DOI: 10.1534/genetics.107.075200

Linkage Disequilibrium Under Skewed Offspring Distribution Among Individuals in a Population

Bjarki Eldon[1] **and John Wakeley**

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received April 28, 2007
Accepted for publication December 21, 2007

ABSTRACT

Correlations in coalescence times between two loci are derived under selectively neutral population models in which the offspring of an individual can number on the order of the population size. The correlations depend on the rates of recombination and random drift and are shown to be functions of the parameters controlling the size and frequency of these large reproduction events. Since a prediction of linkage disequilibrium can be written in terms of correlations in coalescence times, it follows that the prediction of linkage disequilibrium is a function not only of the rate of recombination but also of the reproduction parameters. Low linkage disequilibrium is predicted if the offspring of a single individual frequently replace almost the entire population. However, high linkage disequilibrium can be predicted if the offspring of a single individual replace an intermediate fraction of the population. In some cases the model reproduces the standard Wright–Fisher predictions. Contrary to common intuition, high linkage disequilibrium can be predicted despite frequent recombination, and low linkage disequilibrium under infrequent recombination. Simulations support the analytical results but show that the variance of linkage disequilibrium is very large.

LINKAGE disequilibrium (LD) refers to the nonrandom association of alleles at different loci (Lewontin and Kojima 1960). Changes in population size, natural selection, population structure, and random drift can all lead to LD. Recombination, or the reciprocal exchange of material between homologous chromosomes, breaks down associations between alleles at different loci. Estimating LD can thus give insight into the forces that have shaped extant genetic diversity. The potential utility of LD for fine-scale mapping of human disease loci has also raised interest in estimating levels of linkage disequilibrium in human populations (Jorde 1995; Lander 1996; Risch and Merikangas 1996). The evolutionary history of many organisms is marked by growth and decline of populations as well as various kinds of subdivision (with or without migration and admixture). As an example, the Icelandic human population has undergone severe bottlenecks, accompanied by recent population growth, in its ~1100-year history (Thorarinsson 1961; Thorsteinsson and Jónsson 1991; Jónsson and Magnússon 1997). Bataillon *et al.* (2006) report extensive linkage disequilibrium in the Icelandic human population and estimate the effective population size N_e to be ~5000, much less than the current census size of ~300,000 (Garðarsdóttir and Sigurjónsson 2006).

Linkage disequilibrium is a function of the frequencies of alleles in the population and LD can be quantified as a function of allele frequencies in a number of ways (*cf.* Hedrick 2000). One commonly used measure of LD is the coefficient D of linkage disequilibrium and is defined as the difference between the observed frequency of a gametic type (haplotype) and the frequency expected on the basis of random association of alleles in gametes (Lewontin and Kojima 1960). The coefficient D can be written as $D = P_{AB}P_{ab} - P_{aB}P_{Ab}$ in which P_{xy} is the frequency of haplotype xy. High absolute values of D correspond to high linkage disequilibrium in the population.

Following Slatkin (1994) we can understand the effects of population history on LD between two diallelic loci by considering the shape of the gene genealogy of a sample without recombination. In this case the alleles at both loci have the same gene genealogy. As an example, consider a population that has recently grown in size. Figure 1a shows a gene genealogy of a sample from a population that has experienced recent expansion. A single neutral mutation has arisen at each locus. The location of the mutations on the gene genealogy determines the level of linkage disequilibrium. Since neutral mutations arise randomly on the genealogy, the shape of the gene genealogy becomes a deciding factor. A gene genealogy of a sample from a recently expanded population is composed mainly of external branches. This happens because most coalescence events occur in the smaller ancestral population. Hence, each mutation is most likely to arise on an external branch and will therefore be present on a single haplotype in the sample. Thus the haplotypes observed in a sample of size n would

[1] *Corresponding author:* 4100 Biological Laboratories, 16 Divinity Ave., Harvard University, Cambridge, MA 02138.
E-mail: eldon@fas.harvard.edu

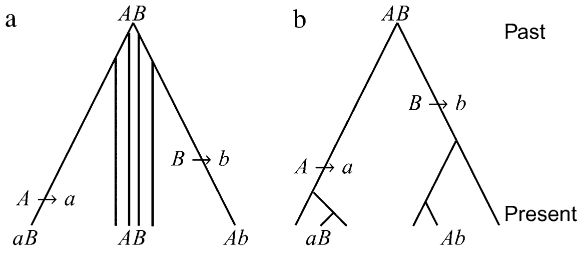

Figure 1.—Gene genealogies of a sample of two completely linked loci (a) from a population that has recently grown in size and (b) from a stable population. One mutation event is assumed at each locus, and the ancestral gametic type is *AB*. The sample in a consists of one each of *aB* and *Ab*, and four *AB*, indicating low linkage disequilibrium. The sample in b consists of the gametic types *aB* and *Ab* in equal proportions, indicating high linkage disequilibrium.

be one each of *aB* and *Ab*, and $n-2$ of *AB*. In this case, $|D| = |((n-2)/n) \times 0 - 1/n^2| = 1/n^2$, which becomes quite small for large n.

For comparison, Figure 1b shows a gene genealogy for a sample from a stable population. Now, the two mutations are more likely to co-occur on an internal branch that has, by definition, more than one descendent in the sample. The result is only two gametic types; *i.e.*, we could observe $n/2$ each of *aB* and *Ab*, and it follows that $D = \frac{1}{4}$, or the largest possible value that D can have. Any ancestral process that increases the chance of events like this will tend to increase LD.

In this work, we consider another commonly used measure of LD, r^2, which is given by

$$r^2 = \frac{D^2}{p_a(1-p_a)p_b(1-p_b)} \tag{1}$$

(Hill and Robertson 1968), in which p_x is the frequency of allele x, and which is commonly used to estimate the statistical association between alleles at two loci. Ohta and Kimura (1971) suggested approximating the expected value of r^2 as a ratio of expected values. This approximation appears good provided the allele frequencies are not too small (Hudson 1985; McVean 2002). Considering the gene genealogy of a sample of size two from two loci and using the results of Strobeck and Morgan (1978) and Hudson (1985), McVean (2002) showed that an approximation to the expected value of r^2 can be written in terms of correlation in coalescence times. Griffiths (1981, 1991) (see also Pluzhnikov and Donnelly 1996; Durrett 2002) gives the covariances in coalescence times based on the standard Wright–Fisher model of reproduction (Fisher 1930; Wright 1931). Since McVean's (2002) approach makes no assumptions about reproduction, predictions about LD can be obtained under different population models.

The present study derives correlations in coalescence times in cases where a single individual can have very large number of offspring with some probability. In terms of reproduction, our model is a special case of the models considered by Sagitov (1999) and Pitman (1999). In terms of genetics, our model is novel because we include the possibility of recombination. Sagitov (1999) and Pitman (1999) did not consider recombination, but proved convergence to an ancestral process that allows for many ancestral lines to reach a common ancestor, or coalesce, at exactly the same instant (or same generation) and that occurs on a shorter timescale than in the standard coalescent (Pitman 1999; Sagitov 1999; Schweinsberg 2000; Möhle and Sagitov 2001). Predictions of patterns of genetic diversity also differ from those under Kingman's coalescent (Eldon and Wakeley 2006; Möhle 2006). Such models may be appropriate for many marine organisms with high fecundities and high mortality in early life stages or type III survivorship curves (Hedgecock 1994).

Depending on parameter values, these models can predict much lower levels of genetic variation than would be expected on the basis of census size. This is often observed in marine species and is quantified using the ratio of effective to census size, N_e/N. Low N_e/N ratios reported in Atlantic cod (*Gadus morhua*; Árnason 2004), red drum (*Sciaenops ocellatus*; Turner *et al.* 2002), and the Pacific oyster (*Crassostrea gigas*; Hedgecock 1994) have been thought to indicate high variance in offspring number (Crow and Kimura 1970; Hedrick 2005). Hedgecock (1994) proposed a "sweepstakes" reproduction model in which lucky individuals may contribute a large number of offspring to the next generation.

Here we show that allowing individuals to have many offspring typically results in low predicted LD in the population. In some cases, however, higher LD than that predicted under the standard Wright–Fisher model is obtained. Low LD can also be predicted under low recombination, and high LD under high recombination, contrary to common intuition. Finally, the different formulas representing different timescales on which recombination and random drift occur can predict the same level of linkage disequilibrium. This implies that it may be difficult to distinguish between the recombination parameter and the parameters controlling the size and frequency of the large reproduction events using sequence data. Our analytical results are for the expected value of r^2, but we have also performed a simulation study that shows that the variance of r^2 is typically very large.

THEORY AND METHODS

Population models: The modified population model considered is a special case of the neutral population models analyzed by Sagitov (1999) and Pitman (1999). A discrete-generation model, it is a modification of the well-known Wright–Fisher (Fisher 1930; Wright 1931) model of reproduction. A modified Moran model (Moran 1958, 1962) of overlapping generations introduced and

TABLE 1

Expected time to coalescence of two lines under the modified Wright–Fisher population model

Case	λ_{coal}	Timescale
$0 < \alpha < 1$	$\phi\omega^2$	N^{α}
$\alpha = 1$	$1 + \phi\omega^2$	$2N$
$\alpha > 1$	1	$2N$

The rate of coalescence is denoted by λ_{coal}.

studied by Eldon and Wakeley (2006) was also considered. Since the results obtained under the modified Moran model (not shown) are compatible with those obtained under the modified Wright–Fisher population model, we consider only the modified Wright–Fisher model and results derived under that model from now on.

Eldon and Wakeley (2006) treated only haploid individuals. Since the present study addresses recombination, a diploid population is assumed. For the moment, consider a diploid population without recombination or mutation. Note that we do not treat mutation explicitly here, but follow McVean (2002) in assuming that the mutation rate per site is small. As usual, N denotes the population size. The Wright–Fisher model is modified as follows. With probability $1 - \varepsilon$ $(0 < \varepsilon < 1)$ the usual Wright–Fisher sampling occurs; *i.e.*, all individuals contribute equally to the next generation via multinomial sampling, and all are replaced by offspring each generation. With probability ε the offspring of a single randomly chosen individual replace a fraction ω of the population, and the other $2N - 1$ individuals share the remaining $1 - \omega$ fraction of reproduction events according to the usual Wright–Fisher sampling. The probability ε of modified Wright–Fisher sampling is taken as $\varepsilon = \phi N^{-\alpha}$ in which both constants ϕ and $\alpha > 0$. In Eldon and Wakeley (2006) $\phi = 1$. The parameter ϕ allows us to adjust the relative rates of coalescence and recombination when both occur on the same timescale.

Our first concern is with the timescale of coalescence under this model. Under the modified Wright–Fisher population model the expected coalescence time of two lines can be much shorter than under the standard coalescent (Table 1). Consider first the case $0 < \alpha < 1$, in which α controls the timescale at which modified sampling occurs $(\varepsilon = \phi/N^{\alpha})$. In this case, the result is a multiple-mergers coalescent, since "ω-events" (a single individual has $2N\omega$ offspring) occur on a shorter timescale (proportional to N^{α} generations) than the standard coalescent (timescale: proportional to N generations). In the case $\alpha = 1$, all the coalescence events occur on the same timescale. When $\alpha > 1$, the ω-events occur on a longer timescale than the standard coalescent and are thus negligible in large populations. In the first two cases (*i.e.*, $0 < \alpha \leq 1$), the expected time to coalescence is a function of ϕ and ω.

The term "x-merger" denotes the event that x ancestral lines derive from a single individual in one time step. Given n ancestral lines, we are interested in coalescent events: $2 \leq x \leq n$. Let $G_{n,x}$ denote the probability of an x-merger among n ancestral lines. In general, for finite N,

$$\begin{aligned} G_{n,x} = {} & \binom{n}{x}(1-\varepsilon)\left(\frac{1}{2N}\right)^{x-1} A(n-x, 2N) \\ & + \binom{n}{x}\varepsilon\Bigg[(1-\omega)^{n}\left(\frac{1}{2N-1}\right)^{x-1} A(n-x, 2N-1) \\ & \qquad + \delta(x, n-1)\, n\omega(1-\omega)^{n-1}\left(\frac{1}{2N-1}\right)^{x-1} \\ & \qquad \times A(n-x-1, 2N-1) \\ & \qquad + \omega^{x}(1-\omega)^{n-x} A(n-x, 2N-1)\Bigg] \end{aligned} \tag{2}$$

in which $A(m, M) = (1 - 1/M)(1 - 2/M) \cdots (1 - m/M)$, $A(0, M) = 1$, and $\delta(\cdot, \cdot)$ is the delta function

$$\delta(y, j) = \begin{cases} 1 & \text{if } 2 \leq y \leq j \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The function $G_{n,x}$ determines the distribution of the size and shape of gene genealogies when there is no recombination.

McVean (2002) obtains an expression for an approximation to the expected value of r^2 in terms of covariances in coalescence times, assuming that the per-site mutation rate at each locus is very small. In so doing, the gene genealogy of a sample of size two at each of two loci is modeled backward in time using a Markov chain. There are three states in the chain, which correspond to three possible configurations of the sample, and are denoted $\mathscr{S}_0$, $\mathscr{S}_1$, and $\mathscr{S}_2$ (see Figure 2). The subscripts 0, 1, and 2 refer to the number of haplotypes the four genetic types share. The $\mathscr{C}_i$ states in Figure 2 represent ancestral configurations that include a common-ancestral type at one or both loci. A prediction about linkage disequilibrium in a sample is obtained by considering the covariances in coalescence times between two loci.

Recombination is included in the model as follows. To illustrate, assume as in Strobeck and Morgan (1978) that four haplotypes labeled AB, Ab, aB, and ab are segregating in the population with frequency p_1, p_2, p_3, and p_4, respectively, at a given time. Each generation every individual contributes a large number of gametes (*i.e.*, haplotypes) to a common gamete pool. Zygotes are formed by selecting two gametes at random from the gamete pool. Given a zygote formed from haplotypes AB and ab, a single haplotype for the next generation is selected from the four possible meiotic products AB, Ab, aB, and ab with probabilities $(1 - c)/2$, $c/2$, $c/2$, and $(1 - c)/2$, respectively, in which c is the per-generation recombination value $(0 \leq c \leq 1)$. Thus the frequency of haplotype i in the next generation is given by $p_i' = p_i + s_i(p_2p_3 - p_1p_4)c$ for $i = 1, \ldots, 4$ in which $s_1 = s_4 = 1$ and $s_2 = s_3 = -1$ (*cf.* Ewens 2004). When a large reproduction event occurs, a single randomly chosen haplotype represents a fraction ω of the common gamete pool and

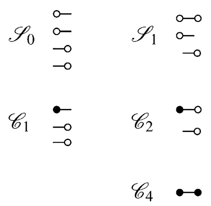

Figure 2.—The states in the ancestral process of two biallelic loci for a sample of size two of two loci from a diploid population. A genetic type at locus 1 is denoted by ○— while —○ denotes a genetic type at locus 2. The common-ancestral genetic types are denoted by ●— and —●.

the other haplotypes compose the remaining $1 - \omega$ fraction of gametes. In this case, the formula for p_i' above holds, but with p_i and $p_{j \neq i}$ replaced by $p_i^* = \omega + p_i(1-\omega)$ and $p_j^* = p_j(1-\omega)$ for $j \neq i$. Note that simultaneous multiple mergers (Schweinsberg 2000; Möhle and Sagitov 2001) are not possible in this simplified model of recombination.

Prediction of linkage disequilibrium: The quantity r^2 given in Equation 1 is a ratio of two nonindependent random variables and has an unknown distribution, but Song and Song (2007) do make an advance in its numerical evaluation.

A first approximation to $E(r^2)$ is the ratio of expectations $E(D^2)/E(p_a(1-p_a)p_b(1-p_b))$ (Ohta and Kimura 1971). We write $\Upsilon = E(D^2)/E(p_a(1-p_a)p_b(1-p_b))$. Considering the gene genealogy of a sample of size two of two loci, McVean (2002) showed that the ratio of expectations Υ can be written in terms of covariances in pairwise coalescence times t_1 and t_2 at each of two loci,

$$\Upsilon = \frac{\mathrm{Cov}(t_1, t_2 \mid \mathscr{S}_0) - 2\,\mathrm{Cov}(t_1, t_2 \mid \mathscr{S}_1) + \mathrm{Cov}(t_1, t_2 \mid \mathscr{S}_2)}{E(t_1)^2 + \mathrm{Cov}(t_1, t_2 \mid \mathscr{S}_0)} \tag{4}$$

in which $\mathscr{S}_0$, $\mathscr{S}_1$, and $\mathscr{S}_2$ denote the three possible configurations in a sample of size two shown in Figure 2. In the case of small samples, Equation 4 can be corrected for the possibility that the same gamete is sampled more than once (Hudson 1985; McVean 2002). In deriving Equation 4, McVean (2002) assumed that the per-site mutation rate is very small.

The covariance terms in Equation 4 have been derived under the standard coalescent (Griffiths 1981, 1991) (see also Pluzhnikov and Donnelly 1996; Durrett 2002). Let ρ denote the scaled recombination rate. Under standard Wright–Fisher sampling, in which time is measured in units of $2N$ generations, $\rho = 2Nc$. Then the corresponding correlations given each sample configuration ($\mathscr{S}_0$, $\mathscr{S}_1$, or $\mathscr{S}_2$; see Figure 2) are given by Equation A1 in the appendix (by replacing η with $\rho/4$). Hence,

$$\Upsilon = \frac{5 + \rho}{11 + 13\rho + 2\rho^2} \tag{5}$$

(McVean 2002), which agrees with the results obtained by Ohta and Kimura (1971) and Weir and Hill (1986) by other methods. Simulations show that Equation 5 provides a good approximation to the average value of r^2 calculated from a sample when the frequency of the minor allele is not too small (Hudson 1985; McVean 2002).

Deriving the covariance terms under a modified population model: The correlations in Equation A1 were obtained on the basis of the standard coalescent (Kingman 1982a,b; Hudson 1983; Tajima 1983), in which only binary mergers are allowed; hence the only parameter is the scaled recombination rate ρ (Griffiths 1981, 1991). The corresponding correlation terms derived under our modified Wright–Fisher population model are shown in the appendix. In this case, and in what follows, we define the scaled recombination parameter to be $\eta = cN^{\beta}$. We assume that η and ϕ are finite; *i.e.*, $\lim_{N\to\infty} cN^{\beta}$ and $\lim_{N\to\infty} \varepsilon N^{\alpha}$ are both finite (recall that the probability of modified Wright–Fisher sampling is $\varepsilon = \phi/N^{\alpha}$). Note that the usual scaling of the recombination rate is obtained by taking $\beta = 1$. Defining the recombination parameter η in this way allows us to investigate effects of order of magnitude differences in timescales of recombination and coalescence. The parameter controlling the timescale of coalescence is α. When $\alpha \geq 1$ coalescence occurs on a timescale proportional to N generations. However, if $0 < \alpha < 1$, the timescale of coalescence is N^{α} generations (Table 1). To obtain a continuous-time limit we rescale time using the coalescent timescale. Thus the rate of coalescence shown in Table 1 is always finite. On the coalescent timescale the rate of recombination is $cN^{\alpha} = \eta N^{\alpha-\beta}$ (we can think of this as our model's analog of the usual parameter ρ). Since η is finite, $\lim_{N\to\infty} \eta N^{\alpha-\beta} = \infty$ if $\alpha > \beta$, *i.e.*, when recombination is an order of magnitude more frequent than coalescence. If $\alpha < \beta$ then $\lim_{N\to\infty} \eta N^{\alpha-\beta} = 0$, and coalescence events are an order of magnitude more frequent than recombination events. Coalescence and recombination occur on the same timescale only when $\alpha = \beta$.

To explain the derivations, recall that Υ can be expressed in terms of covariances (see Equation 4) and since $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$ for any two random variables X and Y, we first obtain the expected values of the products of the pairwise coalescence times t_1 and t_2 at the two loci. That is, the main work involves obtaining $E(t_1 t_2 \mid \mathrm{S}_i)$ for $i = 0, 1, 2$ (see Figure 2). Let s_{ij} denote the transition probability over a single generation between states in the ancestral process for two biallelic loci with recombination between them, under a standard population model (*e.g.*, the Wright–Fisher model). Denote by m_{ij} the transition probability per generation under a modified sampling scheme. Then under a modified population model the transition probability p_{ij} over one generation from state i to state j is

$$p_{ij} = (1 - \varepsilon)s_{ij} + \varepsilon m_{ij}. \tag{6}$$

The gene genealogical process we are describing is that for a sample of size two of two loci, or a total of four genetic types. The transition probabilities take the general form of Equation 6 in which s_{ij} and m_{ij} are given in the APPENDIX, and the seven possible states are given in Figure 2.

By considering the corresponding Markov jump chain, we can follow DURRETT (2002) to obtain $E(t_1 t_2 \mid \mathscr{S}_i)$ in which t_1 and t_2 denote the time until coalescence of the two genetic types at loci 1 and 2, respectively. Conditioning on the first change in the genealogical history of the sample we obtain the set of recursions given in Equation 7,

$$\begin{aligned} E(t_1 t_2 \mid \mathscr{S}_0) &= q_{\mathscr{S}_0\mathscr{S}_1}E(t_1 t_2 \mid \mathscr{S}_1) + q_{\mathscr{S}_0\mathscr{S}_2}E(t_1 t_2 \mid \mathscr{S}_2) \\ &\quad + E(J \mid \mathscr{S}_0)E(t_c)S_0 + E(J^2 \mid \mathscr{S}_0) \\ E(t_1 t_2 \mid \mathscr{S}_1) &= q_{\mathscr{S}_1\mathscr{S}_0}E(t_1 t_2 \mid \mathscr{S}_0) + q_{\mathscr{S}_1\mathscr{S}_2}E(t_1 t_2 \mid \mathscr{S}_2) \\ &\quad + E(J \mid \mathscr{S}_1)E(t_c)S_1 + E(J^2 \mid \mathscr{S}_1) \\ E(t_1 t_2 \mid \mathscr{S}_2) &= q_{\mathscr{S}_2\mathscr{S}_0}E(t_1 t_2 \mid \mathscr{S}_0) + q_{\mathscr{S}_2\mathscr{S}_1}E(t_1 t_2 \mid \mathscr{S}_1) \\ &\quad + E(J \mid \mathscr{S}_2)E(t_c)S_2 + E(J^2 \mid \mathscr{S}_2), \end{aligned} \tag{7}$$

in which $q_{\mathscr{S}_i\mathscr{S}_j} = p_{ij}/(1 - p_{ii})$, $S_0 = 2(q_{\mathscr{S}_0\mathscr{S}_1} + q_{\mathscr{S}_0\mathscr{S}_2}) + q_{\mathscr{S}_0\mathscr{C}_1} + q_{\mathscr{S}_0\mathscr{C}_2}$, $S_1 = 2(q_{\mathscr{S}_1\mathscr{S}_0} + q_{\mathscr{S}_1\mathscr{S}_2}) + q_{\mathscr{S}_1\mathscr{C}_1} + q_{\mathscr{S}_1\mathscr{C}_2}$, and $S_2 = 2(q_{\mathscr{S}_2\mathscr{S}_0} + q_{\mathscr{S}_2\mathscr{S}_1}) + q_{\mathscr{S}_2\mathscr{C}_1} + q_{\mathscr{S}_2\mathscr{C}_2}$. Since J denotes the time until the process moves out of a particular state, $E(J \mid \mathscr{S}_i) = 1/(1 - q_{\mathscr{S}_i\mathscr{S}_i})$ and $E(J^2 \mid \mathscr{S}_i) = (1 + q_{\mathscr{S}_i\mathscr{S}_i})/(1 - q_{\mathscr{S}_i\mathscr{S}_i})^2$. The quantity t_c is the time until two lines coalesce (ignoring recombination), for which $E(t_c) = 1/G_{2,2}$, and $G_{n,x}$ given in Equation 2 is the probability that x lines of n coalesce in one generation. Note that Equation 4 can also be written as

$$\Upsilon = 1 + \frac{E(t_1 t_2 \mid \mathscr{S}_2) - 2E(t_1 t_2 \mid \mathscr{S}_1)}{E(t_1 t_2 \mid \mathscr{S}_0)}. \tag{8}$$

Let $E(t_1 t_2 \mid \mathscr{S}_i)$ denote the solution of Equation 7 for $i = 0, 1, 2$. The continuous-time limit prediction of linkage disequilibrium is then given by

$$\Upsilon = 1 + \lim_{N\to\infty} \frac{E(t_1 t_2 \mid \mathscr{S}_2) - 2E(t_1 t_2 \mid \mathscr{S}_1)}{E(t_1 t_2 \mid \mathscr{S}_0)}. \tag{9}$$

The limit in Equation 9 will depend on the different values of α and β, the parameters controlling the timescale of coalescence and recombination, respectively. The different continuous-time limit predictions of linkage disequilibrium obtained from Equation 9 are given in Table 2, and the correlations are given in the APPENDIX.

RESULTS

Correlations in coalescence times are strongly affected by demography, in this case extreme differences in reproductive success among individuals in a population. The correlations obtained under different assumptions about the two parameters, α and β, that control the timescales of recombination and random drift are shown in the APPENDIX. When found to be functions of ω (the fraction of the population replaced by offspring of a single individual) and ϕ (recall that the probability of modified Wright–Fisher sampling is given by $\varepsilon = \phi/N^{\alpha}$), the correlations ascend to 1 as ω and ϕ increase.

Different predictions of LD: Eleven different predictions (Equation 9) of linkage disequilibrium are identified (Table 2), depending on α and β (recombination is scaled as $\eta = cN^{\beta}$). These results are summarized in Figure 3 on the parameter space spanned by α and β. The timescale in a standard Wright–Fisher diploid population is $2N$ generations. Thus when $\alpha > 1$ ω-coalescence events are an order of magnitude less frequent than binary mergers, and the ancestral process is Kingman's coalescent. The coalescent timescale when $0 < \alpha < 1$ is in units of N^{α} generations, and the ancestral process is dominated by ω-coalescence events allowing for multiple mergers. Kingman's coalescent and ω-coalescence events occur on the same timescale (proportional to N generations) when $\alpha = 1$. Recombination occurs on a faster timescale (by an order of magnitude) than any type of coalescent event on the region of the (α, β) parameter space represented by zero ($\beta < 1$ except $0 < \alpha < \beta < 1$) in Figure 3. Thus this region is labeled as "frequent recombination" in Figure 3. The remaining part of the (α, β) parameter space is labeled as "infrequent recombination" to remind us of the longer timescale (for this part of the parameter space) on which recombination occurs relative to the corresponding coalescent timescale.

Just three possible limiting behaviors—0, $\frac{5}{11}$, and $\Upsilon(\omega)$—occupy nearly all of the (α, β) parameter space (Figure 3). Linkage equilibrium ($r^2 = 0$) is predicted when $0 < \beta < \alpha < 1$ or when $0 < \beta < 1$ and $\alpha \geq 1$—this is the region of the (α, β) parameter space occupied by 0 in Figure 3. In these cases, the rate of recombination is an order of magnitude higher than any type of coalescence event ($\lim_{N\to\infty} N^{\alpha}c = \lim_{N\to\infty} \eta N^{\alpha-\beta} = \infty$) and hence we do not expect to see any disequilibrium. In contrast, high linkage disequilibrium is predicted when both α and $\beta > 1$, which is the region occupied by $\frac{5}{11}$ in Figure 3. Here the timescale of coalescence is proportional to N generations, the ancestral process is Kingman's coalescent, and recombination occurs on a timescale that is an order of magnitude longer than the coalescent timescale; *i.e.*, $\lim_{N\to\infty} Nc = \lim_{N\to\infty} \eta N^{1-\beta} = 0$. Note that $\frac{5}{11}$ is the value obtained from Equation 7 when $\rho = 0$. Finally, when $0 < \alpha < \beta$ and $\alpha < 1$, the prediction of linkage disequilibrium is a function of ω (the fraction of the population replaced by offspring of a single individual)—the region occupied by $\Upsilon(\omega)$ (Table 2) in Figure 3. The ancestral process occurs on a timescale of N^{α} generations and is characterized by ω-coalescence events allowing for multiple mergers. Since $\beta > \alpha$ it follows that recombination occurs on a timescale that is longer (by an order of magnitude) than the coalescent timescale ($\lim_{N\to\infty} \eta N^{\alpha-\beta} = 0$).

TABLE 2

The continuous-time limit predictions of linkage disequilibrium Υ under a modified Wright–Fisher population model

Timescale		
Coalescence	Recombination	Limit process
$\alpha > 1$	$\beta > 1$	$\dfrac{5}{11}$
	$\beta = 1$	$\dfrac{5+\eta}{11+13\eta+2\eta^2}$
	$\beta < 1$	0
$\alpha = 1$	$\beta > 1$	$\Upsilon_3(\phi, \omega)$
	$\beta = 1$	$\Upsilon_2(\eta, \phi, \omega)$
	$\beta < 1$	0
$\alpha < 1$	$\beta > 1$	$\Upsilon(\omega)$
	$\beta = 1$	$\Upsilon(\omega)$
$0 < \alpha < \beta < 1$		$\Upsilon(\omega)$
$0 < \alpha = \beta < 1$		$\Upsilon_1(\eta, \phi, \omega)$
$0 < \beta < \alpha < 1$		0

The different formulas are the limit (9) obtained given different values of α and β. See text for explanation of symbols. Note that $\Upsilon_4(\eta) = (5+\eta)/(11+13\eta+2\eta^2)$.

$$\Upsilon(\omega) = \frac{(1-\omega)^2(5-4\omega)}{11-22\omega+16\omega^2-4\omega^3}$$

$$\Upsilon_1(\eta, \phi, \omega) = \frac{\phi(1-\omega)^2\omega^2(\phi(5-4\omega)\omega^2+\eta)}{\phi^2(11-22\omega+16\omega^2-4\omega^3)\omega^4+\phi\eta(13-18\omega+8\omega^2)\omega^2+2\eta^2}$$

$$\Upsilon_2(\eta, \phi, \omega) = \frac{(5+\eta+(5-4\omega)\phi\omega^2)(1+(1-\omega)^2\phi\omega^2)}{11+2\eta^2+2\phi\omega^2(11-11\omega+4\omega^2)+\phi^2\omega^4(11-22\omega+16\omega^2-4\omega^3)+(13+\phi\omega^2(13-18\omega+8\omega^2))\eta}$$

$$\Upsilon_3(\phi, \omega) = \frac{\phi^2(1-\omega)^2(5-4\omega)\omega^4+\phi(10-14\omega+5\omega^2)\omega^2+5}{11+2\phi(11-11\omega+4\omega^2)\omega^2+\phi^2(11-22\omega+16\omega^2-4\omega^3)\omega^4}.$$

When α and β are equal and ≤ 1, recombination and coalescence occur on the same timescale of $N^\alpha \leq N$ generations, and more complicated behaviors are observed. The prediction of LD is then a function of all three parameters η, ϕ, and ω, and two different limit processes, $\Upsilon_1(\eta, \phi, \omega)$ and $\Upsilon_2(\eta, \phi, \omega)$ (Table 2), arise. The limit process $\Upsilon_1(\eta, \phi, \omega)$ is obtained when $0 < \alpha = \beta < 1$, represented by the thick diagonal line labeled ① in Figure 3, and the limit process $\Upsilon_2(\eta, \phi, \omega)$ represented by ② in Figure 3 results when $\alpha = \beta = 1$.

When $\alpha = 1$, the ancestral process is a mixture of the standard coalescent and a multiple-mergers coalescent process, since both occur on a timescale proportional to N generations. Taking $\alpha = 1$ and restricting recombination to a longer timescale ($\beta > 1$) results in the limit process $\Upsilon_3(\phi, \omega)$ (Table 2), which occupies the part of the (α, β) parameter space represented by the thick vertical line labeled ③ in Figure 3. When $\alpha > 1$ the standard coalescent process results, the timescale is proportional to N generations, and the population follows the usual Wright–Fisher reproduction framework. Scaling recombination in the usual way by taking $\beta = 1$ under the usual Wright–Fisher reproduction ($\alpha > 1$) results in the standard prediction $\Upsilon_4(\eta)$ (Table 2) of LD represented by the thick horizontal line labeled ④ in Figure 3.

Nonmonotonic behavior of $E(r^2)$: Interestingly, while linkage disequilibrium decreases monotonically as recombination increases, the predictions Υ_1 and Υ_2 are nonmonotonic functions of ϕ and ω. Figure 4a shows $\Upsilon_1(\eta, \phi, \omega)$ as a function of ϕ ($1 \leq \phi \leq 10$) and ω ($0 < \omega < 1$) when $\eta = 1$, and Figure 4b shows $\Upsilon_2(\eta, \phi, \omega)$ under the same conditions for η, ϕ, and ω as in Figure 4a. A comparison of the two graphs in Figure 4 shows that

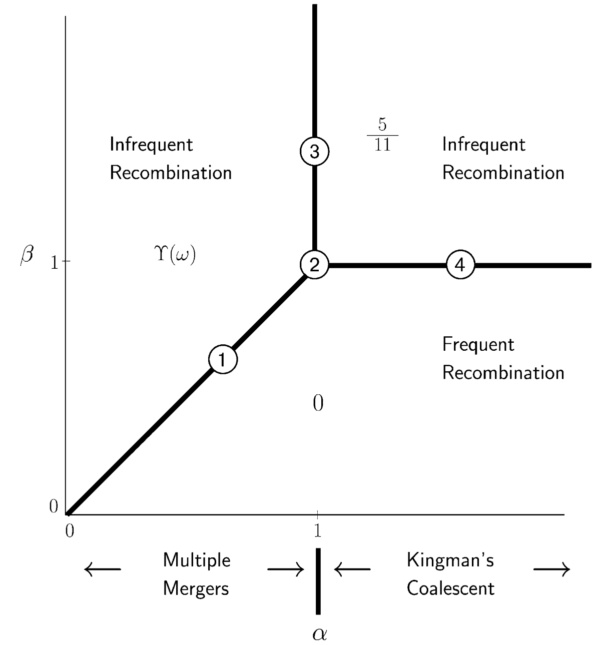

Figure 3.—Predictions of linkage disequilibrium from Table 2 on the (α, β) parameter space. The circled numbers refer to Υ_j ($j = 1, \ldots, 4$) as shown in Table 2. For explanation of symbols see text.

$\Upsilon_1(\eta, \phi, \omega)$ and $\Upsilon_2(\eta, \phi, \omega)$ behave very similarly as functions of ϕ and ω. The nonmonotonic trends emerging from Figure 4 are twofold. First, for any value of ϕ, $\Upsilon_1(\eta, \phi, \omega)$ and $\Upsilon_2(\eta, \phi, \omega)$ ascend as ω goes from 0 to ~0.4 and then descend as ω goes from ~0.4 to 1. The other nonmonotonic trend is that, depending on ω, $\Upsilon_1(\eta, \phi, \omega)$ and $\Upsilon_2(\eta, \phi, \omega)$ are either increasing or decreasing functions of ϕ. Thus, when multiple mergers and recombination occur on the same timescale of $\leq 2N$ generations, the prediction of linkage disequilibrium is nonmonotonic in the parameters that control the rate (ϕ) and size (ω) of multiple mergers.

The nonmonotonic behavior of Υ_1 and Υ_2 is interesting because the correlation in coalescence times increases monotonically with ϕ and ω. Let $\sigma_i = \mathrm{corr}(t_1, t_2 \mid \mathscr{I}_i)$, $i = 0, 1, 2$. Writing $\Upsilon = \sigma_0/(1 + \sigma_0) + \sigma_2/(1 + \sigma_0) - 2\sigma_1/(1 + \sigma_0)$ and looking at each term separately one can see that, relative to $1 + \sigma_0$, σ_2 rises most steeply over the same interval of ω for which Υ is increasing (Figure 5).

The ancestral process occurs on a timescale of N^{α} generations when $0 < \alpha < 1$ and is characterized by ω-coalescence events allowing for multiple mergers. Thus when recombination occurs on a longer timescale (even if $0 < \beta < 1$, *cf.* Figure 3) the prediction of linkage disequilibrium depends only on ω [for this range of the (α, β)-parameter space]. Figure 6a shows $\Upsilon(\omega)$ is a decreasing function of ω. When ω is very small, high LD is predicted since the gene genealogy resembles the one obtained under standard Wright–Fisher reproduction (see Figure 1b). As ω increases, the gene genealogy starts to resemble more the one shown in Figure 1a, *i.e.*, consisting mostly of external branches because of multiple mergers, leaving little opportunity for LD to establish.

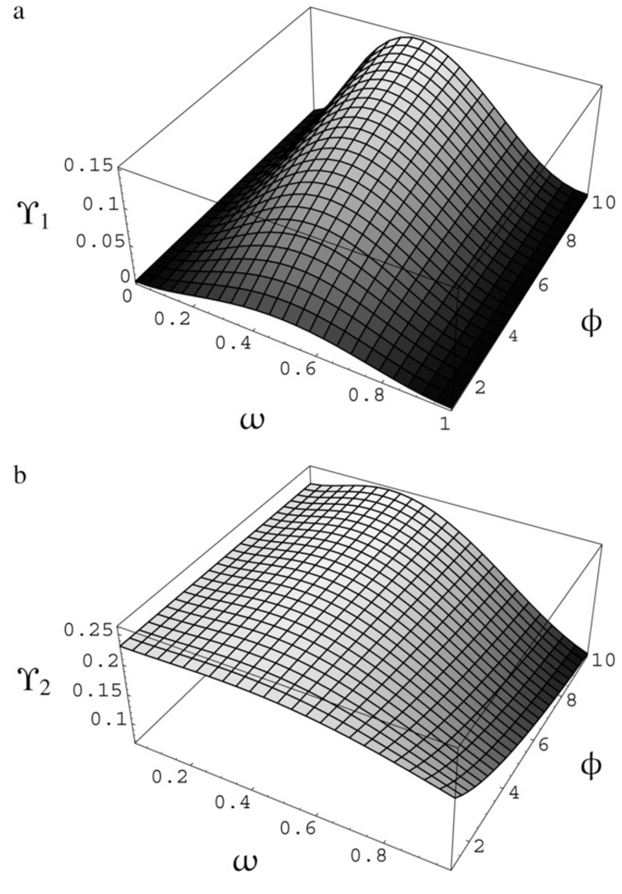

Figure 4.—Predicted levels of linkage disequilibrium (a) $\Upsilon_1(\eta, \phi, \omega)$ and (b) $\Upsilon_2(\eta, \phi, \omega)$, from Table 2 as a function of ϕ and ω when $\eta = 1$. For explanation of symbols see text.

In Figure 4, $\eta = 1$, which gives $\Upsilon_4(1) \approx 0.231$. Thus, for the range of ϕ chosen, predicted levels of linkage

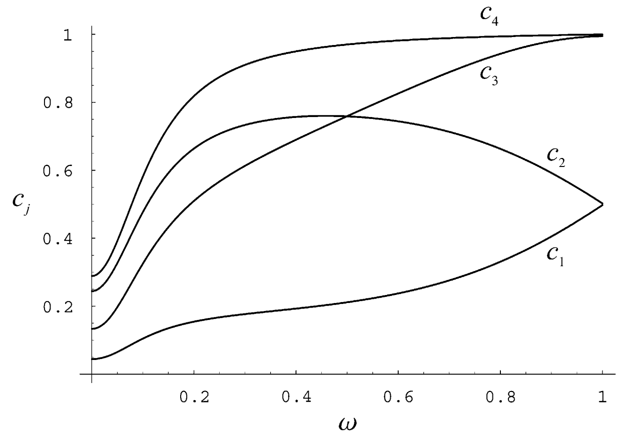

Figure 5.—Correlations in coalescence times σ_i, relative to $1 + \sigma_0$, as functions of ω when $\eta = 1$, $\phi = 100$, and $\beta = \alpha = 1$ (see appendix). Line c_1, $\sigma_0/(1 + \sigma_0)$; line c_2, $\sigma_2/(1 + \sigma_0)$; line c_3, $2\sigma_1/(1 + \sigma_0)$; line c_4, $(\sigma_2 + \sigma_0)/(1 + \sigma_0)$. For explanation of symbols see text.

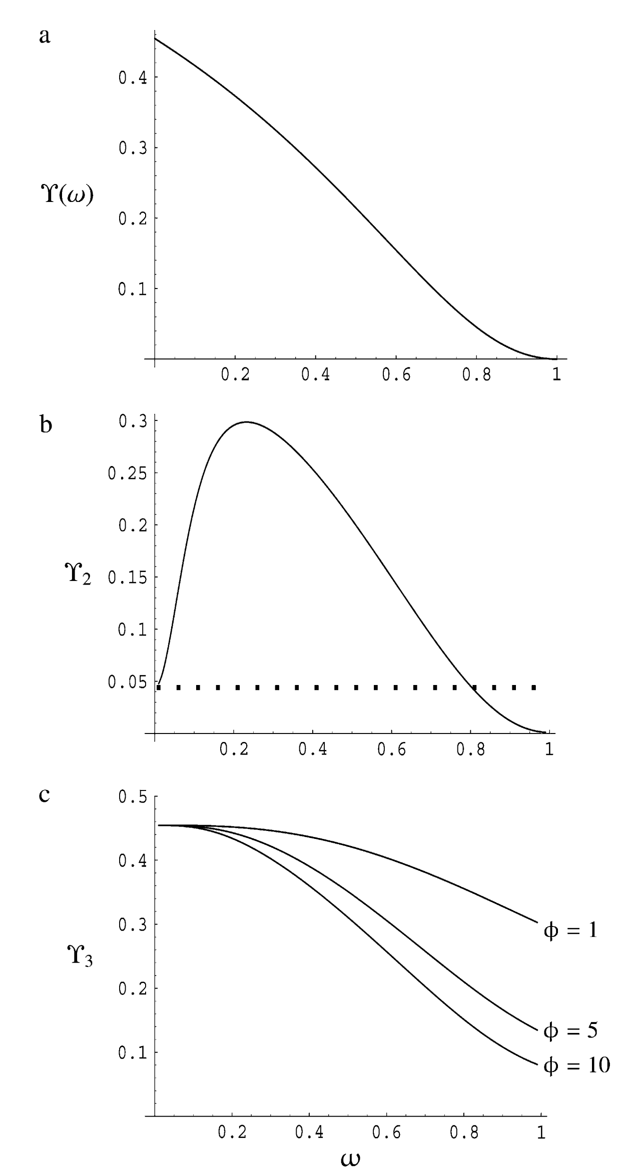

Figure 6.—Predicted levels of linkage disequilibrium (a) $\Upsilon(\omega)$, (b) $\Upsilon_2(\eta, \phi, \omega)$, and (c) $\Upsilon_3(\phi, \omega)$ from Table 2 as a function of ω. In b $\eta = 10$ and $\phi = 1000$. The horizontal dashed line in b represents the predicted level of linkage disequilibrium under the standard Wright–Fisher model ($\Upsilon_4(\eta)$) when $\eta = 10$. For explanation of symbols see text.

disequilibrium under modified Wright–Fisher sampling are generally less than those predicted under standard Wright–Fisher reproduction, given that recombination occurs on the same timescale as coalescence ($0 < \alpha = \beta \leq 1$). When recombination occurs on a shorter timescale than coalescence, Υ is zero. By taking $\alpha = \beta$ when $\beta \leq 1$ the effects of recombination are countered by setting the timescale of modified sampling equal to that of recombination.

Higher than expected LD: For a certain range of the parameter space, higher predicted levels of linkage disequilibrium can also be obtained under modified sampling compared to the standard model. Figure 6b shows $\Upsilon_2(\eta, \phi, \omega)$ as a function of ω with $\phi = 1000$ and $\eta = 10$. The linkage disequilibrium predicted under the standard Wright–Fisher model ($\Upsilon_4(\eta)$) when $\eta = 10$ is shown for reference (Figure 6b, dashed line). For low and high values of ω the level of LD predicted under modified sampling is similar to or less than that predicted under the standard model. For intermediate values of ω, much higher levels of LD are predicted under modified sampling. The interpretation of Figure 6b is that high linkage disequilibrium can result from random sampling in a population with high variance in offspring number. Contrast this with Figure 6c, which shows a graph of $\Upsilon_3(\phi, \omega)$ (Table 2) as a function of ω for three different values of ϕ. In this case of low recombination ($\beta > 1$), Υ_3 is a decreasing function of ω and ϕ. For high values of ω and/or ϕ, the model predicts low linkage disequilibrium even under low recombination. Taken together, a population with highly skewed offspring distribution can have high LD in regions with high recombination and low LD in regions with low recombination.

Another interesting, and cautionary, aspect of our results is that the same prediction of linkage disequilibrium can be obtained with different combinations of parameters. For example, given a standard population model, $\Upsilon_4(\eta) = \frac{1}{3}$ if $\eta = \left(\sqrt{33} - 5\right)/2 \approx 0.372$. However, for any $\beta > \alpha$, as long as $\alpha < 1$, $\Upsilon(\omega) = \frac{1}{3}$ when $\omega \approx 0.2831$ $\left[\text{or } \left(23 - 49/\sqrt[3]{937 - 48\sqrt{330}} - \sqrt[3]{937 - 48\sqrt{330}}\right)/24 \text{ to be exact}\right]$. This implies that it may be difficult to distinguish between η, ϕ, and ω using sequence data.

The distribution of r^2 in simulations: The expected value of r^2 was the focus of the analytical work above. However, Hudson's (1985) analysis of the distribution of r^2 by simulation reveals a high variance of this measure of LD. To obtain quantitative estimates of the variance of r^2 we performed Monte Carlo simulations under a symmetric two-allele mutation model as described by Hudson (1983) with the modification of allowing more than two lines to coalesce at the same time. The program we wrote to perform the simulations correctly predicts the correlations listed in the appendix. A version in C is available upon request.

Table 3 shows results from simulations obtained under two different ancestral processes: one in which the rate of coalescence (x-merger) is given by $\lambda_{n,x} = \phi\binom{n}{x}\omega^x(1-\omega)^{n-x}$ for $2 \leq x \leq n$ and is labeled as "multiple mergers" in Table 3 and the other in which the rate of coalescence is given by Equation 10 and obtained when large reproduction events occur on a timescale of $2N$ generations. The entries in Table 3 are the mean of r^2 and the corresponding standard deviation in parentheses. In nearly all cases the standard deviation is larger than the mean, indicating a high variance in the empirical distribution of r^2. Figure 7 shows the sampling distribution of r^2 under the same two

TABLE 3

Mean and standard deviation (in parentheses) of r^2 for a sample of size 25 (10^4 iterations) under the "multiple-mergers" process ($\eta = 0$) and the mixture-distribution process ($\eta = 1$)

		Ancestral process	
ω	θ	Multiple mergers	Mixture distribution
0.1	1	0.055 (0.073)	0.106 (0.143)
	0.1	0.073 (0.095)	0.107 (0.180)
	0.01	0.149 (0.195)	0.159 (0.238)
			0.230[a] (0.299)
	0.001	0.251 (0.352)	
		0.409[a] (0.373)	
	Predicted	0.416	0.232
0.5	1	0.073 (0.104)	0.162 (0.232)
	0.1	0.119 (0.196)	0.164 (0.269)
	0.01	0.125 (0.275)	0.183 (0.354)
			0.205[a] (0.279)
	0.001	0.176 (0.370)	
		0.212[a] (0.341)	
	Predicted	0.214	0.232

In each case $\phi = 1$. The predicted values are $\Upsilon(\omega)$ for the multiple-mergers process and $\Upsilon_2(1, 1, \omega)$ for the mixture-distribution process (see Table 2).

[a] Low-frequency (<10%) variants excluded.

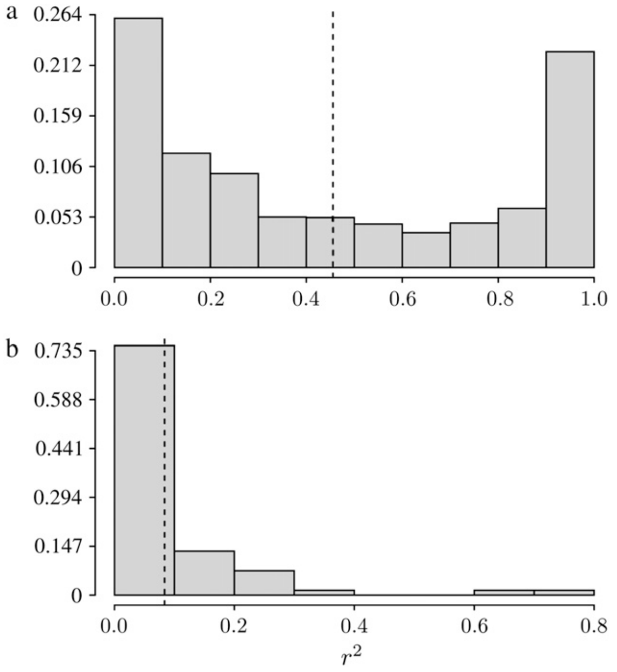

FIGURE 7.—The sampling distribution of r^2 for sample size 25 after 5×10^4 runs under (a) a "multiple-merger" coalescent process ($0 < \alpha < 1$) and no recombination ($\eta = 0$) and (b) a mixture-distribution coalescent process ($\alpha = 1$) with $\eta = 1$. In a, $\theta = 0.001$ and low-frequency variants (<10%) are excluded, while in b $\theta = 0.01$ and all variants are included. In both cases $\omega = 0.1$ and $\phi = 1$. The vertical dashed lines represent the mean $\overline{r^2}$ of each sampling distribution: (a) $\overline{r^2} = 0.452$ with standard deviation 0.377; (b) $\overline{r^2} = 0.084$ with standard deviation 0.132.

coalescent processes as in Table 3. The distributions are similar to those obtained by HUDSON (1985) and reflect the high sampling variance of r^2. Following MCVEAN (2002), our analytical results assume that the population mutation rate θ is small. The analytical predictions for $E(r^2)$ we have derived are in good agreement with simulations for biologically reasonable values of ω (Table 3) as long as θ is small: not $>10^{-2}$ in the case of the mixture-distribution coalescent process or 10^{-3}–10^{-4} if the ancestral process is the multiple-mergers process. For comparison, a value of ω of ~8% was estimated for Pacific oysters (*C. gigas*; ELDON and WAKELEY 2006).

DISCUSSION

Limit predictions (as $N \rightarrow \infty$) about linkage disequilibrium were obtained under a skewed offspring distribution among individuals in a population, in which the offspring of a single (randomly chosen) individual can number on the order of the population size. It is shown that the reproduction parameters ω and ϕ, which control the size and frequency of the large reproduction events, are as important as the recombination rate in predicting levels of LD in a population with highly fecund individuals. Primarily, low LD is predicted due to the star-like shape of the gene genealogy. This can occur, in some cases, despite a low recombination rate and can thus give false evidence for the presence of a recombination hotspot. High LD can also be predicted despite a high recombination rate (see discussion below), *i.e.*, even in the presence of a recombination hotspot. The present results are qualitatively similar to the effects of a recent strong selective sweep on the LD between two neutral loci linked to the selected locus (MCVEAN 2007).

For example, in the model we have described the actual timescale of the ancestral process depends on the parameter α, which determines the frequency of highly fecund individuals (Table 1; the probability of modified Wright–Fisher sampling is $\varepsilon = \phi/N^{\alpha}$). We have studied the effects of a highly skewed offspring distribution on correlations in coalescence times between two loci. The findings show that predictions of linkage disequilibrium are strongly affected by the different timescales on which recombination and random drift operate, as well as the fraction ω of the population replaced by offspring of a single individual. Thus allowing recombination and reproduction to occur on separate, and sometimes very different, timescales uncovers a fundamental way in which LD may be shaped by high variance in offspring number.

For a given α, correlations in coalescence times were shown to be increasing functions of ϕ (recall that the probability of modified Wright–Fisher sampling is $\varepsilon =$

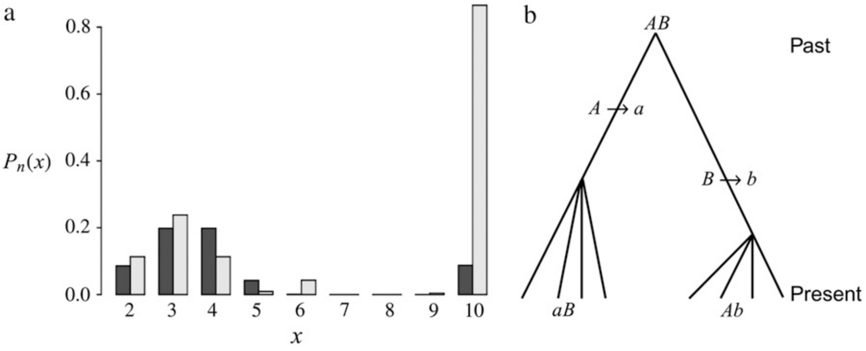

FIGURE 8.—Rates of coalescence and gene genealogy under a multiple-mergers coalescent. (a) The probability distribution of the different types of coalescence events, given by $P_n(x) = \lambda_{n,x}/\sum_x \lambda_{n,x}$ $(2 \leq x \leq n)$ in which $\lambda_{n,x} = \binom{n}{2}\delta_2(x) + \phi\binom{n}{x}\omega^x(1-\omega)^{n-x}$ [$\delta_2(x) = 1$ if $x = 2$ and zero otherwise] in units of $2N$ generations (*i.e.*, $\alpha = \beta = 1$) when the number of ancestral lines $n = 10$ and $\phi = 1000$. Solid bars, $\omega = 0.5$; shaded bars, $\omega = 0.99$. (b) A gene genealogy of a sample of two completely linked loci given a multiple-mergers ancestral process similar to the one described by solid bars in a. One mutation event is assumed at each locus, and the ancestral gametic type is *AB*.

ϕ/N^α) and ω (the fraction of the population replaced by offspring of a single individual). The dependence of the correlations on ϕ and ω can be explained through the effects ϕ and ω have on the shape of the gene genealogy of a sample of two loci on two chromosomes. If the ancestral gametic type (state $\mathscr{C}_4$ in Figure 2) is reached from any of the $\mathscr{S}_i$ sample configurations in Figure 2, with high probability in a large population, the coalescence times would be highly correlated. Reaching the common-ancestral chromosome ($\mathscr{C}_4$) from any of the $\mathscr{S}_i$ states requires only a single multiple merger and is possible in a large population given the population models considered in this report. As ω increases (*i.e.*, tends to 1), the fraction of the population replaced by offspring of a single individual tends to 1. Hence, the gene genealogy assumes a star-like shape composed almost entirely of external branches (similar to the gene genealogy on the right in Figure 1), if ω-coalescence events are frequent enough. The rate of ω-events is determined by ϕ. It follows that as ϕ and ω increase, the more star-like the gene genealogy becomes, and the common-ancestral type $\mathscr{C}_4$ is often reached via a single coalescence event from any given configuration $\mathscr{S}_i$, resulting in high correlation in coalescence times.

Given a highly skewed offspring distribution, predicted levels of LD can range from high to low irrespective of recombination. Consider the case when ω-coalescence events occur on a shorter timescale than the standard coalescent; *i.e.*, $\alpha < 1$. In this case the rate ϕ of ω-coalescence events is much greater than the rate $N^{\alpha-1}$ of the standard coalescent. In a standard Wright–Fisher population predictions of linkage disequilibrium depend only on the recombination rate (η). Common intuition says that when $\eta \gg 1$ there will not be much LD. However, if $0 < \alpha < \beta < 1$, nonzero linkage disequilibrium can occur when $\eta \gg 1$ if the timescale of coalescence is also short due to multiple mergers (Table 2). Given the modified model of reproduction considered in this study, predicted levels of LD depend on ω only when $0 < \alpha < \beta < 1$. For low values of ω the model predicts high LD (*i.e.*, close to $\frac{5}{11}$), and as ω tends to 1 predicted levels of LD descend toward zero. Second, when $\eta \ll 1$ (*i.e.*, $\beta > 1$), predicted levels of LD also only depend on ω in the same way as when $0 < \alpha < \beta < 1$. Thus the model can predict low levels of LD (if $\omega \approx 1$) even when $\eta \ll 1$ and common intuition says that LD should be high. The relation between recombination and linkage disequilibrium may not be as straightforward in organisms with highly fecund individuals as standard theory predicts.

Linkage disequilibrium as predicted by Υ does not change monotonically with ω when recombination and multiple mergers occur on the same timescale of $N^\alpha \leq N$ generations ($0 < \alpha = \beta \leq 1$; see Figure 4). When values of ω are not too close to 1, high LD can result as Figure 6b shows. Again following SLATKIN (1994) a consideration of the gene genealogy of a sample of two completely linked loci can explain the high predicted LD when $0 < \alpha = \beta \leq 1$. The modified Wright–Fisher model considered in the present study allows many lines to reach a common ancestor (coalesce) in the same instance. Thus given n ancestral lines, any number of lines from 2 to n can reach a common ancestor each time a coalescence event occurs. To explain the high predicted LD when $0 < \alpha = \beta \leq 1$ we consider the probability $P_n(x) = \lambda_{n,x}/\sum_x \lambda_{n,x}$ (Figure 8) of each type of coalescence event (x-merger) for $2 \leq x \leq n$ given 10 ancestral lines ($n = 10$). On a timescale of $2N$ generations the rate $\lambda_{n,x}$ of coalescence of x lines of n $(2 \leq x \leq n)$ is given by

$$\lambda_{n,x} = \binom{n}{2}\delta_2(x) + \phi\binom{n}{x}\omega^x(1-\omega)^{n-x} \qquad (10)$$

[in which $\delta_2(x) = 1$ if $x = 2$ and zero otherwise] obtained from Equation 2. Under the standard coalescent only two lines can reach a common ancestor (2-merger) each time a coalescence event occurs. Figure 8 (a, solid bars) shows, however, that when $\phi = 1000$ (in accordance with Figure 6b) and $\omega = 0.5$ the most likely coalescence event is that half the lines reach a common ancestor and that a 2-merger is among the least likely events. Similar patterns are obtained for lower values of ϕ and ω. The high probability of multiple mergers directly affects our prediction of the shape of the gene genealogy of a sample. Figure 8b shows the gene genealogy of a sample of two completely linked loci when $\omega = 0.5$ and most multiple mergers are more likely than 2-mergers (or n-mergers). The gene genealogy consists of short external branches and a long internal branch at each locus. Assuming a single mutation at each locus, both mutations most likely occur in the internal branch, leading to high LD. The probability distribution of multiple mergers for $\phi = 1000$ and $\omega = 0.99$ is shown for reference (Figure 8, shaded bars). When the offspring of a single individual frequently replace almost all the population, the most likely coalescence event is an n-merger, *i.e.*, all ancestral lines reaching a common ancestor. The gene genealogy of a sample then becomes star shaped, similar to the gene genealogy in Figure 1a. Mutations are then most likely to occur in an external branch, resulting in low linkage disequilibrium.

Analytical results presented here rely on the approximations of Ohta and Kimura (1971), who suggested using a ratio of expectations σ_d^2 as a predictor for r^2, and McVean (2002), who assumed a small mutation rate. In addition, the analytical predictions are only for the average value of r^2. These issues were addressed using simulations. The results show that the expected value of r^2 is in agreement with the analytical results when θ is small and the offspring of a single individual replace a modest fraction (ω) of the population ($0 < \omega \leq 0.5$; Table 3). However, the sampling variance of r^2 is quite high (Table 3 and Figure 7), similar to that found for the standard coalescent (Hudson 1985). In the analysis of data, the variance of r^2 can be reduced by averaging values over very many pairs of loci.

We thank the International Centre for Mathematical Sciences for supporting a workshop on Mathematical Population Genetics, March 2006, during which part of this work was done.

LITERATURE CITED

Árnason, E., 2004 Mitochondrial cytochrome *b* variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. Genetics **166:** 1871–1885.

Bataillon, T., T. Mailund, S. Thorlacius, E. Steingrimsson, T. Rafnar *et al.*, 2006 The effective size of the Icelandic population and the prospects for LD mapping: inference from unphased microsatellite markers. Eur. J. Hum. Genet. **14:** 1044–1053.

Crow, J. F., and M. Kimura, 1970 *Introduction to Population Genetics Theory.* Harper & Row, New York.

Durrett, R., 2002 *Probability Models for DNA Sequence Evolution.* Springer, New York.

Eldon, B., and J. Wakeley, 2006 Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics **172:** 2621–2633.

Ewens, W. J., 2004 Theoretical introduction in *Mathematical Population Genetics,* Vol. I, edited by S. S. Antman, J. E. Marsden, L. Sirovich and S. Wiggins. Springer, New York.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

Garđarsdóttir, O., and B. Sigurjónsson (Editors), 2006 *Population* (Statistical Series, Vol. 91). Statistics Iceland, Reykjavík, Iceland (in Icelandic).

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 Proceedings on the Symposium on Applied Probability* (Monograph Series, Vol. 18), edited by I. V. Basawa and R. L. Taylor. IMS Lecture Notes, Institute of Mathematical Statistics, Hayward, CA.

Hedgecock, D., 1994 Does variance in reproductive success limit effective population sizes of marine organisms?, pp. 1222–1344 in *Genetics and Evolution of Aquatic Organisms,* edited by A. Beaumont. Chapman & Hall, London.

Hedrick, P. W., 2000 *Genetics of Populations,* Ed. 2. Jones & Bartlett, Sudbury, MA.

Hedrick, P. W., 2005 Large variance in reproductive success and the N_e/N ratio. Evolution **59:** 1596–1599.

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 the neutral allele model with intergenic recombination. Theor. Popul. Biol. **23:** 183–201.

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

Jónsson, G., and M. S. Magnússon (Editors), 1997 *Hagskinna: Icelandic Historical Statistics.* Statistics Iceland, Reykjavík, Iceland (in Icelandic).

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

Kingman, J. F. C., 1982a The coalescent. Stoch. Proc. Appl. **13:** 235–248.

Kingman, J. F. C., 1982b On the genealogy of large populations. J. Appl. Probab. **19A:** 27–43.

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

Lewontin, R. C., and K. Kojima, 1960 The evolutionary dynamics of complex polymorphisms. Evolution **14:** 450–472.

McVean, G., 2007 The structure of linkage disequilibrium around a selective sweep. Genetics **175:** 1395–1406.

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

Möhle, M., 2006 On the number of segregating sites for populations with large family sizes. Adv. Appl. Probab. **38:** 750–767.

Möhle, M., and S. Sagitov, 2001 A classification of coalescent processes for haploid exchangeable population models. Ann. Appl. Probab. **29:** 1547–1562.

Moran, P. A. P., 1958 Random processes in genetics. Proc. Camb. Philos. Soc. **54:** 60–71.

Moran, P. A. P., 1962 *Statistical Processes of Evolutionary Theory.* Clarendon Press, Oxford.

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.

Pitman, J., 1999 Coalescents with multiple collisions. Ann. Probab. **27:** 1870–1902.

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

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

Sagitov, S., 1999 The general coalescent with asynchronous mergers of ancestral lines. J. Appl. Probab. **36:** 1116–1125.

Schweinsberg, J., 2000 Coalescents with simultaneous multiple collisions. Electron J. Probab. **5:** 1–50.

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

Song, Y. S., and J. S. Song, 2007 Analytic computation of the expectation of the linkage disequilibrium coefficient r^2. Theor. Popul. Biol. **71:** 49–60.

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

Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

THORARINSSON, S., 1961 Population changes in Iceland. Geogr. Rev. **51:** 519–533.

THORSTEINSSON, B., and B. JÓNSSON, 1991 *Íslands Saga.* Sögufélag, Reykjavík, Iceland (in Icelandic).

TURNER, T. F., J. P. WARES and J. R. GOLD, 2002 Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). Genetics **162:** 1329–1339.

WEIR, B. S., and W. G. HILL, 1986 Nonuniform recombination within the human β-globin gene cluster. Am. J. Hum. Genet. **38:** 776–778.

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

Communicating editor: M. W. FELDMAN

APPENDIX

Transition probabilities: Here the transition probabilities from the noncoalescent states $\mathscr{S}_0$, $\mathscr{S}_1$, and $\mathscr{S}_2$ (Figure 3) in the ancestral process are stated. These are in discrete time. The states are based on two biallelic loci in a sample of size two from diploid individuals. Only one crossover is allowed between the loci when a recombination event occurs. A transition probability from state i to state j is denoted $P(i \rightarrow j)$. And of course $\sum_j P(i \rightarrow j) = 1$. As an example, consider $P(\mathscr{S}_0 \rightarrow \mathscr{S}_0)$ or the probability that all four alleles stay on separate chromosomes over one generation. Standard Wright–Fisher sampling occurs with probability $1 - \varepsilon$ and modified sampling with probability ε. The two separate probabilities can be obtained by adopting a balls-in-boxes approach. Going one generation back in time, the four balls (alleles) occupy $2N$ boxes (chromosomes) at random. To stay in state $\mathscr{S}_0$, no recombination is involved, and all the balls must occupy different boxes. Under standard sampling, this happens with probability $(1 - 1/(2N))(1 - 2/(2N))(1 - 3/(2N))$. Under modified sampling, a single randomly chosen box is occupied by a ball with probability ω (the ω-box), while the other $2N - 1$ boxes are each occupied by a ball with probability $1/(2N - 1)$. Under modified sampling, the four balls stay in separate boxes in two ways. First, they can all ignore the ω-box with probability $(1 - \omega)^4$ and then must all occupy different boxes with probability $(1 - 1/(2N - 1))(1 - 2/(2N - 1))(1 - 3/(2N - 1))$. Second, a single randomly chosen ball occupies the ω-box with the binomial probability $4\omega(1 - \omega)^3$, while, of the remaining three balls, each occupies a single box with probability $(1 - 1/(2N - 1))(1 - 2/(2N - 1))$. Taken together,

$$\begin{aligned} P(\mathscr{S}_0 \rightarrow \mathscr{S}_0) = {} & \left(1 - \frac{3}{2N}\right)\left(1 - \frac{1}{N}\right)\left(1 - \frac{1}{2N}\right)(1 - \varepsilon) \\ & + \varepsilon\left(1 - \frac{3}{2N - 1}\right)\left(1 - \frac{2}{2N - 1}\right)\left(1 - \frac{1}{2N - 1}\right)(1 - \omega)^4 \\ & + 4\varepsilon\left(1 - \frac{2}{2N - 1}\right)\left(1 - \frac{1}{2N - 1}\right)\omega(1 - \omega)^3. \end{aligned}$$

The other transition probabilities are obtained similarly:

$$\begin{aligned} P(\mathscr{S}_0 \rightarrow \mathscr{S}_1) = {} & \frac{2(1 - 1/N)(1 - 1/(2N))(1 - \varepsilon)}{N} \\ & + \varepsilon\left(\frac{4(1 - 2/(2N - 1))(1 - 1/(2N - 1))(1 - \omega)^4}{2N - 1} + \frac{8(1 - 1/(2N - 1))\omega(1 - \omega)^3}{2N - 1}\right) \\ & + 4\varepsilon\left(1 - \frac{1}{2N - 1}\right)\omega^2(1 - \omega)^2 \end{aligned}$$

$$\begin{aligned} P(\mathscr{S}_0 \rightarrow \mathscr{S}_2) = {} & \frac{(1 - 1/(2N))(1 - \varepsilon)}{2N^2} \\ & + \varepsilon\left(\frac{2(1 - 1/((2N) - 1))(1 - \omega)^4}{(2N - 1)^2} + \frac{4\omega^2(1 - \omega)^2}{2N - 1}\right) \end{aligned}$$

$$\begin{aligned} P(\mathscr{S}_0 \rightarrow \mathscr{C}_1) = {} & \frac{(1 - 1/N)(1 - 1/2N)(1 - \varepsilon)}{N} \\ & + \varepsilon\left(\frac{2(1 - 2/(2N - 1))(1 - 1/(2N - 1))(1 - \omega)^4}{2N - 1} + \frac{4(1 - 1/(2N - 1))\omega(1 - \omega)^3}{2N - 1}\right) \\ & + 2\varepsilon\left(1 - \frac{1}{2N - 1}\right)\omega^2(1 - \omega)^2 \end{aligned}$$

$$\begin{aligned}P(\mathscr{S}_0\to\mathscr{C}_2) &= \frac{(1-1/(2N))(1-\varepsilon)}{N^2}\\ &\quad+\varepsilon\left(\frac{4(1-1/(2N-1))(1-\omega)^4}{(2N-1)^2}+\frac{4\omega(1-\omega)^3}{(2N-1)^2}+4\omega^3(1-\omega)\right)\end{aligned}$$

$$P(\mathscr{S}_0\to\mathscr{C}_3) = (1-\varepsilon)\frac{1-1/(2N)}{4N^2}+\varepsilon\left(\frac{4(1-1/(2N-1))(1-\omega)^4}{(2N-1)^2}+\frac{4\omega(1-\omega)^3}{(2N-1)^2}+4\omega^3(1-\omega)\right)$$

$$P(\mathscr{S}_0\to\mathscr{C}_4) = (1-\varepsilon)\left(\frac{1}{2N}\right)^3+\varepsilon\left(\frac{(1-\omega)^4}{(2N-1)^3}+\omega^4\right)$$

$$\begin{aligned}P(\mathscr{S}_1\to\mathscr{S}_0) &= (1-\varepsilon)\left(1-\frac{3}{2N}\right)\left(1-\frac{1}{N}\right)\left(1-\frac{1}{2N}\right)c\\ &\quad+\varepsilon c\left(1-\frac{3}{2N-1}\right)\left(1-\frac{2}{2N-1}\right)\left(1-\frac{1}{2N-1}\right)(1-\omega)^4\\ &\quad+4\varepsilon c\left(1-\frac{2}{2N-1}\right)\left(1-\frac{1}{2N-1}\right)\omega(1-\omega)^3\end{aligned}$$

$$\begin{aligned}P(\mathscr{S}_1\to\mathscr{S}_1) &= (1-\varepsilon)\left(\left(1-\frac{1}{N}\right)\left(1-\frac{1}{2N}\right)(1-c)+\frac{2(1-1/N)(1-1/(2N))c}{N}\right)\\ &\quad+\varepsilon\left(\left(1-\frac{2}{2N-1}\right)\left(1-\frac{1}{2N-1}\right)(1-c)(1-\omega)^3+3\left(1-\frac{1}{2N-1}\right)(1-c)\omega(1-\omega)^2\right)\\ &\quad+\varepsilon c\left(\frac{4(1-2/(2N-1))(1-1/(2N-1))(1-\omega)^4}{2N-1}+\frac{8(1-1/(2N-1))\omega(1-\omega)^3}{2N-1}\right)\\ &\quad+4c\varepsilon\left(1-\frac{1}{2N-1}\right)\omega^2(1-\omega)^2\end{aligned}$$

$$\begin{aligned}P(\mathscr{S}_1\to\mathscr{S}_2) &= (1-\varepsilon)\left(\frac{(1-1/(2N))(1-c)}{2N}+\frac{(1-1/(2N))c}{2N^2}\right)\\ &\quad+\varepsilon(1-c)\left(\frac{(1-1/(2N-1))(1-\omega)^3}{2N-1}+\frac{\omega(1-\omega)^2}{2N-1}+\omega^2(1-\omega)\right)\\ &\quad+\varepsilon c\left(\frac{2(1-1/(2N-1))(1-\omega)^4}{(2N-1)^2}+\frac{4\omega^2(1-\omega)^2}{2N-1}\right)\end{aligned}$$

$$\begin{aligned}P(\mathscr{S}_1\to\mathscr{C}_1) &= \frac{(1-\varepsilon)(1-1/N)(1-1/(2N))c}{N}\\ &\quad+\varepsilon c\left(\frac{2(1-2/(2N-1))(1-1/(2N-1))(1-\omega)^4}{2N-1}+\frac{4(1-1/(2N-1))\omega(1-\omega)^3}{2N-1}\right)\\ &\quad+2\varepsilon c\left(1-\frac{1}{2N-1}\right)\omega^2(1-\omega)^2\end{aligned}$$

$$\begin{aligned}P(\mathscr{S}_1\to\mathscr{C}_2) &= (1-\varepsilon)\left(\frac{(1-1/(2N))(1-c)}{N}+\frac{(1-1/(2N))c}{N^2}\right)\\ &\quad+2\varepsilon(1-c)\left(\frac{(1-1/(2N-1))(1-\omega)^3}{2N-1}+\frac{\omega(1-\omega)^2}{2N-1}+\omega^2(1-\omega)\right)\\ &\quad+\varepsilon c\left(\frac{4(1-1/(2N-1))(1-\omega)^4}{(2N-1)^2}+\frac{4\omega(1-\omega)^3}{(2N-1)^2}+4\omega^3(1-\omega)\right)\end{aligned}$$

$$\begin{aligned} P(\mathscr{S}_1 \rightarrow \mathscr{C}_3) = {} & (1-\varepsilon)\left(\frac{(1-1/(2N))(1-c)}{N} + \frac{(1-1/(2N))c}{N^2}\right) \\ & + \varepsilon 2(1-c)\left(\frac{(1-1/(2N-1))(1-\omega)^3}{2N-1} + \frac{\omega(1-\omega)^2}{2N-1} + \omega^2(1-\omega)\right) \end{aligned}$$

$$P(\mathscr{S}_1 \rightarrow \mathscr{C}_4) = (1-\varepsilon)\left(\frac{1-c}{4N^2} + \frac{c}{8N^3}\right) + \varepsilon(1-c)\left(\frac{(1-\omega)^3}{(2N-1)^2} + \omega^3\right)$$

$$\begin{aligned} P(\mathscr{S}_2 \rightarrow \mathscr{S}_0) = {} & (1-\varepsilon)\left(1-\frac{3}{2N}\right)\left(1-\frac{1}{N}\right)\left(1-\frac{1}{2N}\right)c^2 \\ & + \varepsilon c^2\left(1-\frac{3}{2N-1}\right)\left(1-\frac{2}{2N-1}\right)\left(1-\frac{1}{2N-1}\right)(1-\omega)^4 \\ & + 4\varepsilon c^2\left(1-\frac{2}{2N-1}\right)\left(1-\frac{1}{2N-1}\right)\omega(1-\omega)^3 \end{aligned}$$

$$\begin{aligned} P(\mathscr{S}_2 \rightarrow \mathscr{S}_1) = {} & (1-\varepsilon)\left(\frac{2(1-1/N)(1-1/(2N))c^2}{N} + 2\left(1-\frac{1}{N}\right)\left(1-\frac{1}{2N}\right)(1-c)c\right) \\ & + \varepsilon\left(\frac{4(1-2/(2N-1))(1-1/(2N-1))(1-\omega)^4}{2N-1} + \frac{8(1-1/(2N-1))\omega(1-\omega)^3}{2N-1}\right. \\ & \qquad \left. + 4\left(1-\frac{1}{2N-1}\right)\omega^2(1-\omega)^2\right)c^2 \\ & + 2\varepsilon(1-c)\left(\left(1-\frac{2}{2N-1}\right)\left(1-\frac{1}{2N-1}\right)(1-\omega)^3 + 3\left(1-\frac{1}{2N-1}\right)\omega(1-\omega)^2\right)c \end{aligned}$$

$$\begin{aligned} P(\mathscr{S}_2 \rightarrow \mathscr{S}_2) = {} & (1-\varepsilon)\left(\left(1-\frac{1}{2N}\right)(1-c)^2 + \frac{(1-1/(2N))c(1-c)}{N} + \frac{(1-1/(2N))c^2}{2N^2}\right) \\ & + \varepsilon\left(\frac{2(1-1/(2N-1))(1-\omega)^4}{(2N-1)^2} + \frac{4\omega^2(1-\omega)^2}{2N-1}\right)c^2 \\ & + \varepsilon\left(\left(1-\frac{1}{2N-1}\right)(1-\omega)^2 + 2\omega(1-\omega)\right)(1-c)^2 \\ & + 2\varepsilon c(1-c)\left(\frac{(1-1/(2N-1))(1-\omega)^3}{2N-1} + \frac{\omega(1-\omega)^2}{2N-1} + \omega^2(1-\omega)\right) \end{aligned}$$

$$\begin{aligned} P(\mathscr{S}_2 \rightarrow \mathscr{C}_1) = {} & \frac{(1-\varepsilon)(1-1/N)(1-1/(2N))c^2}{N} \\ & + \varepsilon c^2\left(\frac{2(1-2/(2N-1))(1-1/(2N-1))(1-\omega)^4}{2N-1} + \frac{4(1-1/(2N-1))\omega(1-\omega)^3}{2N-1}\right) \\ & + 2\varepsilon c^2\left(1-\frac{1}{2N-1}\right)\omega^2(1-\omega)^2 \end{aligned}$$

$$\begin{aligned} P(\mathscr{S}_2 \rightarrow \mathscr{C}_2) = {} & (1-\varepsilon)\left(\frac{(1-1/(2N))c^2}{N^2} + \frac{2(1-1/(2N))(1-c)c}{N}\right) \\ & + \varepsilon\left(\frac{4(1-1/(2N-1))(1-\omega)^4}{(2N-1)^2} + \frac{4\omega(1-\omega)^3}{(2N-1)^2} + 4\omega^3(1-\omega)\right)c^2 \\ & + 4\varepsilon c(1-c)\left(\frac{(1-1/(2N-1))(1-\omega)^3}{2N-1} + \frac{\omega(1-\omega)^2}{2N-1} + \omega^2(1-\omega)\right) \end{aligned}$$

$$
\begin{aligned}
P(\mathscr{S}_2 \rightarrow \mathscr{C}_3) = (1-\varepsilon)\left(\frac{(1-1/(2N))c^2}{N^2} + \frac{2(1-1/(2N))(1-c)c}{N}\right) \\
+ \varepsilon\left(\frac{4(1-1/(2N-1))(1-\omega)^4}{(2N-1)^2} + \frac{4\omega(1-\omega)^3}{(2N-1)^2} + 4\omega^3(1-\omega)\right)c^2 \\
+ 4\varepsilon(1-c)\left(\frac{(1-1/(2N-1))(1-\omega)^3}{2N-1} + \frac{\omega(1-\omega)^2}{2N-1} + \omega^2(1-\omega)\right)c
\end{aligned}
$$

$$
\begin{aligned}
P(\mathscr{S}_2 \rightarrow \mathscr{C}_4) = (1-\varepsilon)\left(\frac{(1-c)^2}{2N} + \frac{c(1-c)}{2N^2} + \frac{c^2}{8N^3}\right) \\
+ \varepsilon\left(\left(\frac{(1-\omega)^2}{2N-1} + \omega^2\right)(1-c)^2 + 2c\left(\frac{(1-\omega)^3}{(2N-1)^2} + \omega^3\right)(1-c) + c^2\left(\frac{(1-\omega)^4}{(2N-1)^3} + \omega^4\right)\right).
\end{aligned}
$$

Correlations in coalescent times: Here the correlations in coalescent times between the two loci are specified for the modified Wright–Fisher population model. These are the correlations given that the ancestral process starts from one of the three noncoalescent states $\mathscr{S}_0$, $\mathscr{S}_1$,or $\mathscr{S}_2$ (Figure 3) and are obtained as a limit process in a large population (*i.e.*, as $N \rightarrow \infty$). In what follows, let $\sigma_0 = \mathrm{corr}(t_1, t_2 \mid \mathscr{S}_0)$, $\sigma_1 = \mathrm{corr}(t_1, t_2 \mid \mathscr{S}_1)$, and $\sigma_2 = \mathrm{corr}(t_1, t_2 \mid \mathscr{S}_2)$. The different cases depend on the values of the recombination-timescale parameter β (recombination is scaled as $\eta = cN^{\beta}$), and the reproduction-timescale parameter α (the probability of modified Wright–Fisher sampling is $\varepsilon = \phi/N^{\alpha}$), and are of course the same as those given for Υ. Note that the correlation terms can be used to obtain Υ using Equation 11 in McVean (2002). Replacing η with $\eta/2$ and ϕ with $\phi/2$ gives the formulas for Υ in Table 2. The factor of 2 comes from the timescale of $2N$ (in the case $\alpha \geq 1$) used in obtaining the correlation terms. Predictions of the models concerning levels of linkage disequilibrium are independent of the different scaling of the parameters.

Only if $0 < \alpha = \beta \leq 1$ the correlation terms are functions of all three parameters η, ϕ, and ω. If $0 < \alpha = \beta < 1$ the correlation terms are

$$
\sigma_0 = \frac{\omega^2(-2\eta^2 + \phi\omega(-6\omega^2 + 11\omega - 8)\eta + \phi^2\omega^2(2\omega^3 - 7\omega^2 + 8\omega - 4))}{\phi^2(6\omega^3 - 25\omega^2 + 36\omega - 18)\omega^4 + \phi\eta(6\omega^3 - 27\omega^2 + 44\omega - 26)\omega^2 + 2\eta^2(\omega^2 - 2)}
$$

$$
\sigma_1 = -\frac{\omega^2(2\eta^2 + 3\phi\omega(2\omega^2 - 5\omega + 4)\eta + \phi^2\omega^2(3\omega^2 - 8\omega + 6))}{\phi^2(6\omega^3 - 25\omega^2 + 36\omega - 18)\omega^4 + \phi\eta(6\omega^3 - 27\omega^2 + 44\omega - 26)\omega^2 + 2\eta^2(\omega^2 - 2)}
$$

$$
\sigma_2 = \frac{\omega^2(-2\eta^2 - \phi(6\omega^3 - 17\omega^2 + 12\omega + 2)\eta + \phi^2\omega^2(6\omega^3 - 25\omega^2 + 36\omega - 18))}{\phi^2(6\omega^3 - 25\omega^2 + 36\omega - 18)\omega^4 + \phi\eta(6\omega^3 - 27\omega^2 + 44\omega - 26)\omega^2 + 2\eta^2(\omega^2 - 2)}.
$$

If $\alpha = \beta = 1$ the correlation terms are $\sigma_0 = a_0/d - 1$, $\sigma_1 = a_1/d - 1$, and $\sigma_2 = a_2/d - 1$ in which

$$
\begin{aligned}
a_0 = (2\phi\omega^2 + 1)(4\phi^2(11 - 22\omega + 16\omega^2 - 4\omega^3)\omega^4 + 4\phi(4\omega^2 - 11\omega + 11)\omega^2 + 8\eta^2 \\
+ 2\eta(2\phi(8\omega^2 - 18\omega + 13)\omega^2 + 13) + 11),
\end{aligned}
$$

$$
\begin{aligned}
a_1 = 4\phi^3(3\omega^2 - 8\omega + 6)\omega^6 + 4\phi^2(6\eta\omega^3 + (3 - 15\eta)\omega^2 + 4(3\eta - 2)\omega + 9)\omega^4 \\
+ \phi((8\eta^2 + 2\eta + 3)\omega^2 + 8(3\eta - 1)\omega + 18)\omega^2 + 3,
\end{aligned}
$$

$$
\begin{aligned}
a_2 = 2(2\phi\omega^2 + 1)(2\phi^2(18 - 36\omega + 25\omega^2 - 6\omega^3)\omega^4 + 9\phi(2 - \omega)^2\omega^2 + 4\eta^2 \\
+ 2\eta(\phi(5\omega^2 - 16\omega + 14)\omega^2 + 7) + 9),
\end{aligned}
$$

$$
\begin{aligned}
d = 4\phi^3(18 - 36\omega + 25\omega^2 - 6\omega^3)\omega^6 + 4\phi^2(27 - 36\omega + 17\omega^2 - 3\omega^3)\omega^4 + 9\phi(\omega^2 - 4\omega + 6)\omega^2 \\
+ 8\eta^2(\phi\omega^2(2 - \omega^2) + 1) + 2\eta(2\phi^2(26 - 44\omega + 27\omega^2 - 6\omega^3)\omega^4 + \phi(52 - 44\omega + 11\omega^2)\omega^2 + 13) + 9.
\end{aligned}
$$

If $0 < \alpha$, $\beta < 1$, and $\beta < \alpha$, the three correlation terms are all equal to $\omega^2/(2 - \omega^2)$. If $0 < \alpha$, $\beta < 1$, and $\alpha < \beta$, the correlation terms are

$$\sigma_0 = \frac{4 - 8\omega + 7\omega^2 - 2\omega^3}{18 - 36\omega + 25\omega^2 - 6\omega^3},$$
$$\sigma_1 = \frac{1}{3 - 2\omega},$$

and $\sigma_2 = 1$. This is always the case when $\alpha < \beta$ or when both parameters are >1. In the case $\alpha = 1$ and $\beta > 1$,

$$\sigma_0 = \frac{2\phi^2(2\omega^3 - 7\omega^2 + 8\omega - 4)\omega^4 + \phi(-7\omega^2 + 8\omega - 8)\omega^2 - 2}{2\phi^2(6\omega^3 - 25\omega^2 + 36\omega - 18)\omega^4 - 9\phi(\omega - 2)^2\omega^2 - 9}$$
$$\sigma_1 = \frac{2\phi\omega^2 + 1}{2\phi(3 - 2\omega)\omega^2 + 3}.$$

If $\alpha > 1$ and $0 < \beta < 1$, all three correlation terms $= 0$.

If $\alpha > 1$ and $\beta = 1$, the results correspond to those obtained under standard Wright–Fisher sampling:

$$\sigma_0 = \frac{2}{9 + 26\eta + 8\eta^2}$$
$$\sigma_1 = \frac{3}{9 + 26\eta + 8\eta^2}$$
$$\sigma_2 = \frac{9 + 2\eta}{9 + 26\eta + 8\eta^2}. \tag{A1}$$

If $\alpha > 1$ and $\beta > 1$, $\sigma_0 = \frac{2}{9}$, $\sigma_1 = \frac{1}{3}$, and $\sigma_2 = 1$.