 2.

TABLE 1

Parameters for the various simulation models

Set	r	q	N_m	N_f	N_e
1	0.5	0.5	9	—	4.2
1r01	0.01	0.5	9	—	4.2
2	0.5	0.5	25	—	12.6
2r01	0.01	0.5	25	—	12.6
3	0.5	0.5	49	—	25.2
3r01	0.01	0.5	49	—	25.2
4	0.5	0.5	49	49	50.1
4r01	0.01	0.5	49	49	50.1
1q2	0.5	0.2	9	—	4.2
1r01q2	0.01	0.2	9	—	4.2
4q2	0.5	0.2	49	49	50.1
4r01q2	0.01	0.2	49	49	50.1

Parameters of different simulation models include recombination rate, r, frequency of a and b genes, q, and dispersal in terms of the number of nearest of neighbors (including self) from which male and female parents are randomly chosen, N_m and N_f, and Wright's neighborhood size, N_e.

Statistical characterization: To characterize the spatial distributions of genotypes during the quasistationary phase (SOKAL and WARTENBERG 1983), the join-count statistics were computed for each simulation at generation 200. For each surface, a vectorized Fortran program (EPPERSON 1990b), run on a CRAY C90 at the San Diego Supercomputing Center, was used for classifying the 10,000 × 9,999/2 or *ca.* 50 million pairs of individual genotypes. Pairs or "joins" were classified according to the two two-locus genotypes and the distances

TABLE 2

Two-locus genotypes, their codes and converted genotypic values for quantitative traits

Code	Genotype	Genotypic value
1	*AABB*	0
2	*AaBB*	1
3	*aaBB*	2
4	*AABb*	1
5	*AaBb*[a]	2
6	*AaBb*[b]	2
7	*aaBb*	3
8	*AAbb*	2
9	*Aabb*	3
10	*aabb*	4

[a] In coupling phase.
[b] In repulsion phase.

classes, D, in multiples of lattice units ($D = 1-138$). With 10 different two-locus genotypes, there are 55 different types of joins. The number of joins between genotypes i and j for distance class D, $n_{ij}(D)$ minus the expected number under the null hypothesis, H_0, of randomly sampling pairs without replacement (from the sample population), divided by the SD under H_0 forms a test statistic, $\mathrm{SND}_{ij}(D)$, which has a closely approximate standard normal distribution under H_0 for large samples (SOKAL and ODEN 1978a; CLIFF and ORD 1981). A negative value less than -1.96 indicates a statistically significant deficit of a type of join at a distance, and positive values indicate excesses. Effects of events near the boundaries are negligible because the populations are very large (SOKAL and WARTENBERG 1983).

Among the various types of joins, there are 12 "equivalence" categories, which are based on symmetries of genotypes (Table 3). For examples, equivalence category 1 includes all four types of joins between double homozygotes (*i.e.*, joins *AABB* × *AABB*, *AAbb* × *AAbb*, *aaBB* × *aaBB*, and *aabb* × *aabb*), and category 6 includes the joins *AABB* × *aabb* and *AAbb* × *aaBB*. We would expect all members of an equivalence category to respond on average equally to changes in process parameters (or in category 4 approximately equally), although stochasticity will cause some differences in a particular simulation run. Thus, a concise way to characterize the structures under various models, is to present the averages of SND's for each equivalence category, for each set. The most striking statistics are those for joins between pairs of identical genotypes. Another important statistic is based on the total number of joins between unlike genotypes, and a SND test statistic for H_0 was formed, again for each distance class. The set of SND statistics for a type of join for various distance classes is termed a SND-correlogram.

Analyses of the spatial autocorrelations of genotypic values of a quantitative trait that might be controlled by the two-locus genotypes, were also conducted. First, the individual genotypes were converted to genotypic values, by adding 0, one, or two (if an individual had genotype *AA*, *Aa*, *aa* respectively at locus *A/a*) to 0, one, or two (if an individual had genotype *BB*, *Bb*, *bb*, respectively, at locus *B/b*). This transformation (Table 2) represents additive effects within and between loci, with equal weighting of loci. Next, the average genotypic value was calculated in 400 contiguous quadrats, each of size 5 × 5 (25 individuals). This is the same quadrat system used for autocorrelations of gene frequencies in previous simulation studies (SOKAL and WARTENBERG 1983; SOKAL *et al.* 1989; EPPERSON 1990a). Then, for every 10-generation increment, these values were used to calculate unweighted Moran's I-statistics (SOKAL and ODEN 1978a), using distance classes in multiples of quadrat lengths (one length equals five times the distance between nearest neighbor individuals, *i.e.*, five lattice units), and I-correlograms were formed.

TABLE 3

Types of joins for each equivalence class (see text)

Equivalance class	Joins between
1	Like (identical) double homozygotes (4)
2	Like (identical) single heterozygotes (4)
3	Like (identical) double heterozygotes (2)
4	Double homozygote and double heterozygote (8)
5	Single heterozygote and double heterozygote (8)
6	Double homozygote and opposite double homozygote (2)
7	Double homozygote and non-opposite, but unlike double homozygote (4)
8	Unlike double heterozygotes (1)
9	Double homozygote and single heterozygote differing by only one gene (8)
10	Double homozygote and single heterozygote with opposite homozygous genotype at one locus (8)
11	Single heterozygote and single heterozygote at same locus, but opposite homozygote at other locus (2)
12	Other unlike single locus heterozygotes (4)

Number of each type in parentheses.

Exact tests for statistical differences among correlograms are not available (*e.g.*, SOKAL and WARTENBERG 1983). This paper is concerned with the effects of process parameters on the populational spatial autocorrelations. Stochastic variation (SLATKIN and ARTER 1991) in the SND- and I-statistics is low throughout the range of studied processes, for the (more informative) short to intermediate distance classes (*e.g.*, Figure 6). Statistical significance of the contrasts described in this paper generally would certainly obtain. However, statistical significance is largely beside the issue, because the simulated populations are much larger than samples collected from real populations. Issues of statistical significance for purposes of making empirical inferences will be more appropriately addressed in further studies where substantial statistical variation is added through sampling.

RESULTS

Join-count statistics for short distances: The join-count statistics for very short distances fall into categories, which reflect the process. Within categories the small observed variation is stochastic, because a priori each member type of join is expected to depend on the process equally. The categories are illustrated with an example for a simulation with equal allele frequencies (*ca.* 0.5), low dispersal, and free recombination, in Table 4. The greatest values occur for joins between

TABLE 4

Example of values of join-count SND statistics for each equivalence class

Equivalence class	Values of SNDs
1	49.62, 53.72, 54.39, 48.88, 51.65
2	28.04, 25.00, 24.79, 27.66, 26.37
3	4.40, 8.44, 6.42
4	−0.78, −7.22, −7.07, −1.03, −2.78, −5.37, −3.51, −6.39, −4.27
5	2.85, 2.42, 3.64, 1.10, 2.06, 2.47, 2.84, 1.27, 2.33
6	−27.31, −35.10, −31.20
7	−12.57, −15.36, −16.43, −13.35, −14.43
8	4.21
9	20.42, 15.43, 20.34, 20.97, 21.32, 21.85, 18.28, 16.42, 19.38
10	−20.47, −22.20, −23.13, −25.72, −24.41, −25.65, −23.13, −20.94, −23.21
11	−19.55, −20.74, −20.15
12	−6.16, −6.31, −2.97, −7.78, −5.80

Listed are the SND values for joins of each type in each equivalence class, for one simulation of model 1, for distance class 1.0 at generation 200. Underlined value is the average for that equivalence class.

like (identical) double homozygotes. Complementary large deficits of joins between opposite double homozygotes are observed. Deficits in joins between nonopposite but unlike (nonidentical) double homozygotes are more moderate. There are moderate excesses of joins between like (identical) single heterozygotes (single homozygotes). In general, SNDs for joins in all categories (3–5 and 8) involving the double heterozygotes have consistently small absolute values. Similarly, values for joins between unlike single heterozygotes were close to 0. Somewhat larger excesses are indicated for joins between double homozygotes and single heterozygotes differing at only one gene (9), and similar size deficits when the single heterozygotes have the opposite homozygous genotype (10 and 11). The results indicate a connection between the single-locus and two-locus distributions, but the dominant spatial autocorrelations are, perhaps surprisingly, those positive autocorrelations between double homozygotes. We note that the rankings among categories with specific types of unlike joins might differ when allele frequencies are changed. Below we will focus on joins between like types.

Patch structure: Remarkably, after ~50 generations, the correlograms for joins between identical double homozygotes indicate the presence of patches of double homozygotes (analogous to patches of single locus homozygotes—EPPERSON 1990a; B. K. EPPERSON, unpublished data), even when there is free recombination. The statistics are very large at short distances, and then decrease, and finally become negative as distances increase to *ca.* 15–30 (Figure 2). It is evident from inspection of the surfaces that large contiguous areas of double homozygotes develop within ~50 generations and then persist (*e.g.*, Figure 1).

SNDs for joins between like (identical) single heterozygotes exhibit a similar pattern, only the values are not as extreme (Figure 3). This indicates that single-locus heterozygotes occur in concentrations. However, these concentrations differ from the patches of double homozygotes, within which there are very high levels of contiguity. Matings and segregation reduce the contiguity in the former. The concentrations are associated with patches of double homozygotes (such that both genotypes have the same single-locus homozygous genotype).

In contrast, the values of SNDs for joins between like double heterozygotes are quite small when there is free recombination (Figure 4), and apparently any patchiness is at most very fine grained. However, weak levels of concentrations apparently extend over large areas, because the *X*-intercepts are large. When recombination is highly restricted, the values are greatly increased, resulting in concentrations similar to those for single heterozygotes.

The excesses between like double homozygotes are larger than those for any other category, indicating that these constitute a major contribution to the large deficits for the total number of unlike joins (Figure 5). Patches of double homozygotes dominate the spatial structure. Restricting recombination only slightly intensifies this structure. It is worth noting that the correlograms generally appear to have a mode at distance class 4, but no biological interpretation of this is obvious.

Stochastic variation in SNDs was generally small. For example, the differences among values for different double homozygotes are shown in Figure 6. These differences are similar to those observed among simulations within the same set. Thus, the results are quite predictable for a given pair of loci.

Effects of process parameters: A marked decrease in SNDs for very short distances occurs when N_e is increased from *ca.* 4 to 12 (Figures 2–4). Steady but lesser decreases occur as N_e is increased further. The greatest decreases are observed for joins between like double homozygotes, and this is interpreted as resulting from greater degrees of mixing of double homozygotes with other genotypes, in areas including those outside the patches of double homozygotes, when dispersal is increased. The *X*-intercepts are somewhat increased, although only slight changes in values for intermediate distances (D equals ~4–20) are observed. This indicates that the patch sizes of double homozygotes are only slightly affected by relative large increases in dispersal. Similar but smaller changes occur in other like joins (Figures 3 and 4).

Reducing recombination to small values only slightly affects the join-counts for like double homozygotes and for like single heterozygotes (Figure 3). In contrast,

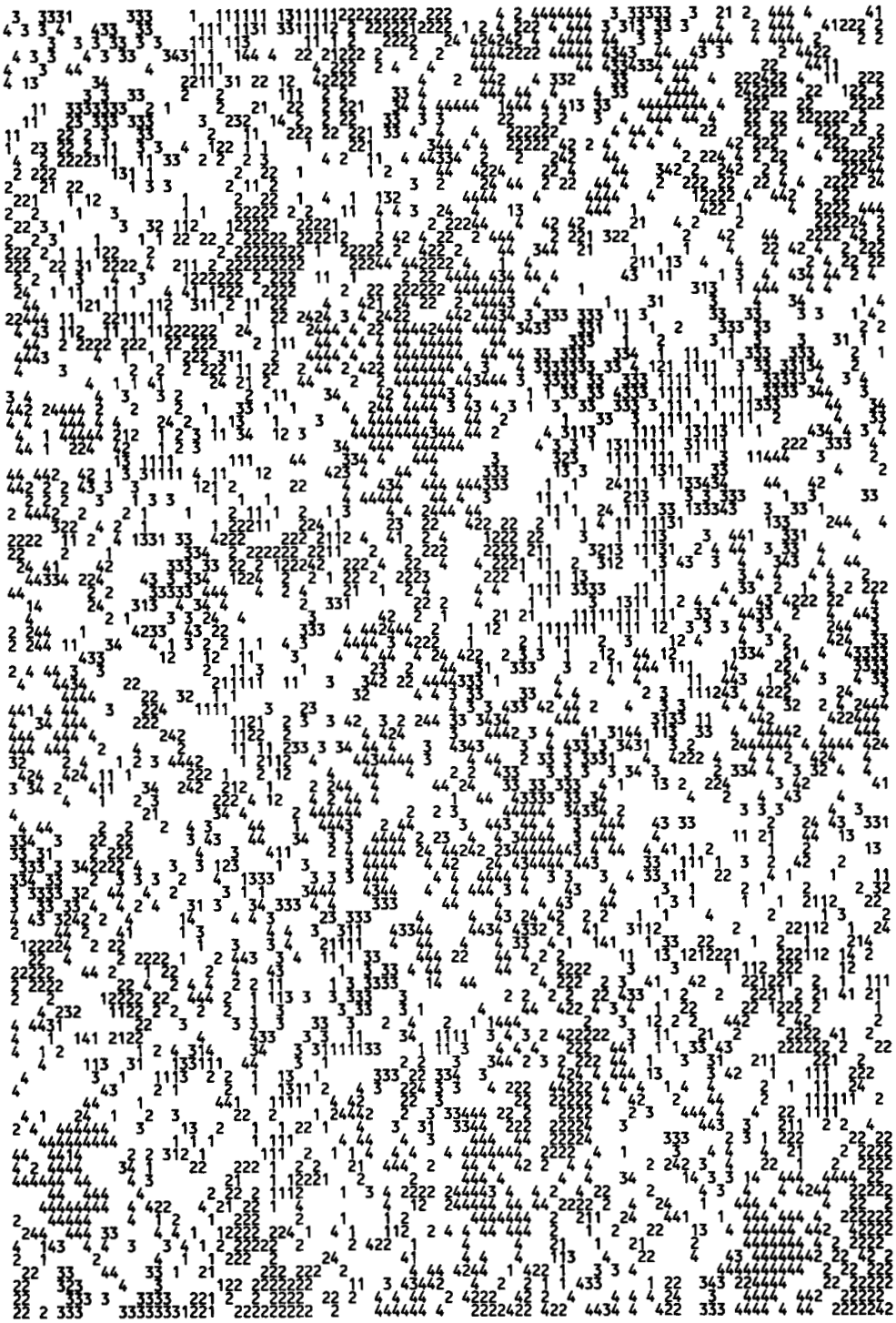

FIGURE 1.—Spatial distributions of the four double homozygotes (Nos. 1–4) for one simulation run of set 1, at generation 200. Blank spaces represent other genotypes.

when $r = 0.01$ the join-count statistics for like double heterozygotes are substantially increased, to values near those for like single heterozygotes (Figure 4). This is not surprising because as linkage becomes tight the gamete types behave more like alleles, under these highly stochastic processes. The SNDs for the total number of unlike joins are somewhat decreased (Figure 5), primarily reflecting increases in the numbers of joins between like double heterozygotes.

Moderate skewing of allele frequencies, in models with dispersal as in sets 1 and 4, scarcely changes the average SNDS for joins between like double homozygotes (compare Figures 7 and 8 with Figure 2). The values are increased for the most common double

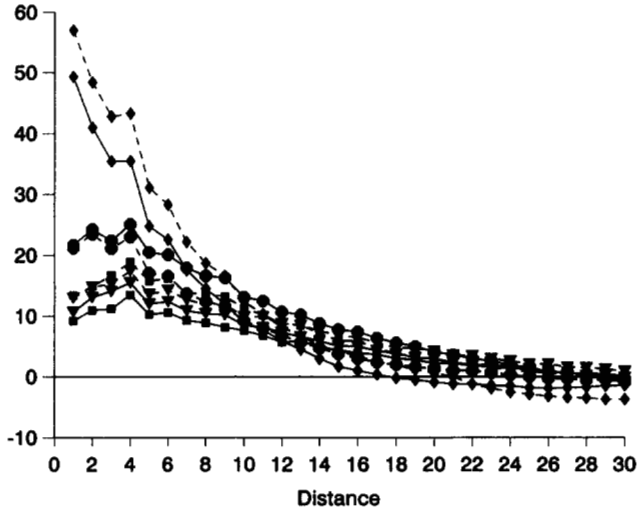

FIGURE 2.—Average SND correlograms for joins between like (identical) double homozygotes for each set, at generation 200: ♦, set 1; ●, set 2; ■, set 3; ▼, set 4 (solid lines); graphs with dashed lines represent the values for sets with equivalent dispersal parameters, but with $r = 0.01$, instead of $r = 0.5$ (*e.g.*, ♦ – – – ♦, set 1r01).

homozygote and decreased for the least common. Similar trends are observed in the two more common single heterozygotes compared with the two less common (Figure 9). Little change was observed in the SNDs for total unlike joins (compare Figure 10 with Figure 5).

Linkage disequilibrium: Average values of linkage disequilibrium for each set were all very close to 0, as expected. The SD of the various linkage disequilibrium coefficients (total, within and between; WEIR 1979) and the average of the absolute values of the coefficients, for each set, are shown in Table 5. All values are small, although those for models with $r =$

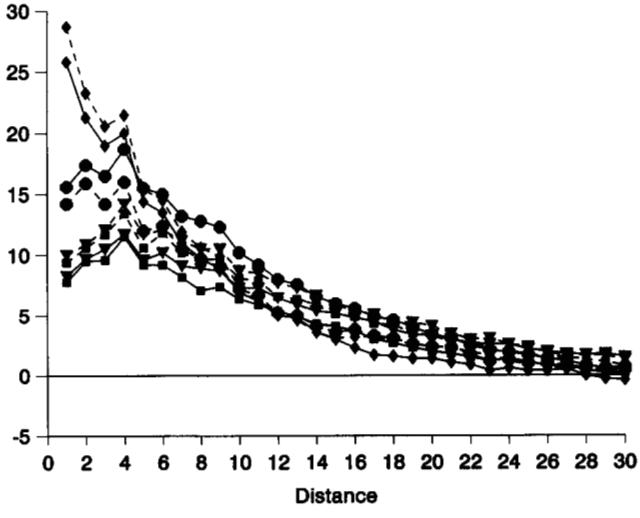

FIGURE 3.—Average SND correlograms for joins between like (identical) single heterozygotes for each set, at generation 200: symbols for sets same as in Figure 2.

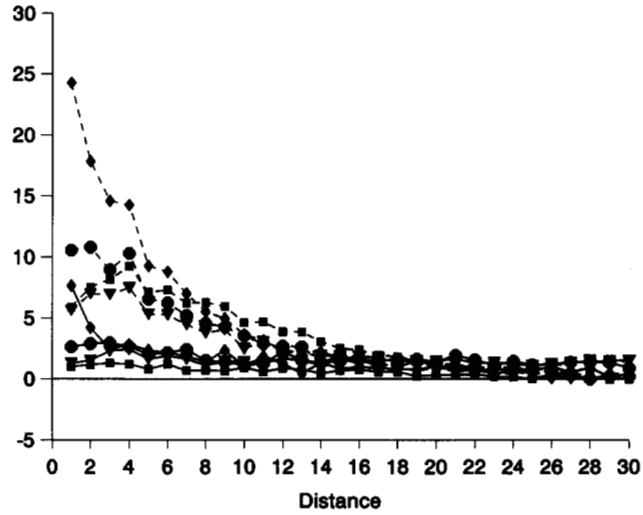

FIGURE 4.—Average SND correlograms for joins between like (identical) double heterozygotes for each set, at generation 200: symbols for sets same as in Figure 2.

0.01 are considerably larger than those with free recombination. Moreover, the values fluctuate positive and negative among simulations within sets, and even during the time course of a single simulation run, especially for runs with free recombination. The within versus between gamete components (WEIR 1979) showed no obvious connection.

Spatial autocorrelations for genotypic values for quantitative traits: Correlograms of I-statistics for average genotypic values for quantitative traits (using the conversions indicated in Table 2) in the quadrats, are very similar for all sets (Figure 11). Moreover, they are very similar to those observed for gene frequencies in single-locus models. Thus, to the resolution of the quadrats, the spatial structure of quantitative traits is robust to changes in dispersal in the low-to-moderate range,

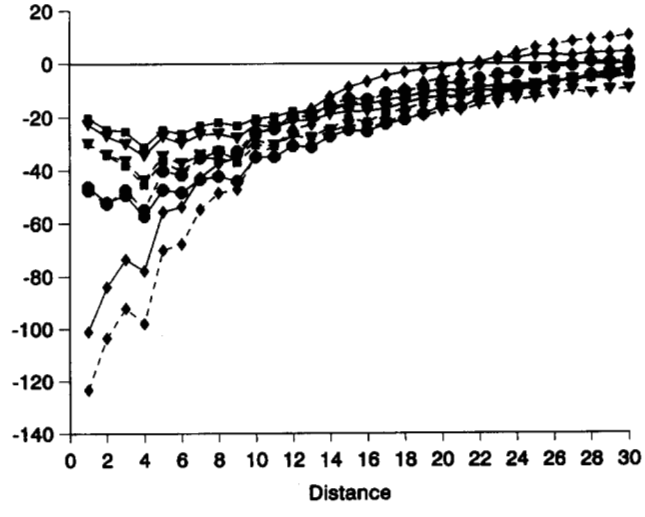

FIGURE 5.—Average SND correlograms for total number of unlike joins for each set, at generation 200: symbols for sets same as in Figure 2.

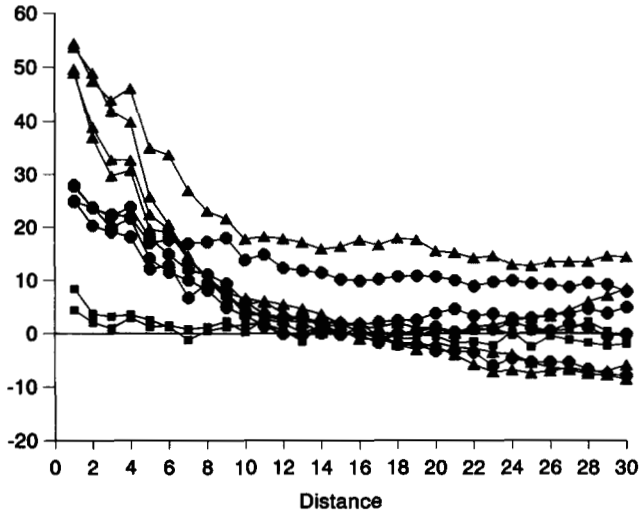

FIGURE 6.—SND correlograms for an individual simulation run of set 1, at generation 200, for like joins for each of the four double homozygotes (▲), four single heterozygotes (●), and two double heterozygotes (■).

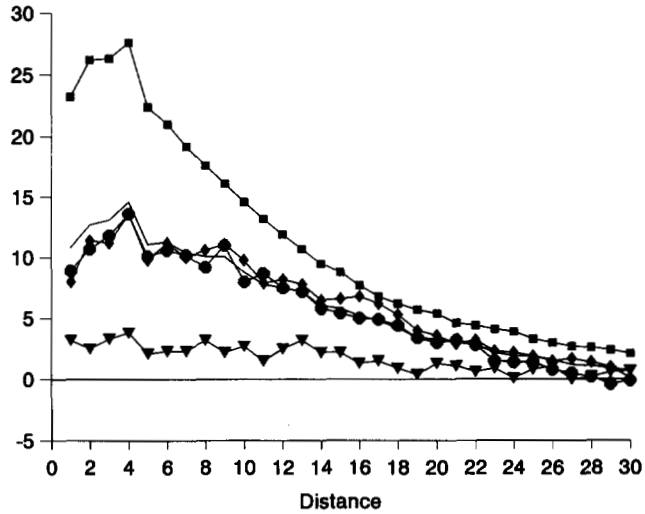

FIGURE 8.—Average SND correlograms for set 4q2 (with $N_m = 49$, $N_f = 49$ and gene frequencies q_a, $q_b = 0.2$) at generation 200, separately for each of the four double homozygotes: ■, *AABB*; ◆, *aaBB*; ●, *AAbb*; ▼, *aabb*; line without symbols is the average among these four.

moderate skewing of allele frequencies (Figure 12), and degree of linkage.

The small changes tied to dispersal follow those observed for Moran's I-statistics for allele frequencies in single-locus models (SOKAL *et al.* 1989). As dispersal is increased from low to moderate, the correlations for short distances generally increase, as do the *X*-intercepts (SOKAL *et al.* 1989).

DISCUSSION

The simulated life cycles of limited dispersal generate spatial distributions of two-locus genotypes with striking structural features and high autocorrelations. The dominant features of the spatial structure are large patches of double homozygotes, which occurs for wide ranges of gene frequencies, and amounts of dispersal, even when loci are unlinked. Here we discuss these and other more subtle aspects, including possible estimators of dispersal based on standing population genetic data; spatial correlations of other genotypes; properties of the total number of unlike joins for multilocus data;

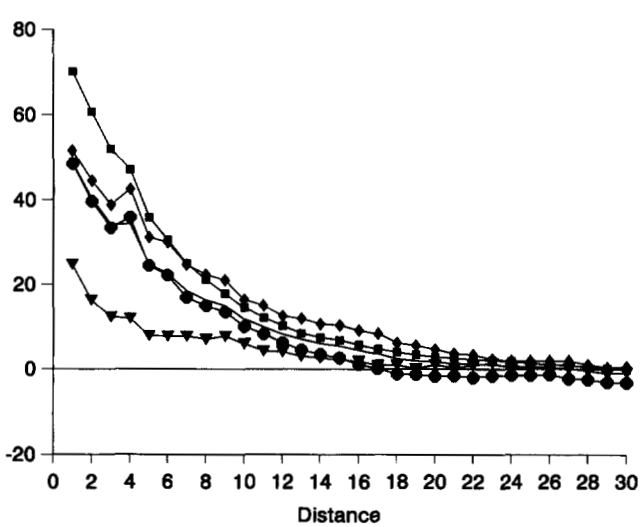

FIGURE 7.—Average SND correlograms for set 1q2 (with $N_m = 9$ and gene frequencies q_a, $q_b = 0.2$) at generation 200, separately for each of the four double homozygotes: ■, *AABB*; ◆, *aaBB*; ●, *AAbb*; ▼, *aabb*; line without symbols is the average among these four.

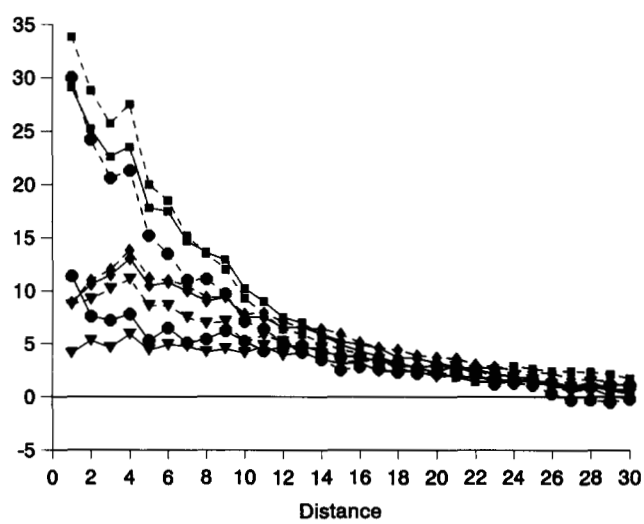

FIGURE 9.—Average SND correlograms for each type of like join between single heterozygotes or between like double heterozygotes for sets with skewed gene frequencies (q_a, q_b = 0.2) at generation 200: symbols ■——■, ■– – –■, ◆ ——◆, and ◆ – – –◆, are for single heterozygotes for sets 1q2, 1r01q2, 4q2, and 4r01q2, respectively; symbols ●——●, ●– – –●, ▼——▼, and ▼– – –▼, are for double heterozygotes for sets 1q2, 1r01q2, 4q2, and 4r01q2, respectively.

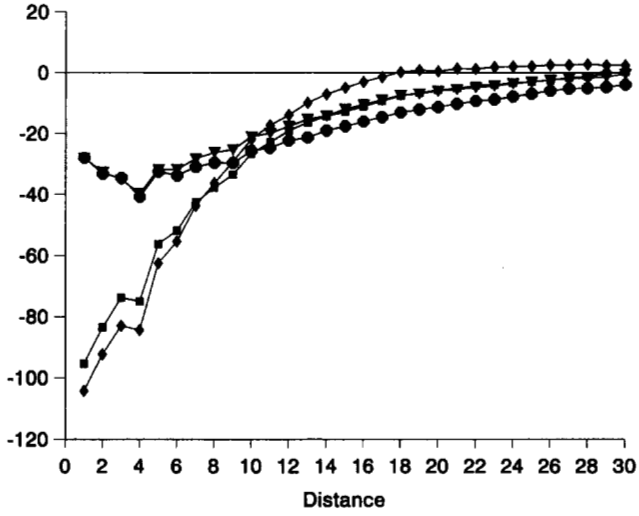

FIGURE 10.—Average SND correlograms for the total number of unlike joins for sets with skewed gene frequencies, at generation 200: ■, set 1q2; ♦, set 1r01q2; ●, set 4q2; ▼, set 4r01q2.

the nature of spatial structural differences caused by changes in gene frequencies, recombination rates, and dispersal; and some implications for spatial correlations of quantitative traits and levels of linkage disequilibrium caused by spatial structure.

Analyses of the SND join-count correlograms, together with inspection of the surfaces, reveal that large patches, or areas of highly contiguous distributions, of double homozygotes develop within *ca.* 50 generations, and then the patchy structure persists as a highly stable or quasistationary state. These patches appear to be functionally analogous to those for single loci (SOKAL and WARTENBERG 1983; B. K. EPPERSON, unpublished data). Remarkably, the tendency of the regional spread of a "lucky" gamete type (*e.g., ab*) in double homozygotes (*aabb*) largely overcomes the decoupling tendencies for recombination, even when the loci are unlinked. The joins for like (identical) double homozygotes occur in great excesses for short distances, but occur in deficits for intermediate distances. The patch structure is reflected in large deficits of joins between opposite double homozygotes (*e.g.*, joins *aabb* × *AABB*, or *AAbb* × *aaBB*), and in lesser deficits of joins between unlike nonopposite double homozygotes, at short distances. The join-count statistics and the *X*-intercepts are similar in size to those for homozygotes in single-locus models (B. K. EPPERSON, unpublished data).

The average join-count correlograms for like double homozygotes are only slightly changed when linkage is tight ($r = 0.01$). The lack of change in patches of double homozygotes complements the lack of efficacy of recombination in decoupling the patterns for each separate locus. The patch structure was largely invariant with respect to the various frequencies of gamete types. When allele frequencies, q_a and p_b, are skewed to 0.2

TABLE 5

Average absolute values of linkage disequilibrium coefficients for various generations for each set

Set	Generation 50	Generation 100	Generation 200
1			
D_g	0.0027 ± 0.0017	0.0036 ± 0.0051	0.0035 ± 0.0050
D_w	0.0032 ± 0.0020	0.0033 ± 0.0048	0.0033 ± 0.0043
D_b	0.0017 ± 0.0022	0.0039 ± 0.0052	0.0050 ± 0.0065
1r01			
D_g	0.0064 ± 0.0093	0.0091 ± 0.0107	0.0111 ± 0.0122
D_w	0.0119 ± 0.0150	0.0100 ± 0.0124	0.0145 ± 0.0135
D_b	0.0070 ± 0.0086	0.0039 ± 0.0047	0.0106 ± 0.0091
2			
D_g	0.0030 ± 0.0034	0.0029 ± 0.0036	0.0060 ± 0.0067
D_w	0.0026 ± 0.0036	0.0051 ± 0.0063	0.0040 ± 0.0034
D_b	0.0029 ± 0.0041	0.0031 ± 0.0032	0.0034 ± 0.0054
2r01			
D_g	0.0077 ± 0.0089	0.0149 ± 0.0242	0.0097 ± 0.0129
D_w	0.0119 ± 0.0153	0.0250 ± 0.0404	0.0164 ± 0.0222
D_b	0.0057 ± 0.0075	0.0114 ± 0.0168	0.0069 ± 0.0098
3			
D_g	0.0016 ± 0.0023	0.0017 ± 0.0023	0.0017 ± 0.0022
D_w	0.0019 ± 0.0027	0.0045 ± 0.0066	0.0014 ± 0.0014
D_b	0.0010 ± 0.0013	0.0034 ± 0.0047	0.0018 ± 0.0017
3r01			
D_g	0.0069 ± 0.0094	0.0073 ± 0.0068	0.0082 ± 0.0100
D_w	0.0118 ± 0.0165	0.0127 ± 0.0137	0.0104 ± 0.0125
D_g	0.0069 ± 0.0085	0.0068 ± 0.0088	0.0054 ± 0.0052
4			
D_g	0.0027 ± 0.0036	0.0026 ± 0.0035	0.0033 ± 0.0041
D_w	0.0040 ± 0.0044	0.0024 ± 0.0037	0.0020 ± 0.0021
D_b	0.0028 ± 0.0033	0.0035 ± 0.0043	0.0028 ± 0.0034
4r01			
D_g	0.0039 ± 0.0050	0.0079 ± 0.0093	0.0077 ± 0.0088
D_w	0.0068 ± 0.0089	0.0176 ± 0.0201	0.0111 ± 0.0136
D_b	0.0030 ± 0.0039	0.0097 ± 0.0110	0.0053 ± 0.0068

Values are means ± SD for the coefficients D_g, gametic (total) disequilibrium; D_w, within gamete disequilibrium; and D_b, between gamete disequilibrium.

(for both loci), the frequencies of *ab* gametes types are reduced (from *ca.* 0.25 to *ca.* 0.04), and the frequencies of *aabb* genotypes are greatly reduced (*e.g.*, at generation 200 the averages ranged from *ca.* 0.004 to 0.0120 depending on the model; thus the number of *aabb* genotypes in populations ranged from *ca.* 40 to 120). Even in these cases substantial patches of *aabb* genotypes occur. Although the SNDs for *aabb* × *aabb* joins are reduced, the SNDs averaged over all four double homozygotes are scarcely changed. Further skewing of allele frequencies might have greater effects. It should be noted that the excess frequencies of homozygotes (which in turn reflects the patchy structure) indicate substantial inbreeding, analogous to the single-locus cases where F_{it} averages about 0.33, 0.15, in models 1 and 2.

For all but the shortest distance classes, the join-count statistics for like double homozygotes scarcely varied,

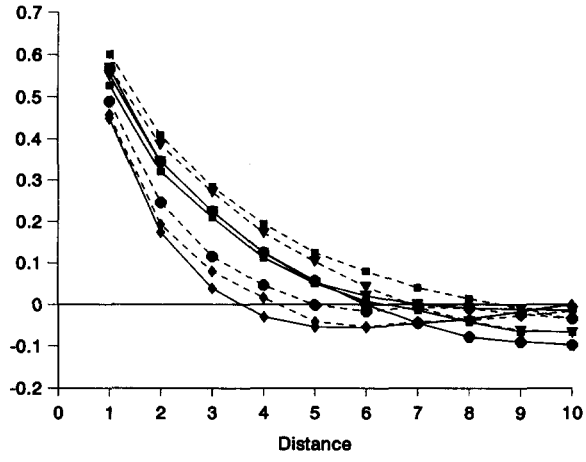

FIGURE 11.—Average I-correlograms for quadrat average genotypic values for quantitative traits for each set, at generation 200: symbols same as in Figure 2.

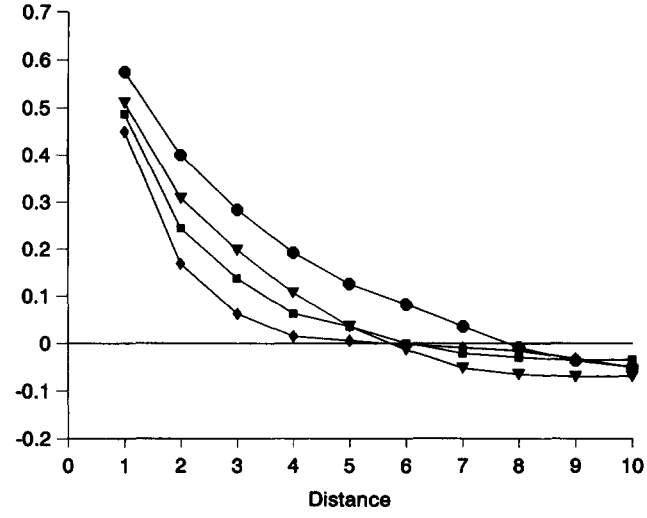

FIGURE 12.—Average I-correlograms for quadrat average genotypic values for quantitative traits for each set with skewed gene frequencies, at generation 200: symbols same as in Figure 10.

over a wide range of amounts of dispersal, N_e from near 0 up to *ca.* 50. However, those for the distances up to three or four were substantially reduced, as N_e increases. This remarkable change in structure is also observed in single-locus autocorrelations (B. K. EPPERSON, unpublished data). Thus when dispersal is increased, there is a greater degree of mixing of double homozygotes with other genotypes at very small spatial scales, and this may primarily occur in areas along the perimeters of patches and elsewhere outside of well-defined patches.

In sum, the results indicate that substantial patches of double homozygotes will be present in populations under a wide range of conditions. Moreover, the magnitudes of the SNDs suggest that the patch structure should be easily detected in samples of moderate size, when these are collected on an appropriate spatial scale (EPPERSON 1990b), in real populations.

The high sensitivity of SNDs for the shortest-distance classes to exact level of dispersal should provide a powerful tool in experimental studies, and suggests that good estimates of dispersal could be obtained solely from samples of standing spatial distributions of multilocus genotypes (see also B. K. EPPERSON, unpublished data).

Spatial patches of double homozygotes can cause striking levels of linkage disequilibrium even among unlinked loci, at some spatial scales. For example, if a sample covered an area containing mostly two patches of opposite double homozygotes (*e.g.*, one of *aabb*, the other of *AABB*), there would be near maximal linkage disequilibrium in the sample. This is in some ways analogous to a two locus "Wahlund effect" (*e.g.*, PROUT 1973), for models of discrete populations which exchange migrants and experience genetic drift (see also CHRISTIANSEN and FELDMAN 1975). Thus substantial disequilibrium can be created from purely stochastic processes, and this could be misinterpreted as evidence of epistatic selection or hitchhiking effects. In contrast, values of linkage disequilibrium in the total simulated populations are near 0 for unlinked loci, and they are very small even when loci are tightly linked. The populations are large enough so that patches (and concentrations) of all combinations of genotypes for each locus are likely to be present in approximately equal numbers. This illustrates how spatial structure-caused linkage disequilibrium depends on spatial scale.

The SND correlograms for other combinations of genotypes group into distinct categories, and these generally reflect weaker autocorrelations than those between double homozygotes. Values of SNDs for joins between like (identical) single heterozygotes indicate that single heterozygotes are substantially concentrated into large regions. However, these regions are not as well defined as patches of double homozygotes; contiguity is relatively low, because matings between heterozygotes produce on average 50% homozygotes (B. K. EPPERSON, unpublished data). These features were evident for the entire spectrum of gene frequencies and recombination rates studied. In addition, increasing dispersal had effects similar to those on the SND correlograms for like double homozygotes. Correlograms for joins between like double heterozygotes indicate that concentrations of like double heterozygotes are weaker still, because of the added effects of recombination. However, when recombination was low, the values of these SND statistics approached those for like single heterozygotes, apparently indicating that then gametes behave similarly as alleles in these highly stochastic models.

The results showed how SNDs for some specific unlike joins, for short distances, generally reflect the

patches and concentrations described above. The most significant of these were the excess of joins between double homozygotes and single heterozygotes differing by only one gene and the deficits for joins in which the heterozygote had the opposite homozygous genotype at the second locus. Evidently this results from juxtaposition of areas of concentrations of single heterozygotes and areas containing patches of double homozygotes. Analogous associations of heterozygotes with homozygotes are observed in single locus distributions (B. K. EPPERSON, unpublished data). The SNDs for other types of specific unlike joins were close to 0. Overall, the great deficits of total numbers of unlike joins are caused primarily by the patches of double homozygotes and to a lesser extent by the concentrations of single-locus heterozygotes.

The multilocus structures in the simulations also reveal some new features of spatial distributions of genetic variation for quantitative traits within populations. The I-correlograms for average genotypic values in quadrats, exhibit high spatial autocorrelations. Much of this is caused by the patches of double homozygotes. When recombination was highly restricted ($r = 0.01$), the forms of the structures for other genotypes were somewhat changed, but the I-correlograms for quantitative traits were virtually unchanged from the free recombination case. This indicates that these correlations probably obtain for a wide range of linkage arrangements.

These results on new features of strong spatial structuring and autocorrelations for traits controlled by two loci, may hold true for a wide variety of quantitative traits, even though these may be selectively neutral, as was the case in the present study. Because of the dominance of the stochastic spread of two-locus double homozygotes into patches, even for unlinked genes, we might expect that for a quantitative controlled by three, four, five, *etc.* loci, that the underlying multilocus structure will be dominated by patches of triple, quadruple, quintuple, *etc.*, homozygotes. Thus the genic effects of the loci on the trait will be strongly autocorrelated. Indeed it seems that spatial correlations for traits controlled by several loci should often be near those observed in the present models. The I-statistics for two-locus traits are very close to those for the analogous statistics calculated for gene frequencies for single loci, using the same quadrat system (SOKAL *et al.* 1989). This correspondence indicates that it should not matter how the degrees of effects on the trait are distributed among the contributing loci. However, one proviso is that if there are very many loci then it seems possible there could be some differences, owing to the improbability of getting patches of multilocus homozygotes for the rare alleles, but this awaits further study. It is worth pointing out that dominance or epistatic interactions (unless extreme) would likely result in only slight changes in spatial correlations. Finally, we may expect that when there is substantial environmental variation (*i.e.*, adding substantial spatially random noise to the phenotypes) weighted against genotypic variation, the spatial correlations of the phenotype would be reduced correspondingly, depending on the heretibility of the trait. Nonetheless, these results suggest there should be substantial spatial autocorrelations for quantitative traits of wide variety of conditions, even though these may be selectively neutral, in many populations with low-to-moderate levels of dispersal. It is noted that selection gradients may substantially change the spatial correlations, according to the migration rates (*e.g.*, ZHIVOTOVSKY and FELDMAN 1993).

This work was supported by National Institute of General Medical Sciences grant GM-48453 to the author and by a grant from the San Diego Supercomputing Center for use of SDSC computing facilities.

LITERATURE CITED

ARGYRES, A. Z., and J. SCHMITT, 1991 Microgeographical genetic structure of morphological and life history traits in a natural population of *Impatiens capensis*. Evolution **45:** 178–189.

CHRISTIANSEN, F. B., and M. W. FELDMAN, 1975 Subdivided populations: a review of one- and two-locus deterministic theory. Theor. Popul. Biol. **7:** 13–38.

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

EPPERSON, B. K., 1990a Spatial autocorrelation of genotypes under directional selection. Genetics **124:** 757–771.

EPPERSON, B. K., 1990b Spatial patterns of genetic variation within plant populations, pp. 229–253 in *Population Genetics and Germplasm Resources in Crop Improvement*, edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER, and B. S. WEIR. Sinauer Associates, Sunderland, MA.

EPPERSON, B. K., 1993 Recent advances in correlation studies of spatial patterns of genetic variation. Evol. Biol. **27:** 95–155.

EPPERSON, B. K., and R. W. ALLARD, 1989 Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. Genetics **121:** 369–377.

EPPERSON, B. K., and M. T. CLEGG, 1986 Spatial autocorrelation analysis of flower color polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). Amer. Nat. **128:** 840–858.

GILLOIS, M., 1966 Le concept d'identite et son importance en genetique. Ann. Genet. **9:** 58–65.

LANDE, R., 1991 Isolation by distance in a quantitative trait. Genetics **128:** 443–452.

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

MALÉCOT, G., 1948 *Les Mathématiques de L'Hérédité*. Masson, Paris.

NAGYLAKI, T., 1994 Geographical variation in a quantitative character. Genetics **136:** 361–381.

PROUT, T., 1973 Appendix to MITTON, J. B., and R. K. KOEHN, Population genetics of marine pelecypods. III. Epistasis between functionally related isozymes in *Mytilas edulis*. Genetics **73:** 487–496.

ROGERS, A. R., and H. C. HARPENDING, 1983 Population structure and quantitative characters. Genetics **105:** 985–1002.

ROHLF, F. J., and G. D. SCHNELL, 1971 An investigation of the isolation-by-distance model. Amer. Nat. **105:** 295–324.

SCHOEN, D. J., and R. G. LATTA, 1989 Spatial autocorrelation of genotypes in populations of *Impatiens pallida* and *Impatiens capensis*. Heredity **63:** 181–189.

SLATKIN, M., and H. E. ARTER, 1991 Spatial autocorrelation methods in population genetics. Am. Nat. **138:** 499–517.

SOKAL, R. R., 1979 Ecological parameters inferred from spatial correlograms, pp. 167–196 in *Contemporary Quantitative Ecology and Related Econometrics*, edited by G. P. PATIL and M. L. ROSEN-

ZWEIG. International Cooperative Publishing House, Fairland, MD.

SOKAL, R. R., 1988 Genetic, geographic, and linguistic distances in Europe. Proc. Nat. Acad. USA **85:** 1722–1726.

SOKAL, R. R., and G. M. JACQUEZ, 1991 Testing inferences about microevolutionary processes by means of spatial atuocorrelation analysis. Evolution **45:** 152–168.

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

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

SOKAL, R. R., and N. L. ODEN, 1991 Spatial autocorrelation analysis as an inferential tool in population genetics. Am. Nat. **138:** 518–521.

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., G. M. JACQUEZ and M. C. WOOTEN, 1989 Spatial autocorrelation analysis of migration and selection. Genetics **121:** 845–855.

TURNER, M. E., J. C. STEPHENS and W. W. ANDERSON, 1982 Homozygosity and patch structure in plant populations as a result of nearest-neighbor pollination. Proc. Natl. Acad. Sci. USA **79:** 203–207.

WAGNER, D. B., Z.-X. SUN, D. R. GOVINDARAJU and B. P. DANCIK, 1991 Spatial patterns of chloroplast DNA and cone morphology variation within populations of a *Pinus banksiana-Pinus contorta* sympatric region. Am. Nat. **138:** 156–170.

WEIR, B. S., 1979 Inferences about linkage disequilibrium. Biometrics **35:** 235–254.

WRIGHT, S., 1943 Isolation by distance. Genetics **28:** 114–138.

ZHIVOTOVSKY, L. A., and M. W. FELDMAN, 1993 Heterogeneous selection in subdivided populations. J. Math. Biol. **31:** 747–759.

Communicating editor: G. B. GOLDING