

The Total Branch Length of Sample Genealogies in Populations of Variable Size

A. Eriksson,^{*,†} B. Mehlig,^{*,1} M. Rafajlovic^{*} and S. Sagitov[†]

^{*}Department of Physics, University of Gothenburg, SE-41296 Gothenburg, Sweden, [†]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom and [‡]Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden

Manuscript received March 29, 2010
Accepted for publication June 17, 2010

ABSTRACT

We consider neutral evolution of a large population subject to changes in its population size. For a population with a time-variable carrying capacity we study the distribution of the total branch lengths of its sample genealogies. Within the coalescent approximation we have obtained a general expression—Equation 20—for the moments of this distribution with a given arbitrary dependence of the population size on time. We investigate how the frequency of population-size variations alters the total branch length.

MODELS for gene genealogies of biological populations often assume a constant, time-independent population size N . This is the case for the Wright–Fisher model (FISHER 1930; WRIGHT 1931), for the Moran model (MORAN 1958), and for their representation in terms of the coalescent (KINGMAN 1982). In real biological populations, by contrast, the population size changes over time. Such fluctuations may be due to catastrophic events (bottlenecks) and subsequent population expansions or just reflect the randomness in the factors determining the population dynamics. Many authors have argued that genetic variation in a population subject to size fluctuations may nevertheless be described by the Wright–Fisher model, if one replaces the constant population size in this model by an effective population size of the form

$$N_{\text{eff}} = \left(\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=0}^{L-1} \frac{1}{N_l} \right)^{-1}, \quad (1)$$

where N_l stands for the population size in generation l . The harmonic average in Equation 1 is argued to capture the significant effect of catastrophic events on patterns of genetic variation in a population: if, for example, a population went through a recent bottleneck, a large fraction of individuals in a given sample would originate from few parents. This in turn would lead to significantly reduced genetic variation, parameterized by a small value of N_{eff} . (See, *e.g.*, EWENS 1982 for a review of different measures of the effective population size and SJÖDIN *et al.* 2005 and WAKELEY

and SARGSYAN 2009 for recent developments of this concept.)

The concept of an effective population size has been frequently used in the literature, implicitly assuming that the distribution of neutral mutations in a large population of fluctuating size is identical to the distribution in a Wright–Fisher model with the corresponding constant effective population size given by Equation 1. However, recently it was shown that this is true only under certain circumstances (KAJ and KRONE 2003; NORDBORG and KRONE 2003; JAGERS and SAGITOV 2004). It is argued by SJÖDIN *et al.* (2005) that the concept of an effective population size is appropriate when the timescale of fluctuations of N_l is either much smaller or much larger than the typical time between coalescent events in the sample genealogy. In these limits it can be proved that the distribution of the sample genealogies is exactly given by that of the coalescent with a constant, effective population size.

More importantly, it follows from these results that, in populations with variable size, the coalescent with a constant effective population size is not always a valid approximation for the sample genealogies. Deviations between the predictions of the standard coalescent model and empirical data are frequently observed, and there are a number of different statistical tests quantifying the corresponding discrepancies (see, for example, TAJIMA 1989, FU and LI 1993, and ZENG *et al.* 2006). The analysis of such deviations is of crucial importance in understanding, for example, human genetic history (GARRIGAN and HAMMER 2006). But while there is a substantial amount of work numerically quantifying deviations, often in terms of a single number, little is known about their qualitative origins and their effect upon summary statistics in the population in question.

¹Corresponding author: Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden.
E-mail: mehlig@fy.chalmers.se

The question is thus to understand the effect of population-size fluctuations on the patterns of genetic variation, in particular for the case where the scale of the population-size fluctuations is comparable to the time between coalescent events in the ancestral tree. As is well known, many empirical measures of genetic variation can be computed from the total branch length of the sample genealogy (the expected number of single-nucleotide polymorphisms, for example, is proportional to the average total branch length).

The aim of this article is to analyze the distribution of the scaled total branch length T_n for a sample genealogy in a population of fluctuating size, as illustrated in Figure 1. For the genealogy of $n \geq 2$ lineages sampled at the present time, the expression $\lfloor NT_n \rfloor$ gives the total branch length in terms of generations. Here $\lfloor Nt \rfloor$ is the largest integer $\leq Nt$, and the scaling factor N is a suitable measure of the number of genes in the population and serves as a counterpart of the constant generation size of the standard Wright–Fisher model.

A motivating example is given in Figure 2, which shows numerically computed distributions $\rho(T_n)$ of the total branch lengths T_n for a particular population model with a time-dependent carrying capacity. The model is described briefly in the Figure 2 legend and in detail in A MODEL FOR A POPULATION WITH TIME-DEPENDENT CARRYING CAPACITY. As Figure 2 shows, the distributions depend in a complex manner on the form of the size changes. We observe that when the frequency of the population-size fluctuations is very small (Figure 2a), the distribution is well described by the standard coalescent result

$$P(T_n \leq t) = (1 - e^{-t/2})^{n-1}, \quad t \geq 0 \quad (2)$$

(HEIN *et al.* 2005). When the frequency is very large (Figure 2e), Equation 2 also applies, but with a different time scaling reflecting an effective population size: t on the right-hand side (rhs) in Equation 2 is replaced by t/c with $c = N/N_{\text{eff}}$. Apart from these special limits, however, the form of the distributions appears to depend in a complicated manner upon the frequency of the population-size variation. The observed behavior is caused by the fact that coalescence proceeds faster for smaller population sizes and more slowly for larger population sizes, as illustrated in Figure 1. But the question is how to quantitatively account for the changes shown in Figure 2.

We show in this article that the results of the simulations displayed in Figure 2 are explained by a general expression—Equation 20—for the moments of the distributions shown in Figure 2. Our general result is obtained within the coalescent approximation valid in the limit of large population size. But we find that in most cases, the coalescent approximation works very well down to small population sizes (a few hundred individuals). Our result enables us to understand and quantitatively describe how the distributions shown in

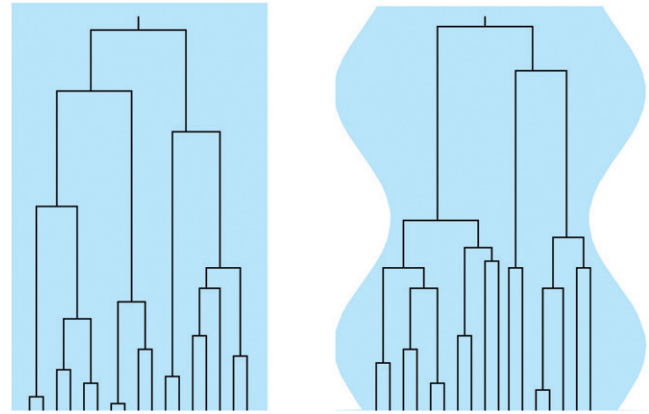


FIGURE 1.—The effect of population-size oscillations on the genealogy of a sample of size $n = 17$ (schematic). Left, genealogy described by Kingman’s coalescent for a large population of constant size, illustrated by the light blue rectangle; right, sinusoidally varying population size. Coalescence is accelerated in regions of small population sizes and vice versa. This significantly alters the tree and gives rise to changes in the distribution of the number of mutations and of the population homozygosity.

Figure 2 depend upon the frequency of the population-size oscillations. It makes possible to determine, for example, how the variance, skewness, and the kurtosis of these distributions depend upon the frequency of demographic fluctuations. This in turn allows us to compute the population homozygosity and to characterize genetic variation in populations with size fluctuations.

The remainder of this article is organized as follows. The next section summarizes our analytical results for the moments of the total branch length. Following that, we describe the model employed in the computer simulations. Then, corresponding numerical results are compared to the analytical predictions. And finally, we summarize how population-size fluctuations influence the distribution of total branch lengths and conclude with an outlook.

COALESCENT APPROXIMATION FORMULAS FOR THE MOMENTS $\langle T_n^k \rangle$

For the purpose of coalescent approximation it is convenient to introduce a “scaled time” t and a “scaled population size” $x(t)$ by writing

$$N_l = Nx \left(\frac{l_s - l}{N} \right) \equiv Nx(t). \quad (3)$$

Here N is a suitable counterpart of the constant generation size of the standard Wright–Fisher model assumed to be large. The population is sampled in generation l_s corresponding to $t = 0$, and the time t is now counted backward in units of N generations, as is common in the coalescent picture. Note that Equation 1 translates into

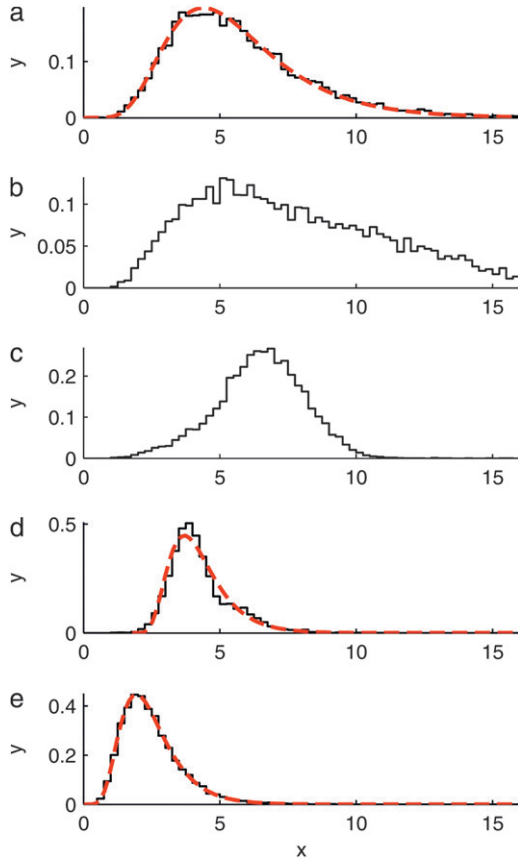


FIGURE 2.—Numerically computed distributions $\rho(T_n)$ of the scaled total branch lengths T_n in genealogies of samples of size $n = 10$. The model employed in the simulations is outlined in A MODEL FOR A POPULATION WITH TIME-DEPENDENT CARRYING CAPACITY. It describes a population subject to a time-varying carrying capacity, $K_t = K_0(1 + \varepsilon \sin(2\pi\nu t))$. The frequency of the time changes is determined by ν , and $l = 1, 2, 3, \dots$ labels discrete generations forward in time. The parameter $N = K_0$ describes the typical population size, which is taken here to be equal to the time-averaged carrying capacity. a–e show $\rho(T_n)$ for populations with increasingly rapidly oscillating carrying capacity. The dashed red line in a shows that in the limit of low frequencies the standard coalescent result, Equation 2, is obtained. The dashed red line in e shows that also in the limit of large frequencies the standard coalescent result is obtained, but now with an effective population size. The dashed red line in d is a two-parameter distribution, Equation 41, derived in COMPARISON BETWEEN NUMERICAL SIMULATIONS AND COALESCENT PREDICTIONS. Further numerical and analytical results on the frequency dependence of the moments of these distributions are shown in Figure 4. Parameter values used: $K_0 = 10,000$, $\varepsilon = 0.9$, and $r = 1$ (see A MODEL FOR A POPULATION WITH TIME-DEPENDENT CARRYING CAPACITY for the exact meaning of the intrinsic growth rate r) and (a) $\nu N = 0.001$, (b) $\nu N = 0.1$, (c) $\nu N = 0.316$, (d) $\nu N = 1$, and (e) $\nu N = 100$.

$$N_{\text{eff}} = N \left(\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \frac{dt}{x(t)} \right)^{-1}. \quad (4)$$

In this section we show how to calculate the moments $\langle T_n^k \rangle$ for the total (scaled) branch length T_n for a given

realization of the curve $x(t)$, making use of results obtained by TAVARÉ (1984).

The starting point is the obvious expression for the total time:

$$T_n = \sum_{j=2}^n j\tau_j. \quad (5)$$

Here τ_j denotes the time during which the genealogy has j ancestral lines. For the population with variable size the times τ_n, \dots, τ_2 all depend upon the sample size n ; however, this dependence is not made explicit, either here or in the following. As shown by GRIFFITHS and TAVARÉ (1994) and TAVARÉ (2004), the joint distribution of the times τ_j can be written in terms of the variables $s_j = \sum_{k=j}^n \tau_k$ for $j \leq n$ ($s_j = 0$ for $j > n$):

$$f(\tau_2, \dots, \tau_n) = \prod_{j=2}^n b_j x(s_j)^{-1} e^{-b_j[\Lambda(s_j) - \Lambda(s_{j+1})]}. \quad (6)$$

Here $b_j = j(j - 1)/2$ and $\Lambda(t) = \int_0^t dt' x(t')^{-1}$ is the “population-size intensity function” defined by GRIFFITHS and TAVARÉ (1994). In a population of constant size, the variables τ_j are mutually independent. In general this is not the case: ZIVKOVIC and WIEHE (2008), for example, calculated $\langle \tau_i \tau_j \rangle$ for a time-varying population (Equations 2 and 3 in their article), using Equation 6.

Given Equation 5, the k th moment of the distribution of T_n is simply

$$\langle T_n^k \rangle = \sum_{\substack{v_2, v_3, \dots, v_n \\ v_2 + v_3 + \dots + v_n = k}} \frac{k!}{v_2! v_3! \dots v_n!} n^{v_2} \dots 2^{v_2} \langle \tau_n^{v_n} \dots \tau_2^{v_2} \rangle, \quad (7)$$

where the variables v_j can assume values between 0 and k (subject to the constraint $v_2 + v_3 + \dots + v_n = k$). In the following we show how the correlation functions of arbitrary order appearing in (7) can be calculated in a very simple manner. Consider first the case $k = 1$. We have

$$\tau_j = \int_0^\infty dt \mathbf{1}_{\{\ell(t)=j\}}. \quad (8)$$

Here $\ell(t)$ denotes the number of lines for a particular realization of the coalescent process at time t in a sample of size $n = \ell(0)$. The indicator function in Equation 8 is unity when $\ell(t) = j$ and zero otherwise. Averaging over realizations gives

$$\langle \tau_j \rangle = \int_0^\infty dt \langle \mathbf{1}_{\{\ell(t)=j\}} \rangle = \int_0^\infty dt f_{nj}(0, t). \quad (9)$$

Here $f_{nm}(t_1, t_2)$ is the conditional probability that n ancestral lines at t_1 coalesce to m lines at time $t_2 > t_1$.

For a constant population size, the coalescent is invariant under time translations, $f_{nm}(t_1, t_2) = g_{nm}(t_2 - t_1)H(t_2 - t_1)$. Here $H(t) = 1$ if $t > 0$ and zero otherwise. The conditional probability $g_{nm}(t)$ was derived by TAVARÉ (1984). For $m \geq 2$ the result is

$$g_{nm}(t) = \sum_{j=m}^n c_{nmj} e^{-b_j t} \tag{10}$$

$$c_{nmj} = (-1)^{j-m} \frac{2j-1}{m!(j-m)!} \frac{\Gamma(m+j-1)}{\Gamma(m)} \frac{\Gamma(n)}{\Gamma(n+j)} \frac{\Gamma(n+1)}{\Gamma(n-j+1)}. \tag{11}$$

In the general case of a variable population size, as shown by GRIFFITHS and TAVARÉ (1994), the conditional probability depends only on the intensity $\Lambda(t_2) - \Lambda(t_1)$ during the time interval $[t_1, t_2]$:

$$f_{nm}(t_1, t_2) = g_{nm}(\Lambda(t_2) - \Lambda(t_1)). \tag{12}$$

Now consider the case $k = 2$. For $i > j$ we have simply

$$\begin{aligned} \tau_i \tau_j &= \int_0^\infty dt_1 \mathbf{1}_{\{\ell(t_1)=i\}} \int_0^\infty dt_2 \mathbf{1}_{\{\ell(t_2)=j\}} \\ &= \int_0^\infty dt_1 \mathbf{1}_{\{\ell(t_1)=i\}} \int_{t_1}^\infty dt_2 \mathbf{1}_{\{\ell(t_2)=j\}}, \end{aligned} \tag{13}$$

because the second indicator function vanishes when $t_2 < t_1$. Averaging over realizations we find

$$\langle \tau_i \tau_j \rangle = \int_0^\infty dt_1 f_{ni}(0, t_1) \int_{t_1}^\infty dt_2 f_{ij}(t_1, t_2). \tag{14}$$

In deriving this result we used the multiplicative rule

$$\langle \mathbf{1}_{\{\ell(t_1)=i\}} \mathbf{1}_{\{\ell(t_2)=j\}} \rangle = f_{ni}(0, t_1) f_{ij}(t_1, t_2). \tag{15}$$

For $i = j$, by contrast, we find

$$\begin{aligned} \tau_j^2 &= \int_0^\infty dt_1 \mathbf{1}_{\{\ell(t_1)=j\}} \int_0^\infty dt_2 \mathbf{1}_{\{\ell(t_2)=j\}} \\ &= 2 \int_0^\infty dt_1 \mathbf{1}_{\{\ell(t_1)=j\}} \int_{t_1}^\infty dt_2 \mathbf{1}_{\{\ell(t_2)=j\}}, \end{aligned} \tag{16}$$

which upon averaging yields

$$\langle \tau_j^2 \rangle = 2 \int_0^\infty dt_1 f_{nj}(0, t_1) \int_{t_1}^\infty dt_2 f_{jj}(t_1, t_2). \tag{17}$$

More general correlation functions are readily obtained in terms of multiple integrals over the functions f_{nm} . Inserting into (7) we see that the combinatorial factors $(\nu_2!)^{-1} \cdots (\nu_n!)^{-1}$ cancel to obtain

$$\begin{aligned} \langle T_n^k \rangle &= k! \sum_{m_1=2}^n \sum_{m_2=2}^{m_1} \cdots \sum_{m_{k-1}=2}^{m_{k-2}} m_1 \cdots m_k \int_0^\infty dt_1 f_{nm_1}(0, t_1) \\ &\quad \cdots \int_{t_{k-1}}^\infty dt_k f_{m_{k-1} m_k}(t_{k-1}, t_k). \end{aligned} \tag{18}$$

Equation 18 provides an explicit expression for the moments of the total branch lengths T_n in populations with population-size variations. The results can be written in a recursive form, particularly convenient for numerical computations,

$$\langle T_n^k(t) \rangle = k \sum_{m=2}^n m \int_t^\infty dt' f_{nm}(t, t') \langle T_m^{k-1}(t') \rangle, \tag{19}$$

with initial conditions $\langle T_m^0(t) \rangle = 1$ for $m \geq 2$ and $\langle T_1^k(t) \rangle = 0$ for $k \geq 1$. Here $T_n(t)$ is the total time corresponding to the genealogy of n sequences sampled at time t in the past given a population-size curve $x(t)$. Note that $t = 0$ corresponds to the present time, so that $T_n(0) \equiv T_n$. In a population of constant size, $T_n(t)$ is independent of t .

Equation 18 or 19 expresses the k th moment of T_n in terms of a $2k$ -fold sum [according to (10) each factor of $f_{n_i m_i}$ contains a sum over j_i]. Equation 18 can be further simplified by explicitly performing the sums over m_1, \dots, m_k . This results in

$$\begin{aligned} \langle T_n^k \rangle &= k! \sum_{j_1=2}^n \cdots \sum_{j_{k-1}=2}^{j_{k-2}} d_{n; j_1, \dots, j_k} \int_0^\infty dt_1 e^{-b_{j_1} \Lambda(t_1)} \int_{t_1}^\infty dt_2 e^{-b_{j_2} [\Lambda(t_2) - \Lambda(t_1)]} \\ &\quad \cdots \int_{t_{k-1}}^\infty dt_k e^{-b_{j_k} [\Lambda(t_k) - \Lambda(t_{k-1})]}. \end{aligned} \tag{20}$$

The coefficients are determined by recursion:

$$d_{n; j} = \sum_{m=2}^j m c_{nmj} = (2j-1)(1 + (-1)^j) \frac{\binom{2n-1}{n-j}}{\binom{2n-1}{n}}, \tag{21}$$

$$d_{n; j_1, \dots, j_k} = \sum_{m=j_2}^{j_1} m c_{nmj_1} d_{m; j_2, \dots, j_k}. \tag{22}$$

For the particular case $k = 1$ our result corresponds to an expression derived by AUSTERLITZ *et al.* (1997) and SLATKIN (1996) and also to the result obtained by summing Equation 1 in ZIVKOVIC and WIEHE (2008). For $k = 2$, the coefficients $d_{n; j_1, j_2}$ are tabulated in Figure A1 in the APPENDIX for small values of n . In general, the nested integrals in Equation 20 cannot be simplified further; their form expresses the correlations of the times τ_j due to population-size variations.

Finally note that for $n = 2$, Equation 18 can be evaluated as follows:

$$\begin{aligned} \langle T_2^k \rangle &= k!2^k \int_0^\infty dt_1 f_{22}(0, t_1) \int_{t_1}^\infty dt_2 f_{22}(t_1, t_2) \cdots \int_{t_{k-1}}^\infty dt_k f_{22}(t_{k-1}, t_k) \\ &= k!2^k \int_0^\infty dt_1 \int_{t_1}^\infty dt_2 \cdots \int_{t_{k-1}}^\infty dt_k f_{22}(0, t_k) = 2^k k \int_0^\infty dt t^{k-1} e^{-\Lambda(t)}. \end{aligned} \tag{23}$$

This representation demonstrates how the expression (18) simplifies when $k > n$.

We conclude this section by briefly describing three different scenarios where our main result (Equation 18) is applicable. First, Sjödin *et al.* (2005) discussed a model where the scaled population size $x(t)$ defined by Equation 3 may assume two values, 1 and x . The population size randomly jumps from 1 to x at rate λ and back at rate λ_x . Initially the population size is $x(0) = 1$. Our result (Equation 18) is directly applicable to a given realization of the random process $x(t)$. We denote the ensemble average over realizations of $x(t)$ by $\overline{\langle T_n^k \rangle}$. By averaging Equation 18 over the corresponding distribution of Λ we find

$$\overline{\langle T_n \rangle} = \sum_{j=1}^n d_{n;j} \frac{\lambda + \lambda_x + b_j/x}{b_j(\lambda/x + \lambda_x + b_j/x)}. \tag{24}$$

Higher moments can be obtained in a similar fashion. This provides explicit expressions for the fluctuations of T_n in the case of slow, fast, and intermediate population-size changes. This model is particularly suited to examine the limit of fast population-size fluctuations $\lambda = \lambda_x \rightarrow \infty$. As expected, the standard Kingman coalescent, Equation 2, is recovered but now with an effective population size $N_{\text{eff}} = N/c$ with $c = (1 + x^{-1})/2$.

Second, intermediate population-size variations over many generations give rise to deviations from the standard Kingman behavior. The deviations are expected to be most significant when the timescale of the size variations is comparable to the times between coalescent events. Such intermediate population-size variations are commonly interpreted as due to a changing environment. In this case it is inappropriate to average over an ensemble of random population size curves $x(t)$. The task is instead to describe the fluctuations of T_n conditional on a particular, externally imposed form of $x(t)$. An example is the question: How does a recent bottleneck influence the distribution of T_n ? To compute the k th moment of T_n , a k -fold integration is required. In general this must be performed numerically. However, in the case of piecewise constant functions $x(t)$ the multiple integrals are straightforward to evaluate. If, on the other hand, the function $x(t)$ is sufficiently “smooth,” the multiple integrals can be evaluated in closed form in the limits of slowly and rapidly varying population sizes as demonstrated below.

Third, in general stochastic population dynamics subject to a slowly changing environment may exhibit both slow changes due to an externally imposed change of the environment (in the form of a time-changing

carrying capacity, for example) and “fast” (generation-to-generation) changes due to the random population dynamics. In the next two sections such a model is introduced and analyzed by means of Equation 18. The analysis is simplified by the observation that the fast size variations are irrelevant when their amplitude remains small. In this case Equation 18 may be evaluated using a deterministic population-size curve that is averaged over the fast changes. In the model discussed in the next two sections this curve is given by the deterministic time dependence of the carrying capacity.

A MODEL FOR A POPULATION WITH TIME-DEPENDENT CARRYING CAPACITY

The purpose of this section is to describe a modified Wright–Fisher model with a fluctuating carrying capacity. This model is used in the numerical simulations of sample genealogies described in the next section. Recall the three key assumptions of the Wright–Fisher model: (a) constant population size, (b) discrete, nonoverlapping generations, and (c) a symmetric multinomial distribution of family sizes. We have adopted the following approach: in our simulations, assumptions b and c are still satisfied, but assumption a is relaxed.

We study a large but finite population of fluctuating size N_l , where $l = 1, 2, \dots$ labels the discrete, nonoverlapping generations forward in time. The model we have adopted is the following: consider a generation l consisting of N_l individuals. The number of individuals in generation $l + 1$ is then given by

$$N_{l+1} = \sum_{j=1}^{N_l} \xi_j, \tag{25}$$

where the random family sizes ξ_j are independent and identically distributed random variables having a Poisson distribution with parameter λ_l (specified below). Consequently the number N_{l+1} is Poisson distributed with mean $N_l \lambda_l$.

This model exhibits a fluctuating population size N_l , rapidly changing from generation to generation. As pointed out in the Introduction, in large populations such fluctuations are averaged over by the ancestral coalescent process and can be captured in terms of an effective population size. The resulting genealogies are simply described by Kingman’s coalescent for a constant effective population size of the form (1) or (4).

Interesting population-size fluctuations occur on larger timescales, corresponding to slow variations of the population size over several generations. Such slow changes are most commonly interpreted as consequences of a changing environment. A natural model for such changes is to impose a finite carrying capacity K_l that varies as a function of generation index l . This is the approach adopted in the following, and we choose

$$\lambda_l = \frac{1+r}{1+rN_l/K_{l+1}} \quad (26)$$

for a certain parameter value $r > 0$. Here K_{l+1} is the carrying capacity in generation $l+1$. If the environmental changes affected the population through fertility variations, K_{l+1} would be replaced by K_l in Equation 26. Equation 26 is chosen so that the population ceases to grow on average when the carrying capacity is reached ($\lambda_l = 1$ for $N_l = K_{l+1}$). When the population size is small and $r \ll 1$, the population growth follows the logistic law, $\lambda_l = 1 + r(1 - N_l/K_{l+1})$, where r is the intrinsic growth rate. The particular form of Equation 26 ensures that $\lambda_l > 0$.

Note that fluctuations of N_l in this model are due to two different sources: rapid fluctuations are caused by the randomness of the family sizes, and slow fluctuations are caused by the time dependence of the carrying capacity. Our choice for the time dependence of K_l is dictated by the following considerations. The aim is to describe the influence of a fluctuating population size upon the statistics of genetic variation. To this end we need to consider the functional form of K_l . A simple choice for K_l is a periodically varying function, such as

$$K_l = K_0[1 + \varepsilon \sin(2\pi\nu l)], \quad \varepsilon \in (0, 1), \quad \nu \in (0, \infty). \quad (27)$$

Note that a more complex dependence of K_l upon l can be obtained from superpositions of such functions with different amplitudes ε and frequencies ν . Here we use simply (27) and investigate how the statistics of genetic variation in a sample depend upon frequency of the fluctuations in K_l .

Figure 3 shows a realization of a curve N_l obtained in this manner (the choice of parameters is given in the Figure 3 legend). Figure 3 clearly exhibits fluctuations in N_l on two timescales, fast and slow. As explained above, the fast fluctuations are irrelevant provided their amplitude is sufficiently small. In this case we expect that the distribution of T_n can be described by a population-size curve that is averaged over the fast fluctuations. In the present model, averaging over the fast fluctuations results in a deterministic population-size curve determined by the carrying capacity (27). This curve is shown in Figure 3 as a dashed line.

Note that conditional on the sequence of population sizes, the genealogy of a set of individuals sampled in generation l can be determined recursively by randomly choosing ancestors in the preceding generations. This is ensured by the assumption that, conditioned on the values of N_l and N_{l+1} , the family sizes follow a symmetric multinomial distribution $\text{Mn}(N_{l+1}; 1/N_l, \dots, 1/N_l)$. The resulting correspondence with the Wright–Fisher rule of reproduction ensures that the genealogies can be determined recursively in the way suggested above.

In the next section we analyze results of three sets of 10,000 computer simulations for this population model

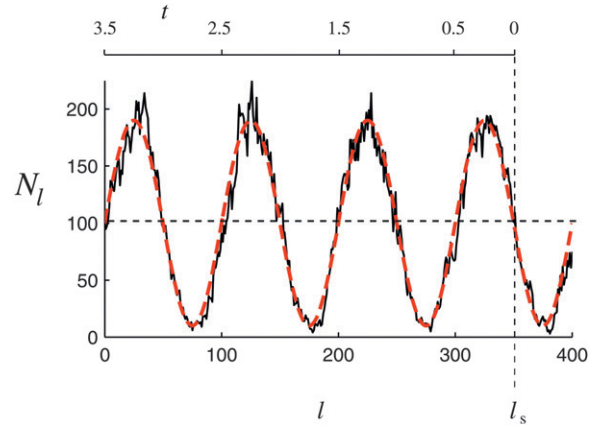


FIGURE 3.—One realization of the curve N_l obtained from simulations of the model described in A MODEL FOR A POPULATION WITH TIME-DEPENDENT CARRYING CAPACITY (black solid line). Choice of parameters: $r = 1$, $K_0 = 100$, $\varepsilon = 0.9$, and $N\nu = 1$, with $N = K_0$. Also shown is an average over the fast fluctuations. The upper horizontal axis illustrates where the population is sampled and how time is counted backward in the coalescent approximation.

with parameters $r = 1$ and $\varepsilon = 0.9$ and for a range of values for ν . The three sets differ in the values for the average carrying capacity that are chosen to be $K_0 = 100$, $K_0 = 1000$, and $K_0 = 10,000$. The population is sampled in generation l_s (see Figure 3), which is chosen so that $2\nu l_s$ becomes an odd natural number. This implies that $K_{l_s} = K_0$ and that the population size was declining toward K_l in the most recent past.

COMPARISON BETWEEN NUMERICAL SIMULATIONS AND COALESCENT PREDICTIONS

In this section we discuss the numerically computed distributions shown in Figure 2 in terms of Equation 18. The shapes observed in Figure 2 are conveniently characterized in terms of their mean $\langle T_n \rangle$, variance, skewness, and kurtosis:

$$\text{var}(T_n) = \langle T_n^2 \rangle - \langle T_n \rangle^2,$$

$$\text{skew}(T_n) = \frac{\langle (T_n - \langle T_n \rangle)^3 \rangle}{\text{var}^{3/2}(T_n)}, \quad (28)$$

$$\text{kurt}(T_n) = \frac{\langle (T_n - \langle T_n \rangle)^4 \rangle}{\text{var}^2(T_n)}. \quad (29)$$

Recall that for a normal distribution the skewness vanishes, and the kurtosis equals three. We can write the skewness and kurtosis in terms of the moments $\langle T_n^k \rangle$ using

$$\begin{aligned} \langle (T_n - \langle T_n \rangle)^3 \rangle &= \langle T_n^3 \rangle - 3\langle T_n^2 \rangle \langle T_n \rangle + 2\langle T_n \rangle^3 \text{ and} \\ \langle (T_n - \langle T_n \rangle)^4 \rangle &= \langle T_n^4 \rangle - 4\langle T_n^3 \rangle \langle T_n \rangle + 6\langle T_n^2 \rangle \langle T_n \rangle^2 - 3\langle T_n \rangle^4. \end{aligned}$$

The moments $\langle T_n^k \rangle$ are evaluated by means of Equations 18 and 19 as functionals of the scaled

population size $x(t)$ followed backward in time. With stochastically fluctuating sizes the scaled population size $x(t)$ also becomes a random process. As Figure 3 indicates, the random fluctuations around the deterministic carrying capacity function are relatively small and we expect that such generation-to-generation fluctuations are irrelevant for the distribution of T_n . We therefore disregard the fast (random) fluctuations of the population sizes and define function $x(t)$ deterministically by

$$x(t) = 1 + \varepsilon \sin(\omega t). \tag{30}$$

This is obtained from an analog of Equation 3 when the population is sampled in generation l_s (as indicated in Figure 3):

$$K_l = Nx \left(\frac{l_s - l}{N} \right). \tag{31}$$

Here (27) was used with $\omega = 2\pi\nu N$, and $N = K_0$. Note that the particular form (30) of $x(t)$ depends upon when the population is sampled. Had the population been sampled at a different time, a different curve $x(t)$ could have resulted, leading in turn to a different distribution $\rho(T_n)$ of T_n , since the distribution depends, for example, upon whether most recently the population was expanding or declining.

Our results are summarized in Figure 4. It shows how the mean, variance, skewness, and kurtosis of the distribution of T_n depend on the scaled frequency ω of the population size variation, Equation 30. Shown are results of numerical simulations of the model described in the previous section (symbols) and results obtained within the coalescent approximation using Equation 19. We observe that the coalescent approximation describes the results of the numerical simulations well, even for small population sizes.

In the numerical simulations we have found that, for very small population sizes, random fluctuations of N_t around the time-dependent carrying capacity K_t become increasingly important. Since we suspected that the small deviations observed in Figure 4a for $K_0 = 100$ were due to such fluctuations, we performed slightly modified simulations imposing a deterministic law upon N_t by forcing $N_t = K_t$ in every generation [where K_t is given by (27)]. Comparison of the corresponding results (not shown) with Figure 4a indicates that the deviations for $K_0 = 100$ at large frequencies are indeed caused by the stochastic fluctuations in the population dynamics underlying Figure 4a. A different interpretation of this effect is the following: when the population size is very small, and when ε is close to unity, the population may exhibit a nonnegligible probability of becoming extinct during the expected time to the most recent common ancestor for a sample of size n . In this

case we have conditioned on the existence of the population during $100K_0$ generations using rejection sampling. In practice this avoids extinction, but it leads to a biased size distribution.

Consider now the frequency dependence of the moments shown in Figure 4. It can be qualitatively and quantitatively understood using Equation 20 together with the following expression for $\Lambda(t)$:

$$\Lambda(t) = \int_0^t \frac{ds}{1 + \varepsilon \sin(\omega s)} = \frac{[\omega t/2\pi + 1/2] - (1/\pi)\arctan(\varepsilon/\sqrt{1-\varepsilon^2}) + (1/\pi)\arctan((\tan(\omega t/2) + \varepsilon)/\sqrt{1-\varepsilon^2})}{(\omega/2\pi)\sqrt{1-\varepsilon^2}}. \tag{32}$$

Here $\lceil z \rceil$ is the smallest integer larger than z . Next we discuss the asymptotical formulas for small and large frequencies ω .

In the limit of $\omega \rightarrow 0$, Equation 32 simplifies to $\Lambda(t) \approx t - \frac{1}{2}\varepsilon\omega t^2$. Inserting this into (20) we find approximately

$$\langle T_n \rangle = 2h_n + 4(g_n - h_n/n)\varepsilon\omega + o(\omega), \quad \omega \rightarrow 0. \tag{33}$$

Here $h_n = \sum_{j=1}^{n-1} j^{-1}$ and $g_n = \sum_{j=1}^{n-1} j^{-2}$. Equation 33 is shown in Figure 4a as a dashed-dotted line. For the variance we find the approximate expression

$$\text{var}(T_n) = 4g_n + 16 \left(f_n - g_n + \frac{h_n - g_n}{n} \right) \varepsilon\omega + o(\omega), \quad \omega \rightarrow 0, \tag{34}$$

with $f_n = \sum_{m=1}^{n-1} (m^{-3} + m^{-2}h_{m+1})$. The limiting value for zero frequency is that of the standard coalescent with constant population size. Equation 34 is shown in Figure 4b as a dashed-dotted line. Similarly the standard results for the constant-size coalescent are obtained for the skewness and for the kurtosis in the limit of $\omega \rightarrow 0$. This limiting behavior is illustrated in Figure 2a, which shows that the distribution of T_n approaches that for Kingman's coalescent for a constant population size in the limit of small frequencies. We note that for $\omega \ll 1$, the population-size dependence is essentially that of a declining population, because the time to the most recent common ancestor is reached before the first maximum in $x(t)$ going backward in time (see Figure 3 and Equation 30).

Of particular interest is the limit of large frequencies, as we now show. As $\omega \rightarrow \infty$, one expects that the coalescent process averages over the population-size oscillations, and the standard coalescent process with a constant effective population size should be obtained. For large but finite frequencies, by contrast, Figure 4a exhibits deviations from the standard coalescent behavior. In the following we analyze the behavior of the moments in this regime. In the limit of large frequencies, Equation 32 simplifies to

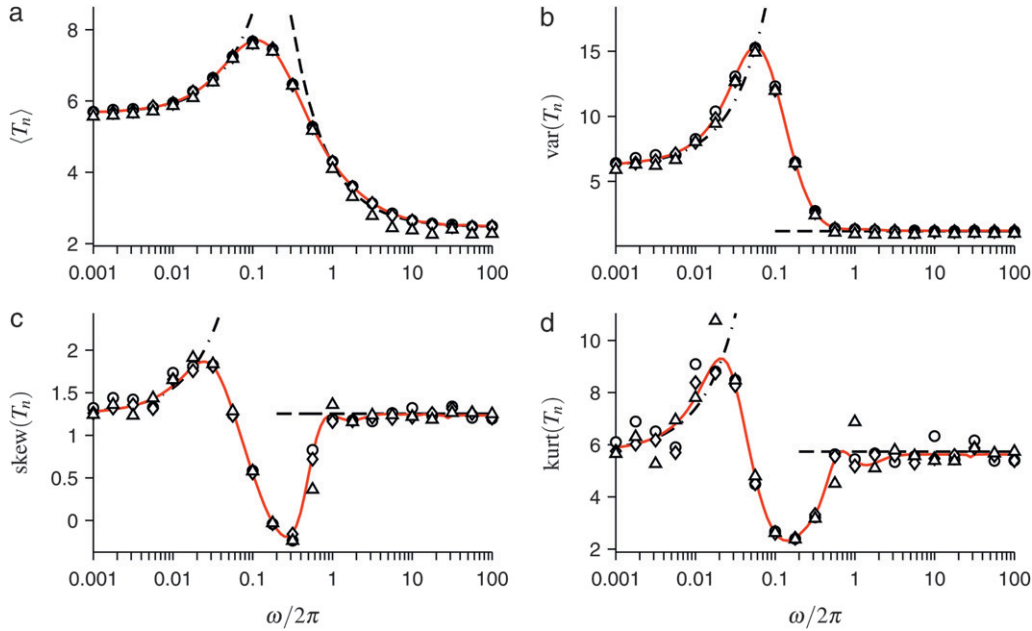


FIGURE 4.—Mean (a), variance (b), skewness (c), and kurtosis (d) of the distribution of T_n for samples of size $n = 10$, as a function of the frequency of the population-size fluctuations. Shown are results of numerical simulations (10,000 simulations with $N = K_0$, $K_0 = 100$, triangles; $K_0 = 1000$, diamonds; and $K_0 = 10,000$, circles) as well as results computed within the coalescent approximation described in COALESCENT APPROXIMATION FORMULAS FOR THE MOMENTS (red solid lines). Black dashed-dotted and dashed lines show the approximations for small frequencies (Equations 33 and 34) and for large frequencies (Equations 39

and 40). The expressions for the limiting behaviors of the skewness and the kurtosis are shown in c and d, but are not given in the text. The remaining parameter values are $r = 1$ and $\varepsilon = 0.9$, as in Figure 2.

$$\Lambda(t) = \frac{t}{\sqrt{1-\varepsilon^2}} - \frac{\arctan(\varepsilon/\sqrt{1-\varepsilon^2})}{\omega\sqrt{1-\varepsilon^2}} + o(\omega^{-1}), \quad \omega \rightarrow \infty. \quad (35)$$

For large frequencies, the function $\Lambda(t)$ is well approximated by a shifted linear function

$$\Lambda(t) \approx tc + \Lambda_0. \quad (36)$$

Here c determines the effective population size according to Equation 4: $N_{\text{eff}} = N/c$ with

$$c = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \frac{dt}{1 + \varepsilon \sin(\omega t)} = \frac{1}{\sqrt{1-\varepsilon^2}}. \quad (37)$$

The parameter c describes the influence of the demographic fluctuations upon the part of the genealogy in the distant past. The small offset

$$\Lambda_0 = -\frac{\arctan(\varepsilon/\sqrt{1-\varepsilon^2})c}{\omega} \approx -\frac{\varepsilon c}{\omega} \quad \text{for } \varepsilon \text{ not too close to unity} \quad (38)$$

describes the influence of demographic changes on the most recent part of the genealogy. Inserting the approximation (36) into (20) we find for large frequencies (and when the amplitude ε is not too close to unity)

$$\langle T_n \rangle = 2h_n c^{-1} + n\varepsilon\omega^{-1} + o(\omega^{-1}), \quad \omega \rightarrow \infty. \quad (39)$$

The first term in (39) is the expected time of Kingman's coalescent for a constant effective population size $N_{\text{eff}} = N/c$. The curve corresponding to (39) is shown as a dashed line in Figure 4a. We infer that corrections to the standard coalescent result are significant when the sample size is large, the amplitude of the size oscillations is not too small, and the frequency ω is of order

unity. This is consistent with the results of SJÖDIN *et al.* (2005).

We now discuss the behavior of the variance shown in Figure 4b. For the second moment we find

$$\langle T_n^2 \rangle \approx 4c^{-2}(g_n + h_n^2) + \frac{4nh_n\varepsilon}{\omega c}. \quad (40)$$

The first term in Equation 40 corresponds to the second moment of T_n in Kingman's coalescent with a constant effective population size. The second term in (40) represents a correction due to finite but large frequencies; it depends in a simple fashion on the effective population-size parameter c and on the sample size n .

Comparing Equations 39 and 40, we arrive at the conclusion that the corresponding correction for the variance $\text{var}(T_n)$ vanishes. This is consistent with the fact that, at large frequencies, the variance of T_n is surprisingly insensitive to changes in frequency (as opposed to the behavior of $\langle T_n \rangle$, see Figure 4, a and b). In fact, the limiting value (shown in Figure 4b as a dashed line) is a very good approximation to $\text{var}(T_n)$ down to $\omega \approx 3$.

Consider now the skewness and the kurtosis shown in Figure 4, c and d. Their behavior is similar to that of the variance: for $\omega > 3$, the skewness and the kurtosis are essentially independent of ω . The results shown in Figure 4 imply that over a large range of frequencies, the distribution of the total branch lengths T_n can be approximated as follows: the distribution is essentially that of the standard Kingman coalescent with an effective population size, but the distribution is shifted such that its mean is given by Equation 39, rather than by $2h_n/c$.

One may wonder when this “rigid shift” occurs. Given Equation 18 it is straightforward to work out the fluctuations of the times τ_j within the approximation (36). We find that for $j < n$, the expected value of τ_j is exactly that of the standard Kingman coalescent with effective population size. But for $j = n$ it is rigidly shifted by $-\Lambda_0/c$. This indicates that the genealogies are essentially those of the standard coalescent, but modified by an initial rigid shift. In the parameter regime discussed here, the distribution of times is expected to be well approximated by a two-parameter family of distributions,

$$P(T_n < z) \approx \left[1 - \exp\left(-\frac{zc + n\Lambda_0}{2}\right) \right]^{n-1} \quad (41)$$

when $zc > -n\Lambda_0$, and $P(T_n < z) \approx 0$ for smaller values of z . The first parameter determines the effective population size. It parameterizes the slope of the function $\Lambda(t)$ at large times and describes the demographic effect on the distant past of the genealogy. The second parameter, Λ_0 describes the influence of the demographic fluctuations on the initial part of the sample genealogy. This parameter can be negative (recent population decline, this is the case shown in Figure 3) or positive (recent population expansion). When $\Lambda_0 > 0$, the distribution $\rho(T_n)$ is rigidly shifted to the left. In this case the approximation (36) is expected to break down when the body of the distribution reaches $T_n = 0$.

Note that the distribution (41) cannot be described by a single parameter (a “generalized effective population size”). The approximation (41) was used to generate the red dashed curve in Figure 2d.

CONCLUSIONS

The aim of this article was to investigate how the frequency of population-size fluctuations determines the shape of the distribution of total branch lengths of sample genealogies and thus of statistical measures of genetic variation.

We performed simulations for a modified Wright–Fisher model of a population subject to a time-periodically varying carrying capacity and determined the distribution of the total branch lengths, shown in Figure 2. We characterized how the shapes of the distributions depend upon the frequency of the population-size fluctuations by computing the frequency dependence of the moments of these distributions. We could explain these dependencies in terms of coalescent approximations. In particular, we derived a general expression—Equation 20—for the moments $\langle T_n^k \rangle$ in populations subject to smooth population changes of otherwise arbitrary form.

Our results show how quickly (or slowly) the standard coalescent result for a constant (effective) population sizes is recovered in the limits of large and small

frequencies. More importantly, our coalescent results allow us to determine how significant deviations are at large but finite frequencies. In this case we argued that at large frequencies, the distribution of T_n is essentially that of the standard Kingman coalescent with an effective population size $N_{\text{eff}} = N/c$, but with a shifted mean value

$$\langle T_n \rangle \approx \frac{2}{c} \sum_{j=1}^{n-1} \frac{1}{j} + \frac{n\varepsilon}{\omega} \quad \text{for } \omega \gg 1. \quad (42)$$

The first term on the rhs corresponds to the result of the standard Kingman coalescent with a constant effective population size. The second term on the rhs is the correction term resulting from the population-size variations (ε is the amplitude of the population-size oscillations, ω is its frequency, and n is the sample size). We infer that corrections to the standard coalescent result are largest when the sample size is large and the amplitude ε of the size fluctuations is not very small. This is consistent with the results of Sjödin *et al.* (2005).

Last but not least we found that the coalescent approximation yields a reliable description of the numerical data, even for very small populations.

We close with a number of remarks. First, Equation 20 is easily generalized to describe the moments of observables that are polynomial functions of the times τ_j . Particularly simple is the case of observables A that are linear functions of the times τ_j , $A_n = \sum_{j=2}^n a_j \tau_j$. In this case the k th moment of A_n is given by Equation 20, but with modified coefficients: the factors m in Equations 21 and 22 are replaced by a_m .

Second, some observables [such as the F -statistic (Fu and Li 1993)] can be written as linear functions of τ_j , but with random coefficients. In this case too it is possible to explicitly compute the moments of the distribution of the observable. These two questions are addressed in a separate article (S. SAGITOV, M. RAFAJLOVIC, B. MEHLIG, and A. ERIKSSON, unpublished results).

Third, Equation 19 allows us to determine in a transparent fashion how the fluctuations of T_n depend upon the time at which the population is sampled. This will make it possible to discuss, for example, how Tajima’s D -statistic or the F -statistic depends upon the time of sampling after a bottleneck, a population expansion, or a decline.

Fourth, population-size fluctuations are sampled nonuniformly by the genealogies: initial coalescent events occur at faster rates and are thus more sensitive to recent size fluctuations. Remote coalescent events, by contrast, occur at slower rates, thus damping the effect of size fluctuations in the distant past. We therefore expect significant deviations from the standard coalescent behavior arising from the most recent history for large sample sizes n . It would be interesting to quantify this expectation by computing the covariances and higher moments of the times τ_j during which the sample genealogy has j lines: first, for large $i \approx n$ and $j \approx n$ we

expect to observe strong correlations between τ_i and τ_j and thus deviations from the standard Kingman coalescent. Second, for small values of i and j we expect the times τ_i and τ_j to decorrelate and to follow the distribution of the standard coalescent (with an effective population size).

Fifth, the model introduced in A MODEL FOR A POPULATION WITH TIME-DEPENDENT CARRYING CAPACITY assumes a carrying capacity that varies sinusoidally, with a single frequency. It turns out, however, that our findings (summarized in Equation 41) are valid for arbitrary time-dependent fluctuations with a sufficiently strong and narrow mode at high frequencies. Examples are linear combinations of high-frequency oscillations or stochastic fluctuations around a constant population size, with a sufficiently narrow frequency spectrum. In this case, too, we expect that $\Lambda(t)$ is well approximated by (36). If this is the case, the distribution of times is of the form (41) when Λ_0 is small.

Taken together, the results derived in this article give a rather complete understanding of the fluctuations of empirical observables due to population size variations. These results will be significant when attempting to disentangle the effects of population-size variations from other factors influencing genetic variation.

Our results raise the question under which circumstances the deviations from standard coalescent behaviors due to population-size fluctuations (Figures 2 and 4) are most likely to strongly affect the interpretation of empirical data. As our analysis indicates, the deviations become substantial when the frequency $\omega = 2\pi\nu N$ is of order unity. Here ν is the frequency of the population size variations, Equation 30, and N is a suitable measure of the population size. In other words, rapid population-size fluctuations will have the strongest effect (other than simply determining the effective population size) in small local subpopulations with restricted gene flow between subpopulations with different fluctuations. The deviations are expected to be smaller at larger spatial scales, because the ancestral process averages over the spatial fluctuations. More generally, we conclude that deviations from standard coalescent behavior are expected for populations subject to an environment that changes as a function of space and time on neither too small nor too large length and timescales. An example for such a population is the marine snail *Littorina saxatilis*. Its habitat on the Northern coast of Bohuslän (Sweden) is fragmented into subpopulations with strongly restricted gene flow between them, and effective population sizes of subpopulations have been found to be very small (K. JOHANNESSON, personal

communication). Starting from the results derived in this article, we hope to determine gene genealogies in such fragmented populations subject to variations of population size in space and time.

We thank two anonymous reviewers for helpful comments and suggestions. Support from Vetenskapsrådet, The Bank of Sweden Tercentenary Foundation, and the Centre for Theoretical Biology at the University of Gothenburg is gratefully acknowledged.

LITERATURE CITED

- AUSTERLITZ, B., B. JUNG-MULLER, B. GODELLE and P. GOUYON, 1997 Evolution of coalescence times, genetic diversity and structure during colonization. *Theor. Popul. Biol.* **51**: 148–164.
- EWENS, W., 1982 The concept of the effective population size. *Theor. Popul. Biol.* **21**: 373–378.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- FU, Y., and W. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GARRIGAN, D., and M. F. HAMMER, 2006 Reconstructing human origins in the genomic era. *Nat. Rev. Genet.* **7**: 669–680.
- GRIFFITHS, R., and S. TAVARÉ, 1994 Sampling theory for neutral alleles in a varying environment. *Philos. Trans. R. Soc. Lond. B* **344**: 403–410.
- HEIN, J., M. H. SCHIERUP and C. WIUF, 2005 *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press, London/New York/Oxford.
- JAGERS, P., and S. SAGITOV, 2004 Convergence to the coalescent in populations of substantially varying size. *J. Appl. Probab.* **41**: 368–378.
- KAJ, I., and S. KRONE, 2003 The coalescent process in a population with stochastically varying size. *J. Appl. Probab.* **40**: 33–48.
- KINGMAN, J., 1982 The coalescent. *Stoch. Proc. Appl.* **13**: 235–248.
- MORAN, P., 1958 Random processes in genetics. *Proc. Camb. Philos. Soc.* **54**: 60–71.
- NORDBORG, M., and S. KRONE, 2003 *Modern Developments in Population Genetics: The Legacy of Gustave Malécot*, pp. 194–232. Oxford University Press, Oxford.
- SJÖDIN, P., I. KAJ, S. KRONE, M. LASCOUX and M. NORDBORG, 2005 On the meaning and existence of an effective population size. *Genetics* **169**: 1061–1070.
- SLATKIN, M., 1996 Gene genealogies within mutant allelic classes. *Genetics* **143**: 579–587.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TAVARÉ, S., 1984 Lines of descent and genealogical processes, and their application in population genetics models. *Theor. Popul. Biol.* **26**: 119–164.
- TAVARÉ, S., 2004 *Ancestral Inference in Population Genetics*, pp. 1–188. Springer, Berlin.
- WAKELEY, J., and O. SARGSYAN, 2009 Extensions of the coalescent effective population size. *Genetics* **181**: 341–345.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- ZENG, K., Y. FU, S. SHI and C. WU, 2006 Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* **174**: 1431–1439.
- ZIVKOVIC, D., and T. WIEHE, 2008 Second-order moments of segregating sites under variable population size. *Genetics* **180**: 341–357.

Communicating editor: J. WAKELEY

APPENDIX: COEFFICIENTS $d_{n;j_1,j_2}$ FOR $n = 2, \dots, 10$

In Figure A1 we give the coefficients $d_{n;j_1,j_2}$ determining the second moment $\langle T_n^2 \rangle$ according to Equation 20 for $n = 2, \dots, 10$. Note that the coefficient for $n = 2$ is consistent with Equation 23.

j_2	j_1	2									
2		4									
j_2	j_1	2	3								
2		6	3								
j_2	j_1	2	3	4							
2		$\frac{36}{5}$	6	$\frac{6}{5}$							
4				$\frac{8}{5}$							
j_2	j_1	2	3	4	5						
2		8	$\frac{60}{7}$	3	$\frac{3}{7}$						
4				4	1						
j_2	j_1	2	3	4	5	6					
2		$\frac{60}{7}$	$\frac{75}{7}$	5	$\frac{9}{7}$	$\frac{1}{7}$					
4				$\frac{20}{3}$	3	$\frac{1}{3}$					
6						$\frac{2}{7}$					
j_2	j_1	2	3	4	5	6	7				
2		9	$\frac{25}{2}$	7	$\frac{27}{11}$	$\frac{1}{2}$	$\frac{1}{22}$				
4				$\frac{28}{3}$	$\frac{63}{11}$	$\frac{7}{6}$	$\frac{7}{66}$				
6						1	$\frac{1}{6}$				
j_2	j_1	2	3	4	5	6	7	8			
2		$\frac{28}{3}$	14	$\frac{98}{11}$	$\frac{42}{11}$	$\frac{14}{13}$	$\frac{2}{11}$	$\frac{2}{143}$			
4				$\frac{392}{33}$	$\frac{98}{11}$	$\frac{98}{39}$	$\frac{14}{33}$	$\frac{14}{429}$			
6						$\frac{28}{13}$	$\frac{2}{3}$	$\frac{2}{39}$			
8								$\frac{16}{429}$			
j_2	j_1	2	3	4	5	6	7	8	9		
2		$\frac{48}{5}$	$\frac{168}{11}$	$\frac{588}{55}$	$\frac{756}{143}$	$\frac{24}{13}$	$\frac{24}{55}$	$\frac{9}{143}$	$\frac{3}{715}$		
4				$\frac{784}{55}$	$\frac{1764}{143}$	$\frac{56}{13}$	$\frac{56}{55}$	$\frac{21}{143}$	$\frac{7}{715}$		
6						$\frac{48}{13}$	$\frac{8}{5}$	$\frac{3}{13}$	$\frac{1}{65}$		
8								$\frac{24}{143}$	$\frac{3}{143}$		
j_2	j_1	2	3	4	5	6	7	8	9	10	
2		$\frac{108}{11}$	$\frac{180}{11}$	$\frac{1764}{143}$	$\frac{972}{143}$	$\frac{36}{13}$	$\frac{9}{11}$	$\frac{405}{2431}$	$\frac{3}{143}$	$\frac{3}{2431}$	
4				$\frac{2352}{143}$	$\frac{2268}{143}$	$\frac{84}{13}$	$\frac{21}{11}$	$\frac{945}{2431}$	$\frac{7}{143}$	$\frac{7}{2431}$	
6						$\frac{72}{13}$	3	$\frac{135}{221}$	$\frac{1}{13}$	$\frac{1}{221}$	
8								$\frac{1080}{2431}$	$\frac{15}{143}$	$\frac{15}{2431}$	
10									$\frac{10}{2431}$	$\frac{10}{2431}$	

FIGURE A1.—Coefficients $d_{n;j_1,j_2}$ occurring in Equation 20 for $n = 2, \dots, 10$. Coefficients for odd values of j_2 vanish.