Genetic Differentiation Within and Between Two Habitats

François Rousset

ABSTRACT

We investigate the usefulness of analyses of population differentiation between different ecological types, such as host races of parasites or sources and sink habitats. To that aim, we formulate a model of population structure involving two classes of subpopulations found in sympatry. Extensions of previous results for Wright’s F-statistics in island and isolation-by-distance models of dispersal are given. It is then shown that source and sinks cannot in general be distinguished by F-statistics nor by their gene diversities. The excess differentiation between two partially isolated classes with respect to differentiation within classes is shown to decrease with distance, and for a wide range of parameter values it should be difficult to detect. In the same circumstances little differentiation will be observed in “hierarchical” analyses between pools of samples from each habitat, and differences between levels of differentiation within each habitat will only reflect differences between levels of gene diversity within each habitat. Exceptions will indicate strong isolation between the different classes or habitat-related divergent selection.

Comparisons of population structure between host races of parasites or other behavioral ecotypes have been conducted to evaluate the potential of such ecological differentiation as a factor of genetic isolation and sympatric speciation (e.g., Feder et al. 1990; Ross and Shoemaker 1993; Duffy 1996; Gibbs et al. 1996). Similar comparisons have also been conducted to evaluate directional “gene flow” from one environment to the other, or to test for “source-sink” functioning (e.g., Dias et al. 1996; Hamilton 1997; Stanton et al. 1997). Source habitats are habitats in which individuals contribute more to the future population than the average individual does, and sink habitats are habitats where the reverse occurs (see Dias 1996, for review).

Attempts to estimate gene flow between the different ecotypes are often based on “hierarchical” analyses of differentiation measured by Wright’s F-statistics between pools of samples from each habitat (Wright 1969), or other forms of comparisons of F-statistics. However, it is not clear what information is brought by such comparisons. Although models of population structure involving different types of individuals or subpopulations have been discussed by several authors, one of the few theoretical works explicitly addressing the biological interpretation of differentiation between two distinct groups of subpopulations is the hierarchical island model of Slatkin and Voelm (1991). In this article, we formulate a model of population differentiation between sympatric subpopulations of two different types. The analysis of this model makes it possible to identify the demographic parameters that determine the patterns of differentiation observable by comparisons of gene diversities or of F-statistics. Then it is possible to assess whether sources and sinks can be distinguished by their levels of gene diversity or by the relative levels of population differentiation between sources, between sinks, and between sources and sinks. It is also possible to assess whether partial genetic isolation between different ecological types can be detected by hierarchical analyses.

The comparisons of population structure for different classes of individuals are often complicated by isolation by distance (e.g., Zink and Barrowclough 1984; Duffy 1996; Hamilton 1997; Stanton et al. 1997). Even if migration occurs preferentially within each class rather than between them, it might be expected that neighboring subpopulations from different classes are less differentiated than more distant subpopulations in the same class. Thus, it is not obvious whether the habitat-related variation in differentiation at a given geographical distance can be interpreted independently of distance. For this reason isolation by distance will be considered in the models. For simplicity, we first present results for island models of dispersal, then consider how they generalize to models of isolation by distance.

THE MODEL

We consider a population on a discrete space, evolving at discrete time intervals. Subpopulations are found at positions on a two-dimensional torus. If $n_y = 1$, this is the one-dimensional model on a circle. Two different classes of haploid individuals are located at each node of the lattice: they form two haploid sub-
In his definition, sources are production and migration. We assume that migration is metrical so that a lattice may be described by a pair of coordinates $(v_{12}, v_{21})$, where $v_{12}$ is the dispersal rate given a transition from class 2 to class 1.

populations of size $2N_1$ and $2N_2$ (see Figure 1). Thus, the two types of subpopulations are assumed to be homogeneously distributed on the same geographical area. The total population size is $2(N_1 + N_2)$. The model assumes discrete generations. From time $t$ and $t+1$ the order of events is reproduction, mutation, and migration. These events are described in the reverse order, going backward in time.

**Migration:** The position or movement of a gene on a lattice may be described by a pair of coordinates $(r_1, r_2)$. Let $q_{k,r_1,r_2}$ be the probability that a gene found in a class-$k$ individual in position $r$ had an ancestor or was itself in a class-$k$ individual in position $r'$ before reproduction and migration. We assume that migration is homogeneous in space for given values of $i$ and $k$; i.e., the probability $q_{k,r_1,r_2}$ depends only on values of the movement, $r = r' - r^*$, and can be written $q_{k}(r)$. Let

$$v_k = \sum_{r_1=0}^{N_1-1} \sum_{r_2=0}^{N_2-1} q_{k}(r) = \sum_{r} q_{k}(r).$$

$v_k$ is the probability that a gene in a class-$k$ individual at $t+1$ was in a class-$k$ individual at $t$ (before reproduction). The $v_k$'s are the (class-)transition rates $q_{k}(r)/v_k$. $v_k$ is the probability that a gene in a class-$k$ individual at $t+1$ moved by $r$ between $t$ and $t+1$ given that its ancestral lineage was in a class-$k$ individual at $t$. There are four possible class-transition events, but as $v_{11} = 1 - v_{12}$ and $v_{21} = 1 - v_{22}$, there are only two linearly independent parameters describing the probabilities of transition between different classes.

The source-sink functioning was formalized by Pulliam (1988). In his definition, sources are compartments that show no net change in population size but send more emigrants than they receive immigrants, the reverse holding for sinks. Habitat 1 is a source when $v_{21}N_1 > v_{12}N_2$, i.e., when $v_{21}/v_{12} > N_1/N_2$. There is source-sink functioning when (and only when) $v_{21}/N_1$ and $v_{12}/N_2$ differ. This corresponds to the case of “nonconservative migration” defined by Nagylaki (1982).

The distinction between sources and sinks can also be expressed in terms of reproductive values (e.g., Fisher 1958; Taylor 1990). The class reproductive values $v_1$ and $v_2$ are defined as the probabilities that the ancestral lineage of some gene was in class 1 or 2 in a distant past (obviously, $v_1 = 1 - v_2$). They can be computed as components of a left eigenvector associated with the dominant eigenvalue of the matrix of class-transition rates (e.g., Taylor 1990, Equation 8). Likewise the individual reproductive values of the two classes are defined as the probabilities that the ancestral lineage of some gene was in a given class 1 or 2 individual in a distant past. They can be computed as elements of a left eigenvector of a demographic “projection” matrix (e.g., Lescie 1945; Caswell 1989; Taylor 1990, Equation 7). In the present model they are obtained from the matrix of class-transition rates $v_k$. One has $e_1 = v_{21}/(v_{21} + v_{12})$, and the individual reproductive values are $v_1 = N_1$ and $v_2 = N_2$, which are proportional to $v_{21}/N_1$ and $v_{12}/N_2$. Hence from the above definition there is source-sink functioning when (and only when) the individual reproductive values differ.

Dispersal is assumed independent and identically distributed in each dimension. Let $z = (x, y)$, and for any function $f(r)$ consider the Fourier series $F(f) = \sum \hat{f}(i)e^{-i\pi i^j}r_j$, where $i = \sqrt{-1}$ and the sum is over all positions on the lattice. Then $\psi_k(z) = F(\phi_k(r)/v_k)$ is the characteristic function of the conditional distribution of dispersal distance for transitions from class $i$ to class $k$. We also assume that the axial dispersal distributions (i.e., dispersal distributions in each dimension) are symmetrical so that $\psi_k(z) = \psi_k(-z)$. In this two-class model we have four dispersal distributions: $\psi_{11}$, $\psi_{21}$, $\psi_{12}$, and $\psi_{22}$.

Let $Q_{ij}$ be the probability of identity in state of two genes, the first sampled in a class-$i$ individual and the second in a class-$j$ individual, at $t$ steps from each other at time $t$. For convenience, we distinguish between $Q_{ij}^{(t)}$ and $Q_{ij}^{(t)}$ at this stage, although they will have identical values. Let $Q_{ij}^{(t)}$ be the probability of identity of a pair of genes that were in classes $i$ and $j$ at time $t$ and are observed before migration but after reproduction and mutation. We assume that the probability of the joint migration events of two genes is simply the product of the probabilities of independent migration events for each gene. Gametes are sampled from an ideally infinite pool, and population regulation is assumed to act after migration. Then for all $k, l$,

$$Q_{ij}^{(t+1)} = \sum_{r} \sum_{r^*} q_{k}(r)q_{l}(r)Q_{k,l}^{(t)}.$$
Gene lineage and that the mutation rate $u$ is identical in both types of subpopulations. Let $Q_{**}$ be the probability of identity of the gene lineages before mutation in the chain of events from $t$ to $t + 1$ and let $\gamma = (1 - u)^2$. Then

$$Q_{**} = \gamma Q_{**}.$$  \hspace{1cm} (3)

**Reproduction:** We assume that the probability that two gametes observed in a class $i$ subpopulation before migration are copies of the same gene from an individual in this subpopulation is $1/(2N_i) < 1$. In other words, the probability of coalescence is $1/(2N_i)$. Thus $Q_{**} = Q_{ii}$ except for the within-subpopulation values,

$$Q_{**} = (1/N_i) + 1 - Q_{jj}/(2N_i).$$  \hspace{1cm} (4)

Let $g_i^{(0)}(z) = \mathcal{F}(Q_{ii}^{(0)}(r))$, and similarly $g_i^{(r)}(z) = \mathcal{F}(Q_{ii}^{(r)}(r))$ and $g_{**}^{(r)}(z) = \mathcal{F}(Q_{**}^{(r)}(r))$. Let $\mathbf{v}$ denote the transpose of some vector $\mathbf{v}$. Consider the vector $g_i^{(0)}(z) = \mathcal{F}(g_{ii}^{(0)}(z))$, $g_i^{(1)}(z)$, and consider $g_i^{(r)}$ and $g_{**}^{(r)}$ defined similarly from the $g_i^{*}$s and $g_{**}^{*}$s. From Equations 3 and 4, $g_i^{(r)} = g_i^{(0)} g_{**}^{(r)}$, where

$$c_i^{(r)} = \mathcal{T}(c_i^{(0)}, c_i^{(1)}, c_i^{(2)}) = \mathcal{T}(1 - (1/Q_{i,0} - 1/Q_{i,2r})/2N_i, 0),$$  \hspace{1cm} (5)

and from Equation 2,

$$g_{i,1}^{(r+1)}(z) = \sum_{i=1}^{2} \sum_{j=1}^{2} \psi_k(z) v_{i,j} \psi_j(z) v_{i,j} g_{i,j}^{(r)}(z).$$  \hspace{1cm} (6)

Hence we obtain the recurrence relationship

$$g_{i,1}^{(r+1)} = \mathbf{A} g_i^{(r)} + c_i^{(r)},$$  \hspace{1cm} (7)

where

$$\mathbf{A} = \gamma \times \begin{pmatrix}
(1 - v_{1,0}) \psi_1 & v_{1,0} \psi_2 & 2v_{2,0}(1 - v_{1,0}) \psi_0 \psi_2 \\
v_{1,0} \psi_1 & (1 - v_{2,1}) \psi_2 & 2v_{2,0}(1 - v_{2,1}) \psi_0 \psi_2 \\
v_{2,1}(1 - v_{2,1}) \psi_2 \psi_1 & v_{1,0}(1 - v_{2,1}) \psi_0 \psi_2 & \psi_1(1 - v_{2,1}) \psi_2 \psi_2 + v_{2,1}(1 - v_{2,1}) \psi_0 \psi_2
\end{pmatrix}.$$  \hspace{1cm} (8)

At equilibrium

$$g = \mathbf{P} \mathbf{c},$$

where $\mathbf{P} = (p_{ij}) = (I - \mathbf{A})^{-1} \mathbf{A}$  \hspace{1cm} (9)

and $\mathbf{I}$ is the identity matrix.

Probabilities of identity are obtained from their generating functions by Fourier inversion techniques (e.g., Gascuet et al. 1998). For a finite lattice of size $n_i \times n_j$, the inverse Fourier transform of some function $f$ is

$$L_i(f(z)) = \frac{1}{n_i n_j} \sum_{q_1=0}^{n_1-1} \sum_{q_2=0}^{n_2-1} \left( \frac{2\pi q_1}{n_i}, \frac{2\pi q_2}{n_j} \right) e^{-2\pi i q_1 z} e^{-2\pi i q_2 z},$$

$$L_i(f(z)) = L_i(g_{ij}(z)) = L_i(p_{ij})(z) + L_i(p_{ij})(z).$$

The $c_i$'s themselves are deduced from the system of linear equations $Q_{ij,0} = L_i(\sigma_i(z))$ for $i = 1, 2$. Results for infinite lattices are obtained as limits results from the finite lattice models when $n_i$ and/or $n_j \to \infty$ without additional assumptions (see Rousset 1999 for details).

**RESULTS**

This section presents analytical approximations for gene diversities and differentiation between pairs of subpopulations under the model presented above. These approximations are derived in the context of a more general model in Rousset (1999). Their accuracy will be assessed through exact numerical computations, using Equations 9 and 10. Mathematica (Wolfram 1991) has been used in all computations.

**Infinite island model:** Consider $n$ populations on a circle. Let the dispersal rate $m_{ij} = \sum_r q_{ij}(r)/v_{ij} < 1$ be the probability that the parent of a class $i$ individual was from a different node, given it was in class $i$ (see Figure 1). We define the island model here as a model where, as in the one-class island model (e.g., Wright 1969), the probability that a class $i$ gene had its ancestral lineage from another node of the lattice, given this lineage was in class $i$, is independent of the distance between parent’s and offspring’s nodes, i.e., $q_{ij}(r) = v_{i,0} m_{ij}(r)$ where $q_{ij}(r) = v_{i,0} m_{ij}(r)$.

The total immigration rate into, e.g., a class $1$ subpopulation is $v_{1,0} m_{11} + v_{1,2}$. The infinite island model is obtained by letting the number of subpopulations $n$ go to infinity.

In the one-class infinite island model, the usual measure of population structure, $F_{ij}(r) = (Q_{i,0} - Q_{i,2r})/(Q_{i,0} - Q_{i,11})$, equals $Q_{i,0}$ under the infinite allele model because the probabilities of identity between different subpopulations, $Q_{i,0}$, are then null. We have a similar result in the two-class model, if we consider the parameters

$$F_{i,1} = \frac{Q_{i,0} - Q_{i,2r}}{1 - Q_{i,2r}},$$

$$F_{i,2} = \frac{Q_{i,0} - Q_{i,2r}}{1 - Q_{i,2r}},$$

where $Q_{i,2r} = (Q_{i,11} + Q_{i,2r})/2$ is the average of probabilities of identity of pairs of genes sampled at distance $r$ within classes. $F_{i,1}$ measures differentiation between subpopulations within class $1$ and $F_{i,2}$ measures differentiation between subpopulations of the two different classes, at distance $r$ in each case. For $r \neq 0$ we have $Q_{i,2r} = 0$ (Rousset 1999, Equation 43) so that $F_{i,1} = Q_{i,0}$ and $F_{i,2} = Q_{i,0}$. Hence the between-class differentiation is the average of the two within-class differentiations.

In an island model with small values of dispersal and transition rates, we expect that $Q_{i,j,0} = 1/(4N_{ij}, M_{ij})$, where $M_{ij}$ is the probability that a parent was in another subpopulation (Beaumont and Nichols 1996). As in the one-class model (Hudson 1990), $Q_{i,j,0}$ is approximately the probability that, going backward in time, there is coalescence of two gene lineages within a class.
i subpopulation before one of their ancestral lineages emigrates from this subpopulation (i.e., immigrants, going forward in time). For example, with small values of dispersal \( m_{22} \) and transition rate \( v_{22} \),

\[
Q_{22} = 1/(1 + 4N_s(v_{22}m_{22} + v_{21})).
\]  

(12)

However, when \( v_{21} < v_{22}m_{22} + m_{12} \), differentiation within class 2 will depend on \( m_{11} \). Then the result of the one-class model, \( Q_{11} = 1/(1 + 4N_s(m_{11})) \), is not easily generalized to either of the within-class differentiations, but it can be generalized if we consider the following parameters. Let \( Q_{w} = wQ_{11} + (1 - w)Q_{22} \), be a weighted average of the probabilities of identities within classes (the single index \( w \) will be used to distinguish such averages from the probabilities \( Q_{ij} \)), and let \( w^* = (e/(N_1)) / (e/N_1 + e/N_2) \). Let

\[
F_w^*(r) = \frac{Q_{w^*} - Q_{w^*}}{1 - Q_{w^*}}.
\]  

(13)

It follows from Proposition 5 in Rousset (1999) that

\[
F_w^*(r) = 1/(1 + 4N_s m_w),
\]  

(14)

where

\[
1 = \frac{\sum_{i} e_i^2}{\sum_{i} N_i} = \frac{v_{21}^2/N_1 + v_{12}^2/N_2}{(v_{21} + v_{12})^2},
\]  

(15)

and

\[
m_w = \sum_{i} \sum_{j} e_{ij} m_{ij}.
\]  

(16)

In general, some detailed knowledge of the demography of the population would be required to estimate \( w^* \) and \( F_w^* \) because their definitions depend on the reproductive values \( e \), in contrast to the parameters defined for pairs of subpopulations (Equation 11), which can be estimated using data from pairs of subpopulations. Parameters involving all subpopulations in the population are not considered, in part because their estimation requires, in principle, having subsamples from all subpopulations. However, “hierarchical” analyses based on the comparison of probabilities of identity within and between pools of subsamples (Wright 1969) are discussed. An analysis based on estimating \( F_w^* \) is not a hierarchical analysis because \( F_w^* \) is defined from the average of the probabilities \( Q_{w^*} \) within each subpopulation rather than from probabilities of identity computed after pooling genes from different subpopulations.

Since \( F_{w^*} = Q_{w^*} \) in the infinite island, infinite allele model, the comparison of the relative values of differentiation in each class is equivalent to the comparison of gene diversities. Intuitively, the probability of identity will be higher in the class with a smaller subpopulation size or in the class with “lower dispersal,” and the differences between the two classes will be stronger the less they exchange migrants. Higher transition rates have the obvious effect of homogenizing the gene diversities.

The individual effects of differences in subpopulation sizes or transition rates can be investigated in the symmetric dispersal model, in which all conditional dispersal distributions (i.e., all \( m_{ij} \)'s) are identical (Figure 2 and appendix). Note that the assumption of symmetric dispersal does not constrain the class-transition rates to particular values. In this model, a large sink may have a higher gene diversity than a small source (e.g., top left of Figure 2), and in some cases even a small sink may have a higher gene diversity than a large source (area below the previous one, where 1 is source, \( N_2 < N_1 \), and \( Q_{110} > Q_{220} \)).

When the two classes differ by a single parameter value (\( N_1 \neq N_2 \) or \( v_{21} \neq v_{12} \)), the consequences are relatively easily predicted (Figure 2). If \( N_2v_{21} = N_1v_{12} \), there is no source and sink and \( Q_{110} < Q_{220} \) for \( v_{12} < v_{21} \). Then the class that receives a higher fraction of its genes from the other class has a lower gene diversity (Equation A4). The fact that gene diversities differ in that case suffices to show that sources and sinks cannot be distinguished simply by their genetic diversities. This is easily seen when all \( m \)'s and \( v \)'s are small, in which case the relative magnitude of the gene diversities will depend on the relative magnitude of \( N_s(v_{21}m_{22} + v_{21}) \) vs. \( N_s(v_{12}m_{11} + v_{12}) \) (this follows from the argument that led to Equation 12). By contrast, which class is source and which is sink will depend on the relative magnitude of \( N_s v_{21} \) vs. \( N_s v_{12} \). This is independent of the dispersal rates from a subpopulation into another subpopulation of the same class, \( m_{11} \) and \( m_{22} \).

**Isolation by distance:** In the isolation-by-distance models, we assume that each axial dispersal distribution has a finite second moment (or average squared dispersal distance) \( \sigma_i^2 \). For example, \( \sigma_i^2 \) is the average squared axial distance (i.e., measured along one dimension) between a gene sampled in class 2 and its parent, given the parent was in class 1.

We determine the increase of differentiation with distance, and we consider how the results of the infinite island model extend to models of isolation by distance. In particular we investigate whether differences in gene diversities determine differences between levels of within-class differentiations as they do in the infinite island model, and whether between-class differentiation is an average of within-class differentiations. We also investigate to what extent the result of the infinite island model, \( F_w^* = 1/(1 + 4N_s m_w) \), approximates the value of differentiation between adjacent populations under isolation by distance, as it does in one-class models with small migration rates (e.g., Kimura and Maruyama 1971).

In the one-class, one-dimensional isolation-by-distance model

\[
\frac{Q_{110} - Q_{11}}{1 - Q_{110}} = \frac{1 - e^{-2\sigma_i^2}}{4N_s \sigma_i^2} + h(r),
\]  

(17)

where \( u \) is the mutation rate and \( h(r) \) is a function
Differentiation Between Two Habitats

Figure 2.—Contour plot of the ratio of gene diversities \((1 - Q_{2i0})/(1 - Q_{1i0})\) for the symmetric dispersal case. Gene diversity is higher in class 1 for the set of parameter values of \(N_1, N_2, v_{12}, \text{and } v_{21}\) below the plain/1.00 line. The relative magnitude of gene diversities in the two classes can be predicted independently of the common dispersal distribution for part of the parameter set, including the three dashed lines, as shown in the appendix. The boundary between \(R_1\) (higher gene diversity in class 2) and \(R_2\) is constrained to be between the diagonal and horizontal lines. The contour lines were computed for an island model of dispersal with rates \(m_{ji} = 1/20\), 100 subpopulations of each class, \(N_1 = 10\) and \(v_{21} = 1/20\), and mutation rate \(u = 10^{-5}\).

bounded as \(u \to 0\) and vanishing as \(r \to \infty\) (Rousset 1997). In the two-class model, for some functions \(h_{ij}\) and \(h_{0i}\),

\[
a_{ijk}(r) = \frac{Q_{ij0} - Q_{ikj}}{1 - Q_{ik0}} = \frac{1 - e^{-2\pi r/\sigma_i}}{4N_e \sigma_i \sqrt{2u}} + h_{ij}(0) - h_0i(r),
\]

where \(N_e\) has been defined in Equation 15,

\[
\sigma_i^2 = \sum_{i}^{2} \sum_{j}^{2} e_{ij} \sigma_{ij}^2,
\]

\(h_0i(0)\) is bounded as \(u \to 0\), and \(h_0i(r) \to 0\) as \(r \to \infty\) (Rousset 1999, Proposition 3). As in the one-class model, Equation 18 implies an approximately linear relationship between differentiation and distance in the limit of low mutation rates, and for \(r\) large enough that \(h_0i(r)\) can be neglected.

In two dimensions

\[
a_{ijk}(r) = \frac{-ln(\sqrt{2u}) - K_0(\sqrt{2ur}/\sigma_i^2)}{4N_e \sigma_i^2} + h_{ij} - h_0i(r)
\]

for \(r > 0\), where \(K_0\) is the Bessel function of second kind and zero order and \(h_{ij} - h_0i(r)\) is bounded when \(r \to \infty\) and \(u \to 0\) with \(\sqrt{ur}\) fixed (Rousset 1999, Proposition 4). As in the one-class model, Equation 20 implies an approximately linear relationship between differentiation and logarithm of distance in the limit of low mutation rates, and for \(r\) large enough that \(h_0i(r)\) can be neglected.

We continue with the one-dimensional model for exposition, with results very similar to those detailed below being obtained in the two-dimensional model. Numerical examples are given for both cases.

Some aspects of differentiation that might in principle be used to investigate the divergence between the two habitats, such as the relative strength of within-habitat and between-habitat differentiations, are a priori dependent on the geographical distance between populations, and their properties may not be summarized in terms of the demographic parameters independently of distance. Then, measures whose values are either less dependent on geographical distance, or have a known relationship to it, should be considered.

Consider first the following measures of population structure,

\[
a_{ij}(r) = \frac{F_{ij} - Q_{ij0} - Q_{ij0}}{1 - F_{ij}} = \frac{Q_{ij0} - Q_{ij0}}{1 - Q_{ij0}}
\]

for \(i = 1, 2\), and

\[
a_{12}(r) = \frac{F_{12} - Q_{12} - Q_{12}}{1 - F_{12}} = \frac{Q_{12} - Q_{12}}{1 - Q_{12}}.
\]

The choice of such parameters is based on the same logic as the choice of the \(F_{ij}\)'s (particularly, they can be estimated using data from pairs of subpopulations), and on the relative simplicity of the results that can
be obtained for these “F/(1 − F)” parameters. From Equation 18,

\[
a_{11}(r) = \frac{1 - Q *_{11,0}}{1 - Q_{11,0}} \times \left( \frac{1 - e^{-\frac{r}{\alpha_e}}}{4N_e\alpha_e^2r} + h_{11}(0) - h_{11}(r) \right),
\]

\[a_{22}(r) = \frac{1 - Q *_{22,0}}{1 - Q_{22,0}} \times \left( \frac{1 - e^{-\frac{r}{\alpha_e}}}{4N_e\alpha_e^2r} + h_{22}(0) - h_{22}(r) \right),
\]

and

\[
a_{12}(r) = \frac{1 - Q *_{12,0}}{1 - Q_{12,0}} \times \left( \frac{1 - e^{-\frac{r}{\alpha_e}}}{4N_e\alpha_e^2r} + h_{12}(0) - h_{12}(r) \right).
\]

We now seek parameters that quantify the differences between \(a_{11}, a_{22}, \) and \(a_{12}\) but the value of which does not depend on the distance-dependent term \(e^{-\frac{r}{\alpha_e}}\).

Within-class differentiation: For the comparison of differentiations within each class, a parameter whose value does not depend on the term \(e^{-\frac{r}{\alpha_e}}\) is \(D(r) = a_{11}(r) - R a_{22}(r)\), where \(R = (1 - Q_{22,0})/(1 - Q_{11,0})\) is the ratio of “gene diversities” in each class. In essence this is simply the difference \(a_{11}(r) - a_{22}(r)\) between the within-class differentiations, corrected by the ratio \(R\) to make it approximately independent of distance. Its value at large distances is obtained by noting that as \(r \to \infty\),

\[
a_{11}(r) \to \frac{Q_{11,0}}{1 - Q_{11,0}} = \frac{1 - Q *_{11,0}}{1 - Q_{11,0}} \times \left( \frac{1}{4N_e\alpha_e^2r} + h_{11}(0) \right).\]

and so on for \(a_{22}(r)\), so that

\[
D(r) \to \frac{1 - Q *_{11,0}}{1 - Q_{11,0}} (h_{11}(0) - h_{22}(0)) = \frac{Q_{11,0} - Q_{22,0}}{1 - Q_{11,0}} = R - 1.
\]

Note that in the infinite island model, \(D = R - 1\). Equation 27 shows that no information that is not already given by either one of the \(a\)’s and by the ratio \(R\) can be obtained from the difference between \(a_{11}\) and \(a_{22}\) at large distances. Moreover, the numerical examples (discussed below) show that in practice \(D(r) = R - 1\) at moderate distances. Hence, the information about demography that could be obtained from \(a_{11}\) and \(a_{22}\) is given by either one of them (or by \(a_{ijلك}\) for any \(i, j, k, l\)) and by \(R\).

Under the island model with small dispersal and transition rates, if \(F_{22} \gg F_{11} = 0\), then \(D = R - 1 = -F_{22}\), hence

\[
D = -1/(1 + 4N_2(v_{22}m_{22} + v_{21})). \quad (28)
\]

The accuracy of this result is numerically assessed under isolation by distance.

Between-class differentiation: For the comparison of between-class differentiation relative to within-class differentiation, we consider \(Z(r) = (1 + R) a_{12}(r) - a_{11}(r) - R a_{22}(r)\). In essence this is a comparison of between-class differentiation \(a_{12}(r)\) to the average within-class differentiation \((a_{11}(r) + a_{22}(r))/2\), corrected by the ratio \(R\) to make it approximately independent of distance. When \(r \to \infty\),

\[
Z(r) \to \frac{2Q_{12,0} - Q_{11,0} - Q_{22,0}}{1 - Q_{11,0}} = 0. \quad (29)
\]

Note that in the infinite island model, \(Z = 0\). Equation 29 shows that the between-class differentiation tends toward an average, weighted according to \(R\), of the within-class differentiations.

Gene diversities: Generally, the class with the higher within-class differentiation is the one with the lower gene diversity (Figure 3), although some exceptions could occur at short distances. Therefore, the relative levels of within-class differentiation convey little additional information relative to that given by \(R\), i.e., by the relative values of gene diversities in each class. As in the island model, we should understand how the different demographic parameters affect the relative values of gene diversities in each class. For the symmetric dispersal model, where all dispersal distributions are identical, the qualitative results noted before and illustrated by Figure 2 for the relationship between subpopulation sizes, transition rates, and gene diversities in the symmetric island model, hold true in general (see appendix). The individual effect of different genes in dispersal distributions cannot be easily summarized. Even in the one-class model the important parameters of the dispersal distribution determining gene diversity have no simple interpretation (Rousset 1997).

Increase of differentiation with distance In the one-dimensional isolation-by-distance model, the increase of \(a_{ijلك}\) with distance is

\[
\frac{da_{ijلك}(r)}{dr} = \frac{e^{-\frac{r}{\alpha_e}}}{4N_e\alpha_e^2r} + \frac{dh_{ijلك}(r)}{dr} = \frac{e^{-\frac{r}{\alpha_e}}}{4N_e\alpha_e^2r}. \quad (30)
\]

If \(R = 1\), the increase is the same for \(a_{11}, a_{12}, b_{12}\). Otherwise, the increase of \(a_{11}, a_{12}, a_{22}\) with distance will differ from each other in the ratios \(1/(1 + R)\), and the evaluation of \(\alpha_e^2\) from \(a_{ijلك}\) would require in principle the detailed knowledge of the demography of the subpopulation necessary to estimate \(w^*\) in \(a_{ijلك}(r)\). The relative differences between the \(a\)’s values (which also depend on \(R\) as discussed above) are often much larger than the relative
Differentiation Between Two Habitats

Symmetric dispersal, $N_2 = N_1, v_{21} = 1/30, v_{12} = 1/100$

![Graph](image1)

$N_2 = N_1/10, v_{21} = 1/100, v_{12} = 0$

![Graph](image2)

Asymmetric dispersal, $N_2 = N_1/2, v_{21} = 1/5, v_{12} = 1/30$

![Graph](image3)

distance $r$ (lattice steps)

Figure 3.—Patterns of differentiation for different cases. Values of the $a_{ij}(r)$’s are shown, with dot size increasing in the order $a_{11}, a_{22}, a_{12}$. One-dimensional examples are shown on the left and two-dimensional ones on the right. See Table 1, cases b, d, f, g, i, and m for parameter values.

Hiearchical analyses: In the framework of Wright’s F-statistics, such analyses compare the probability of identity within pools of subsamples, averaged over the different classes, to the probability of identity between pools. When an increasing number of subsamples are pooled, each within-pool probability approaches an average of the $Q_{ii}$’s, and the probability of identity between pools is close to $Q_{12r}$, for diverse values of $r > 0$.

The hierarchical analysis generally conducted in experimental studies compares the genetic variation within pools ($1 - Q_{i2}$, if equal weight is given to both classes) to the genetic variation between pools ($1 - Q_{12}$). One parameter describing this comparison is

$$F' = (Q_{i2} - Q_{12}) / (1 - Q_{12}).$$  \hspace{1cm} (31)

In an infinite island model, $F'$ is null for any strictly positive transition rate (as all $Q_{1r} = 0$ for $r > 0$). This result is in agreement with the results of Slatkin and
Voel (1991) for the finite island model. Similar results are obtained at increasing geographical distances under isolation by distance, because $F^\prime$ decreases to zero at increasing distances.

**Numerical examples:** In Table 1, the numerical accuracy of the different analytical approximations is investigated. The following points may be seen from this table. For the island model, the number of migrants into subpopulations averaged over the different classes, $M = 2 N_e (v_{12} + v_{13}) + 2 N_e (v_{12} + v_{12})$, is seen to be a very inaccurate descriptor of differentiation in comparison with approximation (14) (Table 1, cases h–k; in this table, the value of $F_{st} / (1 - F_{st})$ is compared to $1/(4N_m)$ and to $1/(2M)$). This is in agreement with previous results (Gaggiotti 1996).

The island model also suggests an approximation for the measure $D$ of differences between genetic differentiations within each class (Equation 28). This approximation is correct within a factor of two in examples where $N_2 \ll N_1$ ($N_2 = 10$, cases a–f–i, and k in Table 1), including some cases not assuming an island model of dispersal.

In isolation-by-distance models, in principle $D(r)$ approaches $R - 1$ more slowly when $v_{21} + v_{12}$ decreases, and some information could be obtained from the rate of approach. However, even under the simple case of symmetric dispersal, this approach is a complex function of the demographic parameters. In the numerical examples we investigated whether this approximation is accurate at distances larger than $r^* = 2 \alpha / \sqrt{v_{21} + v_{12}}$. This choice is guided by considerations detailed in the appendix. The value of $D(r^*)$ was compared to $D(\infty) = R - 1$. The relative error is generally <0.2.

We have seen that the between-class differentiation approaches an average of within-class differentiations as distance increases. This is quantified by the approach of $Z(r)$ to 0, which is also slower when $v_{21} + v_{12}$ decreases. Figure 3 shows that the excess differentiation between classes with respect to this average may be difficult to detect, even for transition rates $=1/50$. Differentiation may be stronger between neighboring sources and sinks than between neighboring sinks or neighboring sources, yet this will not be the case for more distant pairs of subpopulations.

Finally, the increase of differentiation with distance has been approximated by the values of the derivatives discussed above (Equation 30 for the one-dimensional model), which are asymptotic values at large distances. The lower $v_{21} + v_{12}$ is, the slower the asymptotic values are approached as distance increases. At short distances, the increase of a within-class differentiation may depend mainly on local processes rather than on the effective parameters. As for $D$, in the numerical examples we investigated whether the asymptotic values are accurate at the distance $r^* = 2 \alpha / \sqrt{v_{21} + v_{12}}$. Table 1 compares the approximate value in Equation 30 (or $dK_2 / (\sqrt{2} u / \sigma_2)$) to its discrete equivalent in the lattice models, $\Delta s_e(r)/\Delta r$ [or $\Delta s_e(r)/\Delta \ln(r)$ in two dimensions], where $\Delta f(r) = f(x + 1) - f(x)$ for the integer $x$ such that $r^* \in [x, x + 1]$. The relative error (given in the “$\Delta s_e$ error” column of Table 1) is generally <0.2.

**DISCUSSION**

The analysis of the models presented here has shown that genetic differentiation is approximately described by “effective” parameters $N_e$, $\sigma^2$, and $m_0$ and by the relative values of gene diversities in the two types of subpopulations. $N_e$ may be understood as the average rate of coalescence of pairs of ancestral lineages (Nordborg 1997; Nagylaki 1998; Rousset 1999), and $m_0$ and $\sigma^2$ may be understood as average dispersal rates of ancestral lineages. The other aspects of the results have implications both for inferences about source-sink functioning and for hierarchical analyses of differentiation.

The relative values of gene diversities in each habitat have no simple relationship with the reproductive value of individuals in each habitat, because the gene diversities depend among other factors on the migration rates between subpopulations of the same types, while the reproductive values of each type are independent of these migration rates. Therefore, source and sink habitats cannot be distinguished by comparison of their gene diversities.

Further, these models show that the differentiation between two habitats is generally intermediate between the genetic differentiations within each habitat, and that genetic differentiation will be generally higher in the habitat with the lower gene diversity than in the habitat with the higher gene diversity. It follows that source and sink habitats cannot be distinguished by analyses of population structure by F-statistics.

The models also show that a partial isolation between the two classes will often not result in an excess differentiation between them and will not be detected by a hierarchical analysis. Some complications may in principle result from isolation by distance: if the two different classes show a strong but incomplete isolation, between-class differentiation may exceed within-class differentiation at short distances, but this excess differentiation will disappear at larger distances. In practice, no excess differentiation between classes may be detectable at distances only a few times the effective dispersal rate $\sigma^2$, even for transition rates $=1/100$ (see Figure 3). Thus, the conclusion of the infinite island model is much more generally valid: the excess between-class differentiation may be low even though the closest pairs of subpopulations of different classes exchange less migrants than the most distant subpopulations within classes.

On the other hand, divergent selection in the two classes on the loci considered, or on closely linked ones, will result in between-class differentiation being higher at all distances than the average of within-class differenti-
### TABLE 1
Comparison of theoretical approximations to exact numerical results

<table>
<thead>
<tr>
<th>Case</th>
<th>Dimension of lattice</th>
<th>Parameters</th>
<th>Dispersal Type</th>
<th>Fw,∞/(1 − Fw,∞) Error</th>
<th>ΔN*</th>
<th>Error Equation 28</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>1</td>
<td>10</td>
<td>1/50, 1/50</td>
<td>g1 (σ2 = 1)</td>
<td>36.36</td>
<td>10</td>
</tr>
<tr>
<td>(b)</td>
<td>1</td>
<td>100</td>
<td>1/30, 1/100</td>
<td>g1 (σ2 = 1)</td>
<td>155</td>
<td>9.6</td>
</tr>
<tr>
<td>(c)</td>
<td>1</td>
<td>100</td>
<td>1/30, 1/100</td>
<td>g1 (σ2 = 1)</td>
<td>155</td>
<td>9.6</td>
</tr>
<tr>
<td>(d)</td>
<td>2</td>
<td>100</td>
<td>1/30, 1/100</td>
<td>g1 (σ2 = 1)</td>
<td>155</td>
<td>9.6</td>
</tr>
<tr>
<td>(e)</td>
<td>2</td>
<td>100</td>
<td>1/30, 1/100</td>
<td>g1 (σ2 = 1)</td>
<td>155</td>
<td>9.6</td>
</tr>
<tr>
<td>(f)</td>
<td>1</td>
<td>100</td>
<td>1/100, 0</td>
<td>g1 (σ2 = 1)</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>(g)</td>
<td>2</td>
<td>100</td>
<td>1/100, 0</td>
<td>g1 (σ2 = 1)</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>(h)</td>
<td>2</td>
<td>100</td>
<td>1/5, 1/10</td>
<td>g1 (m1 = 0.0166)</td>
<td>64.3</td>
<td>—</td>
</tr>
<tr>
<td>(i)</td>
<td>2</td>
<td>100</td>
<td>1/5, 1/10</td>
<td>g1 (m1 = 0.00166)</td>
<td>64.3</td>
<td>—</td>
</tr>
<tr>
<td>(j)</td>
<td>2</td>
<td>100</td>
<td>1/30, 1/100</td>
<td>g1 (m1 = 0.0176)</td>
<td>155</td>
<td>—</td>
</tr>
<tr>
<td>(k)</td>
<td>2</td>
<td>100</td>
<td>1/100, 0</td>
<td>g1 (m1 = 0.0199)</td>
<td>100</td>
<td>—</td>
</tr>
<tr>
<td>(l)</td>
<td>1</td>
<td>50</td>
<td>1/5, 1/30</td>
<td>g1 (m1 = 0.857)</td>
<td>128.9</td>
<td>3.5</td>
</tr>
<tr>
<td>(m)</td>
<td>2</td>
<td>50</td>
<td>1/5, 1/30</td>
<td>g1 (m1 = 0.857)</td>
<td>128.9</td>
<td>3.5</td>
</tr>
</tbody>
</table>

*See text for definitions of the different parameters. One-dimensional examples were computed for 5000 × 1 lattices and two-dimensional examples were computed for 400 × 400 (cases g and k), 100 × 100 (cases h–j), or 200 × 200 lattices otherwise. The mutation rate was u = 10^-5 and N1 = 100 in all examples. In all examples we assumed a modified geometric axial distribution of parent-offspring distance, where the probability of migration by 1 steps is m1 = (1 - q)q^-1 (1 - m0) / 2 (details in Rousset 1997, appendix X). It approaches an island model when q → 1. Parameter values of dispersal distributions were as follows: g1, 1 - m0 = q = 1/10, axial kurtosis γ2 = 7; g2, 1 - m1 = 1/6, q = 1/2, γ2 = 22; g3, q = 999/1000, m1 = m2 = 1/100, m3 = m0 = 1/200; g4, q = 999/1000, m1 = m2 = 1/1000, m3 = m0 = 1/2000; g5, class 2 individuals do not disperse and class 1 individuals disperse according to g1.

*The exact value is for adjacent nodes on the lattice, 1/(4Nem2) is the approximation for the island model (see Equation 14), and 1/(2M) is the value that would be expected if the average number of migrants into subpopulations determined differentiation (see text for details).

The "error" is D(r*)/(R - 1) - 1 under isolation by distance, evaluated at the least integer distance not smaller than r*, and it is D(1)/(R - 1) - 1 in the island model.
lations at all examples. Examples of populations made of two types of subpopulations in the same geographical area and showing such a pattern of differentiation include host races of the fruit fly Rhagoletis pomonella, which show a substantial excess differentiation between races at some loci (Feder et al. 1990) in spite of large transition rates (6% in Feder et al. 1994). There is independent evidence that selection is responsible for the patterns of differentiation at these loci (Feder et al. 1997). Other examples may be found in Ross and Shoemaker (1993) or Chevillon et al. (1998). Because the reverse observation (that between-class differentiation should be intermediate between within-class differentiations) is a robust expectation of the neutral model for a wide range of parameter values, the comparison of within- and between-class differentiations may be an efficient way to detect habitat-related divergent selection at some loci under investigation or at linked loci.

I thank M. Lacoux, Y. Michalakis, and O. Ronce for comments on various versions of this manuscript. The reviewers of a previous manuscript and J. Hey also provided some very helpful comments. This work was supported by the Service Commun de Biostatistique de Montpellier and grants from the Région Languedoc-Roussillon (963223) and Centre National de la Recherche Scientifique (ACC SV3 9503037). This is paper 98-119 of the Institut des Sciences de l’Évolution.

LITERATURE CITED


Communicating editor: J. Hey

APPENDIX

Let λ1, λ2, and λ3 be the eigenvalues of A. Consider the spectral representation of P (e.g., Karlin and Taylor 1975): P = XD−1, where D is a diagonal matrix with elements

\[ d_i = \frac{\gamma \lambda_i}{1 - \gamma \lambda_i} \]

(A1)

X contains right eigenvectors, and X−1 contains left eigenvectors of A.

If all dispersal distributions are identical with characteristic function ψ, then it is found that for λ = (1 − v_11 − v_12)ψ_i, d_i = γλ_i^1/(1 − γλ_i^1), and

\[ d_i = \frac{\gamma \lambda_i}{1 - \gamma \lambda_i} \]

(A1)
\[
\mathbf{P} = \frac{1}{(\nu_{12} + \nu_{21})^2} \\
\left( \begin{array}{c}
\nu_1 d_1 + 2\nu_1 \nu_{21} d_2 + \nu_1 d_3 & \nu_1 (d_1 - 2d_2 + d_3) \\
\nu_1 (d_1 - 2d_2 + d_3) & \nu_1 (d_1 - 2d_1 + d_2) \\
\nu_2 (\nu_{21} d_2 - d_3) & \nu_2 (\nu_{21} d_1 - d_3) \\
+ \nu_2 (d_2 - d_1) & + \nu_2 (d_2 - d_3)
\end{array} \right)
\] (A2)

Equation 5 shows that the third column of \(\mathbf{P}\) is irrelevant. From Equations 5 and 9, the ratio of gene diversities is
\[
R = \frac{1 + L_0(p_{11} - p_{21})/(2N_1)}{1 + L_0(p_{22} - p_{12})/(2N_2)}
\] (A3)

where the \(p_{ij}\)'s are elements of \(\mathbf{P}\). In the symmetric dispersal model, \(p_{11} - p_{21} = (\nu_{21}(2d_2 - d_1) + \nu_{21} d_3)/(\nu_{21} + \nu_{12})\) and \(p_{22} - p_{12} = (\nu_{12}(2d_2 - d_3) + \nu_{21} d_3)/(\nu_{21} + \nu_{12})\). Assuming \(\nu_{21} > \nu_{12}\) throughout, \(p_{11} - p_{21} > p_{22} - p_{12} \geq 0\) and \(p_{11} - p_{21} < (p_{22} - p_{12})\nu_{12}/\nu_{21}\). Hence, (i) \(R > 1\) when \(N_1 = N_2\); (ii) if \(\nu_{21} = \nu_{12}\) the ratio is \((2 + L_0(d_2)/N_2)/(2 + L_0(d_2)/N_2) > 1\) when \(N_2 > N_1\); and (iii) finally, there is no source and sink when \(N_1 = N_2\nu_{21}/\nu_{12}\). Then

\[
R = \frac{2N_2 + L_0(p_{11} - p_{21})\nu_{12}/\nu_{21} < 1.}{2N_2 + L_0(p_{22} - p_{12})}
\] (A4)

Hence in this case there is no source or sink, and identity is higher in class 2, which has a smaller subpopulation size, as expected.

In the symmetric dispersal model \(Z(r)\) simplifies to
\[
Z(r) = \left( \frac{1}{2N_1} + \frac{R}{2N_2} \right) L_0(d_2)
\]
\[
= \left( \frac{1}{2N_1} + \frac{R}{2N_2} \right) \frac{(1 - \nu_{21} - \nu_{12})^2}{2(\nu_{21} + \nu_{12}) + 1 - \psi^2}. \] (A5)

The decrease of the inverse transform \(L_0(1/[2(\nu_{21} + \nu_{12}) + 1 - \psi^2])\) with distance is a function of \(\nu_{21} + \nu_{12}\) (see Malecot 1950; Nagylaki 1976; Sawyer 1977; \(\nu_{21} + \nu_{12}\) plays the role of the mutation rate in these models). The decrease of \(L_0(d_2)\) is also a function of \(\nu_{21} + \nu_{12}\) \(\alpha_s\) in the symmetric dispersal model, and a complex expression could be obtained for \(D(r)\) involving \(L_0(d_2)\) and \(L_0(d_3)\). Such an expression is not used here, but we investigated the accuracy of asymptotic results for the increase of differentiation with distance and the value of \(D(r)\) at distance \(2\alpha_s/\sqrt{\nu_{21} + \nu_{12}}\) in the numerical examples.