Copyright © 1997 by the Genetics Society of America

The Coalescent Process With Selfing

Magnus Nordborg* and Peter Donnelly†

* *Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637-1573 and*
† *Department of Statistics, University of Oxford, Oxford OX1 3TG, United Kingdom*

Manuscript received August 15, 1996
Accepted for publication March 28, 1997

ABSTRACT

A method for estimating the selfing rate using DNA sequence data was recently proposed by Milligan. Unfortunately, a number of errors make interpretation of his results problematic. In the present paper we first show how the usual coalescent process can be adapted to models that include selfing, and then use this result to find moment estimators as well as the likelihood surface for the selfing rate, s, and the scaled mutation rate, θ. We conclude that, regardless of the method used, large sample sizes are necessary to estimate s with any degree of certainty, and that the estimate is always highly sensitive to recent changes in the true value.

THERE is long-standing interest in estimating the degree of self-fertilization, or selfing, in partially selfing populations. Several methods have been suggested, most of them based on allozyme frequency data (BROWN 1990). Recently, MILLIGAN (1996) suggested a method using DNA sequence data based on the difference in coalescence time for alleles sampled within and between an individual.

In this article, we first show in general how the standard neutral coalescent process can be used for models with partial selfing. We use this result to derive alternatives to MILLIGAN's estimator and to compare the properties of the estimators. Finally, we discuss the usefulness of DNA data for estimating the selfing rate.

THE COALESCENT PROCESS WITH PARTIAL SELFING

Consider a Wright-Fisher model of N diploid hermaphrodites, with N very large. Define the selfing rate s as the fraction of offspring that is produced by self-fertilization. The remaining fraction $1 - s$ is produced via random mating. For a single locus with two alleles, it is easy to show that the genotypic proportions quickly converge to $p^2 + pqF$, $2pq(1 - F)$, and $q^2 + pqF$, where

$$F = \frac{s}{2 - s} \qquad (1)$$

is the equilibrium inbreeding coefficient (HALDANE 1924).

Now consider the dynamics of a single neutral allele when N is finite. Under the assumption that $s \gg 1/N$, it has been argued, using the diffusion approximation, that the theory for random mating populations holds with variance (and inbreeding) effective population size of

$$N_e = \frac{1}{1 + F} N = \frac{2 - s}{2} N \qquad (2)$$

(LI 1955; WRIGHT 1969; POLLAK 1987).

We will describe the analogous result for the standard n-coalescent (KINGMAN 1982a,c; TAVARÉ 1984). More precisely, we will demonstrate that, apart from its initial behavior, the coalescent with partial selfing is identical to the coalescent for random mating if time is rescaled by a factor corresponding to the variance effective population size (2). This similarity should not be unexpected given the close correspondence between variance effective population size and the time scale of the coalescent (KINGMAN 1982b).

Take the case of two alleles first (in the context of the coalescent, we will use "allele" to denote a region of DNA within which recombination can be ignored). With random mating, we simply trace their ancestry until a single allele that is their common ancestor is found. Going backward in time, we have a Markov process with two states: two alleles, or a single common ancestral allele (this state is of course absorbing). With selfing the same process has three states: two alleles *in distinct individuals*, two distinct alleles *in the same individual*, and a single common ancestral allele.

It is easy to see that the transition matrix (with the states in the order described) for this process is

$$\mathbf{P} = \begin{bmatrix} 1 - \frac{1}{N} & \frac{1}{2N} & \frac{1}{2N} \\ (1 - s) + O\left(\frac{1}{N}\right) & \frac{s}{2} + O\left(\frac{1}{N}\right) & \frac{s}{2} + O\left(\frac{1}{N}\right) \\ 0 & 0 & 1 \end{bmatrix}. \qquad (3)$$

Consider the first state: two alleles in distinct individuals in a given generation. The two alleles will remain in

Corresponding author: Magnus Nordborg, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637-1573. E-mail: magnus@darwin.uchicago.edu distinct individuals for a random amount of time that is geometrically distributed with expectation N. When the process leaves this state (because the two alleles find themselves in the same ancestral individual), the common ancestral allele is found, and thus the process jumps to the third state, with probability $^1/_2$, else the process jumps to its second state: two distinct alleles in the same individual. When the process is in its second state, it follows from the transition matrix (3) that it remains there for a random number of generations that is geometrically distributed with expectation $s/(2-s) + O(1/N)$. When it leaves the second state, a common ancestor will be found, and so the process will jump to its third state, with probability

$$\frac{s/2}{s/2 + (1-s)} + O\left(\frac{1}{N}\right) \approx \frac{s}{2-s}, \tag{4}$$

otherwise it returns its first state: two alleles in distinct individuals.

In summary, the process remains in the first state for an amount of time that is geometrically distributed with expectation of $O(N)$, after which either a common ancestor is found with probability $^1/_2$ or the process jumps to the second state. The process remains in the second state for an amount of time that is geometrically distributed with expectation of $O(1)$, after which a common ancestor is found with approximate probability $s/(2-s)$, or else the process returns to the first state.

Recall that in the coalescent, time is measured in units of $O(N)$ generations. With this time-scaling, any time spent in the second state is negligible. Thus in the coalescent approximation to models with selfing, the second state becomes instantaneous. If the process starts in this state, it will leave it instantaneously for either the first or the third state, with respective probabilities $1 - s/(2-s)$ and $s/(2-s)$. If the process starts in the first state, we will never "see" any time spent in the second state. A proportion $1 - s/(2-s)$ of transitions to the second state will instantaneously return to the first state, while the remaining proportion $s/(2-s)$ will move instantaneously to the third state. Thus, with time measured in units of N generations, the process moves from the first state to the third state with rate

$$^1/_2 + {}^1/_2 \frac{s}{2-s} = \frac{1}{2-s}. \tag{5}$$

Alternatively, if time is measured in units of $(2-s)N$ generations (that is, $2N_e$, where N_e is given by Equation 2), the process waits in the first state for an exponentially distributed amount of time with mean one before coalescing (*i.e.,* moving to the third state). If the two alleles of interest are sampled from the same individual, the process starts in the second state, but moves instantaneously to either the first or third state, with the probabilities given above. Thus, except for this initial behavior, the process behaves like the usual coalescent for a sample of two alleles.

The preceding argument can easily be extended to a sample of n alleles. In this case, more states are possible since one needs to keep track of the number of ancestral alleles paired within individuals in each generation. For instance, we may have all n alleles in separate individuals, $n-2$ alleles in separate individuals and one individual with two alleles, $n-4$ alleles in separate individuals and two individuals with two alleles, *etc.* When time is measured in units of $O(N)$ generations, however, all states involving one or more pairs of alleles within the same individual are instantaneous, for reasons analogous to those given above. Again, the effect of selfing is to speed up the coalescence rate. In particular, with n alleles in distinct individuals and time measured in units of $(2-s)N$ generations, the transition to $n-1$ alleles in distinct individuals occurs after an exponentially distributed amount of time with mean $2/(n(n-1))$, again as in the usual coalescent.

Thus, in the presence of selfing and with time measured in units of $(2-s)N$ generations, if we start with all alleles in distinct individuals, this will remain the case, and it is only necessary to track the number of ancestral alleles. Further the process describing the ancestry of the sampled alleles is the usual coalescent. Another way of thinking of this result is that with time measured in units of $2N$ generations, the ancestry is described by a version of the usual coalescent in which the coalescence rates are increased by a factor of $2/(2-s)$. For a formal proof of this result, see MÖHLE (1996).

As we saw with two alleles, the other effect of selfing is to change the initial behavior of the process. Suppose time is measured in units of $O(N)$ generations, and the n sampled alleles consist of $2k$ alleles sampled in pairs within individuals and $n-2k$ alleles sampled each from distinct individuals. Independently, for each of the k pairs, the two alleles will instantaneously coalesce with probability $s/(2-s)$, otherwise they will instantaneously jump to distinct individuals. Following these instantaneous transitions, there will be a random number, $n-X$, of ancestral alleles, all in distinct individuals, where X has a binomial distribution with parameters k and $s/(2-s)$. Thereafter, the process behaves as described in the preceding paragraph.

Simulating samples from models with selfing: As is the case for the model with random mating, the coalescent provides a convenient and efficient simulation tool. The following algorithm describes how to simulate a sample of n alleles, $2k$ of which were sampled in pairs within individuals.

1. Simulate a binomial random variable X with parameters k and $s/(2-s)$.

2. With the usual coalescent algorithms for random mating and the mutation mechanism under consideration, simulate a sample of size $n-X$ alleles. The effect of selfing in increasing coalescence rates needs to be

allowed for. One approach is to use the urn scheme of DONNELLY and TAVARÉ (1995, p. 412), with $\sigma^2 = 2/(2-s)$ and $\theta = 4N\mu$, where μ is the mutation rate per gene per generation. Alternatively, a scheme such as that described in HUDSON (1990) can be used with θ redefined to take the value $2(2-s)N\mu$.

3. Take the $n - X$ alleles simulated in step 2 and write them in random order as $Y_1, Y_2, \ldots, Y_{n-X}$. Form X diploids, which are automatically homozygous, as $(Y_1, Y_1), (Y_2, Y_2), \ldots, (Y_X, Y_X)$. Use the remaining $n - 2X$ alleles from the simulated sample to form the remaining $k - X$ diploids and the $n - 2k$ alleles sampled each from a distinct individual. Thus the final sample will be

$$(Y_1, Y_1), \ldots, (Y_X, Y_X), (Y_{X+1}, Y_{X+2}), \ldots, (Y_{2k-X-1}, Y_{2k-X}), Y_{2k-X+1}, \ldots, Y_{n-X}. \quad (6)$$

Other factors, such as geographic structuring or recombination, can be included in the simulation by modifying step 2 above. Either the overall coalescence rates need to be increased by a factor of $2/(2-s)$, or all other rates need to be decreased by this factor, in coalescent simulations for the appropriate models without selfing.

COALESCENT-BASED ESTIMATES OF THE SELFING RATE

The expected coalescence time for a pair of alleles sampled from the same individual differs from that for a pair of alleles sampled from separate individuals, and this difference depends on the selfing rate, s. MILLIGAN (1996) recently exploited this fact for estimating s.

Let T_w be the coalescence time for a pair of alleles within an individual. Analogously, let T_b be the coalescence time for a pair of alleles sampled from two separate individuals. It follows easily from the discussion of the previous section that

$$ET_w = 1 - s, \quad (7)$$

$$ET_b = \frac{2-s}{2}, \quad (8)$$

with time measured in units of $2N$ generations. These expectations were also derived by MILLIGAN (1996) using discrete-time recursions for the genealogy of a pair of alleles.

Let S_w (S_b) be the number of sites that distinguish a pair of alleles sampled within an individual (between individuals). Assuming the infinite-sites model of neutral sequence evolution, we have

$$ES_w = (1-s)\theta, \quad (9)$$

$$ES_b = \frac{2-s}{2}\theta, \quad (10)$$

where $\theta = 4N\mu$, and μ is the neutral mutation rate. Based on this, MILLIGAN (1996) suggested the following method-of-moments estimators for s and θ:

$$\tilde{s}_S = \frac{2(S_b - S_w)}{2S_b - S_w}, \quad (11)$$

$$\tilde{\theta}_S = 2S_b - S_w. \quad (12)$$

These equations are identical to Equations 12 and 13 in MILLIGAN (1996), ignoring terms of $O(1/N)$.

Unfortunately, MILLIGAN's study of the usefulness of these estimators suffers from two serious problems.

Estimators based on homozygosity: The first problem concerns the estimators with which MILLIGAN compares his estimators. For θ he uses

$$\hat{\theta} = \frac{1-F}{F}, \quad (13)$$

where F is "the frequency of individuals homozygous for alleles identical by descent" (p. 622), whereas for s he uses

$$\hat{s} = \frac{2F}{1+F}. \quad (14)$$

If F denotes the sample homozygosity, and we assume random mating and either the infinite-sites or the infinite-alleles model, then the expression (13) does indeed serve as an estimator of θ, however it has poor statistical properties (STEWART 1976; DONNELLY and TAVARÉ 1995). The expression (14), on the other hand, is derived from the classical equation (1), where F, WRIGHT's "fixation index," measures the deviation from Hardy-Weinberg proportions in a single-locus, two-allele model without mutation (BROWN and ALLARD 1970). Thus, F has quite different meanings in (13) and (14). The choice of the estimator (14) is based on a misinterpretation of F in (14) as the sample homozygosity under the infinite-sites model (effectively confusing "identity by descent" with "identity in state"). It is not surprising then, that, as MILLIGAN finds, this estimator of s is extraordinarily poor.

The homozygosity estimators used for comparison by MILLIGAN are therefore not appropriate. Reasonable method-of-moments estimators based on homozygosity can be found as follows.

Let H_w (H_b) be the proportion of homozygous pairs of alleles within (between) individuals. It is easy to show that, under the infinite-sites model, we have

$$EH_w = \frac{2+\theta s}{2+(2-s)\theta}, \quad (15)$$

$$EH_b = \frac{2}{2+(2-s)\theta}, \quad (16)$$

leading to the following method-of-moments estimators for s and θ:

$$\tilde{s}_H = \frac{2(H_w - H_b)}{1 + H_w - 2H_b}, \quad (17)$$

$$\tilde{\theta}_H = \frac{1 + H_w - 2H_b}{H_b}. \qquad (18)$$

Properties of the estimators: The second problem in MILLIGAN's study concerns the evaluation of the properties of the estimators. Rather than simulating "real" genealogical samples, MILLIGAN simulates multiple realizations of the coalescent for a *pair* of alleles and then estimates the properties of larger samples by drawing from this distribution. He correctly notes that this assumes independence of all pairs of alleles, which is not true because the alleles in a sample are related by a common genealogy, but states (p. 624) that "the effect of that larger genealogy is to reduce the actual variation between distinct pairs of alleles and therefore to reduce the variation in the estimates derived from those pairs." This statement is not correct. In fact, exactly because of shared genealogy, distinct pairs of alleles in the sample are highly positively correlated. As a consequence, the variance of the estimators will be larger, and probably substantially so, than would be the case were all pairs independent. There is also another problem in that, as we show below, the estimators quite often show considerable bias when used on correctly simulated samples, or fail to yield meaningful estimates at all.

To assess accurately the properties of the estimators, we used the method described above to simulate samples from a selfing population with infinite-sites mutation. We simulated all combinations of n = 2, 10, 50, 100, or 200, s = 0, 0.2, 0.4, 0.8, or 1, and θ = 0.1, 1, 5, or 10. For each combination of parameters, 10^4 samples were simulated. For each simulated sample, the estimators described above were calculated. The results for $n = 200$ were very similar to those for $n = 100$ and are not shown.

The results from our simulations, summarized in Figures 1–5, suggest behavior rather different from that reported by MILLIGAN. In fact, the estimators have poor properties for a large range of parameter values. Take the estimators of s, first. It is clear from inspection of (17) that $\tilde{s}_H$ yields a negative estimate of the selfing rate if $H_w < H_b$. Likewise, (11) shows that $\tilde{s}_S$ gives a negative estimate if $S_w > S_b$. One might think that this should only happen rarely, at least for reasonable sample sizes, but as illustrated by Figure 1, which shows the frequency of simulated samples that give negative estimates of s for various real parameter values, this is not so. It is natural to take a negative estimate of s to be zero (and this was done in our simulations), even at the cost of introducing bias. Note also the frequency of samples that yield no estimate of s at all because $S_w = S_b = 0$ (Figure 1, right column). The estimators of θ are much better behaved in these respects. The homozygosity estimator $\tilde{\theta}_H$ cannot be used when $H_b = 0$, but this was observed only very rarely (<0.3% of the samples) for $n = 10$ and never for $n = 50$ or greater.

The bias of the estimators of s is shown in Figure 2. It is clear that both estimators have similar properties, although the bias of the homozygosity estimator $\tilde{s}_H$ is always smaller. In general, the bias is worse the smaller the value of θ. Of course, increasing the sample size always helps, albeit very slowly. The standard deviations of $\tilde{s}_S$ and $\tilde{s}_H$ are compared in Figure 3. Again, both estimators have similar properties in general, although $\tilde{s}_H$ is clearly superior. The bias and standard deviation of the estimators of θ are shown in Figure 4. The homozygosity estimator, $\tilde{\theta}_H$, is unquestionably inferior, except perhaps for small values of s and high θ, in which case it is more biased but has smaller variance.

Maximum-likelihood estimates: For moderate to large sample sizes, maximum-likelihood estimates should be superior to those based on moments. The structure of the process described above allows the likelihood in the model with selfing to be written in terms of likelihoods for models with random mating. Loosely speaking, one simply has to allow for the extra randomness involved in initial instantaneous coalescences.

Suppose the data, $\mathcal{D}$, consists of sequences from n diploid individuals. Write m for the number of homozygous individuals and $\mathcal{D}^*$ for the $2(n - m)$ alleles sampled in the non-homozygous individuals. Among the m homozygous individuals, write k for the number of different alleles ($k \le m$). Label these alleles $A_1, \ldots, A_k$ and write m_i, $i = 1, 2, \ldots, k$ for the number of A_iA_i homozygotes. In other words,

$$\mathcal{D} = \underbrace{A_1A_1 \cdots A_1A_1}_{m_1 \text{ times}} \underbrace{A_2A_2 \cdots A_2A_2}_{m_2 \text{ times}} \cdots \underbrace{A_kA_k \cdots A_kA_k}_{m_k \text{ times}} \mathcal{D}^*. \qquad (19)$$

Re-parameterize, and write $\psi = (1 - s/2)\theta$. Then the likelihood $l(s, \psi, \mathcal{D})$ for s and ψ with data $\mathcal{D}$ is

$$l(s, \psi, \mathcal{D}) = \left(1 - \frac{s}{2 - s}\right)^{n-m} \sum_{i_1=0}^{m_1} \sum_{i_2=0}^{m_2} \cdots \sum_{i_k=0}^{m_k} \prod_{j=1}^{k} B\left(m_j, \frac{s}{2 - s}, i_j\right) l_c(\psi, \tilde{\mathcal{D}}_{i_1,\ldots,i_k}), \qquad (20)$$

where $B(n, p, x)$ is the probability that a Binomial(n, p) random variable will take the value x, $\tilde{\mathcal{D}}_{i_1,\ldots,i_k}$ is a data set formed from $\mathcal{D}$ by taking $\mathcal{D}^*$ and $i_j + 2(m_j - i_j)$ copies of allele A_j, $j = 1, \ldots, k$, *i.e.*,

$$\tilde{\mathcal{D}}_{i_1,\ldots,i_k} = \underbrace{A_1 \cdots A_1}_{i_1 \text{ times}} \underbrace{A_1A_1 \cdots A_1A_1}_{m_1 - i_1 \text{ times}} \cdots \underbrace{A_k \cdots A_k}_{i_k \text{ times}} \underbrace{A_kA_k \cdots A_kA_k}_{m_k - i_k \text{ times}} \mathcal{D}^*, \qquad (21)$$

and $l_c(\xi, \mathcal{D})$ is the usual coalescent likelihood function with data $\mathcal{D}$ and mutation parameter $\theta = \xi$. Computer programs for evaluating this function are available (GRIFFITHS and TAVARÉ 1994a–c; KUHNER *et al.* 1995), so it suffices to write a "front-end" that implements the summation in (20).

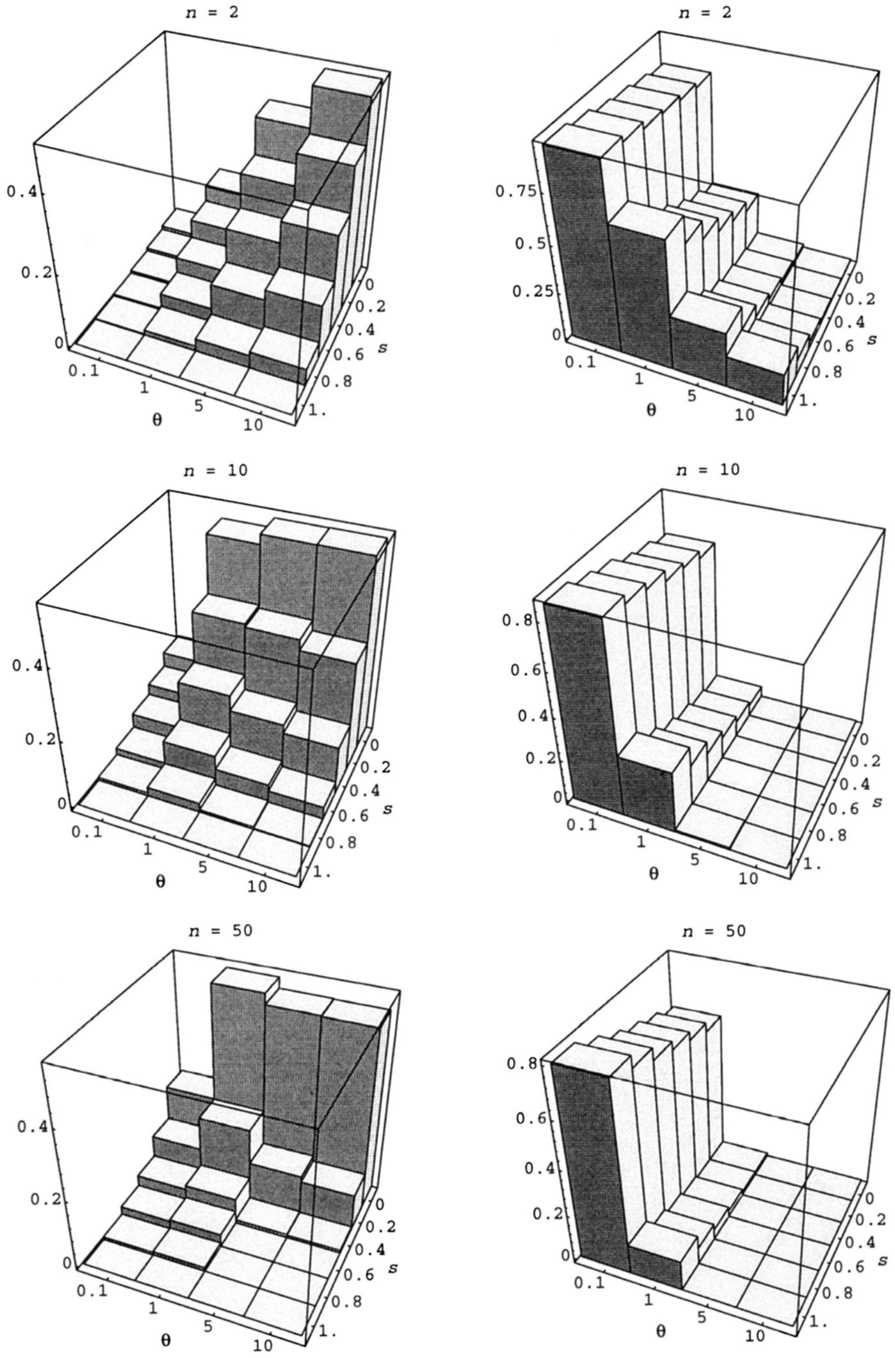

FIGURE 1.—The fraction of simulated samples that yield negative estimates (left column) or no estimate (right column) using $\tilde{s}_S$. The homozygosity estimator $\tilde{s}_H$ behaves similarly.

We wrote such a program, using code kindly provided by R. C. GRIFFITHS to calculate l_c. Figures 6 and 7 show examples of likelihood surfaces produced using our program. Notice that the surface is much flatter in the s than in the θ direction, indicating that the estimate of s is considerably more uncertain. Unfortunately, evaluating expression (20) is extremely time-consuming because it involves a large number of calculations of l_c, each of which is quite time-consuming. For example, for typical samples of size $n = 20$, evaluation took on the order of days to weeks on a workstation. We are therefore unable to compare the properties of this estimator with those given above using simulated data, except when n and θ are both very small, in which case the maximum-likelihood estimator performs very poorly. In fact, for small n it performs worse than the moment estimators described above. On the other hand, the method *is* fast enough to be used on real data sets. While we cannot asses its properties directly, background statistical theory suggests that it should be preferred to the moment estimators in such cases. In addition, the method gives the entire likelihood surface, which is, of course, much more informative than a point estimate. In fact, according to the likelihood principle

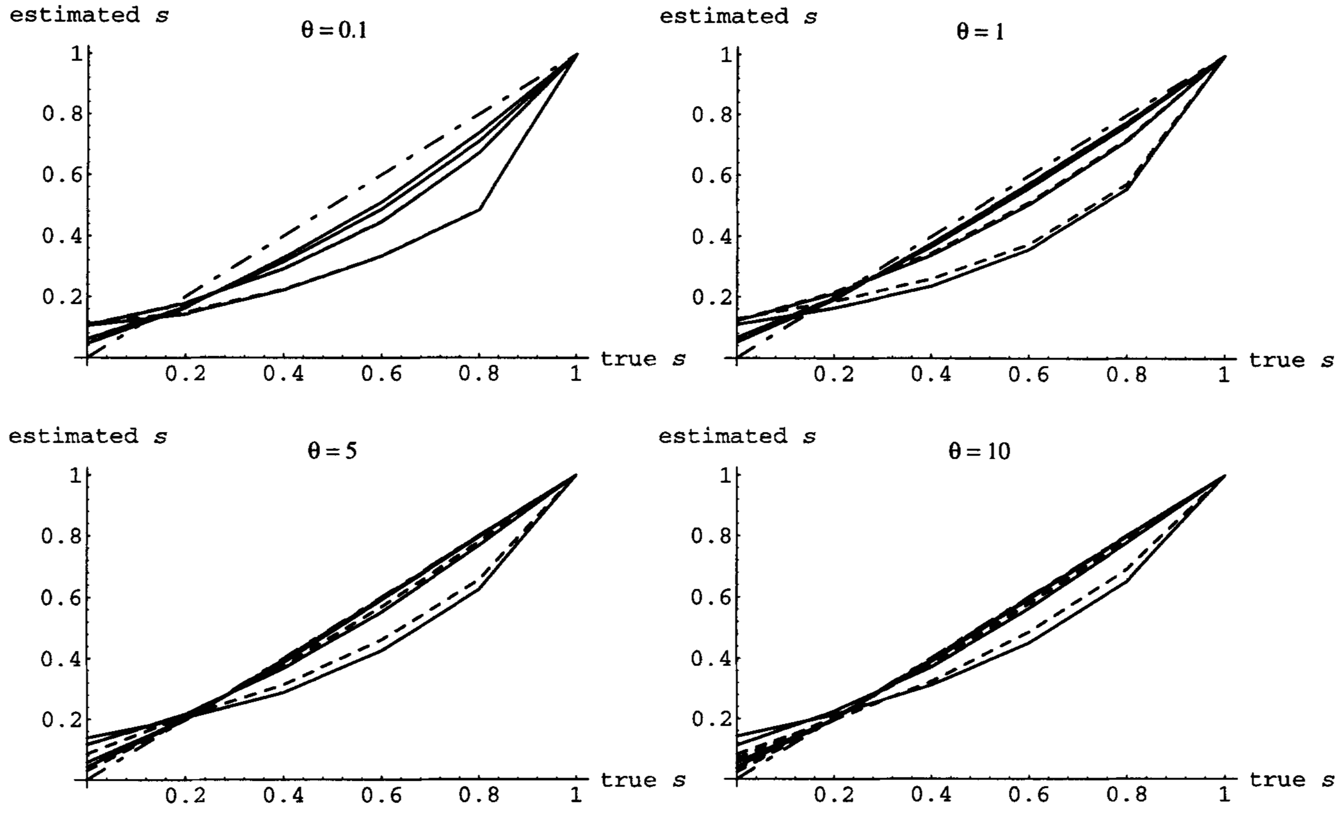

FIGURE 2.—Expectation of the moment estimators of s as a function of the actual s for various combinations of θ and n. The straight line gives the unbiased expectation; the solid lines, results for $\tilde{s}_S$; and the dashed lines, results for $\tilde{s}_H$. For both estimators, results are plotted for $n = 2$ (largest bias), 10, 50, and 100 (smallest bias).

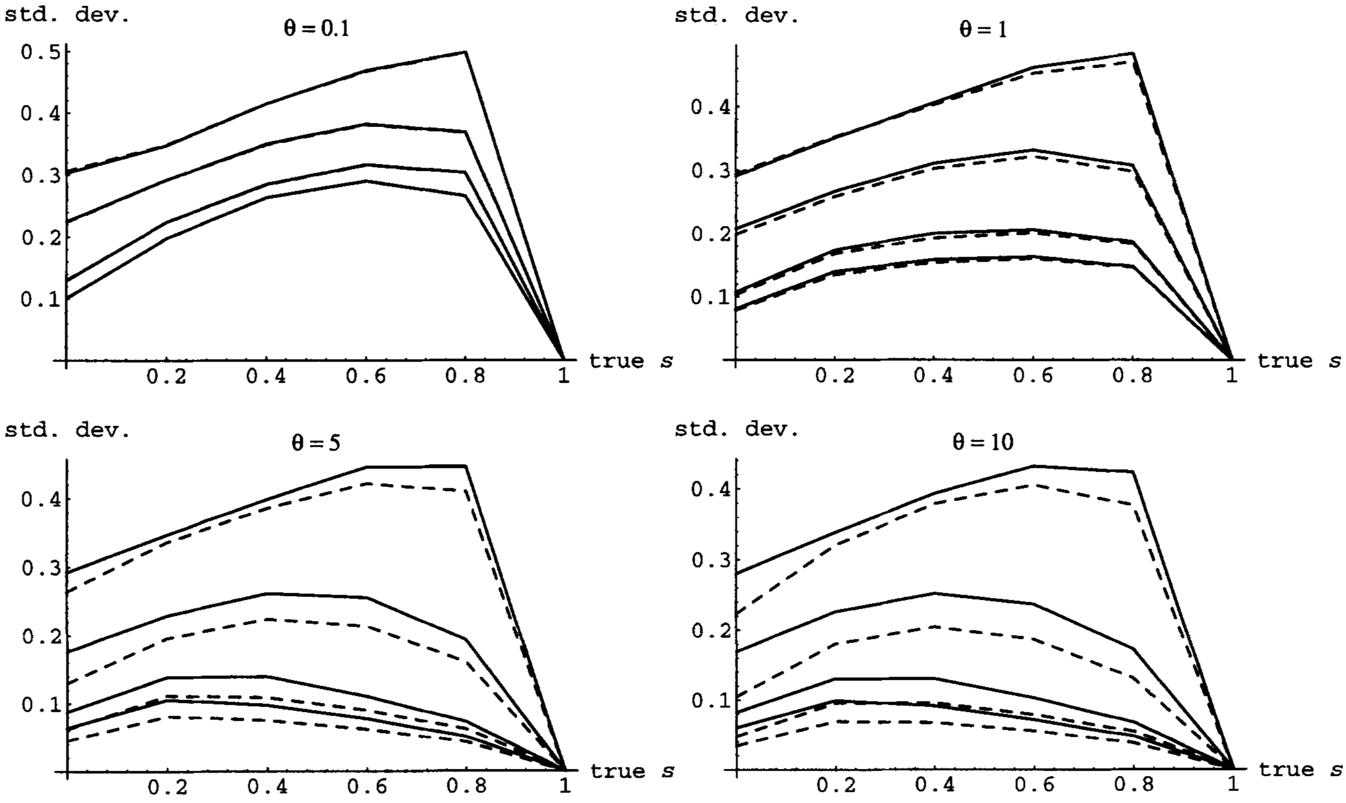

FIGURE 3.—Standard deviations of the moment estimators of s as functions of the actual s values for various combinations of θ and n. The solid lines give the results for $\tilde{s}_S$, and the dashed lines give those for $\tilde{s}_H$. For both estimators, results are plotted for $n = 2$ (largest values), 10, 50, and 100 (smallest values).

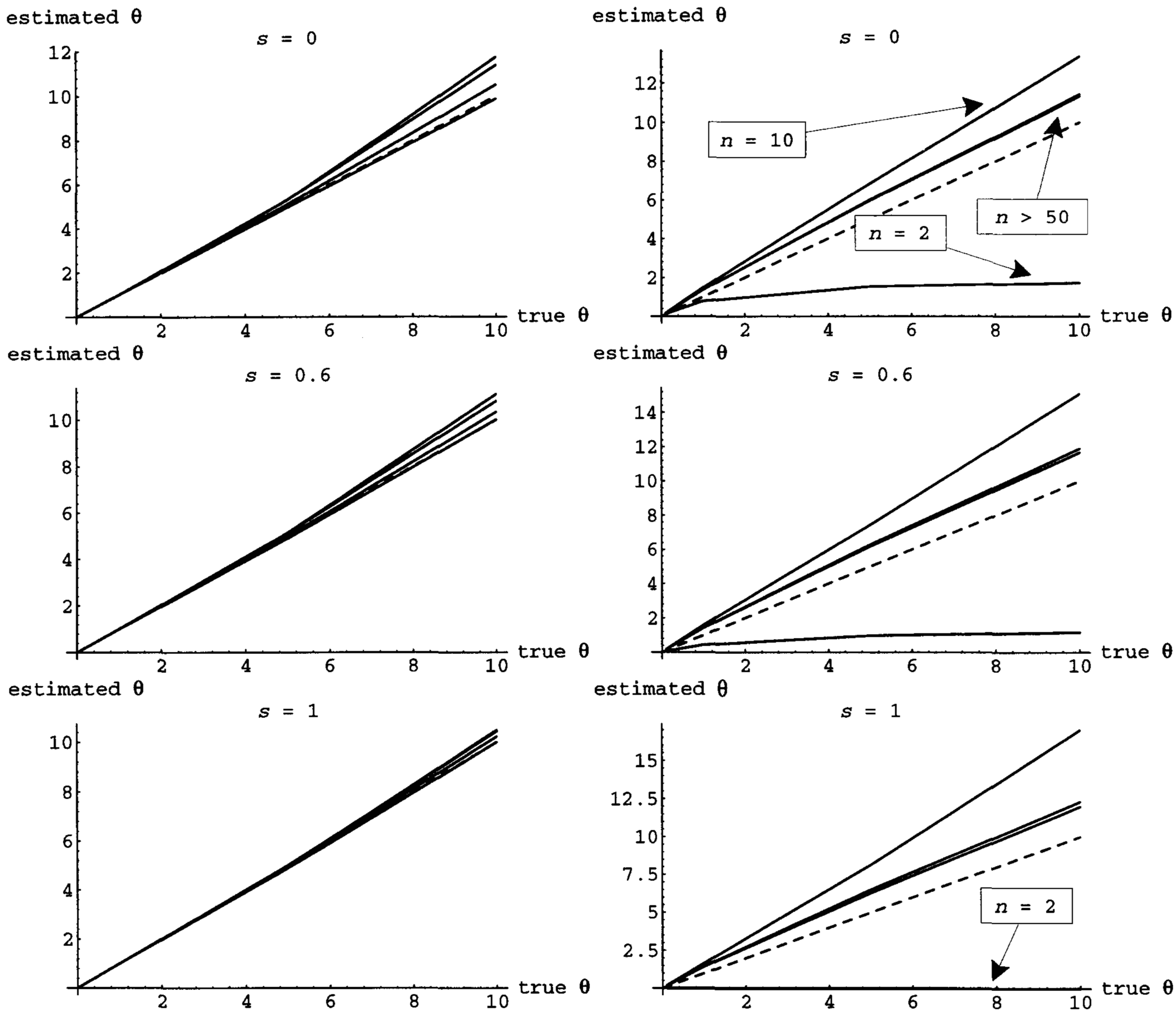

FIGURE 4.—Expectation of the moment estimators of θ as a function of the actual θ for various combinations of s and n. The left column gives the results for $\tilde{\theta}_S$ and the right those for $\tilde{\theta}_H$. The dashed line gives the unbiased expectation. Results for $s = 0.4$ and $s = 0.8$ are not shown but are very similar. For both estimators, results are plotted for $n = 2$ (largest bias), 10, 50, and 100 (least bias). Note that when $n = 2$, $\tilde{\theta}_H$ exhibits negative bias rather than positive bias (and that it equals zero when $n = 2$ and $s = 1$).

(BERGER and WOLPERT 1988), the likelihood surface contains all of the information about the parameters in the data.

DISCUSSION

The coalescent with selfing: We have shown that partial selfing can be incorporated into a coalescent framework without difficulty, essentially because the ancestral process decomposes into two different processes, a "slow" one that consists of common ancestor events among individuals in a population, and a "fast" one that consists of common ancestor events among alleles within individuals.

Estimation: Using this theory, we have been able to assess correctly the properties of the estimators for s and θ proposed in MILLIGAN (1996) and above. MILLIGAN estimated variance and bias by repeatedly drawing from the distribution of pairwise coalescent times without regard for the positive correlations induced by the underlying genealogy of a real sample. This is inappropriate, and we note the almost complete lack of correspondence, for example, between his estimates of the variance of $\tilde{s}_S$ (Figures 4 and 5, p. 624) and ours (Figure 3).

The reason that the estimators suggested in MILLIGAN (1996) and above do not provide much information about the long-term mating system is that the only direct information about s comes from the initial, "fast," coalescent events for alleles within individuals. These events are, of course, analogous to the deviations from Hardy-Weinberg proportions classically used to esti-

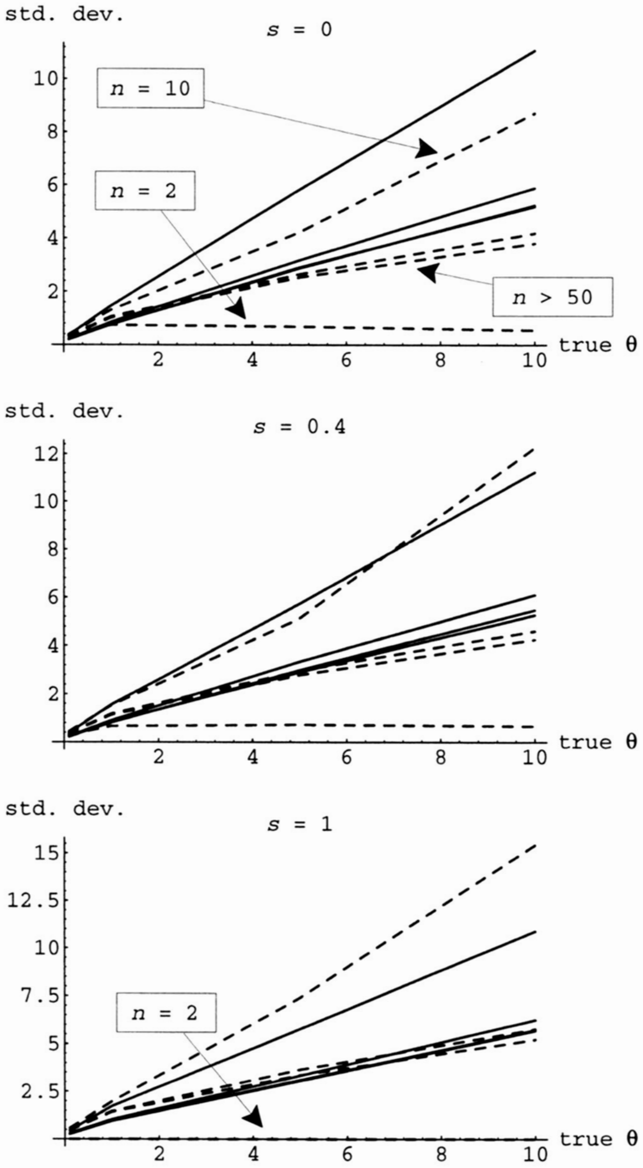

FIGURE 5.—Standard deviation of the moment estimators of θ as a function of the actual θ for the same combinations of s and n used in Figure 4. The solid lines give the results for $\tilde{\theta}_S$, and the dashed lines give those for $\tilde{\theta}_H$.

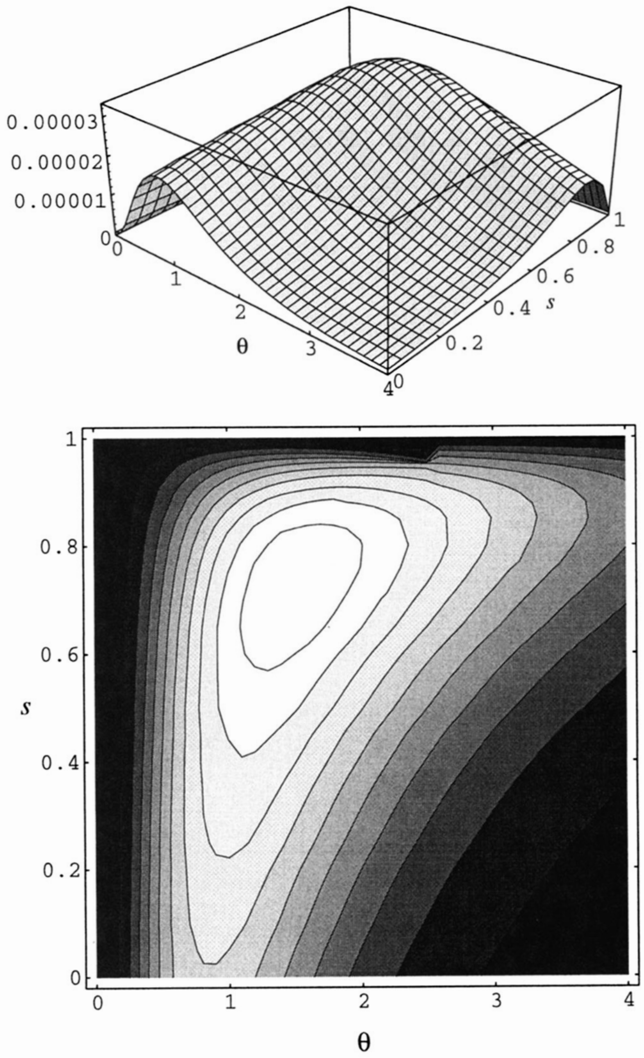

FIGURE 6.—Likelihood surface for s and θ for a simulated sample of size $n = 10$. The actual parameter values are $s = 0.9$, $\theta = 2$. For this sample, $\tilde{s}_S = 0.96$, $\tilde{\theta}_S = 2.72$, and $\tilde{s}_H = 0.93$, $\tilde{\theta}_H = 7.00$.

mate s (BROWN and ALLARD 1970; BROWN 1990). When we estimate s and θ jointly, we are estimating parameters of the two very different processes into which the coalescent with partial selfing decomposes: the "fast" process, which provides information about s only, and the "slow" one, which provides information about the parameter $\psi = (2 - s)\theta/2$, in which θ and s are confounded. This can be seen from the likelihood function (20), which loosely consists of a sum of "binomial-like" probabilities multiplied by coalescent likelihoods. Knowing this, the observed behavior of the estimators becomes easy to understand. Since information about s is available from the fast, binomial-like process, we might expect the variance of the estimate to decrease in the typical (n^{-1}) fashion of independent observations with increased sample size. This is precisely what we see in Figure 3, which gives the standard deviations of the moment estimators of s. The situation is rather different for estimation of θ. Even in the random-mating case, it is well known that increasing the sample size does not, in general, provide much extra information for estimating the scaled mutation rate, here ψ (DONNELLY and TAVARÉ 1995). Indeed, estimators based on pairwise measures, such as the moment estimators discussed here, are not necessarily even consistent. Estimation of θ is more difficult than estimation of ψ because of the confounding with s and the initial randomness induced by the fast process. Figure 5 shows that, as expected, the standard deviations of the moment estimators of θ do not decrease much with increased n.

We have shown how to derive the likelihood surface in s and θ. For moderate sample sizes, likelihood-based inference should be superior to the moment methods, although this is difficult to assess numerically. For small sample sizes, maximum-likelihood estimates are not necessarily reliable and may be worse than the moment

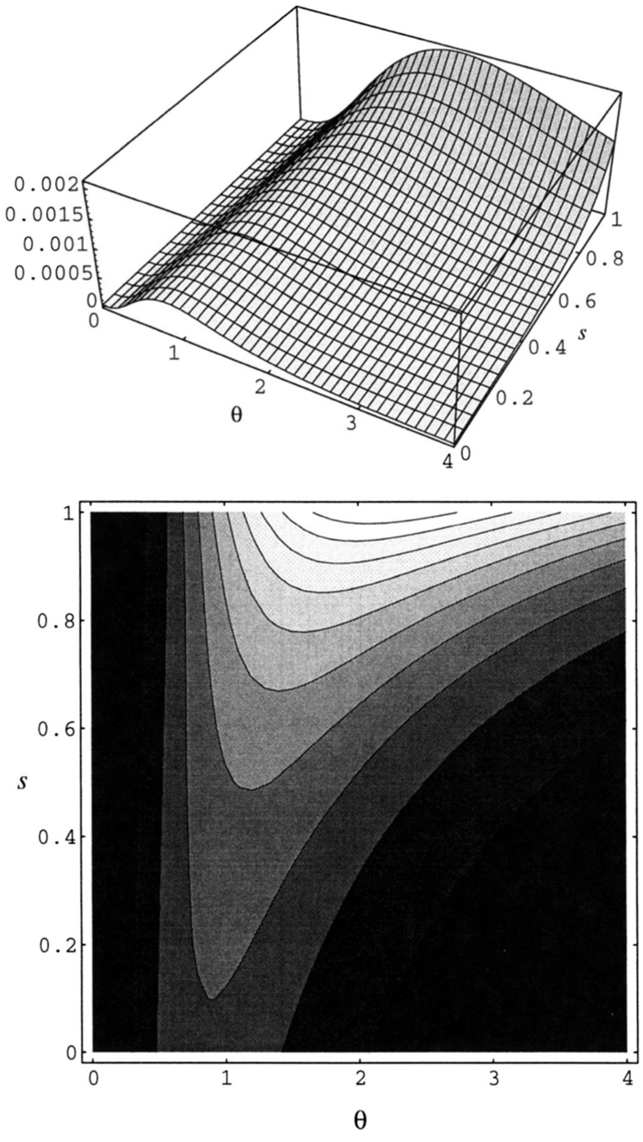

FIGURE 7.—Likelihood surface for s and θ for a simulated sample of size $n = 10$. The actual parameter values are $s = 0.9$, $\theta = 2$, as in Figure 6. For this sample, $\tilde{s}_S = 1$, $\tilde{\theta}_S = 3.73$, and $\tilde{s}_H = 1$, $\tilde{\theta}_H = 1.75$.

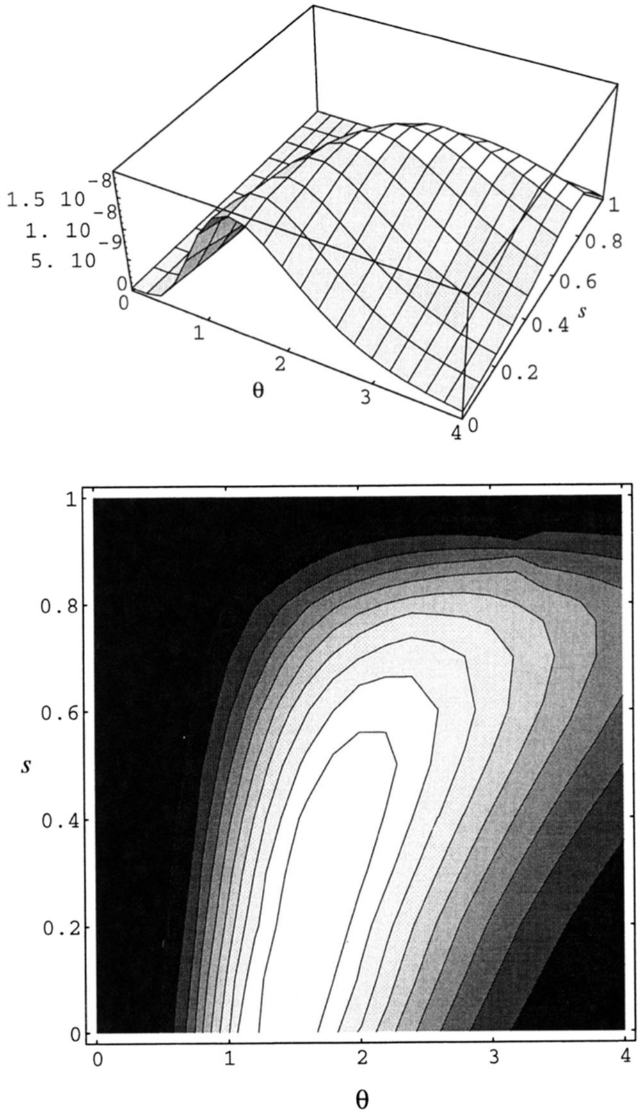

FIGURE 8.—Likelihood surface for s and θ for a simulated sample of size $n = 20$. The actual parameter values are again $s = 0.9$, $\theta = 2$. For this sample, $\tilde{s}_S = 0.98$, $\tilde{\theta}_S = 7.03$, and $\tilde{s}_H = 0.92$, $\tilde{\theta}_H = 4.56$.

estimates (Figures 6–8). For the likelihood surfaces we examined, estimation of θ with s assumed known is robust to the value of s, and conversely. Furthermore, reporting of the likelihood surface is considerably more informative than simply providing point estimates. For asymmetric surfaces (*e.g.*, Figure 6), even standard interval estimates could be quite misleading.

Robustness: MILLIGAN assumes a constant population size and notes that his results are insensitive to the exact value of that constant size. It does not follow, as he claims (p. 620 and p. 626), that the results are insensitive to the assumption of constant population size. In fact, an extension of the argument given above should show that the standard theory for the coalescent with varying population sizes (DONNELLY and TAVARÉ 1995) (with modifications analogous to those above) will apply to the coalescent with partial selfing as well. We have not assessed the estimators in this more general setting, but dependence on the population size process would certainly be expected.

MILLIGAN also argues that his method estimates the "long-term mating system," rather than being "based on the segregation of alleles during a single generation of mating." In fact, the opposite is true. Most of the information about s comes from the association of alleles within individuals, and this information only reflects the last few generations. This is easily seen, for instance, by considering how data from the coalescent with partial selfing is simulated (see above). Going backward in time, pairs of alleles sampled within individuals either coalesce instantly, or become part of a standard, haploid coalescent with an appropriate timescale. This point is further illustrated by Figure 9, where the expectation of $\tilde{s}_S$ is plotted against the actual value of s under the last few generations. It is clear that the

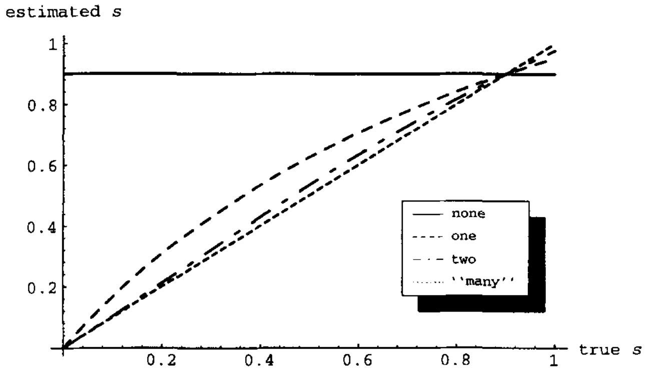

FIGURE 9.—Sensitivity of the estimated selfing rate to changes in s. The "long-term" value of s is 0.9, and curves show the expectation of $\hat{s}_S$ given that the actual value of s was that on the abscissa in the preceding few generations, where "few" is 0, 1, 2, or "many" (*i.e.*, when the long-term value is the one on the abscissa).

estimated s almost exclusively reflects the value of s in the preceding generation, and that the "long-term" mating system is largely irrelevant to the estimate. Note that the same argument applies to the other estimators discussed above.

It can be shown that in a model in which the selfing rate varies independently from generation to generation the behavior of the slow process in the coalescent with selfing is as described above, with s replaced by the mean of the distribution of selfing rates. The behavior of the fast process, and hence of the estimators, on the other hand, is quite sensitive to varying selfing rates. While the actual value of an estimator for s will be heavily dependent on the actual values of the selfing rate over the last few years, the sampling properties will depend sensitively on the entire distribution of possible values for the selfing rate over the same period.

All results derived here assume a single neutral locus. It is worth pointing out that selection on linked loci may have a very strong effect on the variability at neutral loci, either in reducing it through processes such as background selection (CHARLESWORTH *et al.* 1993) and selective sweeps (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989), or in increasing it through some form of balancing selection (STROBECK 1983; HUDSON and KAPLAN 1988; KAPLAN *et al.* 1988; NORDBORG *et al.* 1996), and that all these effects will be stronger under selfing (NORDBORG *et al.* 1996).

Conclusion: Coalescent-based estimates do not provide a magic bullet when it comes to estimating the selfing rate. As we have seen, the problem of recent fluctuations in the degree of selfing is in no sense avoided. Furthermore, collecting data to estimate both s and θ involves a contradiction: to estimate s, we want simple data (the genotype) from large number of individuals, whereas to estimate θ we need detailed data (a sequence or several sequences) from just a few individuals (PLUZHNIKOV and DONNELLY 1996). Because DNA sequence data provide more certain information about the genotype than do data based on classical markers such as allozymes, it is certainly preferable, *ceteris paribus* (in particular, the sample size needs to be roughly similar), but if moment estimators of s are of primary interest, not much is gained.

We thank BRIAN CHARLESWORTH, DEBORAH CHARLESWORTH, MARTIN MÖHLE, and especially TOM NAGYLAKI for helpful discussions and comments on the manuscript. M.N. was supported by U.S. National Science Foundation grant DEB 92-17683 to DEBORAH CHARLESWORTH. P.D. was supported in part by U.S. National Science Foundation grant DMS 95-05129, the Block Fund of the University of Chicago, and U.K. Engineering and Physical Sciences Research Council Advanced Fellowship B/AF/1255.

LITERATURE CITED

BERGER, J. O., and R. L. WOLPERT, 1988 *The Likelihood Principle.* Institute of Mathematical Statistics, Hayward, CA.

BROWN, A. H. D., 1990 Genetic characterization of plant mating systems, pp. 145–162 in *Plant Population Genetics, Breeding, and Genetic Resources,* edited by A. H. D. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WEIR. Sinauer Associates, Sunderland, MA.

BROWN, A. H. D., and R. W. ALLARD, 1970 Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. Genetics **66:** 133–145.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

DONNELLY, P., and S. TAVARÉ, 1995 Coalescents and genealogical structure under neutrality. Annu. Rev. Genet. **29:** 401–421.

GRIFFITHS, R. C., and S. TAVARÉ, 1994a Ancestral inference in population genetics. Stat. Sci. **9:** 307–319.

GRIFFITHS, R. C., and S. TAVARÉ, 1994b Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. B **344:** 403–10.

GRIFFITHS, R. C., and S. TAVARÉ, 1994c Simulating probability distributions in the coalescent. Theor. Popul. Biol. **46:** 131–159.

HALDANE, J. B. S., 1924 A mathematical theory of natural and artificial selection. Part II. Proc. Camb. Phil. Soc., Biol. Sci. **1:** 158–163.

HUDSON, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–43 in *Oxford Surveys in Evolutionary Biology,* edited by D. FUTUYMA and J. ANTONOVICS. Oxford University Press, Oxford.

HUDSON, R. R., and N. L. KAPLAN, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

KAPLAN, N. L., T. DARDEN and R. R. HUDSON, 1988 The coalescent process in models with selection. Genetics **120:** 819–829.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking" effect revisited. Genetics **123:** 887–899.

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

KINGMAN, J. F. C., 1982b Exchangeability and the evolution of large populations, pp. 97–112 in *Exchangeability in Probability and Statistics,* edited by G. KOCH and F. SPIZZICHINO. North-Holland Publishing Company, Amsterdam.

KINGMAN, J. F. C., 1982c On the genealogy of large populations. J. Appl. Prob. **19A:** 27–43.

KUHNER, M. K., J. YAMATO and J. FELSENSTEIN, 1995 Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics **140:** 1421–1430.

LI, C. C., 1955 *Population Genetics.* University of Chicago Press, Chicago.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitchhiking effect of a favourable gene. Genet. Res. **23:** 23–35.

MILLIGAN, B. G., 1996 Estimating long-term mating systems using DNA sequences. Genetics **142:** 619–627.

MÖHLE, M., 1996 Coalescent results for diploid population models and the coalescent with selfing. Technical Report 433, Department of Statistics, The University of Chicago, Chicago.

NORDBORG, M., B. CHARLESWORTH and D. CHARLESWORTH, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. R. Soc. Lond. B **263:** 1033–1039.

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

POLLAK, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117:** 353–360.

STEWART, F. M., 1976 Variability in the amount of heterozygosity maintained by neutral mutations. Theor. Popul. Biol. **9:** 188–201.

STROBECK, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics **103:** 545–555.

TAVARÉ, S., 1984 Line-of-descent and genealogical processes, and their applications in population genetic models. Theor. Popul. Biol. **26:** 119–164.

WRIGHT, S., 1969 *Evolution and the Genetics of Populations,* Vol. 2. University of Chicago Press, Chicago.

Communicating editor: R. R. HUDSON