doi:10.1111/j.1558-5646.2009.00891.x

A STRUCTURED COALESCENT PROCESS FOR SEASONALLY FLUCTUATING POPULATIONS

Max Shpak,[1,2] John Wakeley,[3,4] Daniel Garrigan,[5,6] and Richard C. Lewontin,[7,8]

[1]*Department of Biological Sciences, University of Texas at El Paso, El Paso, Texas 79968*

[2]*E-mail: mshpak@utep.edu*

[3]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138*

[4]*E-mail: wakeley@fas.harvard.edu*

[5]*Department of Biology, University of Rochester, Rochester, New York 14627*

[6]*E-mail: daniel.garrigan@rochester.edu*

[7]*Museum of Comparative Zoology, Harvard University, Cambridge Massachusetts, 02138*

[8]*E-mail: lewontin@oeb.harvard.edu*

Received July 30, 2009
Accepted October 16, 2009

Many short-lived organisms pass through several generations during favorable growing seasons, separated by inhospitable periods during which only small hibernating or estivating refugia remain. This induces pronounced seasonal fluctuations in population size and metapopulation structure. The first generations in the growing season will be characterized by small, relatively isolated demes whereas the later generations will experience larger deme sizes with more extensive gene flow. Fluctuations of this sort can induce changes in the amount of genetic variation in early season samples compared to late season samples, a classical example being the observations of seasonal variation in allelism in New England *Drosophila* populations by P.T. Ives. In this article, we study the properties of a structured coalescent process under seasonal fluctuations using numerical analysis of exact state equations, analytical approximations that rely on a separation of timescales between intrademic versus interdemic processes, and individual-based simulations. We show that although an increase in genetic variation during each favorable growing season is observed, it is not as pronounced as in the empirical observations. This suggests that some of the temporal patterns of variation seen by Ives may be due to selection against deleterious lethals rather than neutral processes.

KEY WORDS: **Coalescence, *Drosophila*, heterozygosity, metapopulations, migration.**

The populations of many organisms are subject to seasonal fluctuations. This is particularly true of short-lived, semelparous species in which individuals forming small refugia populations hibernate or estivate during inhospitable periods of the year, and subsequently emerge to found the first of several generations at the start of the growing season. Because these founding demes are small and relatively isolated, gene flow in the first generations will be limited and local genetic drift will be strong. With abundant resources during the growing season, the population size of each deme (consisting of descendants of each refugium) is expected to increase, and along with it, the extent of dispersal and gene flow between the demes.

As with many insects in temperate climates, the North American populations of *Drosophila melanogaster* follow the pattern above, and serve as a well-studied model organism with seasonal population dynamics that are in many ways representative of the model. In a series of publications spanning four decades, P.T. Ives and his colleagues (e.g., Ives 1945, 1954; Band and Ives 1961; Ives and Band 1986) undertook a detailed analysis of geographic and temporal variation of allelism (defined by the frequencies of distinct segregating lethal alleles) of *Drosophila* at various locations in the United States, particularly in the vicinity of South Amherst, Massachusetts. Among the phenomena observed in the populations was a decline of allelism for lethals in populations

sampled just before the first frosts (late autumn in New England) when compared to populations sampled in the early spring, implying a seasonal increase in heterozygosity.

In each population sampled by Ives, lines containing lethal alleles were identified and crossed with one another to determine whether the lethal alleles were identical in the F2 generation. This served as a proxy measure for the degree of allelism, and therefore the inbreeding coefficient for each deme. Ives discovered that although the frequency of lethal alleles increased between the first generations of the early spring to the last generations in the late fall, allelism declined significantly. For example, populations surveyed in early June of 1969 showed 16.1% allelism whereas those sampled in late October had 0.3% allelism (Ives 1970). The extent of change in allelism varied from year to year (e.g., 4.9% vs. 0.6% in 1966, 34.1% vs. 4.0% in 1968), but all seasonal changes in allelism were statistically significant.

Ives interpreted these results in the following way: the increase in lethals and semilethals was primarily a consequence of the accumulation of new mutations in each generation of the growing season. Given a generation time of approximately 3 weeks in *Drosophila*, there would be up to 10 generations between the spring and fall in New England and other temperate areas, with potentially more generations per season in the southern populations. In addition, this would have been facilitated by population growth, which would decrease the efficacy of selection against lethals and semilethals due to reduced inbreeding.

The question of declining allelism could be accounted for by the life cycle, described above, of adult flies overwintering in isolated refugia then emerging in the spring to form small demes. The subsequent generation(s) of flies would suffer from increased inbreeding, with relatively high levels of lethal allelism. With each generation in the growing season, the size of each deme increases (conservatively, by a factor of 2 to 3 in flies). Larger population sizes lead to increased gene flow for a number of reasons. First, larger populations exchange a greater absolute number of migrants even if the probability of migration remains constant. Perhaps more importantly, the migration rates themselves increase with the size of each deme as its boundary approaches and starts to overlap with other demes. Furthermore, larger deme sizes lead to greater population densities, which are known to induce greater dispersal in organisms as a consequence of resource competition. In the extreme case, it is reasonable to assume that toward the end of the growing season, the network of demes approaches the structure of a single, effectively panmictic population.

The life cycles of many other organisms in seasonal environments may differ from *Drosophila* in the details. Some may overwinter as seeds, eggs, or larvae rather than as adults, some may have longer or shorter generation times, or different values of intrinsic population growth. However, the general phenomenon of small refugia leading to isolated but subsequently expanding and overlapping populations may be quite common for a variety of organisms in environments where temperature or precipitation is subject to sharp seasonal changes. Therefore, it is important to investigate the dynamics of changes in allele frequency induced by seasonal life cycles, particularly with respect to deviations from the predictions of commonly used time-homogeneous neutral models.

This article focuses on the degree to which evolution in fluctuating populations differs from conventional neutral models. In general, models of neutral evolution assume nearly constant population size. Because increases in population size or migration rate in metapopulations may lead to changes in allele frequency that are not consistent with such neutral models, it is critical to take these differences into consideration. Otherwise, the changes in allele frequency distribution induced by the population dynamics may be spuriously attributed to the action of natural selection.

The analysis below will focus on the coalescence times and expected heterozygosities that arise as a result of the seasonal expansion and migration of demes founded from small refugia. It will be argued based on qualitative comparison with numerical models that the pronounced differences in allelism (and consequently, of effective population size, inbreeding coefficients, and heterozygosity) observed by Ives and others appear unlikely to be a consequence of seasonal fluctuations in population size and migration rate alone.

In the following section, we present an idealized model of seasonal metapopulation dynamics that takes into consideration a number of contingencies in a migration pattern. Because our interest lies in determining the degree to which seasonal fluctuations lead to deviations from the predictions of the neutral theory in constant populations, the analysis will emphasize how the expected coalescence times within and between demes, as well as intra and interdemic heterozygosity, are influenced by seasonal variation in deme size and migration rates.

The Model

The population processes we consider can be represented by and idealized scheme of population growth, seasonal decline, migration, and mutation in a finite island model (e.g., Wright 1931; Maruyama 1970a,b). As in a conventional island model, we assume that the demes are equivalent to one another in terms of both population size and migration rate. In other words, there is no explicit distance or spatial structure, so that any particular deme has an equal probability of exchanging a migrant with every other deme in the metapopulation. However, in contrast to the usual island model, here it is assumed that deme sizes and ultimately migration rates are subject to seasonal changes, as shown schematically in Figure 1.

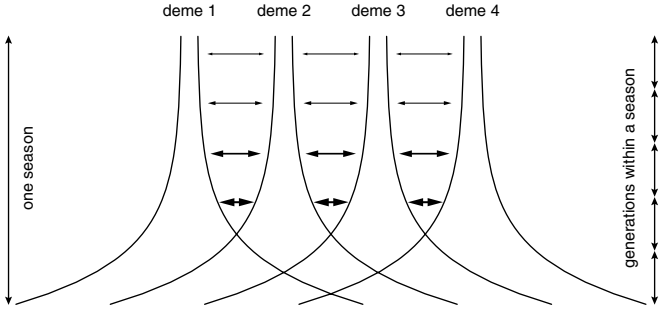

Figure 1. The schematic illustrates the population dynamics assumed in the Ives model, in which initially small and isolated demes expand during the course of a season to form a large metapopulation network with extensive gene flow between subpopulations.

In this "Ives model," there are k generations in every growing season that are separated by a larger break, due to winter, drought, or any other environmental condition not conducive to the organism. For each of the D demes, the population size is set to N_g in the gth generation, where N_g is assumed to be minimal at $g = 1$ and maximal at $g = k$. The migration rate m_g is similarly subject to change across generations, and we typically consider it to be increasing with population size.

As with the standard structured coalescent models of metapopulations (Slatkin 1987, 1991; Hey 1991; Wakeley and Aliacar 2001; Wakeley 2004), the reverse time process for a two allele sample in a diploid population (taken from the same or different demes) can be conceived in terms of three states. In state 1, the two alleles are in the same deme but are not yet coalesced. In state 2, the two alleles are in different demes. In state 3, the two alleles have coalesced, meaning that the most recent common ancestor of the two alleles in the sample has been reached.

In a metapopulation with D demes, each generation's process of coalescence during a "growing season" is a Markov chain defined by a matrix of the form

$$\mathbf{M}^{(g)} = \begin{pmatrix} a\left(1-\frac{1}{2N_g}\right) & 1-a & \frac{a}{2N_g} \\ b\left(1-\frac{1}{2N_g}\right) & 1-b & \frac{c}{2N_g} \\ 0 & 0 & 1 \end{pmatrix} \tag{1a}$$

in which the superscript (g) represents a particular generation within a season, and m_g and N_g are the migration rates and deme sizes in generation g. The coefficients of the matrix are defined as

$$a = \left((1-m_g)^2 + \frac{1}{D-1}m_g^2\right),$$
$$b = \left(2m_g(1-m_g)\frac{1}{D-1} + m_g^2\frac{D-2}{D-1}\frac{1}{D-1}\right).$$

As with the standard structured coalescent model, the entries M_{ij} represent the transition probabilities from state i to state j in the previous generation.

The entries in the matrix in equation (1a) are derived under the assumption that migrant individuals cannot return to their home deme. If members of the migrant pool can do so, we can obtain the corresponding matrix by substituting $m_i(D-1)/D$ in place of m_i, to account for the fact that only a fraction $1 - 1/D$ actually leave the source deme (Latter 1973; Takahata and Nei 1984). Thus

$$\mathbf{M}^{(g)} = \begin{pmatrix} \alpha\left(1-\frac{1}{2N_g}\right) & 1-\alpha & \frac{\alpha}{2N_g} \\ \beta\left(1-\frac{1}{2N_g}\right) & 1-\beta & \frac{\beta}{2N_g} \\ 0 & 0 & 1 \end{pmatrix}, \tag{1b}$$

where

$$\alpha = \left((1-m_g)^2 + \frac{1}{D}\left(1-(1-m_g)^2\right)\right),$$
$$\beta = \frac{1}{D}\left(1-(1-m_g)^2\right).$$

Although perhaps not a realistic model for *Drosophila* dispersal, the matrix in equation (1b) gives a reasonable description for organisms with "broadcast" reproduction, such as many plants and marine invertebrates. Furthermore, the difference between the matrices in equations (1a) and (1b) is trivial when D is large.

Briefly, the terms in the first matrix rows are defined as follows: M_{11} is the probability that two alleles in the same deme now were also together in one deme in the previous generation, M_{12} is the probability that two alleles in the same deme now were in two different demes in the previous generation, and M_{13} the probability that two alleles in the same deme now are derived from a common ancestor in the previous generation (i.e., they coalesced). The second row terms define transition probabilities for alleles currently in different demes, having either come from a single deme in the previous generation without coalescing, having been in separate demes, or having coalesced. The factor of $1/D$ in equation (1b) represents the probability of that a migrant came from a particular deme. In the third row, coalescence is defined to be an absorbing state because we are interested in the ancestry of the sample only back to the most recent common ancestor.

Samples may be taken in any generation, but it is convenient to represent the overall process on a seasonal (or annual) rather than a generational timescale. Thus, time 0 means the end of the current growing season, and times 1, 2, etc., correspond to the ends of previous seasons. Over the course of a k-generation growing season, we have the transition matrix

$$\mathbf{\Pi_D} = \mathbf{M}^{(k)}\mathbf{M}^{(k-1)}\cdots\mathbf{M}^{(2)}\mathbf{M}^{(1)}, \tag{2}$$

where the subscript D signifies the dependence on the number of demes. In the Section "Separation of Timescales," we consider the limit $D \to \infty$.

The entries of the matrix $\mathbf{\Pi_D}$ represent the backward-time transition probabilities over all k generations (one season or one year) rather than over a single generation. The matrix $\mathbf{\Pi_D}$ can be derived by multiplying $\mathbf{M}^{(2)}\,\mathbf{M}^{(1)}$ and showing by induction that the entries shown below describe the dynamics over all k time steps.

$$\mathbf{\Pi_D} = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} \\ 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

Assuming that D is large and ignoring terms of with factors of the order D^{-1}, the first row transition probabilities are

$$\Pi_{11} = \prod_{g=1}^{k}(1-m_g)^2\left(1-\frac{1}{2N_g}\right) + O(D^{-1}),$$

$$\Pi_{12} = \sum_{g=1}^{k}\left(1-(1-m_g)^2\right)\prod_{j=1}^{g-1}(1-m_j)^2\left(1-\frac{1}{2N_j}\right) + O(D^{-1}),$$

$$\Pi_{13} = \sum_{g=1}^{k}(1-m_g)^2\frac{1}{2N_g}\prod_{j=1}^{g-1}(1-m_j)^2\left(1-\frac{1}{2N_j}\right) + O(D^{-1}).$$

The interpretation of these leading $O(1)$ terms in the entries of the first row in $\mathbf{\Pi_D}$ is as follows: Π_{11} is the probability of no migration and no coalescence for either alleles in a single deme sample over k generations, Π_{12} is the probability that a migration event occurred in some particular generation g but not in any other generation whereas Π_{13} is the probability that in some particular generation g of the k, a coalescent event occurred.

For alleles sampled from different demes, the second row entries are

$$\Pi_{21} = \frac{1}{D}\sum_{g=1}^{k}\left(1-(1-m_g)^2\right)\left(1-\frac{1}{2N_g}\right) \times \prod_{j=1}^{g-1}(1-m_j)^2\left(1-\frac{1}{2N_j}\right),$$

$$\Pi_{22} = 1 - \Pi_{21} - \Pi_{23},$$

$$\Pi_{23} = \frac{1}{D}\sum_{g=1}^{k}\left(1-(1-m_g)^2\right)\left(\frac{1}{2N_g} + \left(1-\frac{1}{2N_i}\right) \times \sum_{j=1}^{g-1}(1-m_j)^2\frac{1}{2N_j}\prod_{r=j+1}^{g-1}(1-m_r)^2\left(1-\frac{1}{2N_r}\right)\right).$$

Note that we keep the $O(D^{-1})$ terms in the second row of $\mathbf{\Pi_D}$. The terms represent exact transition probabilities, and include such second-order terms as Π_{23}, the coefficient for simultaneous migration and coalescence. We need to retain these $O\left(D^{-1}\right)$ terms because the ancestral process depends on samples being able to escape from state 2 so that they can coalesce. This is part of our separation of timescales argument, which we present in detail below.

The entry Π_{21} gives the probability that a migration event occurred specifically in the gth generation of the k, such that both alleles are in the same deme in the previous season but have not coalesced. Π_{23} is the probability that the two alleles sampled from different demes coalesce within the span of a single season. This can occur only if one or both alleles is a migrant in some generation (hence the sums over g) and the two alleles came from the same deme (hence the factor $1/D$); once in the same deme, the two alleles either coalesced in that generation (hence the probability $1/2N_g$) or in some subsequent generation (hence the sum over j and product over r). The entry Π_{22}, which is most easily computed as the difference between unity and the remaining terms in this row, is the probability that two alleles sampled in separate demes remain so through the duration of k generations. Note that they are required to remain in separate demes only with respect to one another; they may migrate.

The transition matrix $\mathbf{\Pi_D}$ can be greatly simplified if $N_g \gg 1$ and $m_g \ll 1$, for every $1 \le g \le k$ (below, we use m and N to mean any m_g and N_g). In this case, all terms of the order m^2, m/N, and N^{-2} can be ignored. The products then reduce to sums as quadratic terms are discarded, and we obtain the following approximation to $\mathbf{\Pi_D}$ for small migration rates and large population sizes.

$$\mathbf{\Pi}_D \approx \begin{pmatrix} 1-2\sum_{g=1}^{k}m_g - \sum_{g=1}^{k}\frac{1}{2N_g} & 2\sum_{g=1}^{k}m_g & \sum_{g=1}^{k}\frac{1}{2N_g} \\ \frac{2}{D}\sum_{g=1}^{k}m_g & 1-\frac{2}{D}\sum_{g=1}^{k}m_g & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

$\mathbf{\Pi}_D$ is distinct from the standard migration-coalescence approximation matrix insofar as the values m_g and N_g may change (e.g., increasing in the "Ives model") from the gth generation to the next. The reason that the term $\Pi_{23} \to 0$ is that this transition involves not only a migration event that brings the two alleles into one deme (which it has in common with the transition from state 2 to state 1) but also requires a coalescent event to occur during that same season. Therefore Π_{23} is approximately N times smaller than Π_{21}, and becomes negligible when N is large. In other words, when m and $1/N$ are small, a sample of two alleles from different demes must, to first-order approximation, go through at least two seasonal cycles to coalesce: one for migration and one for coalescence.

SEPARATION OF TIMESCALES

The exclusion of D^{-1} terms in the first row and their inclusion in the second row of equation (4) illustrates a key aspect of the

ancestral process that emerges when the number of demes is very large: the separation of timescales into a fast and a slow component. The slow timescale has a relaxation time proportional to D/m generations due to Π_{21}, whereas the fast timescale has a relaxation time proportional to N or $1/m$ generations, due to the magnitude of the entries in the first row of $\mathbf{\Pi_D}$. Recall that $\mathbf{\Pi}_D$ gives the dynamics over k generations. The slow dynamics depend on two alleles being placed (backward in time) into the same deme (forward in time we would say they came from the same deme) whereas the fast dynamics involve coalescence for alleles already in the same deme, and migration that does not place two allele into the same deme.

We now consider the limit $D \to \infty$, so that this separation of timescales becomes extreme. The method we use was made rigorous by Möhle (1998, 2000). Its interpretation for samples of size two from (time-homogeneous) metapopulations is discussed in Wakeley (2004, 2008 Ch. 6). The matrix $\mathbf{\Pi_D}$ in equation (2) can be decomposed into the sum of two matrices plus some small errors when D is large. Specifically,

$$\mathbf{\Pi_D} = \mathbf{A} + \frac{1}{D}\mathbf{B} + O\left(D^{-2}\right),$$

in which the matrices

$$\mathbf{A} = \lim_{D\to\infty} \mathbf{\Pi_D},$$

$$\mathbf{B} = \lim_{D\to\infty} D\left(\mathbf{\Pi_D} - \mathbf{A}\right),$$

specify the dynamics on the fast timescale and the slow timescale, respectively.

The matrices $\mathbf{A}$ and $\mathbf{B}$ are messy to write down in general, but under the assumption that all N_g are sufficiently large and all m_g are sufficiently small, they can be approximated (recall eq. 4). Thus

$$\mathbf{A} \approx \begin{pmatrix} 1 - 2m_e - \frac{1}{2N_e} & 2m_e & \frac{1}{2N_e} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{5}$$

where m_e and N_e are the effective migration rates and the effective population sizes, over one whole season, resulting from the changes in deme size. We have

$$m_e = \sum_{g=1}^{k} m_g, \quad N_e = \frac{1}{\sum_{g=1}^{k} \frac{1}{N_g}}.$$

The effective migration rate is proportional to its arithmetic mean (multiplied by the number of generations per season), whereas the effective population size is similarly proportional to the harmonic mean. The matrix $\mathbf{A}$ therefore includes the probabilities of all events that have a nonnegligible probability of being observed over the course of a single season.

The matrix defining the slow process, that is, events that are only likely to be observed on the longer timescale of D/m_e seasons, is approximated by

$$\mathbf{B} \approx \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ 2m_e & -2m_e & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{6}$$

Note that the terms in the first row above were omitted in equation (4). As we show later, their effect on the ancestral process becomes negligible when D is large. The reason for this is that, for each of these events, there is a nonnegligible probability in the first row of the single-season matrix $\mathbf{A}$. For these events, the entries in $\mathbf{A}$ will dominate the ancestral process when D is large.

We will now derive the matrix that characterizes the ancestral process over the long timescale. The result will be heuristic because we will use $\mathbf{A}$ and $\mathbf{B}$ as given in equations (5) and (6), thus employing the large N_e, small m_e approximation informally, rather than by taking additional limits. Below, we will test our results against numerical calculations and simulations.

Because of the fast–slow dynamic, the transitions induced by $\mathbf{A}$ over one season will approach a steady-state distribution when viewed over the timescale of the slow process ($\propto D/m_e$ seasons) when D is large. Thus, the transitions captured in $\mathbf{A}$ affect the ancestral process only via the limiting matrix

$$\mathbf{P} = \lim_{r\to\infty} \mathbf{A}^r$$

which gives the net effect of the fast dynamic over the long timescale. Applying the previous definitions, we have

$$\mathbf{P} = \begin{pmatrix} 0 & \frac{2m_e}{2m_e + \frac{1}{2N_e}} & \frac{\frac{1}{2N_e}}{2m_e + \frac{1}{2N_e}} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & p_m & p_c \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{7}$$

where p_m and p_c are, respectively, the probabilities that a sample in the state 1 (two alleles in the same deme) will enter the slow process represented by $\mathbf{B}$ either in state 2 by a migration event or in state 3 by a coalescence event. These are also the entries of the first eigenvector for the transition matrix $\mathbf{A}$. Note that it will take on the order of $(2\,m_e + 1/2\,N_e)^{-1}$ seasons for either a migration event or a coalescent event to occur.

It was shown by Möhle (1998) that, given the above decomposition, time can be rescaled by D and the limit taken so that

$$\mathbf{V}(t) = \lim_{D\to\infty} (\mathbf{\Pi}_D)^{Dt} = \mathbf{P}e^{\mathbf{PBP}t} \tag{8}$$

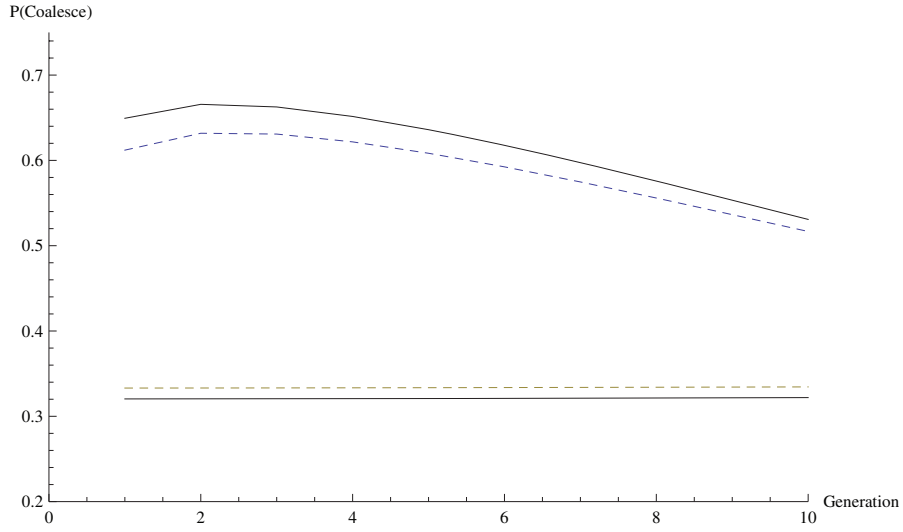

Figure 2. A comparison of coalescence probabilities computed from our Möhle approximation with those calculated from iteration of exact matrices over 10 generations, with migration rates and deme sizes in each season given by the vectors m and N (as defined in eqs. 12 and 13). The two lower curves compare the approximate and exact interdemic coalescence times Q_{23} (solid) versus $\widetilde{Q}_{23}$ (dashed). The two upper curves compare estimated and exact intrademic coalescence times, Q_{13} (solid) versus $\widetilde{Q}_{23}$ (dashed) for the same parameters.

gives the transition probabilities over a time of length t, now measured in units of D seasons. Taking the product, we have

$$\mathbf{PBP} = \begin{pmatrix} 0 & p_m\,(b_{21}\,p_m + b_{22}) & p_m\,(b_{21}\,p_c + b_{23}) \\ 0 & b_{21}\,p_m + b_{22} & b_{21}\,p_c + b_{23} \\ 0 & 0 & 0 \end{pmatrix}, \quad (9)$$

in which the b_{ij} terms are the entries of $\mathbf{B}$. It can be seen that this matrix product, and thus also $\mathbf{V}(t)$, depends solely on the second row terms of $\mathbf{B}$. This is because the slow dynamic principally consists of the process of waiting for a pair of alleles in different demes to come together into a single deme whereas other transitions operate on a much faster scale and are summarized by the matrix $\mathbf{P}$.

In all, from equations (8) and (9), when D is large we may approximate repeated iterations of the exact transition matrix $\mathbf{\Pi} = \mathbf{M}^{(k)} \cdots \mathbf{M}^{(1)}$ with

$$\mathbf{V}(t) = \begin{pmatrix} 0 & \dfrac{4N_e m_e}{4N_e m_e + 1} e^{-t\frac{2m_e}{4N_e m_e + 1}} & 1 - \dfrac{4N_e m_e}{4N_e m_e + 1} e^{-t\frac{2m_e}{4N_e m_e + 1}} \\ 0 & e^{-t\frac{2m_e t}{4N_e m_e + 1}} & 1 - e^{-t\frac{2m_e}{4N_e m_e + 1}} \\ 0 & 0 & 1 \end{pmatrix}. \quad (10)$$

The transition probabilities for some arbitrary generation j within a season, assuming that t seasons have already elapsed, can be estimated by the product of the exact matrices for the first j generations, then using $\mathbf{V}(t)$ above. Defining the transition matrices for the entire season as $\mathbf{\Pi}$ (eq. 2) and up to the jth generation as

$$\mathbf{W}^{(j)} = \mathbf{M}^{(j)}\mathbf{M}^{(j-1)} \cdots \mathbf{M}^{(1)}. \quad (11)$$

The entries of the exact probability matrix

$$\widetilde{\mathbf{Q}}^{(j)}(t) = \mathbf{W}^{(j)}\mathbf{\Pi}^t$$

can be compared to that of our Möhle approximation (which assumes that D is large, but also that all m_g are small and all N_g are large) weighted by exact transition matrices in the jth generation,

$$\mathbf{Q}^{(j)}(t) = \mathbf{W}^{(j)}\mathbf{V}(t).$$

We made calculations of matrix coefficients over $k = 10$ generations, with an initial migration rate of 0.001 that increases to 0.01 in the second generation and subsequently increases arithmetically over units of 0.005. We assumed $D = 100$ demes, each initially containing five diploid individuals, after which the deme size increases by a factor of 3 in every generation. The vectors characterizing migration rates and deme sizes over the 10 generations are thus

$$\mathbf{m} = \{0.001, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05\}, \quad (12)$$

$$\mathbf{N} = \{5, 15, 45, 135, 405, 1215, 3645, 10935, 32805, 98415\}. \quad (13)$$

These migration fractions and population size values will be used in later sections for comparison of predicted coalescence times and heterozygosities with simulation and numerical results.

The upper curves in Figure 2 compare estimated intrademic coalescence probabilities Q_{13} to the exact probabilities computed

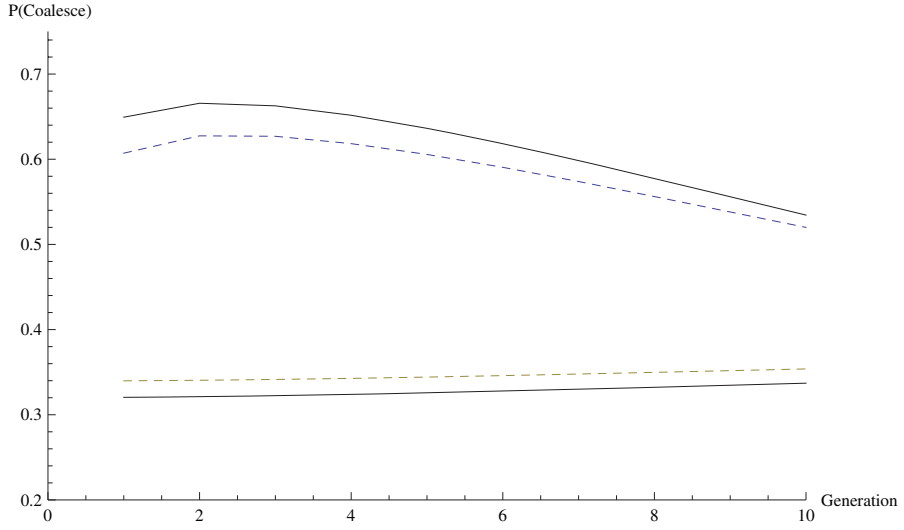

Figure 3. As in Figure 2, but for $D = 10$ demes, illustrating the greater discrepancy between the analytical estimates and numerical calculations from the exact model when D is an order of magnitude smaller.

from equation (1), $\widetilde{Q}_{13}$, which are plotted for $j = 1, \ldots, 10$ generations to show their changes during the course of a season. The two lower curves make the same comparison for coalescence probabilities of interdemic samples, that is, Q_{23} versus $\widetilde{Q}_{23}$. The calculations are done over the 10 generations following $t = 200/D$, corresponding to 200 seasons.

As Figure 2 shows, with these **m** and **N**, the seasonal fluctuations for intrademic coalescence probabilities Q_{13} have a range in magnitude of the order 0.1 whereas those for interdemic samples Q_{23} are two orders of magnitude smaller, at 0.001. This is due to the fact that for an interdemic coalescence to occur, the allele pair must first have migrated to the same deme, which occurs with a probability that scales with $1/D$ and thus is relatively unlikely to occur in a single season when D is very large (here 100). The transition probabilities for intrademic samples, on the other hand, are strongly influenced by the sizes of demes in the current season.

The close correspondence between our Möhle approximation and the exact transition probabilities under a fluctuating deme size holds even for the moderately high migration rates and a very small initial population size, provided that the effective rates m and N are sufficiently small and large, respectively. The other critical assumption in the approximation is that D is large enough to allow a timescale separation between the process of migration (that takes the sample from state 2 to state 1) and process of migration and coalescence from state 1. In Figure 3, the same quantities are compared as in the previous figure, but for only $D = 10$ demes. Due to the smaller number of demes, the seasonal fluctuations in Q_{23} are more pronounced, being only a single order of magnitude smaller than those for Q_{13}.

Even though the estimated coalescence probabilities are not quite as close to the exact values as they were for the 100 deme population in Figure 2, they still provide reasonable first-order estimates. With these caveats, we now use our Möhle approximation to compute the expected time to most recent common ancestors and heterozygosities in a two-allele sample for a structured coalescent with seasonal fluctuations, that is, for the Ives model.

Coalescence Times and Heterozygosity

The entries of the matrix $\mathbf{V}(t)$ give the transition probabilities measured in units of D seasons. The term $V_{13}(t)$ gives the probability that an intrademic sample of two alleles has coalesced by time t (i.e., Dt seasons) in the past, and V_{23} gives the same for an interdemic sample. The probability densities are

$$\beta_{ij}(t) = \frac{dV_{ij}(t)}{dt},$$

from which we can compute the expected times to coalescence,

$$E[\widetilde{T}_w] = \int_0^\infty t\beta_{13}(t)dt = 2N_e \tag{14a}$$

for a within-deme sample, and

$$E[\widetilde{T}_b] = \int_0^\infty t\beta_{23}(t)dt = 2N_e + \frac{1}{2m_e} \tag{14b}$$

for a between-deme sample. The second (and higher) moments of coalescence times can be computed similarly, by integrating higher powers of t. The expectations in equations (14a, b) have the same form as those computed from standard models of migration and drift in Nei and Feldman (1972) and Li (1976), apart from having effective migration rates and effective population sizes in

place of constant census values. Note also that a factor of kD is needed to convert these expected times into generations.

For a sample of two alleles taken at an arbitrary generation during the growing season, coalescence times must also incorporate the ancestral process during the season in which the sample was taken. The expected coalescence times for a sample taken in generation g within a season can be calculated recursively over generations using the exact transition matrix $\mathbf{M}^{(g)}$. We have

$$\begin{aligned} E\left[T_w^{(g)}\right] = M_{13}^{(g)} + M_{12}^{(g)}\left(E\left[T_b^{(g-1)}\right]+1\right) \\ +M_{11}^{(g)}\left(E\left[T_w^{(g-1)}\right]+1\right), \end{aligned} \tag{15a}$$

$$\begin{aligned} E\left[T_b^{(g)}\right] = M_{23}^{(g)} + M_{22}^{(g)}\left(E\left[T_b^{(g-1)}\right]+1\right) \\ +M_{21}^{(g)}\left(E\left[T_w^{(g-1)}\right]+1\right), \end{aligned} \tag{15b}$$

Because time in equations (15a, b) is measured in units of generations, we have as the recursion base case $E[T_w^{(0)}] = kDE[\widetilde{T}_w^{(0)}]$ and $E[T_b^{(0)}] = kDE[\widetilde{T}_b^{(0)}]$, using the expectations given in equations (14a, b). The transition probability $M_{ij}^{(g)}$ is the entry in the ith row and jth column of $\mathbf{M}^{(g)}$, defined by either by equation (1a) or equation (1b) depending on the mode of migration.

Under the infinite-sites mutation model (Kimura 1969), the expected number of pairwise differences π is simply the twice product of the per generation mutation rate μ at the locus and the number of generations since the two alleles diverged from their most recent common ancestor. For interdemic and intrademic samples, respectively, we have

$$\pi_w(g) = 2\mu E\left[T_w^{(g)}\right], \quad \pi_b(g) = 2\mu E\left[T_b^{(g)}\right],$$

which we compare to numerical calculations and simulation results below.

Under the infinite-alleles mutation model (Kimura and Crow 1964), the relevant measure of neutral genetic variation for a sample of size two is the probability of identity by descent (IBD). Following the notation above, the probabilities of IBD for intrademic samples and interdemic samples are denoted F_w and F_b. These quantities can be computed using our Möhle-approximation transition matrix in equation (10). For an intrademic sample, we have

$$\begin{aligned} F_w &= \frac{1}{4N_e m_e + 1} + \frac{4N_e m_e}{4N_e m_e + 1}\int_0^\infty e^{-2\mu_e D t}\beta_{23}(t)dt \\ &= \frac{4DN_e m_e \mu_e}{4DN_e m_e \mu_e + D\mu_e + m_e}, \end{aligned} \tag{16a}$$

where we use $\mu_e = k\mu$ to denote the mutation rate over an entire season, which we assume is small. The term $1/(4N_e m_e + 1)$ in the first line is the probability of coalescence rather than migration for an intrademic sample, or $\mathbf{P}_{1,3} = p_c$ in equation (7). It captures the fact that on the timescale of D seasons, an intrademic sample may coalesce essentially instantaneously, only entering the slow process in the case of migration rather than coalescence. Note that this term, $1/(4N_e m_e + 1)$, does not include mutation. Thus, because the duration of the fast phase of the ancestry of an intrademic sample will be on the order of $(2m_e + 1/2N_e)^{-1}$ generations, we have assumed implicitly above that $\mu_e \ll (2m_e + 1/2N_e)^{-1}$. For any given μ_e, m_e, N_e, and D, equation (16a) will be involve some level of error, which we illustrate using numerical calculations and simulation results below. Interdemic samples enter the slow process directly, and we obtain

$$F_b = \int_0^\infty e^{-2\mu_e D t}\beta_{23}(t)dt = \frac{m_e}{4DN_e m_e \mu_e + D\mu_e + m_e}. \tag{16b}$$

by integrating the probability density of interdemic coalescence.

An independent, heuristic confirmation of these results can be obtained by solving for the expected heterozygosities from a single generation structured coalescent (e.g., Hudson 1991, 1998) and substituting effective migration rates, mutation rates, and deme sizes,

$$\begin{aligned} H_w &= \frac{2\mu_e}{2\mu_e + 2m_e + \frac{1}{2N_e}} + \frac{2m_e\frac{1}{D}}{2\mu_e + 2m + \frac{1}{2N_e}}H_w \\ &\quad + \frac{2m\frac{D-1}{D}}{2\mu_e + 2m + \frac{1}{2N_e}}H_b, \\ H_b &= \frac{2\mu_e}{2\mu_e + 2m_e} + \frac{\frac{2m_e}{D}(D-1)}{2\mu + 2m_e}H_b + \frac{\frac{2m_e}{D}}{2\mu + 2m_e}H_w. \end{aligned}$$

Solving and substituting, we obtain

$$H_w = \frac{4DN_e m_e \mu_e + 4DN_e \mu_e^2}{4DN_e m_e \mu_e + D\mu_e + m_e + 4DN_e \mu_e^2} \tag{17a}$$

$$H_b = \frac{4DN_e m_e \mu_e + D\mu_e + 4DN_e \mu_e^2}{4DN_e m_e \mu_e + D\mu_e + m_e + 4DN_e \mu_e^2}, \tag{17b}$$

which become equivalent to equations (16a, b) if one discards terms with factors with μ_e^2, and also using $H_w = 1 - F_w$ and $H_b = 1 - F_b$.

As with the coalescence times, the probabilities of IBD computed over the long timescale can be weighted by the transition probabilities in a given generation. The recursions for generation g are the weighted IBD values, multiplied by the probability of no mutation in the previous generation, so that

$$F_w^{(g)} = (1-\mu)^{2g}M_{13}^{(g)} + M_{12}^{(g)}F_b^{(g-1)} + M_{11}^{(g)}F_w^{(g-1)}, \tag{18a}$$

$$F_b^{(g)} = (1-\mu)^{2g}M_{23}^{(g)} + M_{22}^{(g)}F_b^{(g-1)} + M_{21}^{(g)}F_w^{(g-1)}, \tag{18b}$$

where the first term is the probability that no mutations have occurred during the first g generations of the season in which

the samples were taken. The heterozygosities for generation g are given by $H_w^{(g)} = 1 - F_w^{(g)}$ and $H_b^{(g)} = 1 - F_b^{(g)}$. From the intrademic and interdemic heterozygosities, one can also estimate the degree of population subdivision using the fixation index F_{ST} (Wright 1951) for generation g, similarly to what has been done in island models with constant population size and migration rates (e.g., Crow and Aoki 1984; Whitlock and McCauley 1999; Rottenstreich et al. 2007).

The expressions for heterozygosities and coalescence times in terms of mutation and migration rates (scaled over a season), effective population sizes, and numbers of demes, provides us with the tools for comparing the properties of coalescent processes in seasonally fluctuating populations to those of null models. In the following section, we will compute these values for metapopulations with varying degrees of seasonal fluctuation to assess whether the seasonality imposes significant deviations from the simple predictions of neutral theory.

Model Populations: Numerical Analysis

In the numerical examples analyzed below, we are not attempting to replicate the population dynamics in temperate *Drosophila* or any other particular system. In any case, the natural history of fruit fly populations in the wild is not sufficiently well known to provide anything beyond extremely rough guesses for the parameters of interest. Instead, we wish to examine the degree to which coalescence times and heterozygosities in model populations show seasonal changes that are qualitatively consistent with empirical observations, that is, whether the magnitude of fluctuations in allelism observed by Ives can be plausibly replicated in the absence of natural selection by changes in the migration rates and deme sizes alone.

To model seasonal fluctuations, we begin with $D = 100$ refugia and use the values of m and N given in equations (12) and (13). Each deme emerges with $N = 5$ diploid organisms. From one generation to the next, the deme size is assumed to grow geometrically by a factor of $R = 3$. Geometric increase is plausible for population growth, but it is much less clear how migration rates should changes through the season. Migration rates might depend on a number of ecological variables, such as increased dispersal due to density-dependent competition, or as a consequence of increased gene flow due to closer deme proximity. For simplicity, an arithmetic increase in migration rates will be assumed. The migration rate in the first generation is set to $m = 0.001$, and is increased by 0.005 in every subsequent generation. In the last generation, the migration rate is $m = 0.05$ and the deme size is $N = 196830$. Note that this final deme size is not given in equation (13). Genetic variation does not depend on it because only a small number (here $N = 5$) of these individuals survive to reproduce at the start of the next season. Finally, we use $\mu = 0.001$.

Following the notation that was used in the derivations, we have $N_e = 3.33$ and $m_e = 0.275$. The effective population size (per generation) is 33.3, that is, N_e multiplied by a factor of $k = 10$. Being proportional to the harmonic mean, N_e is more sensitive to small values than the effective migration rate, m_e, which is proportional to the arithmetic mean. However, for initially low values of either quantity, any estimate of coalescence times or heterozygosities at the end of a growing season may be very different than those estimated in other generations. Here, we will illustrate the degree to which estimates made at the beginning or the end of the season deviate from those in intermediate generations.

Following equations (14a, b), intrademic and interdemic samples of two alleles have expected coalescence times (measured in generations, and for the parameter values above), respectively,

$$E[T_w] = 2DN_e = 666.67, \quad E[T_b] = 2DN_e + \frac{D}{2m_e} = 851.17.$$

The coalescence times in a particular generation are evaluated from equations (15a, b) by weighting the seasonal estimates with migration probabilities in the jth generation.

In a model of migration where migrants cannot return to their deme of origin (i.e., the matrix in equation 1a) the coalescence times $E[T_w]$ and $E[T_b]$ are computed for each of the 10 generations of a season. Multiplying the coalescence times by a factor of $2\mu = 0.002$, gives the expected pairwise differences π_w and π_b. These are plotted in Figure 4, where the two upper curves show the (minimal) seasonal changes in π_b predicted from the exact model and the approximation whereas the two lower curves show the substantial increase in π_w predicted during the course of a season.

The numerical analysis was performed using *Mathematica* (Wolfram 1999), the scripts are available from the first author upon request.

The upper trajectories in the figures represents the coalescence calculated by exact iteration of the transition matrices $\mathbf{M}^{(i)}$, computed as weighted sums. The matrix $\mathbf{W}$ in equation (11) gives the exact probabilities for a specific generation after a given number of seasons t. The within and between deme coalescence times $E[\widetilde{T}]$ (on the seasonal timescale) are obtained numerically from the transition matrices as:

$$E[\widetilde{T}_w] = \sum_{t=1}^{\infty} t[W_{13}(t) - W_{13}(t-1)],$$

$$E[\widetilde{T}_b] = \sum_{t=1}^{\infty} t[W_{23}(t) - W_{23}(t-1)],$$

where $W_{ij}(t) = \mathbf{W}_{ij}^t$. Here, the coalescence times $\widetilde{T}$ are measured on a seasonal timescale to convert to generations, we have $T = k\widetilde{T}$. Defining $kE[\widetilde{T}] = E[T^{(0)}]$, we can obtain coalescence times

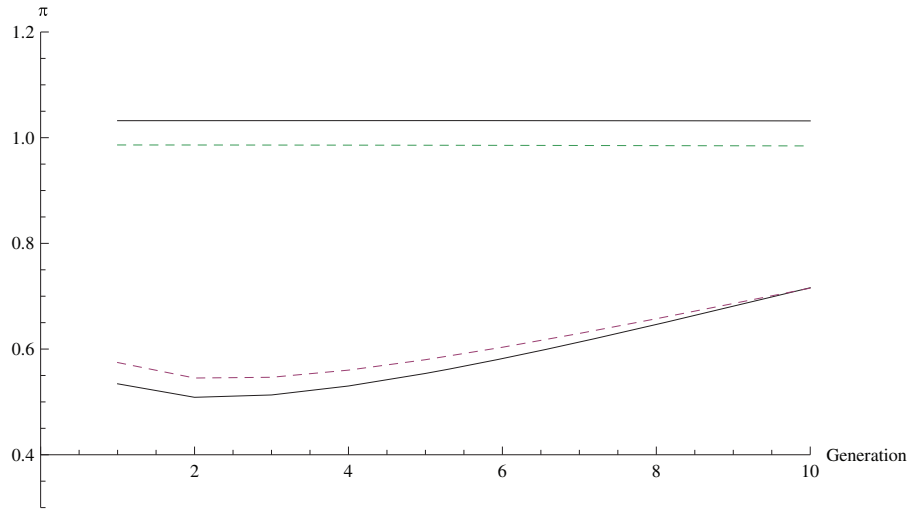

Figure 4. The two lower curves show the average pairwise genetic distances for intrademic samples π_w over 10 generations with population parameters given by m and N with $D = 100$ and $\mu = 0.001$ per generation. Of the two lower curves, the values obtained from the exact matrices are slightly larger than those from the Möhle approximation. The two upper curves show the same for interdemic samples π_b, with exact values (dashed) very slightly larger than the approximated pairwise distances (solid).

in generation g recursively from $M^{(g-1)}$, following equations (15a, b).

Interestingly, the increase in π_w with every generation is not monotonic. This is due to the fact that the maximal loss of genetic variation takes place when the first small populations of overwintering or estivating individuals emerge, mate, and reproduce. These first individuals inherit the genetic variation from the peak population before the bottleneck whereas subsequent generations become inbred, then recover as population size increases.

In comparing the trajectories for interdemic and intrademic coalescence times in Figures 3 and 4, it is apparent that the seasonal changes to migration rates and deme sizes affect the within deme coalescence times far more than the between deme coalescence times. Indeed, the difference between the maximal and minimal values of π_w are an order of magnitude greater than the same differences for π_b. The seasonal range in values for π_w during a course of a season is of the order 0.1 for these parameters, as opposed to a difference of 0.001 for π_b. The significantly greater seasonal changes for intrademic sample values is due to the fact that fluctuations in π_w are driven principally by M_{12}, which is of the order of the migration rate, so that as the migration rate increases, the fraction of individuals in a deme that were originally in an interdemic sample increases (and consequently so does the expected coalescence time). In contrast, the input from single deme allele pairs to π_b is determined by M_{21}, which is of the order of the migration rate divided by D. Therefore, the contribution of intrademic coalescence times to the expected value of π_b changes only negligibly with an increasing migration rate and deme size.

The heterozygosities computed from our Möhle approximation are also compared to values computed by numerically from the exact transition matrices,

$$\widetilde{H}_w = 1 - \sum_{t=1}^{\infty}(1-\mu_e)^{2t}[W_{13}(t) - W_{13}(t-1)],$$

$$\widetilde{H}_b = 1 - \sum_{t=1}^{\infty}(1-\mu_e)^{2t}[W_{23}(t) - W_{23}(t-1)].$$

As with the analytical approximations, the heterozygosity in generation g can be computed recursively from the $\mathbf{M}^{(g-1)}$ terms analogously to equations (18a, b).

The expected heterozygosities are shown in Figure 5, which plots the generational changes in the value of H predicted from equations (16), under the assumption of no return migration. The dashed lines show the heterozygosities calculated numerically, directly by matrix multiplication of the exact transition matrices. As with the coalescence times, the fluctuations of expected intrademic heterozygosities are an order of magnitude greater than those for interdemic samples.

In this case, the discrepancy between realized and predicted values of H_w is considerable—more than 3%—whereas the different values for interdemic heterozygosities are much closer to each other. This is principally due to the approximation used for the first term on the right-hand side of equation (16a), which requires a (seasonal) effective mutation rate much smaller than the effective migration rate. In these examples, effective mutation rate is not very small compared to the effective migration rate ($\mu_e = 0.01$ vs. $m_e = 0.275$), hence the discrepancy between our Möhle approximation and numerics even for $D = 100$. There is a nonnegligible probability that a mutation occurs prior to the first migration or coalescence event. Note that the discrepancy will

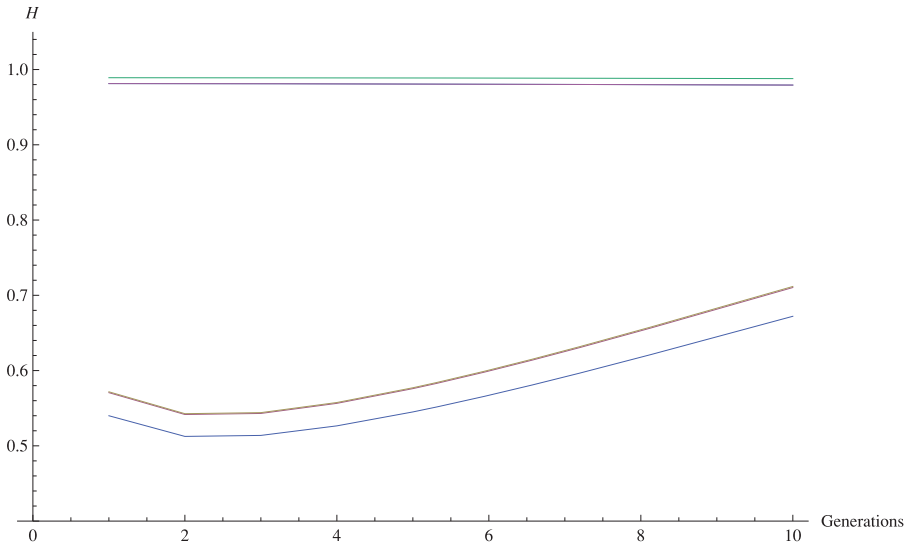

Figure 5. **This figure plots expected heterozygosities for the same population parameters as in Figure 4. The three nearly superimposed upper curves compare seasonal interdemic heterozygosities H_b computed from the Möhle approximation, the exact matrices, and the heuristic approximation from equation (17b). The three lower curves consist of two nearly identical trajectories (thick and dashed lines, respectively) representing heterozygosity H_w calculated from equations (17a) and from the exact transition matrices. The lower (solid) "outlier" curve shows seasonal H_w estimated from the Möhle approximation. The consistent deviation of the order of 0.05 is due to lack of sufficient timescale separation between mutation and migration/coalescence.**

be less significant for $E[\pi_w]$ and $E[\pi_b]$ because these depend linearly on the expected coalescence times, and the fast phase of the ancestry is a small fraction of the total as long as D is large (regardless of μ).

COMPARISON TO INDIVIDUAL-BASED SIMULATIONS

As an independent assessment of the predictions made from the analytical model, individual-based simulations of the Ives population dynamic were run. The C++ code was a slightly modified version of the ISITES program (Garrigan et al. 2009), which efficiently simulates the processes of mutation, migration and genetic drift in an infinite sites model by storing the genealogical history of the population. A documented copy of the modified C++ code is available from the first author upon request.

To replicate the model population represented in the Ives process there are D demes, each with a fixed initial number of individuals N_1 (at the beginning of every growing season, the size of each deme is reset to N_1). Over the course of the k generations in a growing season, the migration rates are specified by a length k input vector whereas the discrete population growth rate parameter R is constant. The mode of migration in the simulations assumes that there is no return migration to the deme of origin. Although the analytical model assumes that there is a fixed fraction of migrants m_i exchanged in each generation, in the simulations the number of migrants between demes is a Poisson random variable with rate parameter $m_i N_i$, so that different demes need not have identical population sizes during generations $2, \ldots, k$. Mutation is also modeled as a Poisson process with rate parameter μ. In addition, the deme sizes in the simulations are not fixed at the values of N_g given in equation (13). Instead, after migration, which leads to stochastic changes in deme size, each deme increases by the factor R.

Because of the computational intensity of this simulation, it is not practically feasible to execute for a large number of demes, so they were run with $D = 10$. The other parameters were as above: $k = 10$, $R = 3$, $\mu = 0.001$, and the arithmetically increasing migration rates are given in equation (12). The simulations were run for 5000 seasons (corresponding to 50,000 generations) in order for coalescence times and heterozygosities to approximate their steady-state values, and results were averaged over 500 independent replicates. Because the goal of this study is to examine the effects of population parameter changes across generations within a season, the output of the simulations was separated by generation ($g = 1, \ldots, 10$) within a season. Average pairwise differences were calculated from the site-frequency spectra, which is stored by default in the ISITES program. Heterozygosities were computed from samples of size 200 (20 from each of the 10 demes) when the deme sizes were sufficiently large to permit it, otherwise, they were computed over all individuals in a deme.

The pairwise differences converge within a few hundred seasons to their equilibrium values, as can be seen in Figure 6, which plots the mean values of π_w and π_b sampled in the 10th generation of every season.

The accuracy of the analytical approximations can be assessed by comparing the seasonal values of π_w and π_b during the final seasons of the simulation to the values predicted by

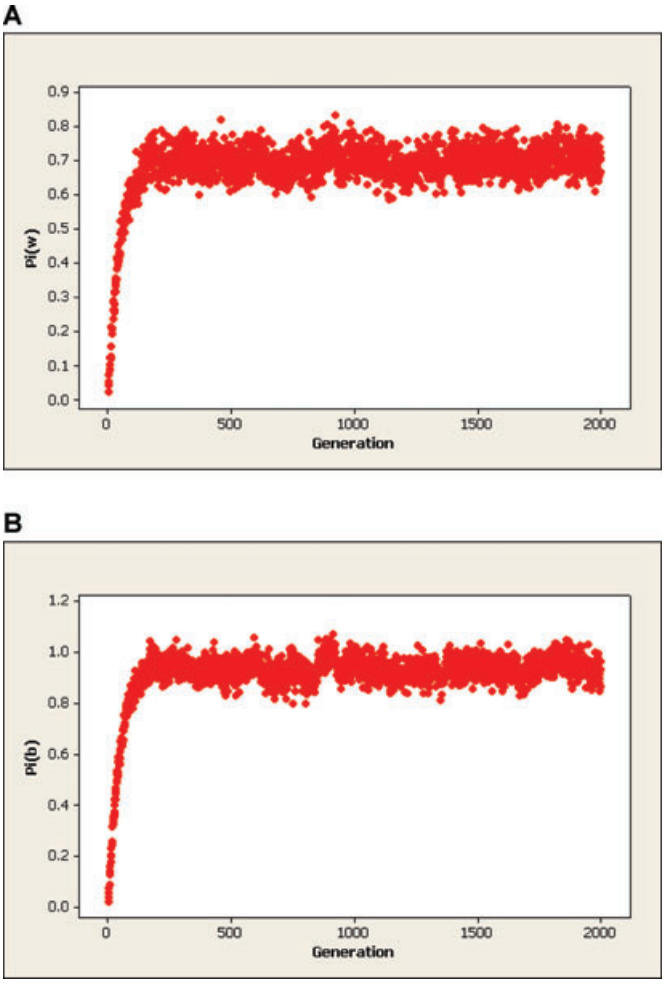

Figure 6. (A) Using the output of individual-based simulations, this figure plots the average intrademic pairwise genetic distances π_w in the last (10th) generation, over the course of 2000 seasons, illustrating the rapid approach to equilibrium values. (B) As in Figure 6A, but for expected pairwise distances π_b in an interdemic sample.

equations (15a, b) for the same set of parameters. Figure 7 plots π_w and π_b calculated from the simulations and compare them both to analytical estimates and brute force calculations of coalescence times based on iterations of the matrix in equation (1a). Figure 8 do the same for average heterozygosities for diallelic intra- and interdemic samples.

Because the simulations assume no return migration to the home deme, the migration rate in every generation for the analytical and numerical estimates is adjusted by multiplying by a factor $D/(D-1)$. In the absence of this migration rate correction, the trajectory for π_b predicted analytically would be greater by an approximate fraction of $1/D$ when compared to the numerical and simulation results. The effect on π_w, on the other hand, is negligible because $E[T_w]$ is independent of the migration rate. With this correction, equations (14)–(16) give qualitatively close approximations to simulated values of π_w.

There are a number of additional factors that contribute to the discrepancies between simulations and analytical predictions. Because of the large number of replicates, deviations due to stochasticity in deme size in the simulations is minimized. Indeed, the standard errors in estimates of both π and H are of the order 0.001. Hence, the principal sources of error are due to our form of the Möhle approximation itself, which, in addition to large numbers of demes, requires $m_g \ll 1$ and $N_g \gg 1$ for all generations g. A model system with much smaller migration rates throughout the life cycle would give a substantially better match between theory and simulations (as was suggested in Fig. 2). However, numerical and simulation results (not shown) indicate that the approximations are robust even for a cases in which migration rates are high and population sizes are large in the final generation of a season, provided that there are sufficiently many initial generations with low migration rates.

Discussion

Seasonal increases in migration rates can lead to marked changes in expected pairwise sequence differences and heterozygosities, at least in intrademic samples. For example, using the range of migration rates and deme sizes in the arrays **m** and **N**, given in equations (12) and (13), π_w and H_w can fluctuate by nearly 0.1, during the course of a growing season. Because population size, mutation and migration rates, and other population parameters are inferred from F_{ST} and π, such estimates will vary depending on the time of year at which a sample is drawn. Such interferences are further complicated by the fact that population parameters cannot be independently estimated, rather, the products Nm and $N\mu$ are inferred, in which both the migration and population size parameters are subject to change.

It was noted both in the simulations and the numerical analysis that the increase in heterozygosities and pairwise distances over the generations was not monotonic (Fig. 7), because the bottleneck generation inherits the allelic diversity from the end of the previous growing season, reaching a minimal heterozygosity one generation after the bottleneck. The effect was especially pronounced when the kth generation approached panmixia. This was not observed by Ives, although this could be a consequence of insufficiently dense sampling during the earliest generations.

In contrast, the genetic diversity in interdemic samples is quite robust to seasonal fluctuations, provided that the number of demes is large. The fact that average pairwise haplotype differences and heterozygosities change minimally about the equilibrium values implies that tests for natural selection based on comparisons between demes is more likely to give consistent results regardless of the time at which the sample is drawn. This suggests that in making tests for neutrality in a population, it may be desirable to take samples from sufficiently distant (with respect to average dispersal) sites.

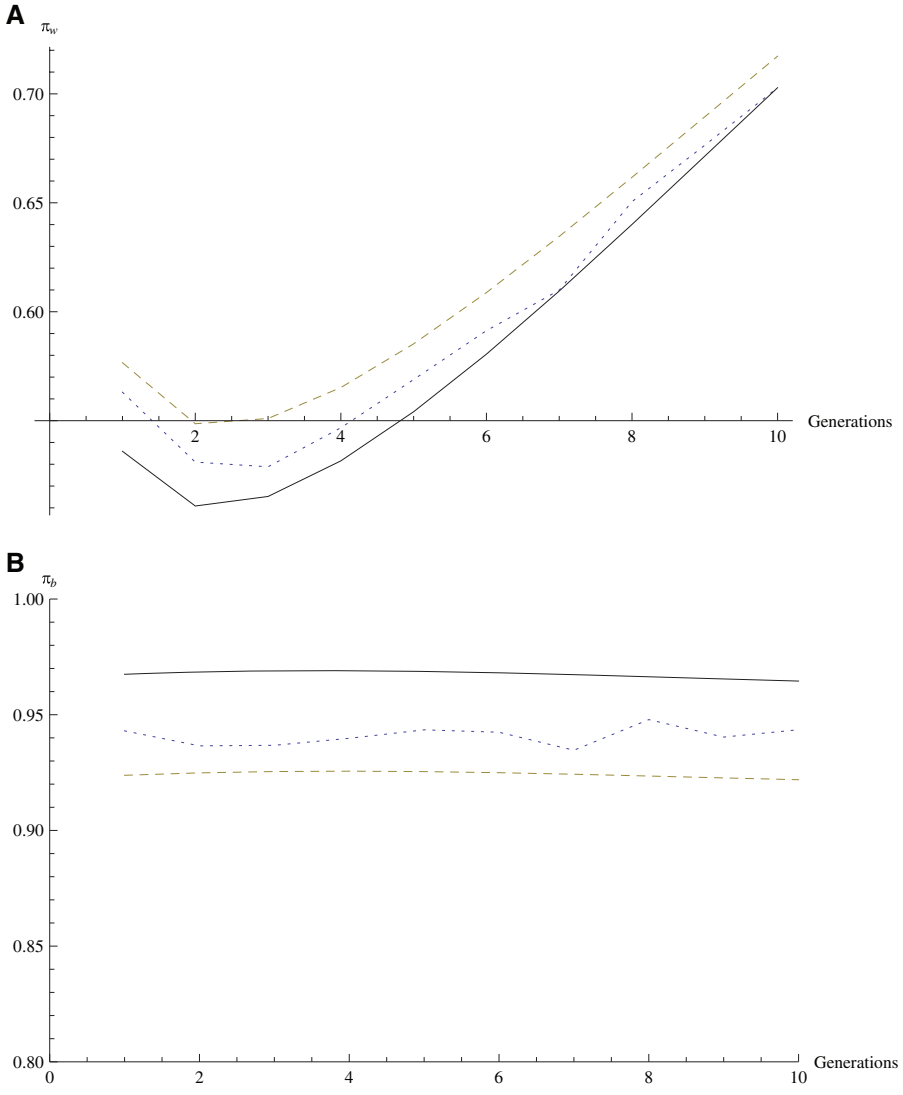

Figure 7. **(A) Average pairwise genetic distances π_w for intrademic samples taken over 10 generations, computed for a seasonal range of migration rates m and deme sizes N when $D = 10$, with a per-generation, per-site mutation rate of 0.001. The solid curve shows Möhle estimates, the dashed curve has values computed numerically from the transition matrices whereas the middle (dotted) trajectory gives the values of π_w from the individual-based simulations. (B) As in Figure 7A, but for pairwise distances π_b in interdemic samples. The dashed curve gives values computed by iteration of transition matrices, the dotted (middle) trajectory shows simulation results, the solid trajectory gives Möhle estimates.**

The other side to these issues is whether the observed fluctuations in genetic diversity during the course of the season can be entirely accounted for by a neutral model that incorporates seasonal fluctuations in deme size and other population parameters. In a review article summarizing data over several decades of systematic study (Ives and Band 1986), estimates of percent allelism in New England *Drosophila* at different times of the year were presented. Ranges in the percent of allelism are from 5% to 10% in June samples and 0.6% to 2.7% in October and November, with the June–July mean at 5.41% and the October–November mean at 1.75%. An earlier article (Ives 1970), gives November values as in excess of 20% and June values as low as 0.3%, and another case study with 34% in November versus 4% in May.

If one takes percent allelism as a proxy measure for homozygosity, the seasonal fluctuations observed by Ives are often in excess of the range of heterozygosity values predicted from the analytical approximation and the simulations alike. This discrepancy could be due an underestimate of the extent to which migration rates and deme sizes increase during the course of a growing season. This seems unlikely, because we assumed almost minimally small initial deme sizes ($N = 5$) and a population growth rate that tripled the population number in every generation. Furthermore, even when the migration rate in the final generation

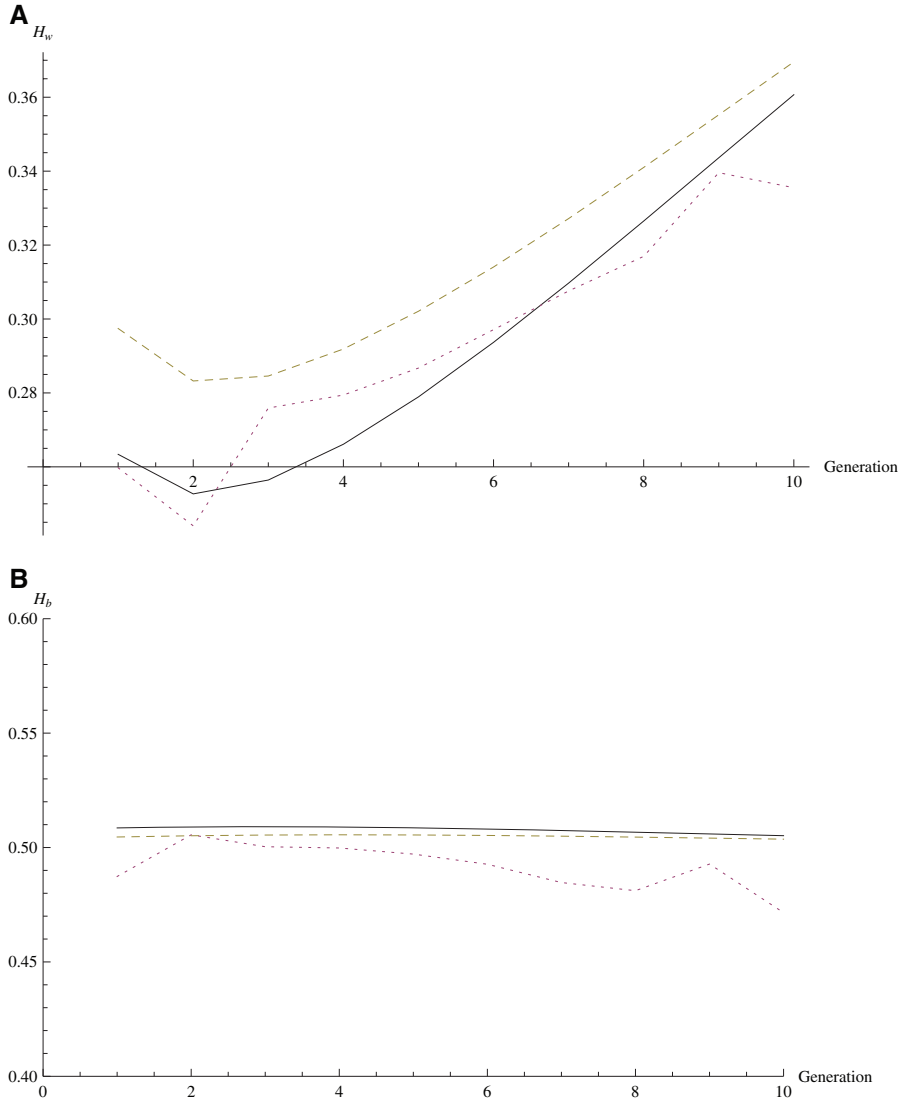

Figure 8. **(A) Average heterozygosities for intrademic samples over 10 generations in a season are shown in this figure. The values of H_w computed from the transition matrices are shown in the upper (dashed) trajectory, those of the Möhle approximation are shown in the lower (solid) trajectory whereas the simulation results are plotted in the dotted (middle) line. (B) As in Figure 8A, but for interdemic heterozygosites H_b The lowermost (dotted) curve shows the simulation results, the estimates from the Möhle approximation (solid) and the values calculated from the exact transition matrices (dashed) are almost identical.**

is very high (unrealistically approaching panmixia), the fluctuations in Figure 8 fall short of the range in allelism (up to 20–30% fluctuations) observed by Ives.

A possible explanation for this discrepancy lies in the nature of the allelic variation in Ives' study. His estimates of homozygosity are based on the frequency of lethal and semilethal alleles. Clearly, this is not neutral genetic variation, it is subject to strong negative selection even if the alleles are completely recessive. The decreases in allelism seen by Ives during a season are at least partly due to natural selection eliminating the recessive lethals and semilethals from the demes at a higher rate than they are introduced by mutation.

Consequently, to appraise the accuracy of the predictions from the structured coalescent model, it would be necessary to conduct a seasonal survey of haplotype variation in *Drosophila* populations using presumably neutral markers, such as microsatellites or noncoding single nucleotide polymorphisms. Genetic variation at such markers would be more likely to show seasonal dynamics qualitatively consistent with the coalescent model predicted in this article.

In spite of these caveats in comparing theory and data, it is likely that seasonal demographic changes can lead to marked differences in the extent of intrademic genetic variation during the course of a season. In Garrigan et al (2009), the Ives model

was used as a canonical example of a nonequilibrium system, in which standard tests for natural selection based on equilibrium neutral theory would give an excess of false positives due to deviations from allele frequency distributions predicted from neutral models at equilibrium. However, because the samples in the study were drawn only in the last generation of a season, it was unclear whether these deviations were due to the Ives dynamic per se, or from the complications of migration in an island model. Equations (15) and (16) suggest that (at least) in diallelic samples, the coalescence times and heterozygosities induced by the Ives model at the end of a season can be very closely approximated by an island model with constant population size and migration rates if these correspond to the N, m induced by the Ives population dynamic, especially when $\mu \ll m$, $1/N \ll 1$. Although this result is for a pair of alleles, it is quite likely that the sampling distribution of multiple alleles at the end of a season following an Ives process will closely resemble the distribution induced by an equivalent fixed parameter model with effective migration rates and deme sizes. This question invites further simulation and numerical study.

If one is not restricted to sampling at the end of a season, an obvious signature of the Ives process should be observable. The numerical and simulation results presented in this article imply a seasonal increase in heterozygosity and pairwise genetic distance, and unpublished results comparing multiallele early and late season samples indicate greater deviations from neutral equilibrium sampling distributions (and an excess of false positive selection test statistics) in the early season. This suggests that studies of genetic variation in seasonal populations may reach entirely different conclusions about population parameters and relevant processes depending on the time of year at which the samples are drawn.

ACKNOWLEDGMENTS

The authors thank G. Cardenas and J. Kugelman for valuable advice on modifying the ISITES code. MS was supported by URI grant 14-64585-9651, was supported by NIH grant GM070543 to RCL and JW.

LITERATURE CITED

Band, H. T., and P. T. Ives. 1961. Correlated changes in environmental and lethal frequency in a natural population of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA. 47:180–185.

Crow, J. F., and K. Aoki. 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proc. Natl. Acad. Sci. USA 81:6073–6077.

Garrigan, D., R. C. Lewontin, and J. Wakeley. 2009. Detecting natural selection from DNA sequence data. Mol. Biol. Evol. doi: 10.1093/molbev/msp209 *In press*.

Hey, J. 1991. A multidimensional coalescent process applied to multiallelic selection models and migration models. Theor. Popul. Biol. 39:30–48.

Hudson, R. R. 1991. Gene genealogies and the coalescent process. Oxford Sury. Evol. Biol. 7:1–44.

———. 1998. Island models and the coalescent process. Mol. Ecol. 7:413–418.

Ives, P. T. 1945. Genetical structure of American populations of *Drosophila melanogaster*. Genetics 30:167–196.

———. 1954. Genetic changes in American populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 40:87–92.

———. 1970. Further genetic studies of the South Amherst population of *Drosophila melanogaster*. Evolution 24:507–518.

Ives, P. T., and H. T. Band. 1986. Continuing studies on the South Amherst *Drosophila melanogaster* natural population during the 1970's and 1980's. Evolution 40:1289–1302.

Kimura, M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61:893–903.

Kimura, M., and J. F. Crow. 1964. The number of alleles that can be maintained in a finite population. Genetics 49:725–738.

Latter, B. D. H. 1973. The island model of population differentiation: a general solution. Genetics 73:147–157.

Li, W.-H. 1976. Distribution of nucleotide differences between two randomly chosen cistrons in a subdivided population: the finite island model. Theor. Popul. Biol. 10:303–308.

Maruyama, T. 1970a. Effective number of alleles in a subdivided population. Theor. Popul. Biol. 1:273–306.

———. 1970b. On the fixation probability of mutant genes in a subdivided population. Genet. Res. 15:221–225.

Möhle, M. 1998. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. Adv. Appl. Prob. 30:493–512.

———. 2000. Ancestral processes in population genetics: the coalescent. J. Theor. Biol. 204:629–638.

Nei, M., and M. W. Feldman. 1972. Identity of genes by descent within and between populations under mutation and migration pressure. Theor. Popul. Biol. 3:460–465.

Rottenstreich, S., M. B. Hamilton, and J. R. Miller. 2007. Dynamics of F_{ST} for the island model. Theor. Popul. Biol. 72:485–503.

Slatkin, M. 1987. The average number of sites separating DNA sequences drawn from a subdivided population. Theor. Popul. Biol. 32:42–49.

———. 1991. Inbreeding coefficients and coalescence times. Genet. Res. 58:167–175.

Takahata, N., and M. Nei. 1984. Fst and Gst statistics in the finite island model. Genetics 107:501–504.

Wakeley, J. 2004. Recent trends in population genetics. J. Hered. 95:397–405.

———. 2008. Coalescent theory: an introduction. Roberts and Company, Greenwood Village, CO.

Wakeley, J., and N. Aliacar. 2001. Gene genealogies in a metapopulation. Genetics 159:893–905.

Wolfram, S. 1999. The mathematica book, 4th ed., Wolfram Media/Cambridge Univ. Press, Cambridge, UK.

Whitlock, M. C., and D. E. Mccauley. 1999. Indirect measures of gene flow and migration. Heredity 82:117–125.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

———. 1951. The genetical structure of populations. Ann. Eugen. 15:323–354.

Associate Editor: J. Hermisson