

Spatial genetic structure and non-equilibrium demographics within plant populations

BRYAN K. EPPERSON

Department of Forestry, Michigan State University, East Lansing, MI 48824, USA (Email: epperson@msu.edu)

Abstract

Spatial genetic structure within plant populations is primarily determined by the effects of limited seed and pollen dispersal. In the absence of selfing, spatial genetic structure and limited pollen flow are generally considered to be the primary controlling factors of the dynamics of biparental inbreeding and inbreeding depression. However, in some cases spatial structure may depend on specific demographic processes. Most theoretical studies have focused on the long-term or equilibrium structure of genetic variation, which allows, if demographic considerations are simple enough, a system to be collapsed into a single standardized measure (such as Wright's neighborhood size) of total pollen and seed dispersal distances. Specific demographic considerations may take many forms and strengths. In this paper we examine some particular cases that illustrate a wide range of effects on spatial genetic structure, and identify some of the most critical parameters. For example, *Cecropia obtusifolia* is a canopy-gap-filling, dioecious species. Perennially, the highly limited availability of habitat forces the removal of seedlings and the associated strong but ephemeral genetic structure, leaving one or a few trees to fill in each canopy gap. The structure among reproductive adults is minor and in line with standard amounts and forms expected for a wind-pollinated species. In contrast, *Silene dioica* is a dioecious, bee-pollinated, gravity-seed-dispersed perennial, and a colonizer or early successional species. In certain regions of Sweden, *S. dioica* grows primarily on uplifting islands that undergo rapid succession, and *S. dioica* exhibits extreme levels of spatial autocorrelation, caused by highly limited seed and pollen flow at a specific spatial scale that corresponds to demographic patches. Moreover, the types of genotypic concentrations in space vary, being dominated by matrilineal relationships. The availability of habitat rapidly increases, then rapidly decreases (with analogous changes in population growth rates), and, combined with a long-life, causes the genetic structure to exist on such a time scale that it appears to be the main form of within-population genetic structure for the species in this region. Clonal reproduction may have marked effects on the spatial genetic structure of plant populations. For monoecious, self-compatible species, clonal groups may increase effective selfing, whereas in dioecious plants any effects on the development of IBD may be more subtle. We examine a number of dioecious plants with a variety of spatial distributions of clones and populations, existing for various time scales. The demographics of clonal structure may interact with spatial-temporal demographics and structure of the genetic variation.

Keywords: mating system, plant demographics, population genetics, population structure, spatial autocorrelation.

Received 16 August 1999; accepted 4 July 2000

Introduction

The spatial distribution of genetic variation has several important influences on plant population genetics, and many of these interact with demographic processes. First,

in many plant populations, various population genetic processes both influence and are influenced by spatial structure; in some cases process and structure may be inseparable (e.g. microenvironmental selection). Thus, theoretical models of population genetics that include a

spatial context can add to our understanding of evolutionary and ecological genetic processes. Second, because spatial distributions are sometimes distinctively altered by various evolutionary factors, we may expect to use spatial patterns to quantify different evolutionary factors, including means for obtaining indirect measures of gene flow or dispersal. This approach is especially powerful where spatial patterns for multiple genetic loci may be compared within the same population, in part because neutral loci are subject to the same dispersal and demographic parameters (i.e. spatial patterns for neutral genes may differ from patterns for loci under natural selection). Third, spatial structure causes genetic correlations among the elements of spatial samples from natural populations. When populations are sampled, the sample is taken over space, whether the sample is taken for securing germplasm for gene conservation purposes or simply for analysis. Spatial correlations imply non-independence among the elements of samples, and this can cause serious biases in estimators of many important population genetic parameters, including the level of inbreeding within the sample. More fundamentally, spatial correlations violate one of the basic assumptions about random samples (Epperson 1993).

In disentangling complex relationships between spatial patterns and the underlying spatial-temporal population genetic and demographic process, considerable attention must be paid to the spatial structure of genetic variation caused for neutral genetic markers. In its simplest form, spatial structure is caused by limits to seed and pollen dispersal, however, the development of structure may also depend on other factors.

In this article we distinguish 'standard' from 'non-standard' genetic isolation by distance (IBD). For selectively neutral loci, the former depends only on seed and pollen dispersal distances, relative to density, for which there are well-developed theoretical models. We will focus on theory and measurement by spatial autocorrelation statistics (Sokal & Oden 1978). Biological examples of standard IBD will be analyzed, followed by several examples where the demographics violate the assumptions of the standard theory, sometimes resulting in different types of spatial structures. Examination of the various cases can lead to insights about those demographic factors, including clonal reproduction, age structure and population growth rates, which may change spatial structure from the standard form.

Standard theory of genetic isolation by distance

The classical theories of Wright and Malécot, developed in the 1940s, considered a single, large, evenly aged population in which individuals were effectively continuously

or uniformly distributed (Wright 1943; Malécot 1948). In other words, densities were assumed to be uniform over space and fixed over time. Although such circumstances may often be reasonably approximated in nature, they imply constraints on various demographic processes. In the classical theory, limits to dispersal of individuals cause individuals to tend to mate with nearby individuals (i.e. mating by proximity). If dispersal is limited enough, then repeated generations of mating by proximity causes matings between kin-related plants, and through this process, the build up of spatial structure, even though the population is continuous. In the standard theory, the process does not depend on details of the demographics. The original standard models were in terms of male and female dispersal, and it was only the total of the two that determined the spatial structure (Wright 1943). In plants, dispersal is not male versus female, rather it is seed versus pollen, however, it appears that as long as neither is too extreme, spatial structure under the standard model depends only on a weighted average of the two (Crawford 1984). Thus, measures such as Wright's neighborhood size (N_e) of the total amount of dispersal standardized for density, capture essentially all effects on spatial structure, under the assumptions of the standard models. Implicit assumptions about demographics such as local unchangeability of densities may fit for some plants but not others. Wright (1943) expressed his results in terms of the development of locally inbred demes, or genetic IBD. Malécot (1948) showed that this process results in adjacent or nearby individuals being kin-related and he showed how the degree of kinship falls off with the distance of separation. In Malécot's analytic results the precise curve depends only on the total amount of dispersal (e.g. Wright's neighborhood size) and another parameter (the outside systematic pressure).

More recently, a series of simulation studies (e.g. Sokal & Wartenberg 1983; Epperson 1990) of the standard model have revealed that a dominant spatial feature is the growth of patches within 10–50 generations; where a patch is loosely defined as a large area of concentrations of alleles. In cases of low dispersal a patch consists of mostly one homozygous genotype, but for higher dispersal levels the homozygosity concentrations are weaker and heterozygosity is higher. Adjacent groups of individuals have highly similar or autocorrelated gene frequencies compared to the rest of the population.

Statistical methods

There are two different types of spatial autocorrelation statistics that can be used to characterize and quantify spatial structure. The first of these are join-count statistics. As an example, consider the spatial distribution of a

single locus for genotypes at a locus with two alleles: there are three genotypes and six different types of combinations in pairs, or 'joins'. For a sample of size n , there are ' n choose two' joins in total. We first form distance classes – ranges of distances that may separate a pair of plants – then classify or 'count' all pairs or joins according to the two genotypes and the distance class of separation. For each type of join for each distance class, we calculate the mean and variance expected under the null hypothesis that genotypes are randomly distributed and form a test statistic or SND of this null hypothesis, which has an approximately standard normal distribution (values greater than 1.96, or less than -1.96, are statistically significant). A set of SND statistics for the different distance classes is called a correlogram. In addition, SND can be calculated for the total number of unlike joins.

For large simulated populations of genotypes at a locus (A/a) the SND values are very large and highly significant. Most distinctively for short distances, large excesses are observed for joins between neighbors that have the same homozygous genotype, either both AA (i.e. $AA \times AA$) or both aa ($aa \times aa$). Highly significant negative SND for $AA \times aa$ joins indicate deficits of joins between neighbors that have opposite homozygous genotypes. The excesses and deficits 'drop off' as the distance between individuals increases, and typically, at some distance they switch sign. This distance is important because it measures the diameter of patches. Typically, there are large patches, each consisting of four to five hundred or more individuals.

The second set of statistics, Moran's I statistics, are calculated for variables of numerical values (Cliff & Ord 1981), particularly allele frequencies in some local areas. Typically, this is done by first computing the allele frequencies among individuals within quadrats. Another popular method is to convert individual genotypes (i.e. quadrat size of one) into allele frequencies for each allele separately (1.0, 0.5, and 0.0 for homozygotes for that allele, heterozygotes and other genotypes, respectively). Then for each allele, Moran's I -statistics are calculated separately using the following procedure: (i) break up the population into quadrats or individuals and map the locations of each, (ii) calculate the allele frequency in each, and (iii) calculate the distances between quadrats or individuals and place each pair of quadrats or individuals into these distance classes. Then, as for the join-counts, for each distance class Moran's I statistic is calculated. Moran's I statistic is the sum of the products of the mean-adjusted gene frequencies in pairs of quadrats separated by distances in distance class k , normalized by dividing by a measure of the spatial variance in gene frequency. It measures correlation as a function of distance.

Spatial statistical measures under standard isolation by distance processes

Starting either from a random spatial distribution of genotypes for multiple alleles or from an initial population that is fixed for an allele, and allowing sufficient input of new alleles, the spatial structure builds rapidly. Large increases in spatial autocorrelations are evident within 10 generations. Within about 30 generations the structure and statistics have nearly reached (and by 50–100 generations the populations have reached) a quasi-stationary distribution. The spatial autocorrelations scarcely change after about 50–100 generations (Sokal & Wartenberg 1983; Epperson 1990). A typical I -correlogram (Fig. 1) has high positive correlations between adjacent individuals. The correlations drop off with distance, and become negative at about 25 or more times the distance between adjacent plants. I -correlograms intersect zero at about the same distance as the SND correlograms for like homozygotes, indicating that both measure patch sizes in the simulations.

Under standard forms of IBD (i.e. unless dispersal is great) spatial autocorrelation statistics should detect genetic structure for well-designed (Epperson 1992, 1993) samples of a few hundred or less individual genotypes (Epperson *et al.* 1999). Moreover, in large populations the structure for any given locus is highly predictable (i.e. there is little stochastic variance among simulations). As sample sizes decrease, statistical error may be added to stochastic variation. Changes in allele frequencies have

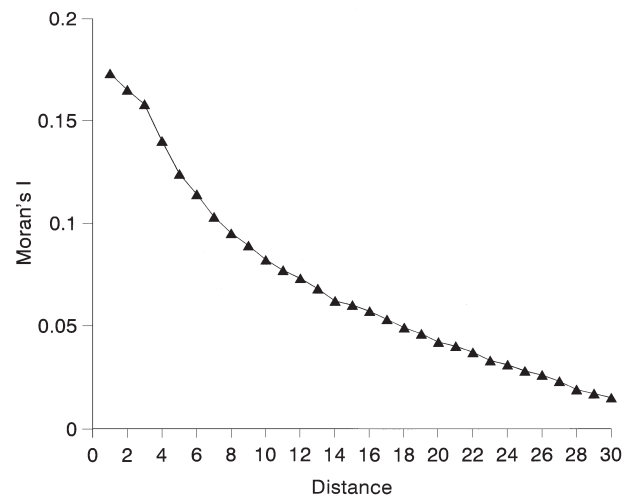


Fig. 1 Typical correlogram for Moran's I statistics for individual genotypes (\blacktriangle) in a simulated population with moderate amounts of dispersal (Wright's neighborhood size=25) under the standard genetic isolation by distance model.

little or no effect unless frequencies are near 1.0 or 0.0. For the SND, the precise level of dispersal has a large effect (Epperson 1995; Epperson & Li 1997; Epperson *et al.* 1999). The most distinct feature is that increasing dispersal causes reductions in SND for joins between identical homozygotes for the shortest distance classes. Complementary effects are seen for joins between opposite homozygotes; greater dispersal causes lower deficits of these joins. The same is seen for the total number of joins between unlike genotypes. The level of dispersal affects Moran's I statistics similarly, unless quadrat sizes are large (Epperson & Li 1996, 1997).

In summary, the results of numerous simulation studies, anchored by analytic results of mathematical models, lead to strong or robust predictions for neutral loci, and it is relatively easy to determine if a study population has dispersal within a broad theoretical range. In contrast, it is notoriously difficult to directly estimate dispersal with great precision. We expect spatial correlations to be very similar for different neutral loci, because generally in a population loci are subject to the same mating and dispersal. These kinds of studies lead to a strong hypothesis testing framework, which is important because loci under the same influences should have similar spatial correlations, and it is usually relatively easy to obtain data for many loci. For example, consider the following common situation where: (i) a plant population has dispersal in the low to moderate range, say $N_e < 100$; (ii) the population is at least 30 or more generations old; (iii) the sample is taken over an appropriate spatial scale; and (iv) the allele frequencies are not too close to zero. In this situation, all neutral loci should have large patch structure and the characteristic SND-correlograms and I-correlograms. In most cases, if sample sizes are large the SND correlograms are also large and reflect the patch structure. If SND are averaged over several loci, it is possible to use SND-correlograms to precisely estimate dispersal (Epperson & Li 1997; Epperson *et al.* 1999). In addition, spatial autocorrelations for the first distance class are very sensitive to dispersal level, and these can be used separately for measuring dispersal, greatly simplifying the statistical procedures. It should be noted that the values of SND also depend on sample sizes, and tables that allow adjustments for sample sizes have been published (Epperson & Li 1997; Epperson *et al.* 1999). It should be noted that SND generally have slightly higher statistical power than I-correlograms. Additional unpublished results show that selection also reduces SND-correlograms in a similar way to the way it reduces I-correlograms, thus if one locus is under (non-microenvironmental) selection, it will have markedly reduced patch structure and greatly reduced autocorrelations. Other distinctive changes in autocorrelations can be caused by some forms of microenvironmental selection

(Sokal *et al.* 1989), although more theoretical and experimental work in this research area is needed.

Moran's I statistics for allele frequencies converted from individual genotypes have very similar properties as those for the SND for like homozygotes, and have similar statistical power. They decrease monotonically with increasing dispersal, and have a similar relationship to allele frequencies. Importantly, the size of the quadrat (i.e. the average number of individuals in a quadrat sample) can be critical. Most notably, when the quadrat size is around 25 the I-correlograms scarcely vary over a fairly wide range of dispersal parameters, from low to moderate dispersal. In fact, within this range as dispersal increases, initially the curves become steeper, then they begin to flatten out. This is unfortunate in one sense, because it means that in such cases I-correlograms cannot provide precise estimates of dispersal; however, predictions for neutral loci are robust to dispersal, and this could be useful in certain circumstances.

Correlograms of Moran's I statistic calculated for the converted gene frequencies differ from those for quadrat sizes of 25, which were used to characterize structure in early simulation studies (Sokal & Wartenberg 1983; Epperson 1990). This difference caused considerable confusion in the many experimental studies that subsequently used Moran's I statistics. In fact, experimental studies that have used converted genotypes have obtained low values, as is expected from IBD theory (Epperson *et al.* 1999). Extensions of these basic results apply to loci with multiple alleles (Epperson *et al.* 1999), and for multilocus genotypes and the genetic value of quantitative traits, subject to some provisos (Epperson 1995).

Non-standard isolation by distance models

It is known that the predictions discussed above are robust despite some forms of deviations from the standard model assumptions, such as some forms of demographic details. Only one theoretical study has focussed on this. Doligez *et al.* (1998) found that age structure and clumping of individuals in a distribution (i.e. non-uniform density) had in most cases relatively little effect on the long-term ('quasi-stationary') spatial distribution of genetic variation.

Ipomoea purpurea

The first example is based on a study of nine populations of the morning glory, *Ipomoea purpurea*, in south-eastern USA (Epperson & Clegg 1986), where populations contain a remarkable variety of flower color types. One locus controls pink versus blue color: two alleles segregate and the allele conferring pink is recessive to the allele conferring

blue. Morning glory is self-compatible and can autopollienate well. Bumblebees account for more than 98% of pollinator visits to morning glory populations in southeastern USA, and they cause a high degree of mating by proximity (Epperson & Clegg 1987). The seeds of morning glory are large and do not disperse far. The combined results of a series of direct experimental measures of gene flow showed that N_e should generally be in the range of 5–15 (Ennos & Clegg 1982). The study populations are generally at least 50 or so generations old, and apparently stable in size. Based on the theoretical results reviewed above, large, sharply defined patches should be present for neutral loci. Each population was sampled on a regularly spaced grid. The results clearly showed remarkable differentiation of pink homozygotes versus the blue phenotypes, and these patterns are revealed in the join-count statistics. There were large excesses of joins between neighbors on the sample grids, where both were pink homozygotes, and a similar excess of blue–blue joins, with a corresponding deficit of blue–pink joins. The distances at which the SND-correlograms switch sign were measured as the diameters of the patches. In accounting for the overall density, which was also measured, the results indicated that the populations consisted of two types of patches, pink and blue, where each patch contained about 300–500 individuals, which was very similar to the simulated populations for neutral loci. Recently, it has been shown that the observed statistical values fit closely to those expected for the quasi-stationary distribution for neutral loci in simulated standard IBD models with N_e in the range of 10–25 (Epperson & Li 1997), which matches the direct measures of dispersal.

Cecropia obtusifolia

Cecropia obtusifolia is a dioecious tree that regenerates in gaps in the canopy of subtropical forests. It is wind-pollinated and fruits are dispersed by bats and other animals. Alvarez-Buylla studied one population in southern Mexico in detail by mapping and genotyping all individuals within an area of several hectares (Alvarez-Buylla & Garay 1994). Among all 111 adult trees in the population, males and females were fairly evenly dispersed, displaying only slight clustering. In contrast, the distribution of all 242 seedlings was very clumped, almost all occurred within six clumps, and each clump was located in a recent gap in the forest canopy. The distribution of the juveniles was similar to that of the adults. Join-count statistics were calculated for eight isozyme loci; separately for the three age-classes (Epperson & Alvarez-Buylla 1997).

Strikingly non-random spatial distributions of genotypes were observed among the seedlings. The distance classes were chosen so that the shortest distance class, 0–12.5 m, (distance class 1) contained no pairs separated

into different gap clumps, and distance class 2, 12.5–37.5 m, contained the remaining within-gap pairs and many between-gap-clump pairs. For the distance class 1, 31 out of 61 (approximately 50%) SND join count statistics were statistically significant at the 5% level. Many were also significant for distance class 2. There was almost no autocorrelation for larger distance classes. In total, for distance classes 1 and 2, there were many very large values that were larger than had been previously observed for any other species, including *Ipomoea purpurea*. Many of the values were based on large numbers of joins. However, the results were quite different in form (as well as magnitude) from those expected under standard IBD theory. The results indicated a very strong concentration of genotypes into gap clumps, and little relationship among gap clumps. The gaps contained far fewer individuals than the patches of genotypes observed in simulations of the standard model. In addition, some types of joins fit the usual excess types expected under IBD, but others do not. For example, there are significant deficits of joins between identical homozygotes and excesses of joins between different heterozygous genotypes for distance classes 1 and 2. For the adults, 15 out of 67 results (22%) were significant for distance class 1, and 11 of 79 (14%) were significant for distance class 2. Very few results were significant at larger distances. The values were generally smaller than those for the seedlings, but more importantly, almost all involved far fewer joins. Most of the significant values involve odd combinations of genotypes. For example, for locus *Pgm-1* there was an excess of joins between the opposite homozygotes 22 and 33, and this reflected concentrations of the rare 22 homozygotes into one of the patches. In addition, more standard associations were found. For locus *Fe-2* the only two 13 heterozygotes are located next to each other, as were three of the four 22 homozygotes. For *Fe-1* all three 23 heterozygotes were adjacent. There can be little doubt that such concentrations of genotypes are not statistical artifacts, and the type of structure differs from the standard model in form, although the differences in terms of numbers of joins are small. Similar results were observed for the juveniles. For both adults and juveniles there was little structure at a very fine spatial scale. Essentially the structure consists of clusters of few individual genotypes.

The structures observed at different age classes are non-standard, but make sense when we consider the demographic details of seed dispersal and population regulation. As noted, *C. obtusifolia* is wind pollinated and seeds are dispersed in fruits by birds and other animals. Fruit dispersal covers long distances (Alvarez-Buylla & Garay 1994). The population size is stable and regeneration only occurs in recently opened gaps in the canopy. Wind pollination would indicate that most seeds on a tree are maternal half-sibs. Animals may disperse one or more

groups (fruits) of seeds from only one tree (or two or more trees) and deposit them at a particular locale. If this locale occurs in a gap the seeds may successfully germinate and grow. Outside recently opened gaps, seeds have been observed but not seedlings. Thus, the seeds that arrive in a gap may consist of a single half-sib array, two essentially unrelated (unrelated because there is little structure among adults) half-sib arrays, and so on. It is likely that on occasion these fruits carry combinations of less frequent alleles, and thus genotypic concentrations occur within one or another of the gaps, causing statistically significant structure for the seedling population as a whole. As the seedlings grow, competition removes all but one on average, but, on occasion two, three or four, may survive. This results in a non-random distribution and a slight structure among the adults. The prediction of the standard IBD model is essentially no structure, based on a high level of seed dispersal and wind pollination. Thus, in contrast to the seed, the distribution of genetic variation in the adults (except for this relict, weak and non-standard form of structure) fits the standard model predictions. The structure among seedlings is strong, but transient, and essentially unimportant with respect to the buildup of IBD.

Pinus contorta

Two natural populations of lodgepole pine, *Pinus contorta*, in the north-eastern corner of Washington state were studied. Each population is continuous and large, with more than 1 million trees, and each is dense, with about 2500 trees per ha, and outcrossing rates are essentially 100% (Epperson & Allard 1984). *Pinus contorta* has among the most highly dispersible pollen and seed even among conifers, which generally have fairly high dispersal. N_e must be greater than 1000 and it could be 10 times larger, therefore, we would not expect any structure in these populations. Each population was sampled on a grid representing about one in every 50 trees in the sample area. Fourteen allozyme loci were assayed, many with more than two alleles. There were many genotypes, and several thousand different join-count statistics were computed for each distance class (Epperson & Allard 1989). Overall, about 5% were significant at the 5% level. Generally, the genotypes in these populations are extremely close to a truly random distribution, and this fits expectations because dispersal is so great. This is consistent with the IBD model, expected from the large amount of dispersal, the stability of the population size, and the lack of any age-structure. Although we cannot be certain that there is no genetic structure more finely scaled than the distance between adjacent samples, it seems unlikely. It is often believed that genetically correlated seed dispersal

(maternal half-sibs in wind pollinated species such as *P. contorta*) could cause fine scale structure, but it is not expected in species with even-aged, stable population dynamics and high seed production and dispersal.

Silene dioica

Perhaps the most informative study about the importance of demographic details in determining nonstandard spatial structures is one conducted by B. E. Giles on *Silene dioica*. *Silene dioica* is a small dioecious perennial, which primarily inhabits islands in certain regions of northern Sweden. In this region, it is insect-pollinated, usually by bumblebees, and has no particular seed dispersal mechanism. The region has been undergoing rapid uplift (approximately 1 cm per year) since the last ice age, and the landscape is rising from the sea from north to south. New islands are continually rising above sealevel. Once above sealevel, the islands undergo rapid ecological successional stages, starting as rocks washed by sea, through small plants, then to mountain ash and alders, and conifers. The successional stages are visible among islands, and within each island as bands of vegetation types corresponding with topographic bands. *Silene dioica* occupies an intermediate position in this succession. Giles has collected detailed information on the initial founder events on different islands including the study population 'Bigstone', for which she also obtained multilocus genotypes for a full census of adults and older seedlings (Giles, unpublished data). There were 320 adult plants, and the locations of all sampled individuals were mapped. Most individuals were in clusters, but there were some isolated 'loners.' Spatial analyses for seven isozyme loci were conducted (Giles and Epperson, unpublished).

Based on observations of demographics, Ingvarsson and Giles (1999) previously described numerous spatial demographic patches (not to be confused with spatial genetic patches) and it is illustrative to consider how such demographic patches arise once an island has been colonized by at least one female and male (before that there was no seed bank). Once a female 'loner' plant becomes established, and if there is one or more males somewhere to mate with, seeds will be produced, which are heavy and generally fall within a circle with a radius equal to the height of the inflorescence, or within approximately 20 cm. This will occur year after year, and it is possible to recognize cohorts for several years. After several years, some offspring become reproductively mature, and daughters can then begin to contribute to the local seed-fall. The seeds aggressively fill in available habitats within 10–30 years. Although individuals can live for fairly long periods (5–10 years), a founding matriarch may die,

although her legacy persists for the 'life' of the patch, and within a few decades the habitat in the patch is no longer suitable. *Silene dioica* must 'move' elsewhere on the island (to a lower, earlier successional topographic zone) or to another island. It was also clear that the island actually consisted of three subpopulations, which had different ages. We analyzed these in detail, and this allowed further insights into the genetic consequences as demographic patches mature and available habitat is filled in. Nonetheless, join-count spatial autocorrelation analysis of the entire island population also showed that there were many of the same spatial features as for the structure of genetic variation.

The main points for observations made for the entire island population were evident from results for the total numbers of unlike joins for genotypes for each locus. First, the average value indicated substantial autocorrelation within distance class 1 (0.0–1.0 m), the upper bound of which is near the average diameter of the demographic patches. Second, there seems to be a 'step' in the curve at about 5–6 m, which is the distance separating the subpopulations. Normally, we would not make much out of such a 'step', but it fits with additional information showing that there is little autocorrelation among patches within subpopulations. Furthermore, analysis of distinct types of joins revealed excesses and deficits of large but non-standard types, reminiscent of the seedling data for *Cecropia obtusifolia*.

The youngest section of the island population was in the north, where *S. dioica* occurs at low densities and there is a relatively high percentage of 'loners.' Two separate spatial analyses were done for each of the three subpopulations: (i) an analysis that included one 'distance' class for pairs that were within the same patches, as defined by the demographic studies, and all other pairs classified according to distance of separation; and (ii) an analysis based on the usual classification of all pairs into classes defined by distance. The results for (i) revealed many large correlations, and many of these clearly indicated odd excesses or deficits of joins that were of nonstandard IBD types. As a set, the results were similar to those observed for (ii). As for the total population, significant results were caused by concentrations of relatively low-frequency genotypes within one or another patch. For example, in the distribution of genotypes for *Acn-1* all three 23 heterozygotes are concentrated into a single patch, which also has a concentration of 12 heterozygotes, suggesting that the founding matriarch may have been a 13 heterozygote (the most common genotype in the north is the 22 homozygote). There was no autocorrelation outside the patches, even for relatively short distances. These results indicate that the demographic patches do indeed represent genetic patches, precisely.

The western section of the island contains the second youngest subpopulation, and compared to the north, the west has higher densities, larger patches, fewer 'loners' and less space between patches. Like the north, large correlations are observed and again these are often odd or nonstandard IBD types, and they also represent concentrations of relatively low-frequency genotypes within one or another patch. For example, for locus *Pgd-1* there was a concentration of 12 heterozygotes in patches where the most common allele was 1, and in general patches contained either only 11 genotypes or approximately equal numbers of 11 and 12 genotypes. In the one case, the matriarch was probably 11 and in the other 12. Moreover there was a remarkable allele frequency difference for this locus, suggesting that the original founder of the western subpopulation may have been a 12 genotype from outside the island. There was some, but little, autocorrelation outside patches, even for relatively short distances. In the west, the demographic patches represented genetic patches almost exactly.

The east is the oldest section of the island population. The autocorrelation results with patch designation revealed some large correlations, and again sometimes odd or nonstandard IBD types. As for the other subpopulations, the results represent collections of relatively low-frequency genotypes within one or another patch. For example for *Pgd-2* there was significant clustering of 23 heterozygotes within patches. However, unlike the north and west, there was also genetic relationships among nearby demographic patches, especially in a certain subregion that happened to be located at the extreme eastern edge of the east subpopulation, where there were, for example, concentrations of 22 homozygotes for *Pgd-2* and 12 heterozygotes for *Acn-1*. Overall, most of the structure is within demographic patches, but not all of it. There was a fair amount of autocorrelation outside patches, even for relatively short distances, although again there was no autocorrelation at larger distances (i.e. beyond distance class 1). Overall, the amount and form of spatial genetic structure was very similar to the younger north and west subpopulations, and the spatial scale was the same. However, in the east, although demographic patches do indeed represent genetic patches, not all relationships were contained within the *a priori* defined demographic patches. There are two factors to consider: (i) possible error in the delineation of demographic patches in the east, and (ii) some changes over time in the degree of the genetic relatedness of the patches, probably caused by pollen flow among adjacent patches. Despite this, the matrilineal nature of patches remains strong and is the main form of spatial structure of genetic variation, presumably because the route of maternal lineages must be through seed, and the recent

dynamics (because density had been high) did not allow new seed to grow anywhere, as available habitat had already been used up.

The kind of structure that is observed currently at all three population stages should persist long into the future, even in the east. In the east, the adults are relatively young and will live for many years, and because available habitat has already been filled, future pollen and seed flow (which is required for the transition to a more standard IBD) can occur only as these die. Moreover, it appears that the east region is at a late successional stage for *S. dioica*, and soon the available habitat will vanish. Thus, it will become increasingly likely that dead *S. dioica* will be replaced by individuals of other species. We think this will cause the genetic structure to continue to be dominated by matrilineages (rather than patrilineages) until *S. dioica* is extinct in this area.

It appears that the form and magnitude of genetic structure observed in all three subpopulations is in fact characteristic of the species. Moreover, it is important to point out that the reproductive contributions of this population to elsewhere on the island, or to new islands, will be dominated by seed produced by this structure, combined with the matings expected based on pollinator movements. Because *S. dioica* is dioecious, this structure should be the main determinant of the mating system. The main difference between *S. dioica* and *Cecropia obtusifolia* is that in *S. dioica* the 'colonizers', which in both cases are perennial maternal groups of seedlings, survive to maturity in higher numbers because the population density is increasing. From our study of *S. dioica*, it appears that, in general, wherever available habitat first rapidly increases then rapidly decreases, which defines demographic parameters, non-standard genetic structure exists on such a time scale that it is the main form of within-population genetic structure for the species.

As noted, there were significant differences in the allele frequencies among the three subpopulations, evidence for three separate founding events of the island (Giles & Goudet 1997). It is important to point out that this, combined with the stability of the demographic patches, has allowed far greater genetic diversity to be maintained on the island in comparison with that expected for a small area under the standard IBD model for species with low levels of dispersal.

Cymbidium goeringii

Cymbidium goeringii (Orchidaceae) is pollinated by bumblebees, and its seeds are small and wind-dispersed. The two study populations (Shinjin and Kwangju) are old, and plant densities are quite low; on average approximately one individual per 6.5 m². Spatial autocorrelation statistics (Moran's I) calculated for isozyme loci (Chung *et al.* 1998)

had either near zero values (at the Kwangju population) indicating a nearly random distribution, or small values (at the Shinjin population) consistent (Epperson *et al.* 1999) with N_e of approximately 120. Pollen flow must be quite limited, implying that seed dispersal is high. It appears that seed flow distances are essentially random across an entire population, because both populations contain only slightly more than 100 individuals. Thus, although the densities are very low and pollen flow is limited, seed flow is apparently large, and the form and magnitude of the spatial genetic structure appears to fit the standard IBD theoretical expectations overall.

Adenophora grandiflora

Another demographic process, clonal reproduction, is thought to have potentially marked effects on the spatial genetic structure of plant populations. For selfing species, clonal groups may increase effective selfing, whereas in dioecious plants any effects on the development of IBD may be more subtle. Other factors may include the spatial distributions of clones and (as in the case of *Silene dioica*) the time scale in which populations exist. Moreover, the demographics of clonal structure may interact with the spatial-temporal demographics of sexual reproduction and the structure of genetic variation. *Adenophora grandiflora* is a monoecious, herbaceous perennial, pollinated by bumblebees, other bees and possibly other pollinators, and it appears to be a predominantly outcrossing species. *Adenophora grandiflora* also reproduces vegetatively, and seeds are reportedly wind-dispersed. An analysis of statistics on the total number of joins between unlike (non-identical) multilocus allozyme genotypes indicated large genotypic spatial autocorrelations – concentrations contained within distances of approximately 15 m (Chung & Epperson 1999) – reflecting the clumping of clones. Indeed, these measures are influenced only by clonal structure in these study populations, because the probability that sexually reproduced individuals have the same multilocus genotype is extremely small. Therefore, pairs of individuals that have the same multilocus genotypes must be clones of each other. It is worth noting that this new way of analyzing the spatial distribution of clones also takes into account the spatial distribution of individuals, unlike simpler statistics based solely on the distances separating pairs of clones.

Clonal effects were separated out by randomly retaining only one individual of each multilocus genotype, and then conducting autocorrelation analysis on the remainder, the so-called 'sexually reproduced' individuals. In both populations, there is significant autocorrelation among the sexually reproduced individuals. Consanguineous matings are caused apparently by limited pollen flow combined with structure, itself caused prob-

ably by both restricted seed flow and restricted clonal 'dispersal' (distances at which clones form or are formed over time). Based on autocorrelations for sexually reproduced individuals (Fig. 2), the predicted values under the standard IBD model correspond to a N_e of about 50. This value is perhaps smaller than expected given that the seeds are wind-dispersed, although pollen flow is probably quite limited. The difference is apparently caused by clonal reproduction. Moreover, a substantial deficit of heterozygotes was observed, compared to Hardy-Weinberg expectations, and this was probably caused by selfing or intracrossing among clones.

Rhus tricocarpha and *Rhus javanica*

It is especially interesting to compare *Adenophora grandiflora* to dioecious species that have substantial clonal reproduction, because effective selfing through intracrossing of clones cannot occur in a dioecious species, and the effects of clonal structure may be more subtle. Both *Rhus tricocarpha* and *Rhus javanica* are dioe-

cious, woody perennials, with insect pollination, and with seed dispersed by gravity and by birds. In two study populations of *R. tricocarpha* in Korea, the frequency of clones was high (51 and 60%). However, contrary to typical expectations clones were often found at long distances from each other, indicating long distances for either primary or secondary clonal reproduction, and there was no autocorrelation among clonal genotypes (Chung *et al.* 1999). Autocorrelation among sexually reproduced individuals was similar to that for the total population. Because, in this case, clonal reproduction did not cause spatial autocorrelation, it cannot contribute to the buildup of IBD. Because there was no autocorrelation among clones, the influence of dioecy could not be determined. However, the effects of population age could be observed. Both populations are very young according to demographic data, existing at the sites for only 26 and 12 years, respectively, and both have some spatial genetic structure. The genetic structure of sexually reproduced trees was greater for the older population, and it appears that both populations are in the early stages of the buildup of IBD, the kind of structure that would be expected for the standard IBD model. However, clonal reproduction should continue to randomize, to some extent, the structure at large spatial scales, because clones spread over large distances. It is also worth noting that in Korea, both *R. tricocarpha* and *R. javanica* are pioneer species that accelerate the establishment of other tree species, which may replace *Rhus* species within a few to several decades.

For two study populations of *R. javanica* (Chung *et al.* 2000), there was no autocorrelation among clonal genotypes for one, but there was slight autocorrelation within 4–6 m for the other. For the first, there was no difference in the structures for the total versus the sexual population (Fig. 3), again because the distribution of clones was random. However, there was substantial structure among sexuals, presumably caused by limited seed and/or pollen flow. For the second population, there was an autocorrelated distribution of clonal genotypes, and this did contribute to total structure. In fact, clones contributed all of the structure, and there was essentially no structure among sexually reproduced individuals. In both populations, clones have not caused the build up of IBD, whereas it appears that limited pollen flow is in the process of creating standard IBD similar to that for *R. tricocarpha*, because the first population is older than the second.

Other species

It is evident from the experimental studies that many aspects of demographic genetic processes may be characterized by variations of correlations at short distances. It is important to point out, however, that some common

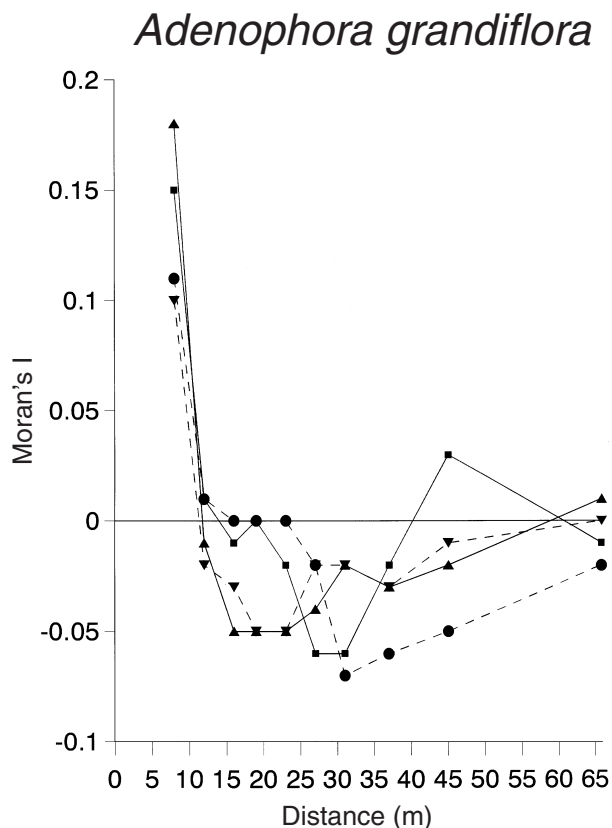


Fig. 2 Average correlogram for Moran's I statistics for two study populations (SEO and ODA) of *Adenophora grandiflora*. Separate analyses were carried out for all genotypes (–tot) and only sexually reproduced genotypes (–sex). ■, SEOtot; ●, SEOsex; ▲, ODAtot; ▼, ODAsex.

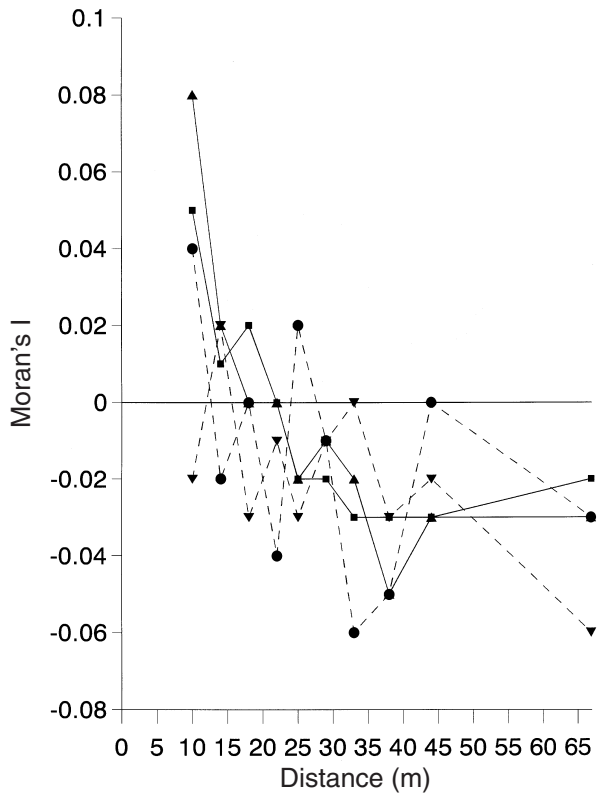


Fig. 3 Average correlogram for Moran's I statistics, for two study populations (SEO and SAN) of *Rhus javanica*. Separate analyses were done for all genotypes (–tot) and only sexually reproduced genotypes (–sex). ■, SEOtot; ●, SANsex; ▲, SANtot; ▼, ODAsex.

observations among many other studies not described in this article do not bear directly on this issue. It has been often observed in other studies that there are relatively large values for the shortest distance class, and little or none for longer-distance classes. This does not necessarily mean that there are only 'short distance' autocorrelations, which would imply either that the population is very young or that the structure is of a non-standard form, possibly because of demographics. Most of the studies do not report the densities, so it is unclear what standardized spatial scale forms the shortest distance class. Moreover, many studies use Wartenberg's SAAP program (Spatial Autocorrelation Analysis Program, Version 4.3; Exter software, Setauket, New York, USA) for converted genotypes, and they may often use the standard option that equalizes the numbers of joins among distance classes, which will generally force the upper bound on the shortest distance class to be large, probably in many cases much larger than necessary, and quite large on the scale of the simulation studies. An appropriate upper bound, which will generally produce good statistical properties yet allow closer observations on the struc-

ture, can be obtained by two rules of thumb: (i) for the upper bound of distance class 1, use the square root of the inverse of the density (or an estimate of the average distance between nearest neighbors) multiplied by $\sqrt{2}$, and (ii) maintain at least 200 or so joins for the first distance class. Compared to the SAAP standard option, the foregoing rule would decrease the upper bound two- to four-fold, which is in the critical range for the capability of characterizing the spatial structure of genetic variation for the most important smaller spatial scales.

General summary

The spatial structure of genetic variation has many substantive effects on the population genetics of plant species. Spatial autocorrelation statistics are useful for describing spatial patterns and the kinship relations and genetic correlations caused by IBD. They can also be used for forming a hypothesis-testing framework for detecting various forms of natural selection, especially by contrasting patterns for different loci. Moreover, they can be used to obtain estimates of the amounts of dispersal. In part, the power of the spatial approach is due to the fact that the spatial patterns may capture the cumulative effects of 50 or more generations. Finally, by carefully examining the shapes of correlograms at the smallest distances, it is possible to differentiate standard IBD processes from those that differ in some cases because of details of demographics, and/or simply a short time since a population founds a site. Combining genetic analyses with information on demographics allows for modeling of the processes that determine the genetic structure of populations. In some cases, but not in others, processes considered to be non-standard and short-term, and therefore perhaps less interesting, may in fact be the dominant form of spatial structure for a species, again because of demographics.

Roles of demographic factors in the formation of standard versus non-standard isolation by distance

In summary, it appears that structure can depend on life stage and whether or not the population is increasing in density. Some key factors seem to be: (i) the degrees to which limits to, and correlation of, seed dispersal produce and maintain substantial spatial structure of genetic variation through groups of kin-related (e.g. maternal half-sib or full sib) individuals, and how the growth of such lineages depends on population regulation, and (ii) the degree and time scale of how this structure is swamped by long-distance pollen dispersal, or the length of time before the structure created by limited seed dispersal is replaced by more standard IBD curves. The amount and

form of genetic variation maintained locally can be affected by demographics. In addition, the effects of clonal reproduction (which can be viewed as a short-term or temporary structure) on the long-term genetic IBD probably depends on: the distances of clonal spread, both primary and secondary; dioecy versus monoecy; and pollen dispersal distances.

References

- Alvarez-Buylla E. R. & Garay A. A. (1994) Population genetic structure of *Cecropia obtusifolia*, a neotropical pioneer tree species. *Evolution* **48**: 437–453.
- Chung M. G., Chung J. M., Chung M. Y. & Epperson B. K. (2000) Spatial distribution of allozyme polymorphisms following clonal and sexual reproduction in populations of *Rhus javanica* (Anacardiaceae). *Heredity* **84**: 178–185.
- Chung M. Y., Chung G. M., Chung M. G. & Epperson B. K. (1998) Spatial genetic structure in populations of *Cymbidium goeringii* (Orchidaceae). *Genes and Genetic Systems* **73**: 281–285.
- Chung J. M., Chung M. G. & Epperson B. K. (1999) Spatial genetic structure of allozyme polymorphisms within populations of *Rhus trichocarpa* (Anacardiaceae). *Silvae Genetica* **48**: 223–226.
- Chung M. G. & Epperson B. K. (1999) Spatial genetic structure of clonal and sexual reproduction in populations of *Adenophora grandiflora* (Campanulaceae). *Evolution* **53**: 1068–1078.
- Cliff A. D. & Ord J. K. (1981) *Spatial Processes*. Pion, London.
- Crawford T. J. (1984) The estimation of neighborhood parameters for plant populations. *Heredity* **52**: 273–283.
- Doligez A., Baril C. & Joly H. I. (1998) Fine-scale spatial genetic structure with nonuniform distribution of individuals. *Genetics* **148**: 905–919.
- Ennos R. A. & Clegg M. T. (1982) Effect of population substructuring on estimates of outcrossing rate in plant populations. *Heredity* **48**: 283–292.
- Epperson B. K. (1990) Spatial autocorrelation of genotypes under directional selection. *Genetics* **124**: 757–771.
- Epperson B. K. (1992) Spatial structure of genetic variation within populations of forest trees. *New Forests* **6**: 257–278.
- Epperson B. K. (1993) Recent advances in correlation studies of spatial patterns of genetic variation. *Evolutionary Biology* **27**: 95–155.
- Epperson B. K. (1995) Spatial structure of two-locus genotypes under isolation by distance. *Genetics* **140**: 365–375.
- Epperson B. K. & Allard R. W. (1984) Allozyme analysis of the mating system in populations of lodgepole pine. *Journal of Heredity* **75**: 212–214.
- Epperson B. K. & Allard R. W. (1989) Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics* **121**: 369–377.
- Epperson B. K. & Alvarez-Buylla E. (1997) Spatial autocorrelation analysis of family structure in multiple life stages of *Cecropia obtusifolia*. *Evolution* **51**: 275–282.
- Epperson B. K. & Clegg M. T. (1986) Spatial autocorrelation analysis of flower color polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). *American Naturalist* **128**: 840–858.
- Epperson B. K. & Clegg M. T. (1987) Frequency-dependent variation for outcrossing rate among flower color morphs of *Ipomoea purpurea*. *Evolution* **41**: 1302–1311.
- Epperson B. K., Huang Z. & Li T.-Q. (1999) Spatial genetic structure of multiallelic loci. *Genetical Research Cambridge* **73**: 2251–2261.
- Epperson B. K. & Li T.-Q. (1996) Measurement of genetic structure within populations using spatial autocorrelations statistics. *Proceedings of the National Academy of Science USA* **93**: 10528–10532.
- Epperson B. K. & Li T.-Q. (1997) Gene dispersal and spatial genetic structure. *Evolution* **51**: 672–681.
- Giles B. E. & Goudet J. (1997) Genetic differentiation in *Silene dioica* metapopulations: Estimation of spatio-temporal effects in a successional plant species. *American Naturalist* **149**: 507–526.
- Ingvarsson P. K. & Giles B. E. (1999) Kin-structured colonization and small-scale genetic differentiation in *Silene dioica*. *Evolution* **53**: 605–611.
- Malécot G. (1948) *Les Mathématiques de L'hérédité*. Masson, Paris.
- Sokal R. R., Jacquez G. M. & Wooten M. C. (1989) Spatial autocorrelation analysis of migration and selection. *Genetics* **121**: 845–855.
- Sokal R. R. & Oden N. L. (1978) Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of Linnean Society* **10**: 199–228.
- Sokal R. R. & Wartenberg D. E. (1983) A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics* **105**: 219–237.
- Wright S. (1943) Isolation by distance. *Genetics* **28**: 114–138.