Molecular Ecology (2008) **17**, 3389–3400 doi: 10.1111/j.1365-294X.2008.03839.x

A heterogeneity test for fine-scale genetic structure

PETER E. SMOUSE,* ROD PEAKALL† and EVA GONZALES*
Department of Ecology, Evolution & Natural Resources, School of Biological and Environmental Science, Rutgers University, New Brunswick, New Jersey 08901-8551, USA, †School of Botany and Zoology, Australian National University, Canberra, ACT 0200, Australia

Abstract

For organisms with limited vagility and/or occupying patchy habitats, we often encounter nonrandom patterns of genetic affinity over relatively small spatial scales, labelled fine-scale genetic structure. Both the extent and decay rate of that pattern can be expected to depend on numerous interesting demographic, ecological, historical, and mating system factors, and it would be useful to be able to compare different situations. There is, however, no heterogeneity test currently available for fine-scale genetic structure that would provide us with any guidance on whether the differences we encounter are statistically credible. Here, we develop a general nonparametric heterogeneity test, elaborating on standard autocorrelation methods for pairs of individuals. We first develop a 'pooled within-population' correlogram, where the distance classes (lags) can be defined as functions of distance. Using that pooled correlogram as our null-hypothesis reference frame, we then develop a heterogeneity test of the autocorrelations among different populations, lag-by-lag. From these single-lag tests, we construct an analogous test of heterogeneity for multilag correlograms. We illustrate with a pair of biological examples, one involving the Australian bush rat, the other involving toadshade trillium. The Australian bush rat has limited vagility, and sometimes occupies patchy habitat. We show that the autocorrelation pattern diverges somewhat between continuous and patchy habitat types. For toadshade trillium, clonal replication in Piedmont populations substantially increases autocorrelation for short lags, but clonal replication is less pronounced in mountain populations. Removal of clonal replicates reduces the autocorrelation for short lags and reverses the sign of the difference between mountain and Piedmont correlograms.

Keywords: fine-scale genetic structure, heterogeneity testing, spatial autocorrelation

Received 17 January 2008; revision received 15 May 2008; accepted 25 May 2008

Introduction

Until the mid-1980s, genetic autocorrelation analysis was used to analyse 'isolation-by-distance' patterns among spatially separated populations, using allele frequencies as the data substrate. The univariate methods that were employed paralleled those used for ecological, anthropological and geographical analyses (see review in Sokal *et al*. 1997). At that point, Sokal and colleagues introduced a multivariate approach, using a Mantel test (Mantel 1967) to compare multi-allele, multilocus genetic distances between pairs of populations with their degrees of geographical separation (Sokal *et al*. 1986), and a variety of multivariate approaches have since become available for genetic analysis. By the mid-1990s, the focus of autocorrelation studies turned to the analysis of 'fine-scale genetic structure' within populations, using multi-allele, multilocus genotypes for individuals (e.g. Heywood 1991; Knowles *et al*. 1992; Berg & Hamrick 1995; Epperson 1995; Loiselle *et al*. 1995; Streiff *et al*. 1998; Hardy & Vekemans 1999; Smouse & Peakall 1999; Rousset 2000; Vekemans & Hardy 2004).

Fine-scale genetic structure is particularly interesting because the details of the mating system (Van Rossum & Triest 2007), propagule dispersal (Peakall *et al*. 2003; Cruse-Sanders & Hamrick 2004; Double *et al*. 2005; Temple *et al*. 2006), as well as ecological (Parker *et al*. 2001; Van Rossum *et al*. 2004; Premoli & Kitzberger 2005) factors vary on a

Correspondence: Peter E. Smouse, Fax: +732 932 1126; Email: smouse@aesop.rutgers.edu

microgeographical scale and affect the pattern of genetic structure. Different situations can be expected to yield different outcomes, so comparative analysis becomes important, but to date, comparisons have been more verbal than formal.

Autocorrelation analysis is also important in arenas other than population genetics. Just by way of recent example, autocorrelation analysis has been used to characterize neighbourhood patterns in health status (Auchincloss *et al*. 2007), patterns of taxonomic composition in community ecology (Bacaro & Ricotta 2007), spatial patterns in soil properties, as that relates to shifting cultivation practices in tropical forests (Diekmann *et al*. 2007), near neighbour effects on individual-tree growth models in forestry (Fox *et al*. 2007), fine-scale variation of foliar defence chemicals in Eucalyptus trees (Andrew *et al*. 2007), and on a very tiny scale, the pattern of spatial biases in micro-array experiments (Koren *et al*. 2007). Context is important to pattern in all of these cases, and the expected decay pattern of autocorrelation with distance differs for such examples. The need for a general nonparametric comparative method is growing.

There is, however, currently no formal test of autocorrelational heterogeneity available that would tell us whether two (or more) sampled patterns are credibly divergent or not. We have two objectives here. (i) We first develop a general nonparametric heterogeneity test, based on elaborations of the autocorrelation method of Smouse & Peakall (1999), which can be used for a wide variety of comparative analyses of fine-scale structure, both those from genetics and those from other arenas. (ii) We then illustrate it with two examples, one animal (Australian bush rat) and one plant (toadshade trillium). The Australian bush rat example involves differences in fine-scale structure associated with fragmentation of a previously continuous habitat, for an organism that is dispersal-restricted. The toadshade trillium example involves a case where divergent fine-scale structures are associated with differential vegetative cloning in two ecologically distinct habitat zones. In both cases, we are able to establish that apparent differences in fine scale structure are statistically credible and biologically meaningful.

Extracting autocorrelation

Our first task is to reprise the essentials of the single-population autocorrelation analysis of genetic fine-scale structure (Smouse & Peakall 1999), generalizing the notation in a way that simplifies the extension to multiple populations. The autocorrelation analysis of Smouse & Peakall (1999) is available in GENALEX 6 (Peakall & Smouse 2006). The procedures described here are freely available in GENALEX 6.2 (Australian National University, Canberra, Australia, www.anu.edu.au/BoZo/GenAlEx/). Briefly, we begin with a single population of N individuals, for each of which we have a battery of L genetic loci. The species can be either haploid or diploid. If haploid, then at any particular locus, we score two individuals as having distance $\delta^2_{ij} = 1$ if they have different alleles and as having distance $\delta^2_{ij} = 0$ if they have the same allele. Alleles are treated as the same or different. This is the Manhattan scoring convention described in Excoffier *et al*. (1992). The genetic distance between the ith and jth individuals is obtained by adding 1's and 0's across the L genetic loci,

$$d_{ij} = \sum_{l=1}^{L} \delta^2_{ij}(l). \qquad \text{(eqn 1)}$$

If the species is diploid, we use genetic distances ranging from $\delta^2_{ij} = 0 - 4$ for each locus, depending on the particular pair of diploid genotypes being contrasted, a Euclidean scoring convention described in detail by Peakall *et al*. (1995) and Smouse & Peakall (1999). The metric is summed across loci, as in eqn (1).

Our particular interest here is in genetic analysis, and both of these measures weight all alleles and loci equally, which we have found to be a robust treatment, but the reader interested in weighting inversely by allele frequencies (in typically multinomial fashion) is referred to Smouse & Peakall (1999) for details. More generally, and of particular relevance for nongenetic data and problems, any multivariate Euclidean distance measure that closes properly in geometric space will suffice for what follows.

Distance and covariance matrices

Having constructed distances between all $N \cdot (N-1)/2$ pairs of individuals, we construct a symmetric genetic distance matrix $\mathbf{D} = \{d_{ij}\}$ of dimension $N \times N$, with '0's down the diagonal and with $d_{ij} = d_{ji}$ in the off-diagonal positions. Gower (1966) showed that there is a one-to-one translation between this $N \times N$ distance matrix (**D**) and a corresponding $N \times N$ genetic variance–covariance matrix (**S**). A diagonal element $\{s_{ii}\}$ of **S** is the 'genetic variance' of the ith individual, its squared deviation from the centre of the gene pool, whereas an off-diagonal element $\{s_{ij} = s_{ji}\}$ of **S** is the 'genetic covariance' between the ith and jth individuals, the pairwise cross-product of those deviations from the centre of the gene pool. Translation of **D** into **S** is described by Smouse & Peakall (1999, following Gower 1966). The important point is that $\mathbf{D} \leftrightarrow \mathbf{S}$; they contain the same information.

We could (in principle) translate the variance–covariance matrix **S** into a correlation matrix $\mathbf{R} = \{r_{ij}\}$, with r_{ij} defined as the genetic correlation between the ith and jth individuals. If we construct a vector from the genetic centre of the population to the ith individual and another to the jth individual, those two vectors are offset by an angle (θ_{ij}).

The correlation $r_{ij} = \cos(\theta_{ij})$ is a scale-free measure of genetic affinity between that pair of individuals, closely related to 'kinship' (Loiselle *et al*. 1995) and 'relationship' (Streiff *et al*. 1998) coefficients. The whole focus of examining genetic fine-scale structure is to determine the decay of genetic affinity/relationship with spatial separation. These interindividual correlations have large variances, of course, because each represents a single pair of genotypes, and they are statistically imprecise. Standard practice in correlation analysis is to use several individuals to calculate average correlation coefficients, and we follow that practice here.

Extracting autocorrelations from the S-matrix

By virtue of 'centring' the **S**-matrix for the population, the sum of all elements in the matrix is zero, as are each of the row (and column) sums. Denoting the diagonal array of variances as $\mathbf{V} = \text{diag}\{\mathbf{S}\}$ and the off-diagonal covariances as $\mathbf{C} = \mathbf{S} - \mathbf{V}$, we can partition **S** into separate variance and covariance matrices, $\mathbf{S} = \mathbf{C} + \mathbf{V}$, which guarantees that sum **C** = –sum **V** = –Tr **V**, inasmuch as **V** is diagonal. With this notation, we can denote the average correlation involving all pairwise comparisons in **S** as

$$r^{(0)} = \frac{\text{sum } \mathbf{C}}{(N-1)\cdot \text{Tr } \mathbf{V}} = \frac{-\text{Tr } \mathbf{V}}{(N-1)\cdot \text{Tr } \mathbf{V}} = \frac{-1}{(N-1)}. \qquad \text{(eqn 2)}$$

The $(N-1)$ multiplier in the denominator derives from each individual being compared with $(N-1)$ other individuals in the collection. By virtue of geometric closure, the average interindividual correlation of the entire data set is $-(N-1)^{-1}$. Our standard null hypothesis is that the average correlation of the entire data set, $r^{(0)} = 0$, so it is convenient to reset the reference at $r^{(0)} = 0$ by adding $+(N-1)^{-1}$ to eqn (2), and that is now general practice.

Lag matrices

The practice in spatial autocorrelation analysis is to allocate particular pairs of individuals to one of a series of distance classes (called 'lags' in Geographic Information Systems literature). One usually defines distance classes (lags) in terms of the physical separation between pairs of individuals, say 0–5 m, 5–10 m, 10–15 m, etc., or as logarithmic distance classes, say 0–1 m, 1–2 m, 2–4 m, etc. (e.g. Rousset 2000; Vekemans & Hardy 2004), but any scheme that makes biological sense for the problem at hand is acceptable. It is important for comparative analysis to define the lags in a comparable way for each of the populations, and in such a way that each of the lags possesses pairs in each of the populations to be compared.

Now, we place each pair of individuals in one of H lags, numbered from closest to most distant $\{h = 1, \ldots, H\}$. We construct an $N \times N$ lag matrix, $\mathbf{X}^{(h)}$ for the pairs of the hth lag, which has $x_{ij} = 1 = x_{ji}$ if the physical distance between the ith and jth individuals is within the hth lag, and = 0 otherwise. The diagonal element x_{ii} is the number of times for which the ith individual is paired with another individual within the hth lag, i.e. the number of off-diagonal '1's in the ith row/column. For example, if the 1st individual is h lags away from the 3rd, 5th and 8th individuals, then $x_{13} = 1 = x_{31}$, $x_{15} = 1 = x_{51}$, $x_{18} = 1 = x_{81}$, and $x_{11} = 3$. All other positions in the 1st row/column have 0's. Each pair of individuals is tallied in only one lag-matrix, $\mathbf{X}^{(h)}$, but any particular individual may be paired with different partners for different lags. The matrix $\mathbf{X}^{(h)}$ tells us which elements to count for the hth lag and weights them by the number of pairs to which they contribute. For the hth lag, we include the ijth pair in the correlation argument only if $x_{ij}^{(h)} = 1$. Formally, we compute lag-specific covariance and variance terms,

$$\varepsilon_{ij}^{(h)} = x_{ij}^{(h)} \cdot s_{ij} \qquad \text{and} \qquad f_{ii}^{(h)} = x_{ii}^{(h)} \cdot s_{ii}. \qquad \text{(eqn 3)}$$

We define a covariance matrix for the hth lag as $\mathbf{E}^{(h)} = \mathbf{C} \otimes \mathbf{X}^{(h)} = \{\varepsilon_{ij}^{(h)}\}$, the inner product of $\mathbf{X}^{(h)}$ and **C**, and also the diagonal variance matrix as $\mathbf{F}^{(h)} = \mathbf{V} \otimes \mathbf{X}^{(h)} = \{f_{ii}^{(h)}\}$. We then construct

$$r^{(h)} = \frac{\text{sum } \mathbf{E}^{(h)}}{\text{Tr } \mathbf{F}^{(h)}} + \frac{1}{(N-1)}, \qquad \text{(eqn 4)}$$

with sum $\mathbf{E}^{(h)}$ and Tr $\mathbf{F}^{(h)}$ computed as the sum of terms from eqn (3). The final term once again contains a bias-adjustment to ensure that the null hypothesis value is $r^{(h)} = 0$. We repeat this procedure for each of our H lags, one at a time. For the special case where all pairs of individuals are compared (with the whole population considered as the 0th lag), sum $\mathbf{E}^{(0)}$ = sum **C** and Tr $\mathbf{F}^{(0)} = (N-1) \cdot$ Tr **V**, which reduces eqn (4) to zero for $h = 0$, the null hypothesis (reference) condition.

Testing autocorrelation

We have a set of N individual genotypes, distributed over space but within the confines of a single population. The null hypothesis is that autocorrelation for each lag is zero, barring sampling variation. If the organism in question exhibits restricted dispersal (relative to the spatial scale of sampling), we often discover that genetically similar individuals are physically clustered over a fine-scale landscape, which generally translates into $r > 0$ for small lags and $r < 0$ for large lags. We anticipate declining correlations with increasing lag if the null hypothesis is not correct. The first question is whether there is any credible evidence that $r^{(h)} \neq 0$ for the hth lag. If we permute the N genotypes randomly and without replacement among sample locations, then for any particular lag, we anticipate

a departure from $r^{(h)} = 0$, though it represents the sampling noise of random genetic placement on our fine-scale landscape. Repeating the permutation procedure, we construct an empiric null distribution for each lag, conditioned on the total collection of sampled map coordinates and recovered genotypes. The variance in the outcome is inversely proportional to the sample size (number of pairs) for any particular lag, and there is little to be gained by defining lags with small sample sizes. The short distance lags of particular interest often suffer from this limitation, so it is common to combine the shortest lags to increase the numbers of pairs (replication) within them.

For each permutation, we compute the set of $r^{(h)}$-estimates, one per lag, as in eqn (4). We permute the data (say) 999 times, and for each permutation, we compute a complete battery of $r^{(h)}$-estimates. We add the realized values from the actual data as a final (1000th) trial, on the (null hypothesis) premise that it too is a spatially randomized genetic collection. Any tendency for a deviation of the data from $r^{(h)} = 0$ is treated as due to random noise, because the null hypothesis is that genotypes are locationally exchangeable. We compute a histogram of random r values, one lag at a time, and compare the observed value with that null distribution. We determine the ranking of the observed $r^{(h)}$-value in the listing of 1000 random $r^{(h)}$-trial values and extract the p value empirically. Where fine-scale structure exists, we typically see $r^{(h)} > 0$ for small lags and $r^{(h)} < 0$ for large lags. Since the average, $r^{(0)} = 0$, some values of $r^{(h)}$ must be negative if others are positive, so we compute:

$$p_k^{(h)} = \Pr[r_{\text{perm}}^{(h)} \geq r_{\text{data}}^{(h)}] \text{ for } r_{\text{data}}^{(h)} > 0 \text{ and}$$
$$p_k^{(h)} = \Pr[r_{\text{perm}}^{(h)} \leq r_{\text{data}}^{(h)}] \text{ for } r_{\text{data}}^{(h)} < 0. \qquad \text{(eqn 5)}$$

We also need a test of whether the collection of observed $r^{(h)}$ values is compatible with the 'no autocorrelation at any distance' hypothesis (H_0: $r^{(1)} = r^{(2)} = \ldots = r^{(H)} = 0$). Attached to each $r^{(h)}$ estimate is an empiric p value (one-tailed probability), so we have constructed a matrix $\mathbf{P} = \{p\text{-values}\}$, of dimensions $1000 \times H$ (one row for each of 1000 data permutations, and H columns — one per lag). For the kth row of the **P**-matrix, representing the kth randomization, we compute a combined probability metric ω (Fisher 1958),

$$\omega_k = -2 \cdot \sum_{h=1}^{H} \log_e(p_k^{(h)}). \qquad \text{(eqn 6)}$$

Fisher's test criterion (eqn 6) is usually compared with a χ^2 (2H d.f.), but in the spirit of nonparametric analysis, we rank 1000 ω values and estimate $\Pr(\omega_k \geq \omega_{\text{data}})$ empirically, yielding a test of the hypothesis that the entire correlogram is 'flat' against the alternative that it is not.

Heterogeneous autocorrelation

Rationale

Correlograms are not invariably homogeneous from one population to the next, and there are many situations where differences in ecology, land management, or breeding system may yield differential spatial patterns (cf. Peakall *et al*. 2003; Double *et al*. 2005). The operative question is whether we can establish that the differences we encounter are biologically and statistically credible, and not just sampling artefacts. We need a heterogeneity test for autocorrelograms.

Method

Consider a series of populations, indexed by $\pi = \alpha, \beta, \ldots$, within which we have $N_\alpha, N_\beta, \ldots, N$ sampled individuals, respectively. We assume that the same genetic loci are used for each population, that the lags are defined in comparable fashion, and that each lag has at least one pair (preferably several) in each population. We compute a set of autocorrelation coefficients for each lag within each population,

$$r_\pi^{(h)} = \frac{\text{sum } \mathbf{E}_\pi^{(h)}}{\text{Tr } \mathbf{F}_\pi^{(h)}} + \frac{1}{(N_\pi - 1)} \text{ for } h = 0, 1, \ldots, H \text{ and } \pi = \alpha, \beta, \ldots, \quad . \qquad \text{(eqn 7)}$$

We also construct a set of pooled within-population estimates, representing the null hypothesis reference frame for autocorrelational homogeneity,

$$r_\bullet^{(h)} = \frac{\sum_{\pi=\alpha} \text{sum } \mathbf{E}_\pi^{(h)}}{\sum_{\pi=\alpha} \text{Tr } \mathbf{F}_\pi^{(h)}} + \frac{1}{(\tilde{N} - 1)} \text{ for } h = 1, \ldots, H. \qquad \text{(eqn 8)}$$

The sample sizes for the different populations are typically not the same, so to ensure that $r_\bullet^{(0)} = 0$ for the entire collection of individuals, we define a weighted average bias correction as

$$\frac{1}{(\tilde{N} - 1)} = \frac{\sum_{\pi=\alpha} \text{Tr } \mathbf{V}_\pi}{\sum_{\pi=\alpha} (N_\pi - 1) \cdot \text{Tr } \mathbf{V}_\pi}. \qquad \text{(eqn 9)}$$

We use eqn (7) to estimate the within-population $r^{(h)}$ values for each of the separate populations and eqn (8) for the null-hypothesis (average) reference. The separate populations will have an array of values that will bracket the average, but the question is 'Are the observed $r^{(h)}$ values from different populations compatible with random sampling from the null hypothesis distribution, or are they not compatible?' We evaluate the distribution of random

departures from the null average by bootstrap resampling, drawing paired samples of the sizes actually encountered for each lag and population. For the hth lag, we have $n_{\pi}^{(h)}$ pairs of individuals from within the πth collection and $n_{\bullet}^{(h)}$ pairs from within the total collection. We randomly draw $n_{\pi}^{(h)}$ pairs from the total collection of hth lag pairs and place it in the πth collection. We repeat this process for each of the $n_{\bullet}^{(h)}$ pairs, sampling with replacement. That procedure ensures that the sampling frame is preserved, i.e. that the $n_{\bullet}^{(h)}$ are preserved. For each random trial, we recompute eqn (8) and tally the results. We repeat this bootstrap resampling process for 999 trials and record the complete set of measured $r^{(h)}$ values, trial by trial, adding the observed values as a final (1000th) trial.

We now have 1000 trial $r^{(h)}$ values for each of populations. We convert the r values to z values (Fisher 1958),

$$z^{(h)} = \frac{1}{2} \cdot \ln\left[\frac{1 + r^{(h)}}{1 - r^{(h)}}\right]. \qquad \text{(eqn 10)}$$

Recalling that $(-1 \le r^{(h)} \le +1)$, that $(-\infty < z^{(h)} < +\infty)$ and that r and z are monotonic, we note that rank order is preserved by this normalizing transformation. We compute the square of a paired-sample t-test, which [suppressing the (h) notation] takes the value

$$t^2_{\alpha\beta} = \frac{(z_\alpha - z_\beta)^2}{\mathrm{Var}\,(z_\alpha - z_\beta)}. \qquad \text{(eqn 11)}$$

For each of the random trials, we compute r_α and r_β, convert them to z_α and z_β, and then compute $(z_\alpha - z_\beta)^2$. We compute an empiric variance from the random trials, $V(z_\alpha - z_\beta)$, and use it for the denominator of eqn (11). We compute eqn (11) for each trial, and then rank our data result, thereby extracting an estimate of an upper tail probability (p value). We have a nonparametric test of homogeneity for the hth lag of the αth and βth collections. Computing a separate $t^2_{\alpha\beta}$ test for each of H lags, we have p values that can be combined into a multilag ω-comparison of populations α and β, via eqn (6), a formal test of heterogeneous autocorrelograms.

Multipopulation testing

It will be useful to have a test of correlational heterogeneity among the entire set of populations. We pack the ω_{ij}-criteria into a matrix $\mathbf{\Omega}$ of the form,

$$\mathbf{\Omega} = \begin{vmatrix} 0 & \omega_{\alpha\beta} & \omega_{\alpha\gamma} & \dots & \omega_{\alpha} \\ \omega_{\beta\alpha} & 0 & \omega_{\beta\gamma} & \dots & \omega_{\beta} \\ \omega_{\gamma\alpha} & \omega_{\gamma\beta} & 0 & \dots & \omega_{\gamma} \\ \dots & \dots & \dots & \dots & \dots \\ \omega_{\alpha} & \omega_{\beta} & \omega_{\gamma} & \dots & 0 \end{vmatrix} \qquad \text{(eqn 12)}$$

For each of our 1000 trials, we have a random ω_{ij} value. We rank them and determine the fraction of trials that have ω_{ij} at least as large as the data result, $p_{ij} = \Pr(\omega_{ij\text{-rand}} \ge \omega_{ij\text{-data}})$. Each element of the matrix $\mathbf{\Omega}$ thus has an associated p value. We use the matrix $\mathbf{\Omega}$ to construct an overall test of multipopulation divergence in whole correlograms. We compute an 'among-populations' test of correlogram homogeneity,

$$\omega_{\mathrm{AP(data)}} = \frac{\mathrm{sum}\,\mathbf{\Omega}_{\mathrm{data}}}{2}. \qquad \text{(eqn 13)}$$

For statistical testing, we also compute eqn (13) for each of the random trials. Finally, we determine the fraction of random trials yielding at least as large an ω_{AP} value as the data,

$$p_{\mathrm{AP}} = \Pr(\omega_{\mathrm{AP(rand)}} \ge \omega_{\mathrm{AP(data)}}), \qquad \text{(eqn 14)}$$

our test of the null hypothesis that the correlograms from all populations are homogeneous.

Australian bush rats

The Australian bush rat (*Rattus fuscipes*) is a common terrestrial rodent of coastal southern Australia. Individuals live for 12–15 months, and there is substantial demographic turnover from year to year (Taylor & Calaby 1988). While little is known about the distance moved by dispersing individuals or the proportions of individuals involved in each dispersal phase, there is some evidence that most dispersal movements are in the range of 200–400 m, with males dispersing more than females (see review in Peakall *et al*. 2003). Home ranges overlap, and densities can exceed 10 animals/ha (Peakall & Lindenmayer 2006).

The genetic data to be used here were drawn from Peakall *et al*. (2003; Peakall & Lindenmayer 2006). These studies form part of the Tumut fragmentation project — a large landscape study of habitat fragmentation, established by Lindenmayer *et al*. (1999). The study area is characterized by a series of remnant native eucalypt forest patches, embedded within a 50 000-hectare exotic conifer plantation, bounded by extensive native eucalypt forest. Bush rats are patchily distributed along moist gullies and creeklines with well-developed native vegetation, both within the eucalypt fragments embedded in the unsuitable conifer plantation and within the extensive eucalypt forests surrounding the plantation.

Peakall *et al*. (2003) concluded that the level of gene flow was sufficiently restricted in bush rats to generate the strong positive signal of local spatial genetic structure they detected at Tumut. More recent evidence indicates that the movements of bush rats may be more restricted in this region than reported in other studies, with mean

Table 1 Autocorrelation *r* values and *p* values and numbers of pairs (*n*) for the Australian bush rat (*Rattus fuscipes*) for each of eight 50-m distance classes and a multiclass test criterion (ω) of the departure from the null hypothesis of $r = 0$; the test is conducted for each of the four locations (T3U3, 1875, CAM and MCMD) separately; populations α = T3U3 and β = 1875 occupy *Eucalyptus* fragments, while populations γ = CAM and δ = MCMD are from continuous *Eucalyptus* forest

Distance class	1	2	3	4	5	6	7	9	ω-test	Estimated
Interval (m)	0–50	51–100	101–150	151–200	201–250	251–300	301–350	351–400	Criterion	*p* value
	$r = 0.103$	0.051	−0.001	0.033	0.006	0.012	−0.030	−0.034		
α = T3U3	$n = 187$	177	131	104	114	119	120	87	52.67	0.001
	$p = 0.001$	0.001	0.489	0.027	0.379	0.230	0.037	0.047		
	0.065	0.021	0.002	−0.013	0.007	−0.009	−0.010	−0.031		
β = 1875	159	217	163	180	302	281	182	152	44.49	0.001
	0.001	0.028	0.428	0.145	0.198	0.173	0.208	0.016		
	0.143	0.017	0.031	−0.014	−0.017	−0.038	−0.005	0.021		
γ = CAM	169	104	99	83	85	97	123	64	43.23	0.001
	0.001	0.062	0.054	0.251	0.190	0.014	0.372	0.187		
	0.222	0.157	0.081	0.040	−0.027	−0.016	−0.062	−0.021		
δ = MCMD	113	84	108	85	92	103	121	143	75.26	0.001
	0.001	0.001	0.001	0.036	0.094	0.202	0.001	0.091		

mark–recapture movement distances (tagged animals) of 35 m and a maximum of 280 m, far short of the maximum distance between traps of up to 1000 m (Peakall *et al*. 2003). The question of interest here is whether the restricted dispersal and associated fine-scale genetic structure is more pronounced at sites within fragmented remnant patches than at sites within continuous forest, and if so, whether this pattern is a consequence of dispersal limitations under fragmented habitat conditions?

For illustrative purposes, we have selected a quartet of population samples: two from isolated eucalypt fragments within the pine plantation (α = T3U3 and β = 1875) and two from extensive forest habitats (γ = CAM and δ = MCMD). We sampled with traps, spaced every 10 m along normal gully habitat, for each of the four sites. Each site was 10–50 m in width, with two trapping transects of length 800–1000 m, and the sample sites were 5–10 km apart. Rats were sexed, aged, and subsequently genotyped for a set of seven nuclear STR loci (C2, E5, FG, CR, PB, TT and PL; see Peakall *et al*. 2003). For illustrative purposes here, we ignore differences in sex and age class, yielding sample sizes of $N_\alpha = 77$, $N_\beta = 65$, $N_\gamma = 63$, and $N_\delta = 60$, respectively. The microhabitat preferences of this organism are such that the habitat is basically one-dimensional, so we expect the autocorrelation to decline more or less linearly with distance. We defined eight lags, of equal distance width, 50 m each, as per Peakall *et al*. (2003). The story unfolds over about 400 m.

The driving question is whether the fine-scale patterns from continuous forest sites (γ and δ), with no habitat impediments to extensive dispersal, differ from those of the forest fragments (α and β), between which intersite migration involves the traverse of large stretches of unfavourable (exotic conifer forest) habitat. The first step is to assess the pattern of autocorrelation within each of the separate populations, and the separate analyses of the four populations yield the results in Table 1 and Fig. 1. Autocorrelation is positive and significant for the 1st lag (0–50 m) in all four populations, and always positive and generally significant for the 2nd lag (51–100 m). With these distance class sizes, the correlogram appears to cross the *X*-axis between 120 and 210 m, indicating that proximal pairs are (on average) more related than are spatially random pairs, while more distant pairs are (on average) less related than are random pairs. Peakall *et al*. (2003) showed that the correlograms and crossing points are somewhat dependent on the widths of the distance classes, but the essential point of positive autocorrelation at short distance is quite general for this organism.

The next step was to determine whether the four populations exhibited heterogeneous autocorrelation patterns. Using the testing procedure described above, we obtained the results in Table 2, from which we discover that the two fragment sites (α, β) are not distinguishable, but that both are significantly different from the first continuous forest site (γ), mostly due to modest correlational differences in the first and last distance classes. Interestingly, all three of these sites are easily distinguishable from the second continuous forest site (δ) over several lags. The pronounced fine-scale structure of MCMD was counter to prior expectation. Post-hoc, we are unable to explain the finding of very much stronger structure within the MCMD continuous site, but the results suggest that sensitivity of fine-scale structure to habitat fragmentation per se may not be the only operative factor. The overall findings of positive local spatial genetic structure for all populations reinforce the

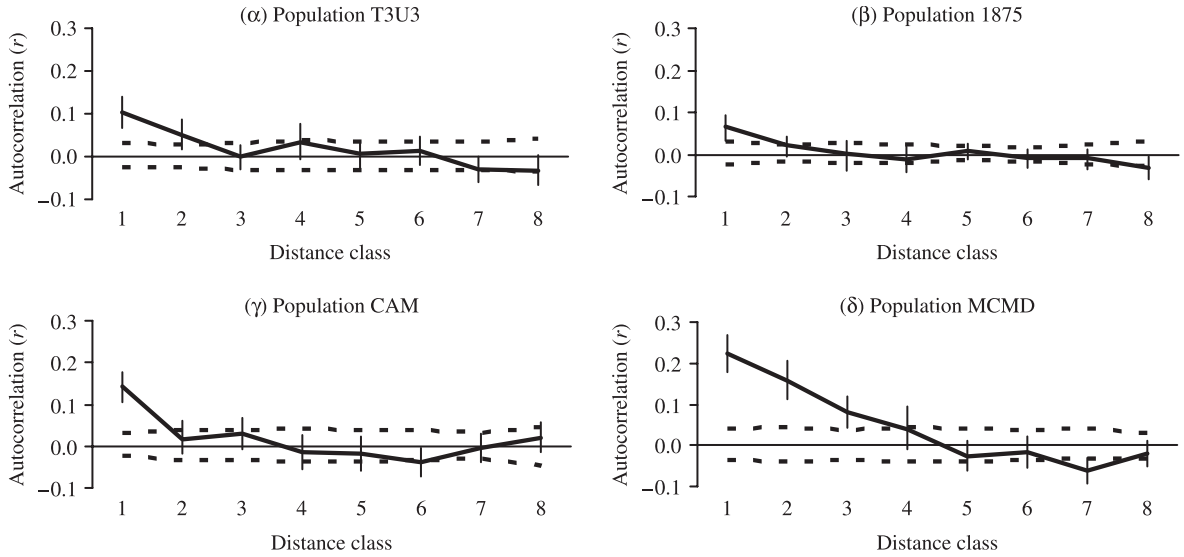

Fig. 1 Fine-scale correlograms for Australian bush rat (*Rattus fuscipes*) populations in fragmented habitat (α = T3U3 and β = 1875) and populations in continuous habitat (γ = CAM and δ = MCMD).

Table 2 Single-class (t^2) and multiclass (ω) test criteria and associated *p* values for six paired-population comparisons of correlogram homogeneity, involving Australian bush rats (*Rattus fuscipes*); two populations (α = T3U3 and β = 1875) are from *Eucalyptus* fragments; two populations (γ = CAM and δ = MCMD) are from continuous *Eucalyptus* forest

Distance class	1	2	3	4	5	6	7	8	ω-test	Estimated
Interval (m)	0–50	51–100	101–150	151–200	201–250	251–300	301–350	351–400	Criterion	*p* value
α vs. β	t^2 = 2.10	1.57	0.01	3.53	0.00+	1.15	0.94	0.01	17.92	0.323
	p = 0.145	0.212	0.912	0.055	0.951	0.284	0.332	0.927		
α vs. γ	2.25	1.40	1.76	2.52	0.85	3.62	1.20	3.85	31.58	0.006
	0.138	0.227	0.181	0.110	0.363	0.051	0.262	0.046		
β vs. γ	8.07	0.02	1.62	0.00+	1.42	1.73	0.06	4.13	26.80	0.036
	0.005	0.888	0.205	0.976	0.252	0.189	0.814	0.044		
α vs. δ	15.05	12.90	10.71	0.06	1.68	1.28	1.97	0.24	51.17	0.001
	0.001	0.001	0.002	0.809	0.190	0.256	0.159	0.617		
β vs. δ	26.14	23.48	10.85	4.21	2.60	0.10	6.48	0.27	60.95	0.001
	0.001	0.001	0.002	0.051	0.106	0.747	0.012	0.601		
γ vs. δ	6.76	18.41	3.62	3.16	0.14	0.68	6.52	2.57	50.59	0.001
	0.007	0.001	0.057	0.073	0.716	0.411	0.012	0.101		

conclusion of Peakall & Lindenmayer (2006) that as a common and widespread species, bush rats may not need well-developed dispersal capability in natural landscapes, but that they may be at risk of irreversible local extinctions in fragmented landscapes, if limited dispersal capability results in a failure to recolonize isolated and depopulated fragments.

Toadshade trillium

Our second case study involves a rhizomatous woodland wildflower, toadshade trillium (*Trillium cuneatum*), a long-lived spring ephemeral that carpets the floors of mature mesic deciduous forests in southeastern North America, both in large continuous and fragmented remnant habitats. This long-lived species exhibits a leaky self-incompatibility system with a mixed reproductive strategy. Both sexual and vegetative reproduction occur, with the balance varying among populations and habitats. Toadshade trillium is insect-pollinated, and pollen is seldom moved more than a few metres (Gonzales *et al*. 2006), and the seeds are gravity and ant-dispersed over very short distances. We have also shown that vegetative reproduction is sometimes pronounced, but that it varies among locations (Gonzales *et al*. in press).

Table 3 An autocorrelation analysis of fine-scale genetic structure for *Trillium cuneatum*: (a) Montane (JK) and Piedmont (EM) sites from Georgia, with *r* values and *p* values for each of eight distance classes and a global test of departure from the null hypothesis of $r = 0$, including tightly clumped clonal ramets (α and β); and (b) the same pair of sites with all clonal copies (except one) removed for each genet (γ and δ)

Distance class	1	2	3	4	5	6	7	8	ω -test	Estimated
Interval (cm)	0–10	11–20	21–40	41–80	81–160	161–320	321–640	641–1280	Criterion	*p* value
α = JK with clones	$r = 0.355$	0.290	0.220	0.183	0.133	0.022	−0.010	−0.038	110.52	0.001
	$n = 56$	123	310	587	1408	4178	10376	9339		
	$p = 0.001$	0.001	0.001	0.001	0.001	0.001	0.001	0.001		
β = EM with clones	0.876	0.663	0.162	0.113	−0.011	−0.006	−0.025	−0.051	91.17	0.001
	515	262	502	677	1467	4016	12702	8539		
	0.001	0.001	0.001	0.001	0.140	0.114	0.001	0.001		
γ = JK no clones	0.337	0.284	0.222	0.193	0.089	0.027	−0.009	−0.036	110.52	0.001
	44	101	268	487	1103	3407	8862	7934		
	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001		
δ = EM no clones	0.245	0.199	0.084	0.077	−0.003	0.036	−0.012	−0.022	97.14	0.001
	76	78	242	363	686	1254	5153	2879		
	0.001	0.001	0.001	0.001	0.402	0.001	0.002	0.001		

Work on another species of *Trillium* (Yamagishi *et al*. 2007) and on clonal American ginseng (e.g. Chung & Epperson 1999) has shown that fine-scale genetic structure is sensitive to habitat disruption. Here, we compare fine-scale structure in a relatively undisturbed, continuous population of *T. cuneatum* at high elevation in the Southern Appalachian Mountains with that from a small, fragmented, and heavily disturbed urban population from the Piedmont of Georgia, USA. The mountain population occupies a site in the wilderness area of the Joyce Kilmer Memorial Forest (JK) in western/mountainous North Carolina, a cool, mesic, and environmentally stable habitat. The population consists of tens of thousands of patchily distributed individuals, spread over hundreds of hectares of continuous hardwood forest. The Piedmont population (EM) occupies a small forest fragment near Emory University in Atlanta, completely surrounded by urban development, subject to ongoing human disturbance and to strong edge effects. The site is hotter and subject to periodic drought. Although this population occupies a small area, the *Trillium* population contains several thousand individuals. Descriptions of contrasting historical and environmental details for these sites are available in Gonzales *et al*. (2006, in press). Our prime interest here is the fact that vegetative spread is quite rare at the mountain (JK) site but frequent at the Piedmont (EM) site. There are two questions: (i) What are the effects of clonal spread on fine-scale genetic structure in these different habitats? (ii) How is clonality and its associated fine-scale structure influenced by different histories of disturbance and environmental conditions?

For illustration, we have selected a single plot from within each population, yielding $N_{JK} = 231$ and $N_{EM} = 240$. We conducted allozyme analyses (9 loci) for all mapped plants (Gonzales *et al*. 2006, in press). Within each population, we analysed spatial autocorrelation in two fashions: (i) with all plants (including clonal ramets), and (ii) after removing clonal replicates (using independent sexual genets). After removing the clonal replicates, the sample sizes were $N_{JK} = 212$ and $N_{EM} = 147$. These populations are spread out across a two-dimensional landscape, for which theory (Rousset 2000) predicts linearly declining genetic affinity with the logarithm of physical distance (over an intermediate range of distances), so we defined distance classes (lags) on a logarithmic scale and held them constant for all analyses.

Fine-scale genetic structure is pronounced for both mountain (JK) and Piedmont (EM) populations, either with or without the clonal replicates (Table 3). The sample sizes are large, and the results are statistically compelling, lag-by-lag and for the whole correlogram. Piedmont populations have higher rates of clonal replication than mountain populations (Gonzales *et al*. in press), and the EM correlogram declines more steeply than that for JK (Fig. 2α vs. 2β). While the observed genetic relatedness among near neighbours in EM ($r_{EM}^{(1)} = 0.876$) is much higher than in JK ($r_{JK}^{(1)} = 0.355$), the difference is smaller after clonal ramets are removed ($r_{EM}^{(1)} = 0.245$ *vs.* $r_{JK}^{(1)} = 0.337$), and the exclusion of the clonal replicates reverses the sign of the difference (Table 3). The JK correlogram (minimally affected by removing a small number of clonal replicates) now exhibits higher values than the EM correlogram (Fig. 2γ vs. 2δ, Table 4).

Decoupling clonal spread from sexual reproduction provides additional resolution on fine-scale genetic structure. The JK site has lower fruiting density, a higher proportion of full-sibs within the fruit and a higher rate of seedling establishment, leading to stronger kinship structure at close

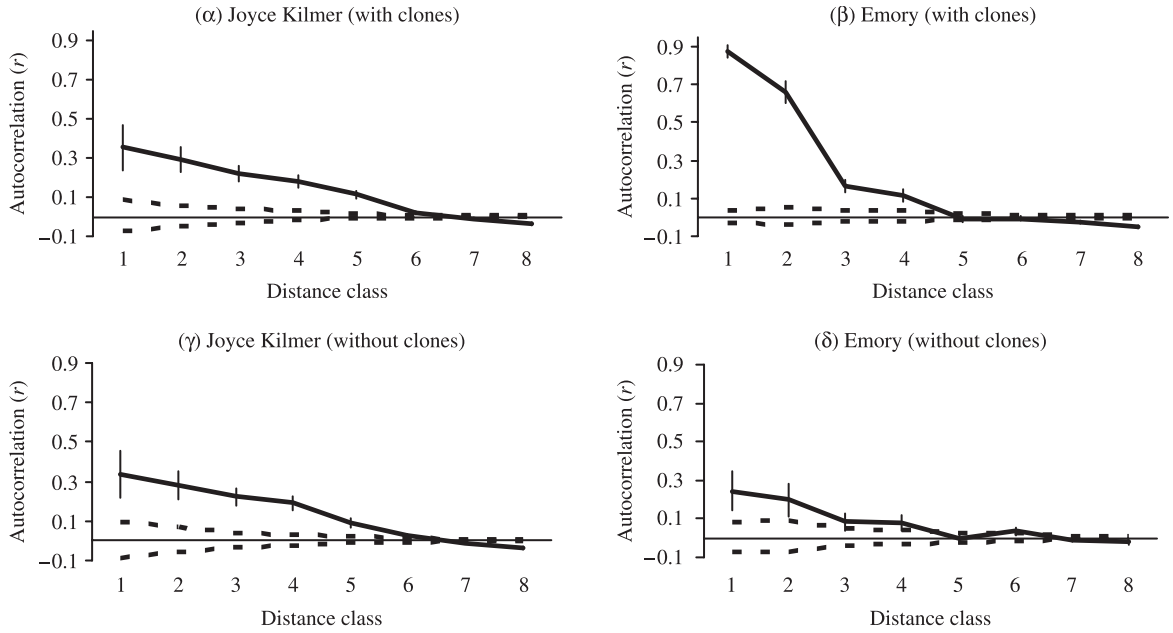

Fig. 2 Fine-scale correlograms for toadshade trillium (*Trillium cuneatum*) populations: (α) location JK (clonal ramets included); (β) location EM (clonal ramets included); (γ) location JK (clonal copies removed); and (δ) location EM (clonal copies removed).

Table 4 Single-distance class (t^2) and multidistance class (ω) test criteria, and associated *p* values, for correlograms in *Trillium cuneatum*: Mountain (α = JK) vs. Piedmont (β = EM) correlograms, clonal replicates included; and Mountain (γ = JK) vs. Piedmont (δ = EM) populations, but with clonal ramets removed; the correlograms are significantly and substantially different for both comparisons, but the relative strengths of the correlograms are reversed without clonal ramets

Distance class	1	2	3	4	5	6	7	8	ω-test	Estimated
Interval (cm)	0–10	11–20	21–40	41–80	81–160	161–320	321–640	641–1280	Criterion	*p* value
α vs. β	$t^2 = 26.99$	49.52	4.68	9.08	68.14	14.89	11.97	7.66	94.61	0.001
with clones	$p = 0.001$	0.001	0.026	0.005	0.001	0.001	0.002	0.011		
γ vs. δ	1.27	2.20	15.38	18.01	26.80	0.69	0.26	3.95	57.13	0.001
no clones	0.257	0.136	0.001	0.001	0.001	0.413	0.593	0.046		

distances. The EM site occupies a narrow strip of a forest, completely exposed to sunlight and to increased wind, both of which contribute to soil desiccation. Thus, edge effects associated with habitat fragmentation exacerbate naturally more stressful environmental conditions in the Piedmont. Although increased pollinator activity leads to fewer correlated matings and higher fruiting density, seedling establishment is rare in the EM population (Gonzales *et al*. in press), contributing to reduced relatedness among sexually produced neighbouring plants. This difference in relatedness among near neighbours in the two habitats is overridden by a higher rate of clonal replication in Piedmont populations, where the sexually produced progeny in close proximity are actually less related than in the mountains.

Discussion

Our objective was to develop a formal test of the hypothesis that the correlograms for two or more populations are heterogeneous, based on elaborations of the autocorrelation method of Smouse & Peakall (1999) and to illustrate it with two interesting example problems. We have developed a nonparametric heterogeneity test for separate patterns of fine-scale genetic structure, either one lag at a time, or for the entire correlogram. The method is general enough to be used for a wide range of comparative problems. While we have illustrated here with genetic data sets, there is nothing inherently genetic about the methodology. We can just as easily apply it to a wide variety of ecological and microgeographical problems for which one can describe

multivariate (multidimensional) distances between all pairs of individuals, constructing an appropriate Euclidean distance matrix **D**, which is substrate for everything that follows.

Other ways of measuring genetic affinity

In narrowly genetic context, there are other measures that can be used for autocorrelation analysis, using the kinship measures constructed by Loiselle *et al*. (1995) and Ritland (1996) or the relationship measures described by Cockerham (1969). Each can be translated into correlational form that will work with methods similar to those deployed here. It develops that kinship, relationship, and correlation are all closely related. Kinship can be measured in a variety of slightly different fashions, not all of which close in Euclidean space, but for the *i*th and *j*th individuals, one can use the form deployed in SPAGEDI (Hardy & Vekemans 2002),

$$k_{ij} = \frac{\sum_{q=1}^{Q}(p_{iq} - p_{\cdot q}) \cdot (p_{jq} - p_{\cdot q})}{\sum_{q=1}^{Q} p_{\cdot q} \cdot (1 - p_{\cdot q})} + \frac{1}{(2N-1)}, \qquad \text{(eqn 15)}$$

where q is the index on alleles, a total of Q alleles for L loci, where p_{iq} (or p_{jq}) = (1, ½, 0), depending on whether the ith (or jth) individual is a homozygote for the qth allele, a heterozygote for the qth allele, or does not possess the qth allele, and where $p_{\cdot q}$ is the global average frequency of the qth allele for the entire collection of N individuals. Where $i = j$ (comparing an individual with itself), we have $k_{ii} > 0$. The appropriate bias correction $(2N - 1)^{-1}$ is built into the pairwise measures. In any case, we can pack the kinship values into an $N \times N$ matrix **K** that is analogous to the **S** matrix. Starting from **K**, one could estimate the autocorrelations of the various lags, in much the same fashion we have described above for **S**.

Hardy & Vekemans (1999) point out that *relationship* metrics are related to *kinship* metrics, specifically (for the diploid case and in our own notation)

$$\rho_{ij} = \frac{2k_{ij}}{1+F}, \qquad \text{(eqn 16)}$$

provided that the inbreeding coefficient F for a single individual is approximately a constant within the population, or alternatively, using a more general Cockerham (1969) result,

$$\rho_{ij} = \frac{2k_{ij}}{\sqrt{1+F_i} \cdot \sqrt{1+F_j}}, \qquad \text{(eqn 17)}$$

if the inbreeding coefficient is not identical among individuals. We pack these relationship measures into a kinship analogue of a correlation matrix **R**, because $\rho_{ii} = 1$. Starting from this **R** matrix, we extract $\rho^{(h)}$ values by averaging the appropriate (within-lag) sets of ρ_{ij} values, and adding an appropriate bias correction to ensure that $\rho^{(h)} = 0$. The correlograms extracted from **S** or **K** or **R** should be three almost indistinguishable versions of the same thing.

There are other autocorrelation metrics and procedures in common usage, prominent among them, Moran's *I* and Join-Count statistics. We direct the reader interested in exploring the wider arena of genetic autocorrelation analysis to the recent review by Epperson (2004, and references cited therein). The nonparametric Smouse-Peakall (1999) form of analysis used here employs a minimum of assumptions, and it has some features that we can exploit to advantage for the heterogeneity analyses, particularly where we wish to extrapolate to wider ecological or geographical scales. All one requires are a Euclidean distance matrix for the response variables and a matrix of pairwise geographical separations. It should also be possible to extend these methods to formal comparison of fine-scale genetic structure patterns for different alleles, different genetic loci, or different sets of genetic markers, but measured on the same individuals. The larger point is that we need to become more formal and overt in our comparative work.

Final note

We have concentrated here on the question of how to determine when two or more correlograms are statistically heterogeneous. The larger context of such analysis is the urge/need to mount comparative studies, searching out both the general and differential patterns exhibited by multiple populations of the same species, both those in similar and those in different settings, reflecting variation in the natural ecology, anthropogenic managerial alternatives, and geographical variation in the mating system. Autocorrelation analysis requires fairly large sample sizes, and seemingly large correlogram differences are sometimes not statistically credible, by virtue of inadequate sample sizes. Conversely, there are cases where tiny correlogram differences are highly significant, by virtue of large sample sizes, but of no practical value, in the context of the problem. What we have constructed here is a productive formalism that has been missing from our comparative arsenal, not a replacement for biological common sense. Used carefully, and in conjunction with a proper biological understanding of the organism and its variable context, we should be able to use it to move the comparative enterprise forward.

Acknowledgements

We wish to thank B. Epperson, D. Grivet, A. Irwin, A. Premoli, V. Sork, B. Wang and the three anonymous reviewers for helpful critique on various drafts of the manuscript. P.E.S. and E.A.G. were supported by NJAES/USDA-17111, NSF-DEB-0211430 and NSF-DEB-0514956. R.P. was supported by Australian Research Council grants C00107400 and DP0451374, Land and Water Australia, and the Australian National University.

References

Andrew RL, Peakall R, Wallis IR, Foley WJ (2007) Spatial distribution of defense chemicals and markers and the maintenance of chemical variation. *Ecology*, **88**, 716–728.

Auchincloss AH, Roux AVD, Brown DG, Raghunathan TE, Erdmann CA (2007) Filling the gaps: spatial interpolation of residential survey data in the estimation of neighborhood characteristics. *Epidemiology*, **18**, 469–478.

Bacaro G, Ricotta C (2007) A spatially explicit measure of beta diversity. *Community Ecology*, **8**, 41–46.

Berg EE, Hamrick JL (1995) Genetic structure of a turkey oak forest. *Evolution*, **49**, 110–120.

Chung MG, Epperson BK (1999) Spatial genetic structure of clonal and sexual reproduction in populations of *Adenophora grandiflora* (Campanulaceae). *Evolution*, **53**, 1068–1078.

Cockerham CC (1969) Variance in gene frequencies. *Evolution*, **23**, 72–84.

Cruse-Sanders JM, Hamrick JL (2004) Spatial and genetic structure within populations of wild American ginseng (*Panax quinqefolium* L., Araliaceae). *Journal of Heredity*, **95**, 309–321.

Diekmann LO, Lawrence D, Okin GS (2007) Changes in the spatial variation of soil properties following shifting cultivation in a Mexican tropical dry forest. *Biogeochemistry*, **84**, 99–113.

Double MC, Peakall R, Beck NR, Cockburn A (2005) Dispersal, philopatry and infidelity: dissecting local genetic structure in superb fairy-wrens (*Mylorus cyaneus*). *Evolution*, **59**, 625–635.

Epperson BK (1995) Spatial structure: correlations for individual genotypes differ from those for local gene frequencies. *Evolution*, **49**, 1022–1026.

Epperson BK (2004) Multilocus estimation of genetic structure within populations. *Theoretical Population Biology*, **65**, 227–237.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction sites. *Genetics*, **131**, 479–491.

Fisher RA (1958) *Statistical Methods for Research Workers*, 13th edn. Hefner, New York.

Fox JC, Bi HQ, Ades PK (2007) Spatial dependence and individual-tree growth models I. Characterising spatial dependence. *Forest Ecology and Management*, **245**, 10–19.

Gonzales E, Hamrick JL, Smouse PE, Dyer RJ (2006) Pollen-mediated gene dispersal within continuous and fragmented populations of a forest understory herb, *Trillium cuneatum. Molecular Ecology*, **15**, 2047–2058.

Gonzales E, Hamrick JL, Smouse PE (in press) A comparison of clonal diversity in mountain and Piedmont populations of *Trillium cuneatum* (Melanthiaceae-Trilliaceae), a forest understory species. *American Journal of Botany*.

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325–338.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecological Notes*, **2**, 618–620.

Heywood JS (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, **22**, 335–355.

Knowles P, Perry D, Foster HA (1992) Spatial genetic structure in two tamarack (*Larix laricinia* (du roi) K. Koch) populations with differing establishment histories. *Evolution*, **46**, 572–576.

Koren A, Tirosh I, Barkai N (2007) Autocorrelation analysis reveals widespread spatial biases in microarray experiments. *BMC Genomics*, **8**, 12.

Lindenmayer DB, Cunningham RB, Pope ML (1999) A large-scale 'experiment' to examine the effects of landscape context and habitat fragmentation on mammals. *Biological Conservation*, **88**, 387–403.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Parker KC, Hamrick JL, Parker AJ, Nason JD (2001) Fine-scale genetic structure in *Pinus clausa* (Pinaceae) populations: effects of disturbance history. *Heredity*, **87**, 99–113.

Peakall R, Lindenmayer DB (2006) Genetic insights into population recovery following experimental perturbation in a fragmented landscape. *Biological Conservation*, **132**, 520–532.

Peakall R, Smouse PE (2006) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecological Notes*, **6**, 288–295.

Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss (*Buchloë dactyloides* (Nutt.) Engelm.). *Molecular Ecology*, **4**, 135–147.

Peakall R, Ruibal M, Lindenmayer DB (2003) Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution*, **57**, 1182–1195.

Premoli AC, Kitzberger T (2005) Regeneration mode affects spatial genetic structure of *Nothofagus dombeyi* forests. *Molecular Ecology*, **14**, 2319–2329.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, **67**, 175–185.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of multi-allele and multi-locus genetic microstructure. *Heredity*, **82**, 561–573.

Sokal RR, Smouse PE, Neel JV (1986) The genetic structure of a tribal population, the Yanomama Indians. XV. Patterns inferred by autocorrelation analysis. *Genetics*, **114**, 259–287.

Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. *Biological Journal of Linnean Society*, **60**, 73–93.

Streiff R, Labbé T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*, **7**, 317–328.

Taylor M, Calaby JH (1988) Mammalian species — *Rattus fuscipes*. *American Society of Mammalogy*, **298**, 1–8.

Temple HJ, Hoffman JI, Amos W (2006) Dispersal, philopatry and intergroup relatedness: fine-scale genetic structure in the white-breasted thrasher, *Ramphocinclus brachyurus*. *Molecular Ecology*, **15**, 3449–3458.

Van Rossum F, Triest L (2007) Fine-scale spatial genetic structure of the distylous *Primula veris* in fragmented habitats. *Plant Biology*, **9**, 374–382.

Van Rossum F, Bonnin I, Fenart S *et al*. (2004) Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Molecular Ecology*, **13**, 2959–2967.

Vekemans X, Hardy OJ (2004) New insights from spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 931–935.

Yamagishi H, Tomimatsu H, Ohara M (2007) Fine-scale spatial genetic structure within continuous and fragmented populations of *Trillium camschatcense*. *Journal of Heredity*, **98**, 367–372.

Peter Smouse is a statistical population geneticist who works primarily on propagule flow and its spatial consequences, in both plant and animal systems. Rod Peakall is an evolutionary biologist whose research on animals and plants integrates tools from ecology and genetics. Eva Gonzales studies population genetic structure and gene flow, with applications to conservation genetics and phylogeography.
