Theory of the Effects of Population Structure and Sampling on Patterns of Linkage Disequilibrium Applied to Genomic Data From Humans

John Wakeley*,[1] and Sabin Lessard†

*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138 and †Département de Mathématiques et de Statistique, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Manuscript received November 4, 2002
Accepted for publication March 13, 2003

ABSTRACT

We develop predictions for the correlation of heterozygosity and for linkage disequilibrium between two loci using a simple model of population structure that includes migration among local populations, or demes. We compare the results for a sample of size two from the same deme (a single-deme sample) to those for a sample of size two from two different demes (a scattered sample). The correlation in heterozygosity for a scattered sample is surprisingly insensitive to both the migration rate and the number of demes. In contrast, the correlation in heterozygosity for a single-deme sample is sensitive to both, and the effect of an increase in the number of demes is qualitatively similar to that of a decrease in the migration rate: both increase the correlation in heterozygosity. These same conclusions hold for a commonly used measure of linkage disequilibrium (r^2). We compare the predictions of the theory to genomic data from humans and show that subdivision might account for a substantial portion of the genetic associations observed within the human genome, even though migration rates among local populations of humans are relatively large. Because correlations due to subdivision rather than to physical linkage can be large even in a single-deme sample, then if long-term migration has been important in shaping patterns of human polymorphism, the common practice of disease mapping using linkage disequilibrium in "isolated" local populations may be subject to error.

We derive the correlation in coalescence times at a pair of loci, with recombination between them, in a sample of two chromosomes at each locus from a subdivided population. Under the assumption that variation is selectively neutral at both loci and that either the infinite-sites mutation model (Watterson 1975) holds or the mutation rates at both loci are very small, (*e.g.*, see Slatkin 1991 and Nielsen 2000), these correlations in coalescence times are easily translated into correlations in heterozygosity. Griffiths (1981) studied the correlation in heterozygosity at a pair of loci in a panmictic population, as did Hudson (1983) and Kaplan and Hudson (1985) who extended the results to cover many linked sites by averaging over pairs of sites. These results have been used to construct estimators of the population recombination rate on the basis of the variance of pairwise differences (Hudson 1987; Wakeley 1997)—which turn out to perform poorly compared to other estimators (Hudson 2001)—and to discuss the trade-off between sequencing longer stretches of DNA and taking a larger sample when the goal is to measure amounts of DNA polymorphism (Pluzhnikov and Donnelly 1996).

Two recent developments, one theoretical and one empirical, provide renewed motivation for studies of the correlation in heterozygosity or coalescence times. On the empirical front, Reich *et al.* (2002) measured the correlations in polymorphism level between many pairs of short loci in samples of two human chromosomes. These are the first genome-wide estimates of such correlations, corrected for ascertainment bias and variation in mutation rate among loci. Thus, we can now directly compare theoretical predictions about the correlation in heterozygosity to observed data. On the theoretical side, McVean (2002) showed that correlations in coalescence times can be used to predict r^2, a commonly employed measure of linkage disequilibirum, defined to be the correlation coefficient between alleles at two loci (Hill and Robertson 1968). Thirty years ago, Ohta and Kimura (1971) proposed an approximation to the expected value of r^2 at a pair of loci by expressing it as a ratio of expected values, which appears to be accurate as long as the frequencies of alleles at the two loci are not too small (Hudson 1985; McVean 2002). Using the results of Strobeck and Morgan (1978) and Hudson (1985) and assuming that the per-site mutation rate at each locus is close to zero, McVean (2002) showed that Ohta and Kimura's (1971) estimate can be expressed as a ratio of covariances in coalescence times, or expected products of coalescence times, at the two loci.

This demonstration of a direct relationship between expected products of coalescence times at pairs of loci and linkage disequilibrium, as measured by r^2, should help to connect the large body of theoretical work on

[1]*Corresponding author:* 2102 Biological Laboratories, 16 Divinity Ave., Cambridge, MA 02138. E-mail: wakeley@fas.harvard.edu

genealogical processes to the ongoing empirical effort to describe and to understand patterns of linkage disequilibrium in the human genome. Human history has been marked by the growth and decline of populations, subdivision both with and without migration, and the admixture of subpopulation samples (Takahata 1995; Harpending *et al.* 1998; Hawks *et al.* 2000), all of which can strongly influence patterns of linkage disequilibrium (Slatkin 1994; Ewens and Spielman 1995; Kaplan *et al.* 1995; Lander and Schork 1996; Kruglyak 1999; Pritchard and Przeworski 2001). It is particularly important to have an accurate picture of these factors in view of the rising importance of linkage disequilibrium mapping as a tool in disease association studies (Jorde 1995; Lander 1996; Risch and Merikangas 1996) because our ability to judge the significance of associations depends on having an accurate picture of human demographic history. Predictions about r^2 are helpful in this because they measure variability in the more commonly used statistics D and D' (Lewontin 1964). For example, the history of migration we study here predicts larger values of r^2 than would be expected in a panmictic species. This means that values of D that would be judged significant if humans were panmictic might be observed more often by chance among human populations connected by migration.

The study of Reich *et al.* (2002) on samples of size two followed another recent study (Reich *et al.* 2001), which examined linkage disequilibrium in a larger sample, but within regions defined by the existence of a high-frequency coding single-nucleotide polymorphism (SNP) and thus subject to ascertainment bias. Both studies reached the same conclusions. First, correlations in heterozygosity and linkage disequilibrium extend farther than predicted by Kruglyak's (1999) model of an unstructured human population that has grown in size, in which it was assumed that the recombination rate does not vary across the genome. Second, the preferred explanation for this pattern is variation in the recombination rate across the genome rather than demographic factors. However, Reich *et al.* (2001) considered only a population bottleneck in an unstructured human population and Reich *et al.* (2002) considered either a bottleneck or the subdivision of humans into two subpopulations. Here, we focus on population subdivision and show that the conclusions of Reich *et al.* (2002) could be a consequence of the fact that only two subpopulations were considered.

It has been known for some time that population subdivision can increase the level of association between alleles in a population, due to covariance in allele frequencies among demes (Nei and Li 1973). Our results are entirely consistent with this previous work and are especially close to those of Ohta (1982) who partitioned the components of linkage disequilibrium in a subdivided population in much the same way that *F*-statistics are partitioned. However, we take a coalescent approach and this allows us to more naturally address questions of sampling over geography. Because the effects of subdivision on samples of multiple chromosomes from each of a number of demes are direct and obvious (Nei and Li 1973), we focus mainly on two possibilities: that all samples come from a single deme (single-deme sample) or that each sample comes from a different deme (scattered sample).

We find that subdivision-induced associations are negligible for scattered samples, but can be very strong in single-deme samples. The dual nature of migration is the source of these inflated levels of association between loci in samples from a single deme compared to what would be expected for a pair of loci with the same rate of recombination in a panmictic population. Restricted migration structures genetic variation among demes so that immigration events bring genetically dissimilar genomes into a deme from outside. Thus, lower rates of migration in the population lead to stronger average levels of association. We also find a strong dependence on the number of demes in the population. Levels of association in single-deme samples become stronger as the number of demes increases even if the migration rate among demes remains constant. We invoke this as at least a partial explanation of the correlations in the data considered by Reich *et al.* (2002). Clearly, variation in recombination rates across the genome, for which there is good direct evidence (Kong *et al.* 2002), must also have had a role in shaping the observed "block structure" of linkage disequilibrium in humans (Gabriel *et al.* 2002). Our results show that (1) demographic factors may not be so easily dismissed if humans are subdivided into more than two demes, and (2) it may be possible to avoid the misleading portions of the linkage disequilibrium structure of the human genome, those that are due to subdivision and migration, by looking at scattered samples.

THEORY AND METHODS

Assume that a sample of two chromosomes is taken at each of two loci and that $T^{(1)}$ is the length of the genealogy at the first locus and $T^{(2)}$ is the length of the genealogy at the second locus. Note that these are equal to twice the coalescence time at each locus. Our goal is to compute

$$\mathrm{Corr}[T^{(1)}, T^{(2)}] = \frac{\mathrm{Cov}[T^{(1)}, T^{(2)}]}{\sqrt{\mathrm{Var}[T^{(1)}]\mathrm{Var}[T^{(2)}]}}, \qquad (1)$$

where the variances and covariances are defined in the usual way. For example,

$$\mathrm{Cov}[T^{(1)}, T^{(2)}] = E[T^{(1)}T^{(2)}] - E[T^{(1)}]E[T^{(2)}]. \qquad (2)$$

The expectations above are with respect to the ancestral (genealogical or coalescent) process at the two loci, which here will also involve recombination and migration. If mutations occur at each locus according to the infinite-sites model of Watterson (1975), with rate $\theta/2$

per time unit, then the covariance in numbers of SNPs at the two loci is simply $\theta^2/4 \times$ Equation 2 (Griffiths 1981; Hudson 1983). For the model defined below we have $\theta = 4NDu$, where N is the deme size, D is the number of demes, and u is the neutral mutation rate per locus-copy per generation; and time is measured in units of $2ND$ generations.

Equation 1 is what Reich *et al.* (2002) estimate using genomic data from humans and call $\rho(\tau_x, \tau_{x+d})$ for a pair of loci separated by d intervening nucleotides. The value of $\rho(\tau_x, \tau_{x+d})$ depends on whether the two copies at each locus are linked on the same two chromosomes, share one chromosome, or are located on two distinct pairs of chromosomes (Strobeck and Morgan 1978), and Reich *et al.* (2002) call these possibilities *cis*, *trans*, and *dis*, respectively. In addition to (1), we use the covariances of tree lengths to compute

$$\frac{\mathrm{Cov}_{cis}[T^{(1)}, T^{(2)}] - 2\,\mathrm{Cov}_{trans}[T^{(1)}, T^{(2)}] + \mathrm{Cov}_{dis}[T^{(1)}, T^{(2)}]}{E[T^{(1)}]E[T^{(2)}] + \mathrm{Cov}_{dis}[T^{(1)}, T^{(2)}]}. \quad (3)$$

This is what McVean (2002), in the context of a panmictic population, calls σ_d^2, following Ohta and Kimura (1971). As mentioned above, Equation 3 gives an approximation to the expected value of a commonly used measure of linkage disequilibrium, r^2, which is defined to be the square of the correlation coefficient between alleles at two loci (Hill and Robertson 1968). Hudson (1985) showed that σ_d^2 accurately predicts r^2 only in a large sample (or the total population) so that (3) will be inaccurate for small samples.

In contrast to the case of a single randomly mating population, in a subdivided population the expected values that go into Equations 1 and 3 will depend on how the sample is distributed among demes. Thus, we cannot simply use *cis*, *trans*, and *dis* here and instead develop an expanded notation (described below) for samples from a subdivided population. We begin with a general statement of the model in the next section and then use this framework to compute (1) and (3) in the finite island model (Wright 1931; Moran 1959; Latter 1973; Maruyama 1974) of subdivision for different sample configurations. We also present some limiting expressions that hold for the two-locus ancestral process in the island model as the number of demes becomes large, a topic we treat in detail elsewhere (Lessard and Wakeley 2003).

The ancestral recombination graph for a pair of sites in a structured population: We assume discrete, nonoverlapping generations in a diploid population structured into D demes, with backward migration rates constant through time. We consider two loci or sites with recombination rate r per generation between them (not to be confused with the r in r^2). Elsewhere (Lessard and Wakeley 2003) we describe the ancestral process for this model, which is a generalization of the single-locus structured coalescent (Wilkinson-Herbots 1998; Nordborg 2001), or an extension of the two-locus ancestral graph (Griffiths 1991) to the case of a structured population. Here, we restate enough of the framework to solve the problem at hand. In the general version of the model, demes may be of different relative sizes, c_i, for $i = 1, \ldots, D$, and migration rates can vary across the population. We let m_{ij} be the proportion of deme i that came from deme j in the previous generation, and let $M_{ij} = 4Nm_{ij}$ be the scaled migration rate. Note that in doing this, we have made the usual coalescent assumption that the migration rate m_{ij} is small and deme size N is large.

In considering the genealogy of a sample of chromosomal segments at the two sites, it is necessary to distinguish three kinds of segments: those ancestral to the sample at site 1 only (type 1), those ancestral at site 2 only (type 2), and those ancestral at both sites (type 3). If deme i contains $n_i^{(1)}$ segments of type 1, $n_i^{(2)}$ segments of type 2, and $n_i^{(3)}$ segments of type 3, then the number and location of the different ancestral segments at any given time can be described by the vector

$$\mathbf{n} = (\mathbf{n}_{1,\ldots,}\mathbf{n}_D), \quad (4)$$

in which

$$\mathbf{n}_i = (n_i^{(1)}, n_i^{(2)}, n_i^{(3)}), \quad (5)$$

for $i = 1, \ldots, D$. In addition, we use

$$n_i = n_i^{(1)} + n_i^{(2)} + n_i^{(3)} \quad (6)$$

to represent the total number of ancestral segments in deme i. The coalescent process for such a sample is a continuous-time Markov chain that remains in state $\mathbf{n}$ for an exponentially distributed length of time with parameter

$$\lambda_{\mathbf{n}} = \sum_i \left\{ \frac{n_i D M_i}{2} + \frac{n_i^{(3)} R}{2} + \frac{n_i(n_i - 1)D}{2c_i} \right\}, \quad (7)$$

where $R = 4NDr$ is the scaled recombination rate, and

$$M_i = \sum_{j \neq i} M_{ij}. \quad (8)$$

The three terms in Equation 7 correspond to all possible migration, recombination, and coalescent events, respectively, that change the numbers and/or locations, $\mathbf{n}$, of ancestral segments. Time is measured in units of $2ND$ generations.

After spending an exponential amount of time in state $\mathbf{n}$, there is a jump to another state $\mathbf{n}'$ with transition probabilities

$$Q_{\mathbf{n}\mathbf{n}'} = \begin{cases} \dfrac{n_i^{(k)} D M_{ij}}{2\lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i^{(k)} + \mathbf{e}_j^{(k)}, \text{ for } j \neq i \text{ and } k = 1, 2, 3 \\ \dfrac{n_i^{(3)} R}{2\lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} + \mathbf{e}_i^{(1)} + \mathbf{e}_i^{(2)} - \mathbf{e}_i^{(3)} \\ \dfrac{n_i^{(1)} n_j^{(2)} D}{c_i \lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i^{(1)} - \mathbf{e}_i^{(2)} + \mathbf{e}_i^{(3)} \\ \dfrac{2n_i^{(k)} n_i^{(3)} D + n_i^{(k)}(n_i^{(k)} - 1)D}{2c_i \lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i^{(k)}, \text{ for } k = 1, 2 \\ \dfrac{n_i^{(3)}(n_i^{(3)} - 1)D}{2c_i \lambda_{\mathbf{n}}} & \text{if } \mathbf{n}' = \mathbf{n} - \mathbf{e}_i^{(3)}, \end{cases} \quad (9)$$

where $\mathbf{e}_i^{(k)}$ designates a vector of all zero D triplets except the ith, which is (1, 0, 0) if $k = 1$, (0, 1, 0) if $k = 2$, and (0, 0, 1) if $k = 3$ (Lessard and Wakeley 2003). From top to bottom, the terms of Equation 9 have the following meanings. The first term represents all possible migration events. These simply move ancestral chromosomes around the population. The second term represents recombination events. These affect the history of the sample only when they occur on a chromosome of type 3, breaking it into a chromosome of type 1 and a chromosome of type 2. The third term represents coalescent events between type 1 and type 2 chromosomes. A type 3 chromosome is created by such a merger, but neither site experiences a common ancestor event. The fourth term represents coalescent events in which just one of the sites has a common ancestor event, and the fifth term represents coalescent events in which both sites have a common ancestor event.

We can use this framework to obtain systems of equations for the quantities required to compute the covariances of genealogical tree lengths at two sites in a structured population. Suppose that the Markov chain is currently in state $\mathbf{n}$, given by Equation 4. Let $T_{\mathbf{n}}^{(1)}$ be the length of the genealogical tree since the most recent common ancestor at the first site and $T_{\mathbf{n}}^{(2)}$ be the corresponding variable for the second site. Conditioning on the first change in the genealogical history of the sample, we have for the expectation of the first variable at equilibrium,

$$E(T_{\mathbf{n}}^{(1)}) = \left\{\frac{\sum_i (n_i^{(1)} + n_i^{(3)})}{\lambda_{\mathbf{n}}}\right\} + \sum_{\mathbf{n}'} Q_{\mathbf{n}\mathbf{n}'} E(T_{\mathbf{n}'}^{(1)}), \quad (10)$$

and similarly for the expectation of the second variable. These expectations depend only on the state at the site under consideration and, therefore, do not depend on the scaled recombination rate R between the two sites. For the expectation of the product of these variables at equilibrium, we have

$$\begin{aligned} E(T_{\mathbf{n}}^{(1)} T_{\mathbf{n}}^{(2)}) = {} & \left\{\frac{\sum_i (n_i^{(1)} + n_i^{(3)})}{\lambda_{\mathbf{n}}}\right\} E(T_{\mathbf{n}}^{(2)}) \\ & + \left\{\frac{\sum_i (n_i^{(2)} + n_i^{(3)})}{\lambda_{\mathbf{n}}}\right\} E(T_{\mathbf{n}}^{(1)}) \\ & + \sum_{\mathbf{n}'} Q_{\mathbf{n}\mathbf{n}'} E(T_{\mathbf{n}'}^{(1)} T_{\mathbf{n}'}^{(2)}), \end{aligned} \quad (11)$$

which will depend on R.

The island model with an arbitrary number of demes: In the island model, the D demes are assumed to be of the same size and the backward migration rates to other demes are all equal. Therefore, we have $c_i = 1$ for all i and $M_{ij} = M/(D - 1)$ for all $j \neq i$. These assumptions make the ancestral graph at two sites symmetric with respect to any permutation of the demes in addition to being symmetric with respect to the two sites. In computing (1) and (3), we need to consider only samples of two chromosomes at each site. This simplifies the state space considerably, and Table 1 lists all the possibilities numbered for simplicity from 1 to 15. We focus below on states 5 and 11, which represent the most common sample configuration for a sample of size two at each of a pair of sites, or loci, in a subdivided population. These are both *cis* comparisons; state 5 is when the two chromosomes are sampled from the same deme and state 11 is when they come from different demes.

TABLE 1

Notation for states in the island model

State 1	if	{1, 1}	or	{2, 2}
State 2	if	{1} {1}	or	{2} {2}
State 3	if	{1, 1, 2, 2}		
State 4	if	{1, 12, 2}		
State 5	if	{12, 12}		
State 6	if	{1, 1, 2} {2}	or	{2, 2, 1} {1}
State 7	if	{1, 12} {2}	or	{2, 21} {1}
State 8	if	{1, 1} {2, 2}		
State 9	if	{1, 2} {1, 2}		
State 10	if	{1, 2} {12}		
State 11	if	{12} {12}		
State 12	if	{1, 1} {2} {2}	or	{2, 2} {1} {1}
State 13	if	{1, 2} {1} {2}		
State 14	if	{12 } {1} {2}		
State 15	if	{1} {1} {2} {2}.		

1 is for a segment ancestral at site 1, 2 for a segment ancestral at site 2, 12 for a segment ancestral at sites 1 and 2, and an ancestral deme is represented by the set of ancestral segments that it contains.

The quantities in Equations 1, 2, and 3 that depend only on the history at a single site have been known for some time; see Hey (1991) and references therein. If we represent the sample states by their numbers, then for site 1 we have

$$E[T_1^{(1)}] = 2, \quad (12)$$

$$E[T_2^{(1)}] = 2 + \frac{2(D-1)}{MD}, \quad (13)$$

$$\mathrm{Var}[T_1^{(1)}] = 4 + \frac{8(D-1)^2}{MD^2}, \quad (14)$$

$$\mathrm{Var}[T_2^{(1)}] = 4 + \frac{8(D-1)^2}{MD^2} + \frac{4(D-1)^2}{M^2D^2}, \quad (15)$$

and the expressions for site 2 are identical.

We use Equation 11 to obtain the other necessary quantities, *i.e.*, the expectations of the product of the tree lengths at two sites, depending on the distribution of the ancestral segments within and between demes. To save space, we let $V_s = E[T_s^{(1)} T_s^{(2)}]$ for every state number s in Table 1 with two segments ancestral at each site. Then, assuming at least four demes, we have the equations given in the appendix, which can be solved analytically using software like Mathematica (Wolfram

1999). The resulting expressions are too lengthy to reproduce here but are available upon request to the authors. For $D < 4$, the equations in the APPENDIX must be modified because some of the states in Table 1 will not exist.

A simpler ancestral process when *D* is large: The unwieldy expressions for the case of arbitrary D can be checked against simpler predictions that hold in the limit as D goes to infinity. When the number of demes is large, the ancestral process for the sample becomes much simpler. In particular, in the island model both with recombination (LESSARD and WAKELEY 2003) and without it (WAKELEY 1998, 1999), the history of a scattered sample is the same as that in a panmictic population of ND diploid organisms, but on a timescale longer by the factor $(1 + 1/M)$. Samples that include multiple segments from single demes are subject to a rapid stochastic sample-size adjustment, the "scattering phase" (WAKELEY 1999), before they enter this much longer panmictic process. Coalescent events during the scattering phase are the source of the greater within-deme than between-deme relatedness predicted under the island model.

The APPENDIX gives approximations for $V_s = E[T_s^{(1)} T_s^{(2)}]$ for the case when D is large. These expressions can also be obtained directly from the limiting (large-D) two-locus ancestral recombination graph (LESSARD and WAKELEY 2003). They illustrate the relative simplicity of the large-D ancestral process, which includes the scattering phase. For example, for V_5 and V_{11} we have

$$V_5 = \frac{M}{M+1} V_{11}. \tag{16}$$

This equation relates the results for single-deme samples (state 5) and scattered samples (state 11) of size two via the scattering phase probability, $M/(M + 1)$, that one or the other of the two segments migrates before they coalesce (WAKELEY 1998). In addition, these values (*e.g.*, V_{11} in the APPENDIX) reflect the fact that the timescale of the coalescent process is increased by the factor $(1 + 1/M)$.

By substituting these large-D approximations into Equation 1 we have

$$\mathrm{Corr}[T_{11}^{(1)}, T_{11}^{(2)}] = \frac{R + 18}{R^2 + 13R + 18}, \tag{17}$$

for a scattered sample in the large-D limit. Equation 17 is identical to the correlation of tree lengths in a sample of two chromosomes at two loci (*cis*) in a panmictic population (GRIFFITHS 1981; HUDSON 1983). In addition, substituting the approximations for V_{11}, V_{14}, and V_{15} from the APPENDIX (for *cis*, *trans*, and *dis*, respectively, for a scattered sample) into expression (3), and simplifying, gives

$$\frac{R + 10}{R^2 + 13R + 22}, \tag{18}$$

which is identical to the prediction, σ_d^2, for r^2 under panmixia (OHTA and KIMURA 1971; MCVEAN 2002). Expected values for samples from the same deme, on the other hand, do reflect the effects of subdivision (see Figure 1b below).

Equations 17 and 18 illustrate a surprising fact about the large-D ancestral process for a scattered sample, namely, that the appropriate recombination parameter continues to be equal to $R = 4NDr$, even as D goes to infinity (LESSARD and WAKELEY 2003). The reason this is surprising is that the timescale of the coalescent process, and thus the time over which mutation and recombination events can occur in the history of the sample, is increased by the factor $(1 + 1/M)$. The number of mutation events in the history of the sample does indeed depend on $\theta(1 + 1/M)$ (WAKELEY 1998), but the number of potentially observable recombination events depends only on R (LESSARD and WAKELEY 2003). While $(1 + 1/M)$ more recombination events do occur in the history, a fraction $1/(1 + M)$ of these are repaired instantaneously because the two resulting segments will initially be present in the same deme and thus will be like a single-deme sample, subject to a scattering phase. What remains is $R \times (1 + 1/M) \times M/(M + 1) = R$. This might have important practical consequences because disease loci can be mapped with greater resolution when recombination is more frequent, as long as enough SNPs are present, and higher numbers of SNPs lead to increased power for a given recombination map (KRUGLYAK 1999). We do not pursue these issues here, but note that NORDBORG and TAVARÉ (2002) discuss this in relation to the effect that partial selfing has on the numbers of SNPs and recombination events (NORDBORG 2000). LESSARD and WAKELEY (2003) detail the similarities and differences between partial selfing and island-model migration.

RESULTS

Population subdivision provides an additional axis for comparison of polymorphism levels and associations/correlations between loci. It introduces the possibility of making within- *vs.* between-deme comparisons [as embodied by WRIGHT's (1951) well-known fixation index F_{ST}], an idea that OHTA (1982) extended to measures of linkage disequilibrium. Thus, the distribution of the sample among demes, as well as the distribution of ancestral genetic material among chromosomes, becomes important. Figure 1 shows that the correlation of tree lengths at two loci depends strongly on how the sample is taken, either from two different demes or from the same deme, as well as on the distance separating the two sites, the demic migration rate, and the number of demes in the population. Note that in Figure 1, a and

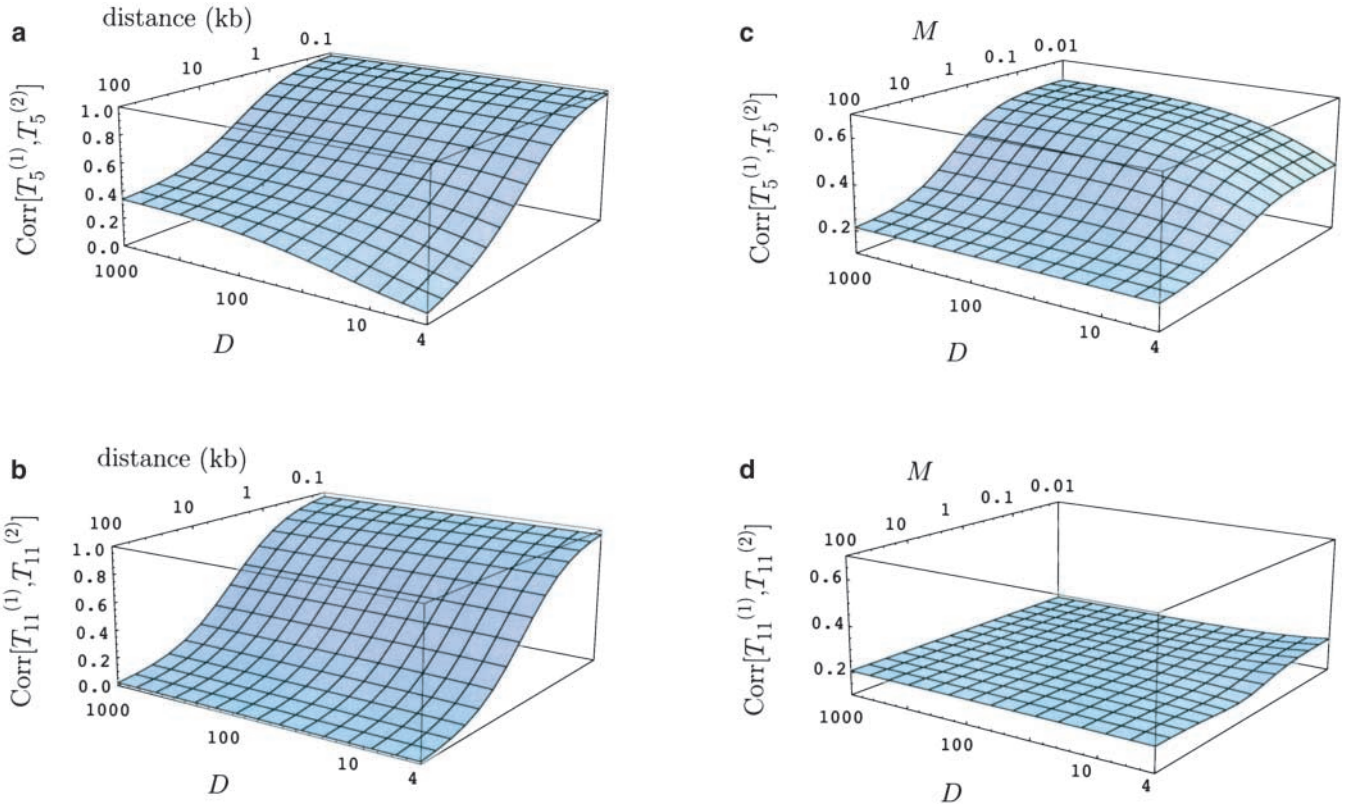

FIGURE 1.—Analytic predictions for the correlation in tree lengths at two loci in samples of two chromosomes at each locus. (a and b) The correlation as a function of the distance between the two sites, measured in kilobases (kb), and the number of demes D in an island model population with $M = 1$. (c and d) The correlation at two sites separated by 10 kb as a function of the migration rate M and the number of demes. As discussed in the text, the population per-site rate of recombination was assumed to be $R = 5.2 \times 10^{-4}$ after REICH *et al.* (2002). a and c are for the case of two chromosomes [*cis* using the terminology of REICH *et al.* (2002)] sampled from two different demes; *i.e.*, they plot Equation 1 for state 11 in Table 1. b and d are for the case of two chromosomes (again, *cis*) sampled from the same deme; *i.e.*, they plot Equation 1 for state 5 in Table 1.

b have a broader range on the vertical axis than c and d. For Figure 1, a–d, the population rate of recombination was assumed to be 5.2×10^{-4}/site, following REICH *et al.* (2002). Thus, the distances in a and b correspond to a range of values of R from 0.052 ($= 0.1$ kb $\times 10^3$ bp/kb $\times 5.2 \times 10^{-4}$/bp) to 52. For c and d, the distance was set to 10 kb, which corresponds to $R = 5.2$.

For a single-deme sample, Figure 1, a and c, the predicted correlation depends on all three quantities: the distance, M, and D. Clearly, the distance between the two sites, which in our model linearly determines the recombination rate between them, has the strongest effect, with shorter distances corresponding to higher correlations. The migration rate is the next most important factor for single-deme samples, with lower migration rates producing stronger correlations in tree lengths. Finally, there is an effect of the number of demes on the correlation of tree lengths at two loci in a single-deme sample, with larger numbers of demes yielding stronger correlations. A large number of demes is qualitatively similar to a small migration rate because in both cases samples share either a very recent common ancestor at both loci due to coalescence within the deme or a very ancient one if a migration event occurs in their recent history, prior to which it may be a long time before another migration event again puts them in the same deme so that they have the chance to coalesce. This subdivision-induced inflation of the correlation is true for pairs of loci at any distance, but the effect is stronger for loci that are farther apart.

For a scattered sample, Figure 1, b and d, only the distance between the two sites strongly affects the correlation, again with shorter distances producing stronger correlations. Surprisingly, neither the migration rate nor the number of demes has a large effect on the correlation of tree lengths for scattered samples. There is some dependence on M and D when D is small, but the magnitude of the change in the correlation there is much smaller than that when the distance between sites is varied. Thus, the correlation in tree lengths for a scattered sample is similar to that in a panmictic population even though the average lengths of the trees change substantially with both M and D; see Equation 13. This constancy for scattered samples is predicted

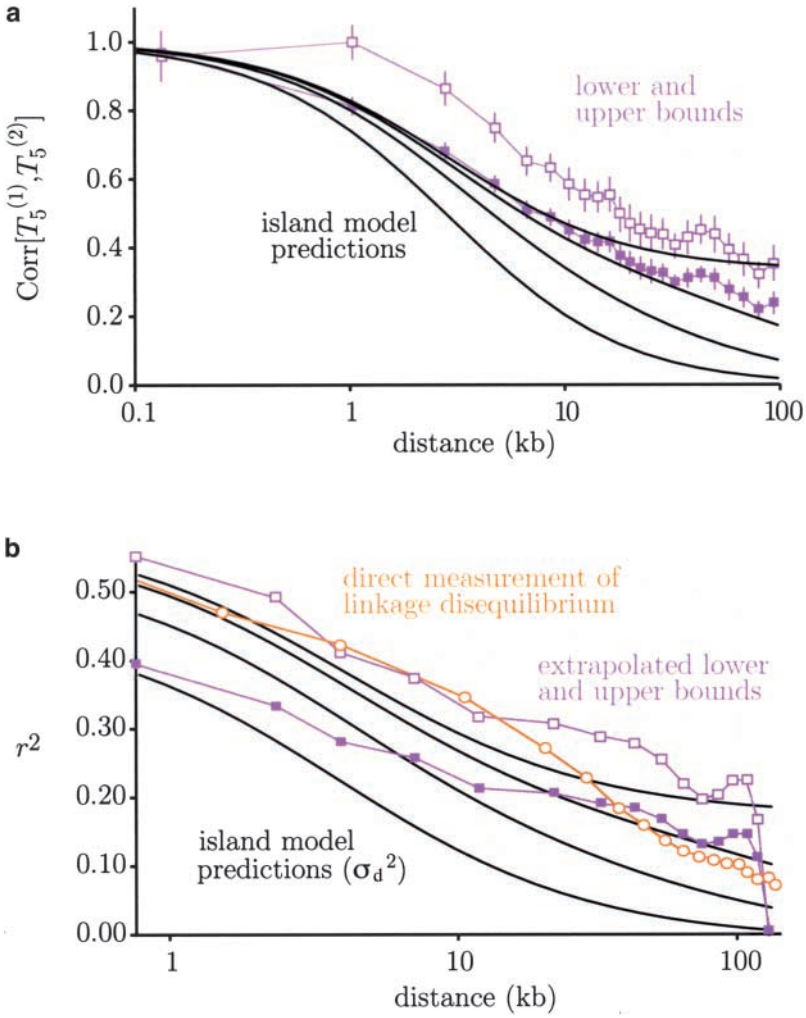

FIGURE 2.—Comparison of the predictions of the island model to the data of REICH *et al.* (2002). The population per-site rate of recombination was set, as in Figure 1, to be $R = 5.2 \times 10^{-4}$ and was constant across the genome. The different curves for the island model predictions are for $D = 1$, 4, 16, ∞, from bottom to top. The curves for $D = 1$ are the ones labeled "theoretical prediction" in REICH *et al.* (2002). In a, $M = 1$ and the prediction curves plot Equation 1 for state 11 in Table 1. In b, $M = 4$ and the prediction curves plot Equation 3 for state 11 in Table 1. Data points are redrawn from Figure 5, a and c, in REICH *et al.* (2002): the open/solid squares in both a and b give the upper/lower bounds for the correlation and for σ_d^2 on the basis of variability in the correction for among-locus variation for mutation rate in comparisons with chimpanzee sequences, and the open circles in b show r^2 calculated from the Utah CEPH data of GABRIEL *et al.* (2002).

for large values of D—see Equation 17 and LESSARD and WAKELEY (2003)—but it appears to hold approximately for small D as well.

The protocol of REICH *et al.* (2002) categorized pairs of sites as *cis*, *trans*, and *dis*, but did not distinguish among the possible sample configurations in Table 1. The chance that both samples at both loci are from the same deme will be greatest for *cis* comparisons and smallest for *dis* comparisons, so the effects of subdivision on the correlations of tree lengths will be strongest for the *cis* samples. More detailed statements are difficult due to the heterogeneity of population origin of the samples used by REICH *et al.* (2002). In addition to these correlations in tree lengths at pairs of loci, REICH *et al.* (2002) plotted r^2 calculated from a separate data set that included only Americans of European ancestry from the CEPH Utah pedigrees (GABRIEL *et al.* 2002). For the sake of illustration, we assume that both sets of data are composed of single-deme samples. Figure 2 shows that for given migration rates for these two sets of data ($M = 1$ and $M = 4$, respectively; see below), increasing the number of demes brings the theoretical predictions closer to the observed data.

Both the pairwise data of REICH *et al.* (2002) and the Utah CEPH data of GABRIEL *et al.* (2002) are subject to ascertainment bias with respect to the process of migration. In the case of REICH *et al.* (2002), an upper bound was exerted on the number of SNPs per read, and in the case of GABRIEL *et al.* (2002) SNPs were discovered in an initial smaller sample and then typed in a larger one and an upper bound was exerted on the number of SNPs per read. REICH *et al.* (2002) corrected for ascertainment when calculating correlations, but neither set of authors considered the effect of ascertainment bias on realized, or estimated, migration rates. We can infer that realized migration rates in the Utah CEPH data are greater than the actual migration rates (WAKELEY *et al.* 2001), but in the pairwise data they should be smaller than actual rates because the protocol truncated only the upper tail of the distribution of the number of SNPs per read. This occurs because migrants tend to differ from residents, so the number of migrant chromosomes detected by this method is less than would be found in a random sample. This is the justification for using different values, $M = 1$ and $M = 4$, for the two data sets in Figure 2. These values were chosen for the

sake of illustration, but they are roughly consistent with estimates of M based on F_{ST} (Cavalli-Sforza *et al.* 1994) when ascertainment bias is taken into account.

DISCUSSION

Human beings do not mate randomly on a global scale. Instead, they are subdivided into local populations, or demes, among which there is substantial gene flow. Substantial here means roughly $M \geq 1$, as appears to be true of many local human populations (Cavalli-Sforza *et al.* 1994). We have shown that the number of demes is important in determining levels of association between alleles and of correlations in genealogical tree lengths at different loci when multiple samples come from single demes; see Figure 1, a and c. In particular, average two-locus correlations in a single-deme sample become stronger both when the migration rate decreases and when the number of demes increases. Thus, the conclusion of Reich *et al.* (2002) that migration rates are not low enough among human demes to explain observed long-distance correlations is at least in part a consequence of the fact that only a two-deme model was considered.

We have also shown that associations due to subdivision are negligible for scattered samples; see Figure 1, b and d. This has consequences for the prospect of disease-locus mapping using patterns of linkage disequilibrium. There is active debate over population choice for linkage disequilibrium mapping studies, based both on the chance that the disease is less heterogeneous and on the knowledge of the demographic history of local populations (Wright *et al.* 1999). For instance, it is popular to focus on recently founded local populations because these are expected to have preserved levels of linkage disequilibrium due to sampling of founders. There is some empirical evidence of this, *e.g.*, Jorgensen *et al.* (2002), but there is also evidence that some local populations do not follow this pattern (Eaves *et al.* 2000). So far, there has been little attention to migration in these discussions. The migration model we studied here predicts that local populations that receive fewer migrants each generation are expected to show greater correlations between loci and higher levels of linkage disequilibrium. When M is small, migrants will differ from residents at many loci simultaneously, and the model predicts an equilibrium level of linkage disequilibrium as a balance among immigration, genetic drift, and recombination. If a sample contains recent migrants, the chance of spurious association increases.

While we have not studied any particular disease model and have considered only levels of association between neutral markers, further study of the role of migration among local populations of humans in establishing genomic patterns of linkage disequilibrium seems warranted. Oversimplified models of human history are not consistent with available data (Przeworski *et al.* 2000; Pluzhnikov *et al.* 2002), and appropriate tools for linkage disequilibrium mapping are necessary. Methods in the spirit of Pritchard *et al.* (2000), which use background levels of association to control for population structure, appear promising. Note, however, that the method of Pritchard *et al.* (2000) assumes no association within demes and so does not appear to be applicable if there has been migration. If the island model with a large number of demes (or a related model; see below) holds for humans, the present analysis, *e.g.*, Figure 1, b and d, suggests that linkage disequilibrium mapping studies should be done using scattered samples because these should have the lowest level of subdivision-induced association. This would not be recommended, however, if the causes of the disease under study differed from population to population.

Of course, the island migration model has shortcomings: it assumes that the population has always been subdivided, that every deme is of the same size and has the same migration rate, and that the number of demes has remained constant. Perhaps most importantly, it lacks true geographic structure because every deme can exchange migrants with every other deme. Many of these problems can be dealt with in the case where the number of demes is large, and extinction/recolonization of demes can also be included (Wakeley and Aliacar 2001). Thus, the results we describe here should hold, at least qualitatively, in a much broader setting. We have not here considered changes in population size or migration rate, but these too can be incorporated relatively easily into the model when the number of demes is large (Wakeley 1999).

The work we have presented here is similar in spirit to the recent work of Vitalis and Couvet (2001a,b) who studied two-locus probabilities of identity under the infinite island model, with the possible addition of partial selfing. Thus, it might be expected that our results, in the limit as D goes to infinity, should match theirs. However, the model considered by Vitalis and Couvet (2001a,b) differs from ours in one very important respect. While we assume that the population mutation rate $\theta = 4NDu$ and the population recombination rate $R = 4NDr$ remain finite as D goes to infinity, Vitalis and Couvet (2001a,b) assume that the demic mutation rate $4Nu$ and the demic recombination rate $4Nr$ are not necessarily small. The population rates of mutation and recombination in their model are, implicitly, infinite. One consequence of this is that their model predicts zero identity linkage disequilibrium when the migration rate for a deme is very large (or, equivalently, for a scattered sample); see Figures 4 and 5 in Vitalis and Couvet (2001b). In addition, their model would predict an infinite number of SNPs, or mutation events, at a locus. On the other hand, the large-D limit we consider predicts zero SNPs at a locus in a single-deme sample if no migration events occur in the history of the sample. Which of these models is most appropriate depends on

whether allelic data or sequence data are analyzed and whether mutation and recombination events that occur in the recent history of a deme are the focus of study or can be ignored because they are far outnumbered by mutation and recombination events that occur earlier in the history of a sample.

We thank David Reich and Stephen Shaffner for helpful discussions of Reich *et al.* (2002) and for making their article available before it was published. Kristin Ardlie and Monty Slatkin provided helpful comments on an earlier version of the manuscript. J.W. was supported by a Career Award (DEB-0133760) and by a grant (DEB-9815367) from the National Science Foundation. S.L. was supported by grants from the Natural Sciences and Engineering Research Council of Canada, the Fonds Québécois de la Recherche sur la Nature et les Technologies, and the Université de Montréal.

LITERATURE CITED

Cavalli-Sforza, L. L., P. Menozzi and A. Piazza, 1994 *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.

Eaves, I. A., T. R. Merriman, R. A. Barber, S. Nutland, E. Tuomilehto-Wolf *et al.*, 2000 The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat. Genet. **25:** 320–323.

Ewens, W. J., and R. S. Spielman, 1995 The transmission/disequilibrium test: history, subdivision, and admixture. Am. J. Hum. Genet. **57:** 455–464.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy *et al.*, 2002 The structure of haplotype blocks in the human genome. Science **296:** 2225–2229.

Griffiths, R. C., 1981 Neutral two-locus multiple allele models with recombination. Theor. Popul. Biol. **19:** 169–186.

Griffiths, R. C., 1991 The two-locus ancestral graph, pp. 100–117 in *Selected Proceeedings of the Symposium on Applied Probability*, edited by I. V. Basawa and R. L. Taylor. Institute of Mathematical Statistics, Hayward, CA.

Harpending, H., M. A. Batzer, M. Gurven, L. B. Jorde, A. R. Rogers *et al.*, 1998 Genetic traces of ancient demography. Proc. Natl. Acad. Sci. USA **95:** 1961–1967.

Hawks, J., K. Hunley, S.-H. Lee and M. Wolpoff, 2000 Population bottlenecks and Pleistocene human evolution. Mol. Biol. Evol. **17:** 2–22.

Hey, J., 1991 A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theor. Popul. Biol. **39:** 30–48.

Hill, W. G., and A. R. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Hudson, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

Hudson, R. R., 1985 The sampling theory of linkage disequilibrium under an infinite allele model without selection. Genetics **109:** 611–631.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159:** 1805–1817.

Jorde, L. B., 1995 Linkage disequilibrium as a gene mapping tool. Am. J. Hum. Genet. **56:** 11–14.

Jorgensen, T. H., B. Degn, A. G. Wang, M. Vang, H. Gurling *et al.*, 2002 Linkage disequilibrium and the demographic history of the isolated population of the Faroe Islands. Eur. J. Hum. Genet. **10:** 381–387.

Kaplan, N. L., and R. R. Hudson, 1985 The use of sample genealogies for studying a selectively neutral *m*-loci model with recombination. Theor. Popul. Biol. **28:** 382–396.

Kaplan, N. L., W. G. Hill and B. S. Weir, 1995 Likelihood methods for locating diseases in nonequilibrium populations. Am. J. Hum. Genet. **56:** 18–32.

Kong, A., D. F. Gubdjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson *et al.*, 2002 A high-resolution recombination map of the human genome. Nat. Genet. **31:** 241–247.

Kruglyak, L., 1999 Prospects for whole genome linkage disequilibrium mapping of common diseases. Nat. Genet. **22:** 139–144.

Lander, E. S., 1996 The new genomics: global views of biology. Science **274:** 536–539.

Lander, E. S., and N. J. Schork, 1996 Genetic dissection of complex traits. Science **265:** 2037–2048.

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

Lessard, S., and J. Wakeley, 2003 The two-locus ancestral graph in a subdivided population: convergence as the number of demes grows in the island model. J. Math. Biol. (in press).

Lewontin, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49:** 49–67.

Maruyama, T., 1974 A simple proof that certain quantities are independent of the geographical structure of population. Theor. Popul. Biol. **5:** 148–154.

McVean, G., 2002 A genealogical interpretation of linkage disequilibrium. Genetics **162:** 987–991.

Moran, P. A. P., 1959 The theory of some genetical effects of population subdivision. Austr. J. Biol. Sci. **12:** 109–116.

Nei, M., and W.-H. Li, 1973 Linkage disequilibrium in subdivided populations. Genetics **75:** 213–219.

Nielsen, R., 2000 Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics **154:** 931–942.

Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial selfing. Genetics **154:** 923–929.

Nordborg, M., 2001 Coalescent theory, pp. 179–212 in *Handbook of Statistical Genetics*, edited by D. J. Balding, M. J. Bishop and C. Cannings. John Wiley & Sons, Chichester, England.

Nordborg, M., and S. Tavaré, 2002 Linkage disequilibrium, haplotype evolution, and the coalescent. Trends Genet. **18:** 83–90.

Ohta, T., 1982 Linkage disequilibrium with the island model. Genetics **101:** 139–155.

Ohta, T., and M. Kimura, 1971 Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. Genetics **68:** 571–580.

Pluzhnikov, A., and P. Donnelly, 1996 Optimal sequencing strategies for surveying molecular genetic diversity. Genetics **144:** 1247–1262.

Pluzhnikov, A., A. D. Rienzo and R. R. Hudson, 2002 Inferences about human demography based on multilocus analyses of noncoding sequences. Genetics **161:** 1209–1218.

Pritchard, J. K., and M. Przeworski, 2001 Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. **69:** 1–14.

Pritchard, J. K., M. Stephens, N. A. Rosenberg and P. Donnelly, 2000 Association mapping in structured populations. Am. J. Hum. Genet. **67:** 170–181.

Przeworski, M., R. R. Hudson and A. DiRienzo, 2000 Adjusting the focus on human variation. Trends Genet. **16:** 296–302.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti *et al.*, 2001 Linkage disequilribium in the human genome. Nature **411:** 199–204.

Reich, D. E., S. F. Schaffner, M. J. Daly, G. McVean, J. C. Mullikin *et al.*, 2002 Human genome sequence variation and the influence of gene history, mutation and recombination. Nat. Genet. **32:** 135–142.

Risch, N., and K. Merikangas, 1996 The future of genetic studies of complex human diseases. Science **273:** 1516–1517.

Slatkin, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. **58:** 167–175.

Slatkin, M., 1994 Linkage disequilibrium in growing and stable populations. Genetics **137:** 331–336.

Strobeck, C., and K. Morgan, 1978 The effect of intragenic recombination on the number of alleles in a finite population. Genetics **88:** 829–844.

Takahata, N., 1995 A genetic perspective on the origin and history of humans. Annu. Rev. Ecol. Syst. **26:** 343–372.

Vitalis, R., and D. Couvet, 2001a Estimation of effective population size and migration rate from one- and two-locus identity measures. Genetics **157:** 911–925.

Vitalis, R., and D. Couvet, 2001b Two-locus identity probabilities

and identity disequilibrium in a partially selfing subdivided population. Genet. Res. **77:** 67–81.

Wakeley, J., 1997 Using the variance of pairwise differences to estimate the recombination rate. Genet. Res. **69:** 45–48.

Wakeley, J., 1998 Segregating sites in Wright's island model. Theor. Popul. Biol. **53:** 166–175.

Wakeley, J., 1999 Non-equilibrium migration in human history. Genetics **153:** 1863–1871.

Wakeley, J., and N. Aliacar, 2001 Gene genealogies in a metapopulation. Genetics **159:** 893–905. (corrigendum: Genetics **160:** 1263–1264).

Wakeley, J., R. Nielsen, S. N. Lui-Cordero and K. Ardlie, 2001 The discovery of single nucleotide polymorphisms and inferences about human historical demography. Am. J. Hum. Genet. **69:** 1332–1347.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wilkinson-Herbots, H. M., 1998 Genealogy and subpopulation differentiation under various models of population structure. J. Math. Biol. **37:** 535–585.

Wolfram, S., 1999 *The Mathematica Book*, Ed. 4. Wolfram Media/Cambridge University Press, Cambridge, UK.

Wright, A. F., A. D. Carothers and M. Pirastu, 1999 Population choice in mapping genes for complex diseases. Nat. Genet. **23:** 397–404.

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

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Communicating editor: W. Stephan

APPENDIX

Using Equations 9 and 11, we obtain the following equilibrium equations for the expected product of tree lengths at two sites in the island model, $V_s = E[T_s^{(1)}\ T_s^{(1)}]$, for states $s = 3, \ldots, 15$ defined in Table 1. To save space, we put $E_1 = E[T_1^{(1)}] = E[T_1^{(2)}]$, which is given by Equation 12, and $E_2 = E[T_2^{(1)}] = E[T_2^{(2)}]$, which is given by Equation 13:

$$V_3 = \frac{2E_1 + 2DV_4 + MDV_6}{(3 + M)D},$$

$$V_4 = \frac{8E_1 + RV_3 + 2DV_5 + MD(2V_7 + V_{10})}{6D + R + 3MD},$$

$$V_5 = \frac{4E_1 + RV_4 + MDV_{11}}{D + R + MD},$$

$$V_6 = \frac{4(D-1)(E_1 + E_2) + MDV_3 + 4D(D-1)V_7 + MD(V_8 + 2V_9) + MD(D-2)(V_{12} + 2V_{13})}{(6D + 3MD - 6 - 2M)D},$$

$$V_7 = \frac{4(D-1)(E_1 + E_2) + MDV_4 + R(D-1)V_6 + MDV_{10} + 2MD(D-2)V_{14}}{(2D + R + 2MD)(D-1)},$$

$$V_8 = \frac{2(D-1)E_1 + MDV_6 + MD(D-2)V_{12}}{(1 + M)(D-1)D},$$

$$V_9 = \frac{2(D-1)E_2 + MDV_6 + D(D-1)V_{10} + MD(D-2)V_{13}}{(1 + M)(D-1)D},$$

$$V_{10} = \frac{8(D-1)E_2 + MD(V_4 + 2V_7) + R(D-1)V_9 + 2D(D-1)V_{11} + 2MD(D-2)V_{14}}{R(D-1) + 2D(D-1) + MD(2D-1)},$$

$$V_{11} = \frac{4(D-1)E_2 + MDV_5 + R(D-1)V_{10}}{R(D-1) + MD},$$

$$V_{12} = \frac{2(D-1)(E_1 + E_2) + MD(V_6 + V_8 + 2V_{13}) + MD(D-3)V_{15}}{(D - 1 + M + MD)D},$$

$$V_{13} = \frac{4(D-1)E_2 + MD(V_6 + V_9 + V_{12}) + D(D-1)V_{14} + MD(D-3)V_{15}}{(D - 1 + MD)D},$$

$$V_{14} = \frac{8(D-1)E_2 + 2MD(2V_7 + V_{10}) + R(D-1)V_{13}}{R(D-1) + 6MD},$$

$$V_{15} = \frac{2(D-1)E_2 + MD(V_{12} + 2V_{13})}{3MD}.$$

Solving the system of equations above and taking the limit as D goes to infinity, we obtain the following approximations for the expected product of tree lengths at the two loci:

$$V_3 \approx \frac{4(1+M)(M^2(22+13R+R^2)+4M(24+13R+R^2)+2(36+14R+R^2))}{M(2+M)(3+M)(18+13R+R^2)},$$

$$V_4 \approx \frac{4(1+M)(36+14R+R^2+M(24+13R+R^2))}{M(2+M)(18+13R+R^2)},$$

$$V_5 \approx \frac{4(1+M)(36+14R+R^2)}{M(18+13R+R^2)},$$

$$V_6 \approx \frac{4(M^2(22+13R+R^2)+2(24+13R+R^2)+M(70+39R+3R^2))}{M(2+M)(18+13R+R^2)},$$

$$V_7 \approx \frac{4(1+M)(24+13R+R^2)}{M(18+13R+R^2)},$$

$$V_8 \approx \frac{4(22+13R+R^2)}{18+13R+R^2},$$

$$V_9 \approx \frac{4(36+14R+R^2+M^2(22+13R+R^2)+2M(24+13R+R^2))}{M^2(18+13R+R^2)},$$

$$V_{10} \approx \frac{4(1+M)(36+14R+R^2+M(24+13R+R^2))}{M^2(18+13R+R^2)},$$

$$V_{11} \approx \frac{4(1+M)^2(36+14R+R^2)}{M^2(18+13R+R^2)},$$

$$V_{12} \approx \frac{4(1+M)(22+13R+R^2)}{M(18+13R+R^2)},$$

$$V_{13} \approx \frac{4(1+M)(24+13R+R^2+M(22+13\mathrm{R}+R^2))}{M^2(18+13R+R^2)},$$

$$V_{14} \approx \frac{4(1+M)^2(24+13R+R^2)}{M^2(18+13R+R^2)},$$

$$V_{15} \approx \frac{4(1+M)^2(22+13R+R^2)}{M^2(18+13R+R^2)}.$$