

Coalescent Times and Patterns of Genetic Diversity in Species with Facultative Sex: Effects of Gene Conversion, Population Structure, and Heterogeneity

Matthew Hartfield,^{*,†,1} Stephen I. Wright,^{*} and Aneil F. Agrawal^{*}

^{*}Department of Ecology and Evolutionary Biology, University of Toronto, Ontario, Canada M5S 3B2, and [†]Bioinformatics Research Centre, University of Aarhus, 8000C Aarhus, Denmark

ABSTRACT Many diploid organisms undergo facultative sexual reproduction. However, little is currently known concerning the distribution of neutral genetic variation among facultative sexual organisms except in very simple cases. Understanding this distribution is important when making inferences about rates of sexual reproduction, effective population size, and demographic history. Here we extend coalescent theory in diploids with facultative sex to consider gene conversion, selfing, population subdivision, and temporal and spatial heterogeneity in rates of sex. In addition to analytical results for two-sample coalescent times, we outline a coalescent algorithm that accommodates the complexities arising from partial sex; this algorithm can be used to generate multisample coalescent distributions. A key result is that when sex is rare, gene conversion becomes a significant force in reducing diversity within individuals. This can reduce genomic signatures of infrequent sex (*i.e.*, elevated within-individual allelic sequence divergence) or entirely reverse the predicted patterns. These models offer improved methods for assessing null patterns of molecular variation in facultative sexual organisms.

KEYWORDS facultative sex; migration; gene conversion; genealogies; heterogeneity in sex

FACULTATIVE sexual species, which have both parthenogenetic and sexual stages in their life cycles, are widespread in nature. They have been a focus for empirical studies of the role of sex in evolution (Goddard *et al.* 2005; D'Souza and Michiels 2008; King *et al.* 2009; Morran *et al.* 2009; Becks and Agrawal 2010; Hojsgaard and Hörandl 2015). The biology of facultative sexual organisms is also a research field with broad applications because many organisms of evolutionary, medical, and agricultural importance undergo both sexual and asexual reproduction (Chang *et al.* 2013; Ankarklev *et al.* 2014; Hand and Koltunow 2014; Yoshida *et al.* 2014). However, our under-

standing of the structuring of genetic diversity in such organisms remains limited.

For effective population genetic analysis of facultative sexual individuals to be made, there needs to be a theoretical basis for predicting how neutral diversity is affected under various demographic and reproductive scenarios. This is so that the two are not confounded, especially because they are strongly intertwined [reviewed by Halkett *et al.* (2005) and Arnaud-Haond *et al.* (2007)]. A classic prediction for organisms with very low rates of sex is that owing to the resulting lack of segregation, diploid asexual organisms tend to accumulate extensive diversity within individuals, because the two alleles from a single individual remain isolated by descent in the absence of sex. This phenomenon has been labeled the *Meselson effect* because it was used by Meselson and colleagues as evidence for a lack of sex in bdelloid rotifers (Mark Welch and Meselson 2000; Butlin 2002) [but see Mark Welch *et al.* (2008) for an alternative explanation for their original findings].

Copyright © 2016 by the Genetics Society of America
doi: 10.1534/genetics.115.178004

Manuscript received May 8, 2015; accepted for publication November 16, 2015;
published Early Online November 18, 2015.

Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178004/-/DC1

¹Corresponding author: Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, Ontario, Canada M5S 3B2.

E-mail: matthew.hartfield@utoronto.ca

One theoretical approach to quantify this effect has been to derive how traditional diversity measures, such as the effective population size N_e and F -statistics, are affected when sex is rare (Balloux *et al.* 2003; Yonezawa *et al.* 2004; de Meeùs and Balloux 2005). Such studies are based on pairwise sample comparisons, with heightened within-individual diversity not becoming apparent unless sex is extremely rare (with frequency less than $1/N$, for N the total population size). Combining these summary statistics with information on clonal or genotypic diversity can improve inference when rates of sex are not too low (Balloux *et al.* 2003; Halkett *et al.* 2005; Arnaud-Haond *et al.* 2007).

An alternative approach is to analyze polymorphisms using coalescent theory to recreate the possible genealogies of samples (Kingman 1982; Wakeley 2009). Basic calculations were carried out by Brookfield (1992) and Burt *et al.* (1996) to determine how partial sex affects diversity in DNA fingerprinting analyses and isolates of a human pathogen, respectively. More complete analyses then were performed by Bengtsson (2003) and Ceplitis (2003) to determine expected coalescent times for pairs of samples taken either from the same or different individuals. These analyses confirmed that the rate of sex has to be less than the reciprocal of the population size for mean coalescent times to be substantially altered compared with the classic case of obligate sex.

With the advent of high-throughput genome sequencing technologies, there are growing opportunities to characterize both the demographic and reproductive histories of partially sexual populations. However, formal theory does not yet exist for genealogies of facultative sexual organisms that take into account demographic factors, gene conversion, and temporal and spatial heterogeneity in the rate of sexual reproduction. The potential importance of gene conversion is especially pertinent because there is growing recognition that this process may be an important homogenizing force in diploid asexual populations (Crease and Lynch 1991; Schön *et al.* 1998; Normark 1999; Schön and Martens 2003; Schaefer *et al.* 2006; Flot *et al.* 2013). In addition, previous theoretical studies have not accounted for cases where the rates of sex change in either space or time. It is well known that several organisms adjust their rate of sex in a stressful or dense environment. This mechanism is well supported both theoretically (Redfield 1988; Hadany and Otto 2007) and empirically (Grishkan *et al.* 2003; Snell *et al.* 2006; Levin and King 2013), but the overall consequences for patterns of genetic diversity are not clear. Previously published theory also suffers from ambiguities or typographical errors (explained later) that make interpretation of existing results difficult. In addition, no algorithm yet exists for how to recreate genealogies for more than two sequences with a range of sexual reproduction rates, which is essential for simulating the distribution of neutral variation under complex scenarios.

Here we outline theory to rectify this issue. We rederive equations of the coalescent time for pairs of samples taken from either different or the same individuals using a struc-

tured coalescent framework (Nordborg 1997), clarifying discrepancies that appear between Bengtsson (2003) and Ceplitis (2003). We further show how additional phenomena, including self-fertilization, gene conversion, population subdivision, and heterogeneity in rates of sex, can be included in the coalescent with partial sex. These results can be used to derive estimators for mutation, migration, and sex rates based on mean pairwise diversity. These initial analyses then will inform on how to create a simulation program in which there exists an arbitrary rate of sex that can be used to efficiently create genealogies for more than two chromosomes.

Materials and Methods

The basic model

We begin by considering a population with an arbitrary rate of sex but is otherwise ideal. Bengtsson (2003) and Ceplitis (2003) derived mean coalescent times for two alleles (two-sample mean coalescent times) from this basic model. However, we reexamine this model to clarify discrepancies between the earlier studies, as well as to lay the groundwork for further extensions.

Model outline

We consider a diploid Wright-Fisher population of size N individuals so that there are $2N$ gene copies in total. Then σ is the rate of sex; that is, with probability σ , an offspring is produced sexually, otherwise, it is produced asexually. For a specific individual, one can imagine that each of its gene copies inhabits a separate genomic deme within the individual, pertaining to the left- and right-hand allele copy respectively (Ceplitis 2003). We choose two of these gene copies and track their genealogy back in time until they reach their common ancestor. These allele pairs can be sampled from either different individuals or the same individual. With obligate sex, this distinction is irrelevant owing to the random segregation and union of gametes. However, if sex is infrequent, then pairs of alleles can remain “packaged” within an individual lineage as a left and right haplotype for long periods of time. This structuring and the low level of exchange between haplotypes when sex is infrequent create a fundamental difference from the standard coalescent.

Two sampled alleles can be found in one of three states: (1) one allele in each of two separate individuals, (2) two separate alleles within the same individual, and (3) coalesced. Ceplitis (2003) argued that it is necessary to divide the first state into two separate states: both alleles in the same deme (either the left or right demes of separate individuals), or one allele in the left deme and one in the right deme. However, it can be shown that this partitioning is unnecessary as long as the appropriate transition probabilities are applied, which average over these possible states. We calculate the mean coalescent time by first determining the transition probabilities among the three states mentioned earlier. We assume that N is large, so terms of order $1/N^2$ and higher are discarded.

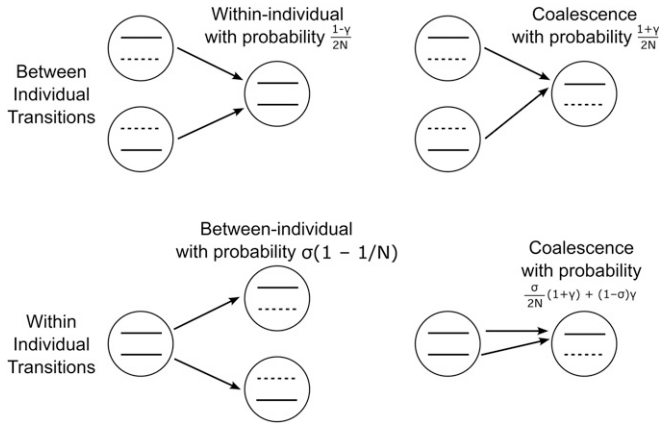


Figure 1 A schematic of the transition probabilities between states for the coalescent with infrequent sex (with rate σ) and gene conversion (rate γ). Only those states where the configurations have changed are shown. Focal samples are shown as solid black lines, whereas dashed black lines represent placeholder homologous alleles that are not sampled. Transitions from the within-individual state to the between-individual state require sex, whereas the reverse transition can occur via sexual or asexual reproduction.

The transition matrix for the process is

$$\mathbf{T} = \begin{bmatrix} 1 - \frac{1}{N} & \frac{1}{2N} & \frac{1}{2N} \\ \sigma \left(1 - \frac{1}{N}\right) & 1 - \sigma \left(1 - \frac{1}{2N}\right) & \frac{\sigma}{2N} \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Further details of all transitions and analysis are outlined in Supporting Information, File S2A; a schematic of the basic transitions is shown in Figure 1.

To find the expected coalescent times, we follow the method of Slatkin (1991). Note that the 2×2 submatrix \mathbf{G} , representing the top-left corner of \mathbf{T} (Equation 1) is a matrix of noncoalescence; that is, it denotes the probability that two samples do not coalesce in a single generation. The probability that two alleles from either state have not coalesced t generations in the past is given by $\mathbf{a}(t) = \mathbf{G}^t \mathbf{a}(0)$, where $\mathbf{a}(0) = [1, 1]^T$ (here the T denotes vector transposition).

Conversely, the probability that two alleles *have* coalesced by time t is $\mathbf{P}(t) = \mathbf{a}(t-1) - \mathbf{a}(t)$. For now, we will focus on the mean coalescent time; Slatkin (1991) showed that it can be calculated by

$$\bar{\mathbf{t}} = (\mathbf{I} - \mathbf{G})^{-1} \cdot \mathbf{a}(0) \quad (2)$$

where $\bar{\mathbf{t}} = (E[t_w], E[t_b])$ is the vector of mean between- and within-individual mean coalescent times, and \mathbf{I} is the identity matrix. The noncoalescence matrix \mathbf{G} is crucial for determining the expected coalescent times. It can be decomposed into the relative contributions of sexual and asexual reproduction as follows:

$$\mathbf{G} = \begin{bmatrix} 1 - \frac{1}{N} & \frac{1}{2N} \\ \sigma \left(1 - \frac{1}{N}\right) & 1 - \sigma \left(1 - \frac{1}{2N}\right) \end{bmatrix} \\ = \sigma \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} \\ 1 - \frac{1}{N} & \frac{1}{2N} \end{pmatrix} + (1 - \sigma) \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} \\ 0 & 1 \end{pmatrix} \quad (3)$$

This decomposition makes it easier to see how other complexities (e.g., selfing and gene conversion) can be included in this framework.

To make subsequent calculations more convenient, we shift to working in continuous time by rescaling time by the population size, *i.e.*, $\tau = 2Nt$. A formal way to rescale in continuous time is to write $\mathbf{T} = \mathbf{A} + \mathbf{B}/2N$; *i.e.*, \mathbf{A} captures transitions over “fast” timescales [which occur with probability $\mathcal{O}(1)$], and \mathbf{B} captures transitions over “slow” timescales [which occur with probability $\mathcal{O}(1/N)$]. Möhle (1998) proved that if the matrix is written in this manner, the continuous time analogue of the transition matrix \mathbf{T} , which we denote \mathbf{Q} , can be written in the form

$$\mathbf{Q} = \mathbf{P}e^{\mathbf{F}\tau} \quad (4)$$

where

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

and

$$\mathbf{F} = \mathbf{PB}\mathbf{P}$$

If we partition \mathbf{T} (Equation 1) while assuming that $\sigma = \mathcal{O}(1)$, then we can show that the continuous-time transition matrix tends to the obligate sex case (File S1A). If $\sigma = \mathcal{O}(1/N)$, then \mathbf{A} is equal to the identity matrix; thus $\mathbf{F} = \mathbf{B}$, where

$$\mathbf{B} = \begin{pmatrix} -2 & 1 & 1 \\ \Omega & -\Omega & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (5)$$

Here $\Omega = 2N\sigma$, which is equal to twice the expected number of sexually reproduced offspring produced per generation. This matrix shows that with rare sex, two samples from the same individual (the state represented by the second row of matrix \mathbf{B}) cannot coalesce. Instead, the samples have to be placed within different individuals before coalescence can occur, which requires at least one bout of sex.

Where possible, subsequent analyses will be performed using the discrete transition matrix \mathbf{T} , and the resulting expressions are derived assuming that N is large and are expressed in units of $2N$ generations; *i.e.*, using the coalescent timescale ($\tau = 2Nt$) and compound parameters [e.g.,

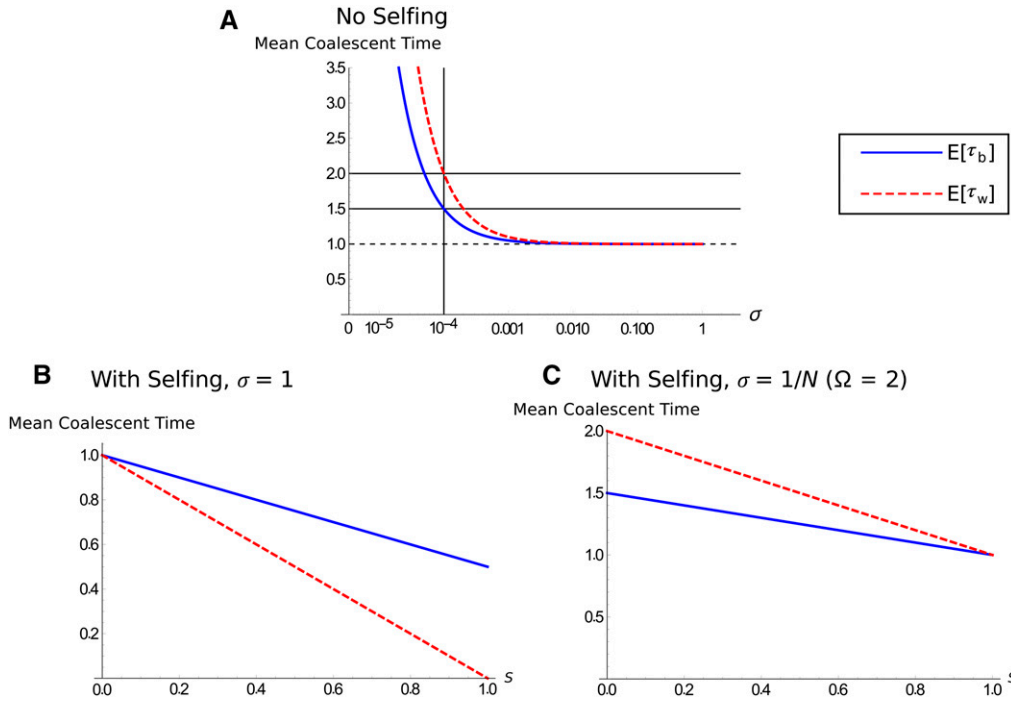


Figure 2 Expected coalescent times with different reproductive systems. The population size for all plots is $N = 10,000$. (A) Expected mean coalescent time according to Equation 6 in coalescent time units as a function of sex σ . The blue line represents between-individual mean coalescent time $E[\tau_b]$, and the red line is within-individual mean coalescent time $E[\tau_w]$. The vertical line represents $\sigma = 1/N$ ($\Omega = 2$), and the horizontal solid lines show the predicted mean coalescent times for this value ($E[\tau_b] \approx 1.5$; $E[\tau_w] \approx 2$ on the coalescent timescale), whereas the horizontal dashed line shows $E[\tau] = 1$. (B and C) Expected coalescent time if selfing is also included, according to Equation 9. Line colors are the same as in A, with (B) $\sigma = 1$ or (C) $\sigma = 1/N$ ($\Omega = 2$).

$\Omega = 2N\sigma = \mathcal{O}(1)$. Nevertheless, we will sometimes use unscaled parameters when investigating rates of sex where $\sigma = \mathcal{O}(1)$.

Mean coalescent times: Using Equations 2 and 3, we obtain the expected coalescent times for two samples taken from different individuals $E[\tau_b]$ and from the same individual $E[\tau_w]$ in units of $2N$ generations

$$\begin{aligned} E[\tau_b] &= 1 + \frac{1}{\Omega} \\ E[\tau_w] &= 1 + \frac{2}{\Omega} \end{aligned} \quad (6)$$

Results are similar in the unscaled case, so they generally hold even if $\Omega \gg \mathcal{O}(1)$. Ceplitis (2003, Equation 2) derived equivalent equations; Bengtsson (2003, Equations 1 and 2) presented results in terms of the eigenvalues of \mathbf{G} , although there appears to be typographical errors in the way these eigenvalues are presented. File S1A shows how equivalent results can be derived using the eigenvalues of \mathbf{G} and how they compare with what is presented in Bengtsson (2003).

Visual inspection of Equation 6 shows some key properties of the within- and between-individual mean coalescent times. They verify that the rate of sex will significantly affect the expected coalescent time only if it is very rare, at least $\mathcal{O}(1/N)$, as found in previous studies. Second, the within-individual expected coalescent time is greater than that between individuals, formalizing the Meselson effect (Mark Welch and Meselson 2000; Butlin 2002). Two samples from within a single individual are, by definition, members of different left/right sides. Sex is therefore required to put them in the same side before coalescence can occur (on average,

two sex events are needed). Only 50% of two-allele samples from different individuals will be on different left/right sides, so the increase in mean coalescent time resulting from low rates of sex between individual samples is only half as large.

Figure 2A demonstrates how the mean coalescent times increase rapidly as σ becomes significantly less than $1/N$. If we set $\sigma = 1/N$ (i.e., $\Omega = 2$) in Equation 6, then $E[\tau_b] = 1.5$ and $E[\tau_w] = 2$. That is, the between-individual mean coalescent times are 1.5 times higher than in Kingman (obligate sexual) coalescent, and the within-individual mean coalescent times are twofold higher. We can also use analytical methods to determine the variance and probability distribution of the coalescent process; this analysis will be left until gene conversion is added so that its effect with rare sex can be quantified.

Multisequence coalescent simulations

The two-sample results are useful for clarifying how neutral drift should operate in facultative sexual populations. However, when analyzing empirical data, one would normally obtain many samples from a population and then estimate neutral diversity based on those samples. For obligate sexual organisms, coalescent simulations have proved to be essential for modeling expected neutral diversity levels caused by complex demographic scenarios, such as population bottlenecks or expansions (Hudson 2002). Equivalent analysis on sequence data from facultative sexual species is not possible using existing coalescent simulation software. This is so because gene samples remain paired within individual lineages if sex is rare, a type of genetic isolation not considered in existing simulation packages. Therefore, it is necessary to create a new coalescent algorithm to account for partial rates

Table 1 Possible events in facultative sexual coalescent simulation

k_j samples split by sex in deme j : $p(k_j x_j, \sigma_j) = \binom{x_j}{k_j} \sigma_j^{k_j} (1-\sigma_j)^{x_j-k_j}$			
Event	Description	Configuration change in a deme	Probability
1	New single samples remain single	$(x_j, y_j) \rightarrow (x_j - k_j, y_j + 2k_j)$	1—all other probabilities
2	One of the paired samples is re-created	$(x_j, y_j) \rightarrow [x_j - k_j + 1, y_j + 2(k_j - 1)]$	$\frac{k_j y_j}{N_d} + \binom{2k_j}{2} \frac{1}{2N_d}$
3	A new single sample coalesces with an existing single sample	$(x_j, y_j) \rightarrow (x_j - k_j + 1, y_j + 2k_j - 1)$	$\frac{k_j y_j}{N_d} + \binom{2k_j}{2} \frac{1}{2N_d}$
4	A new single sample coalesces with an existing paired sample	$(x_j, y_j) \rightarrow (x_j - k_j, y_j + 2k_j - 1)$	$\frac{2k_j(x_j - k_j)}{N_d}$
5	Two existing single samples form a new paired sample	$(x_j, y_j) \rightarrow [x_j - k_j + 1, y_j + 2(k_j - 1)]$	$\binom{y_j}{2} \frac{1}{2N_d}$
6	Two existing single samples coalesce	$(x_j, y_j) \rightarrow [x_j - k_j, y_j + 2k_j - 1]$	$\binom{y_j}{2} \frac{1}{2N_d}$
7	Two remaining paired samples doubly coalesce asexually	$(x_j, y_j) \rightarrow (x_j - k_j - 1, y_j + 2k_j)$	$\binom{x_j - k_j}{2} \frac{1}{N_d}$
8	One remaining paired sample coalesces with a single sample	$(x_j, y_j) \rightarrow (x_j - k_j, y_j + 2k_j - 1)$	$\frac{(x_j - k_j) y_j}{N_d}$
9	Paired sample coalesces via gene conversion	$(x_j, y_j) \rightarrow (x_j - k_j - 1, y_j + 2k_j + 1)$	$\gamma(x_j - k_j)$
10	A sample migrates to another deme	$(x_j, y_j) \rightarrow (x_j - k_j, y_j + 2k_j)$ (before migration)	$m(x_j + y_j + k_j)$

The listed probabilities for events 1–4 are conditional on sex having occurred and are otherwise zero.

of sex. Our goal is to develop an algorithm that would remain valid under all rates of sex, including high rates [as opposed to purely under the low sex limit $\Omega = 2N\sigma \approx \mathcal{O}(1)$].

The simulation procedure considers arbitrary rates of sex, as well as gene conversion and migration via an island model (which will be treated analytically in later sections). The main challenge is to consider how samples change state over time (through coalescence, migration, splitting, or forming pairs within individuals) while allowing an arbitrary rate of sex. Most coalescent simulation algorithms assume that each possible event is rare so that when an event does arise, only one event occurs. However, if there are several paired samples and the rate of sex is not very low ($N_T\sigma \gg 1$, for N_T the total population size), many paired samples can split up into different individuals in a single generation; furthermore, an additional action (e.g., coalescence) is then possible in that same generation. The simulation accounts for this issue by determining the probability that a specific number of paired samples are split by sex in a generation and then designating what possible actions subsequently occur, given these splits.

We formalize this algorithm as follows: let there be x_j paired samples and y_j single samples (i.e., where only one of the two alleles is ancestral to the sample) in deme j , for $j \in [1, \dots, d]$; we can also denote the configuration for deme j as (x_j, y_j) . Let $x_T = \sum_{j=1}^d x_j$ and $y_T = \sum_{j=1}^d y_j$; then the total number of ancestral allele copies over all demes is $n = 2x_T + y_T$, with the configuration for all demes denoted (x_T, y_T) . If the rate of sex in deme j is σ_j (to account for spatial heterogeneity in rates of sex), then the probability that k_j of x_j samples are split by sex, with each part of the paired sample assigned to different adults back in time, is drawn from a binomial distribution

$$p(k_j|x_j, \sigma_j) = \binom{x_j}{k_j} \sigma_j^{k_j} (1-\sigma_j)^{x_j-k_j}$$

This event changes the configuration for all demes by producing $2k_j$ new single samples in that deme: $(x_j, y_j) \rightarrow (x_j - k_j, y_j + 2k_j)$. The k_j splits are chosen independently in each deme, without replacement, from the existing paired samples. This process creates a vector $\mathbf{k} = (k_1, k_2, \dots, k_d)$ of paired samples that split in each deme.

Once this intermediate step is complete, there can be 10 other outcomes, which are outlined in Table 1. All additional outcomes (events 2–9 in Table 1) involve events that are of order $1/N_d$ and are thus assumed to be rare (including the gene conversion rate γ and the migration rate m), so only up to one additional event can arise. Hence, when an event occurs, it involves at least one paired sample splitting by sex, followed by an additional event. Alternatively, no sex occurs, in which case only events 5–10 in Table 1 can happen.

Building on this logic, we unite the preceding actions into a coalescent algorithm as follows:

1. In a single generation, an event occurs if either at least one paired sample is split by sex or no sex occurs, but one of events 5–10 in Table 1 happens regardless. The probability of no paired samples splitting via sex in any of the demes is $p(k_T = 0) = \prod_{j=1}^d (1-\sigma_j)^{x_j}$. If no sex occurs, we need to consider the possibility that one of events 5–10 in Table 1 has happened. Let p_{E0} be the sum of the probabilities of events 5–10 (given that $k_T = 0$) across all demes represented by the sample.
2. The total probability of any event occurring in a generation is $p_{\text{sum}} = [1 - p(k_T = 0)] + p(k_T = 0)p_{E0}$. Standard

probability theory (Wakeley 2009) tells us that the time until an event occurs is geometrically distributed with mean p_{sum} . Hence, p_{sum} is used to draw the time until the next event arises, which is then scaled by $2N_T$ generations so that it is on the coalescent timescale.

3. Once the time to the next event is drawn, it involves sex with probability

$$\frac{1 - p(k_T = 0)}{[1 - p(k_T = 0)] + p(k_T = 0)p_{E0}} \quad (7)$$

If sex does occur, the number of sexual events is drawn independently from each deme from a binomial distribution, discarding cases where no paired samples split by sex across all the demes represented by the sample.

4. Given the vector of paired samples affected by sex \mathbf{k} , one then calculates the probability of subsequent events occurring in each deme using the probabilities outlined in Table 1. Let $p_{E,j}$ be the probability that event E occurs in deme j (if $E = 1$, where split samples remain split, then this event occurs over all demes by definition). The probability that event E occurs in any deme is $p_E = \sum_{j=1}^d p_{E,j}$. Therefore, the probability that event E is chosen is $p_E / \sum_{e=1}^{10} p_e$, and the probability that this event occurs in deme j is $p_{E,j} / \sum_{i=1}^d p_{E,i}$. These probabilities are used to form multinomial distributions, from which the next event and deme of occurrence then can be drawn.
5. The configurations of all demes are then changed based on the outcome. If it is a migration event, an individual is chosen from the deme of event occurrence, and its migration target is chosen at random from the $d - 1$ other demes.
6. The process is repeated until all samples coalesce into a single common ancestor. Mutations are then added to each branch of the genealogy assuming a Poisson distribution with mean $(1/2)\theta\tau_i$, where $\theta = 4N_T\mu$, and τ_i is the length of branch i .

We have implemented this algorithm into a simulation program for R (R Development Core Team 2014), as shown in File S3, or it can be downloaded from <http://github.com/MattHartfield/FacSexCoalescent>. The simulation performs the preceding process, and for each run it outputs the coalescent times and a genealogy in Newick tree format (Felsenstein 2004). It also drops mutations along the genealogy according to the infinite-sites model, which can be used to determine how accurate summary statistics are at inferring the rate of sex and demography. We have rigorously tested this simulation to ensure that it replicates two-sample behavior. Outputs are similar to those presented by *ms* (Hudson 2002) to make it easier to compare simulation outputs between the two programs, if needed. Note that *ms* uses a timescale of $4N$ units, but our program uses a coalescent timescale of $2N$.

In addition, we wrote a variant of the program that does not require population size to be defined as an input and operates

solely on scaled coalescent time. This program is faster than the full program but requires that all parameters, including the rates of sex, are assumed to be on the same order as $1/N$ [i.e., $\Omega = 2N\sigma = \mathcal{O}(1)$]. In this case, instead of first resolving the number of sexual reproductions in a generation, a single bout of sex is defined as one of the possible outcomes were an action to arise. To consider any rate of sex, all simulation results presented here were produced using the full simulation, which can take values of σ between 0 and 1.

Results

Simulation results of summary statistics

Using the full simulation, we investigated how traditional population genetics parameters are affected under low rates of sex to determine whether there is a risk that they show spurious signs of selection or demography. We ran our coalescent algorithm while varying the rate of sex and analyzed the mutational outputs to calculate summary statistics. One thousand trees were run for each point for all simulation results throughout. Confidence intervals were calculated using a normal-distribution approximation; similar intervals were produced if 1000 bootstrap samples were used.

Results are shown in Figure 3 for $N = 10,000$, $\theta = 5$ with 25 paired samples, although results are nearly identical if 50 single samples were simulated instead (File S2A). We need to specify a fixed population size because we cannot explicitly simulate high rates of sex [i.e., $\sigma = \mathcal{O}(1)$] in diploids using only the scaled parameter ($2N\sigma$). As the rate of sex decreases below $1/N$, the number of polymorphic sites shoots up owing to the increased overall coalescent times, especially within individuals, creating greater diversity (Figure 3A). This leads to an increase in traditional estimates of θ that are much higher than the true mutation rate (Figure 3B). With rare sex, Tajima's D also increases to high values (Figure 3C), while Fay and Wu's H drops despite a small increase for $\sigma \approx 1/N$ (Figure 3D). Both of these behaviors are caused by an increase in the number of intermediate-frequency variants, especially within individuals, as sex becomes rare. Finally, the number of unique haplotypes increases (Figure 3E) and the number of unique genotypes decreases (Figure 3F) as σ drops. Both plots reflect how the lack of sex creates new haplotypes within individuals but homogenizes individual genotypes, in line with theory previously reported by Balloux *et al.* (2003). These results make clear that if one does not account for rare rates of sex, spurious signatures of selection or demography can arise that can confound analyses. Deeper branches also can lead to an overestimation of the mutation rate or effective population size.

Sex with selfing

We can extend the preceding analysis to consider a case where an individual that reproduces sexually can self-fertilize with probability S and outcross with probability $1 - S$. The coalescent for obligate sexuality with selfing has been well

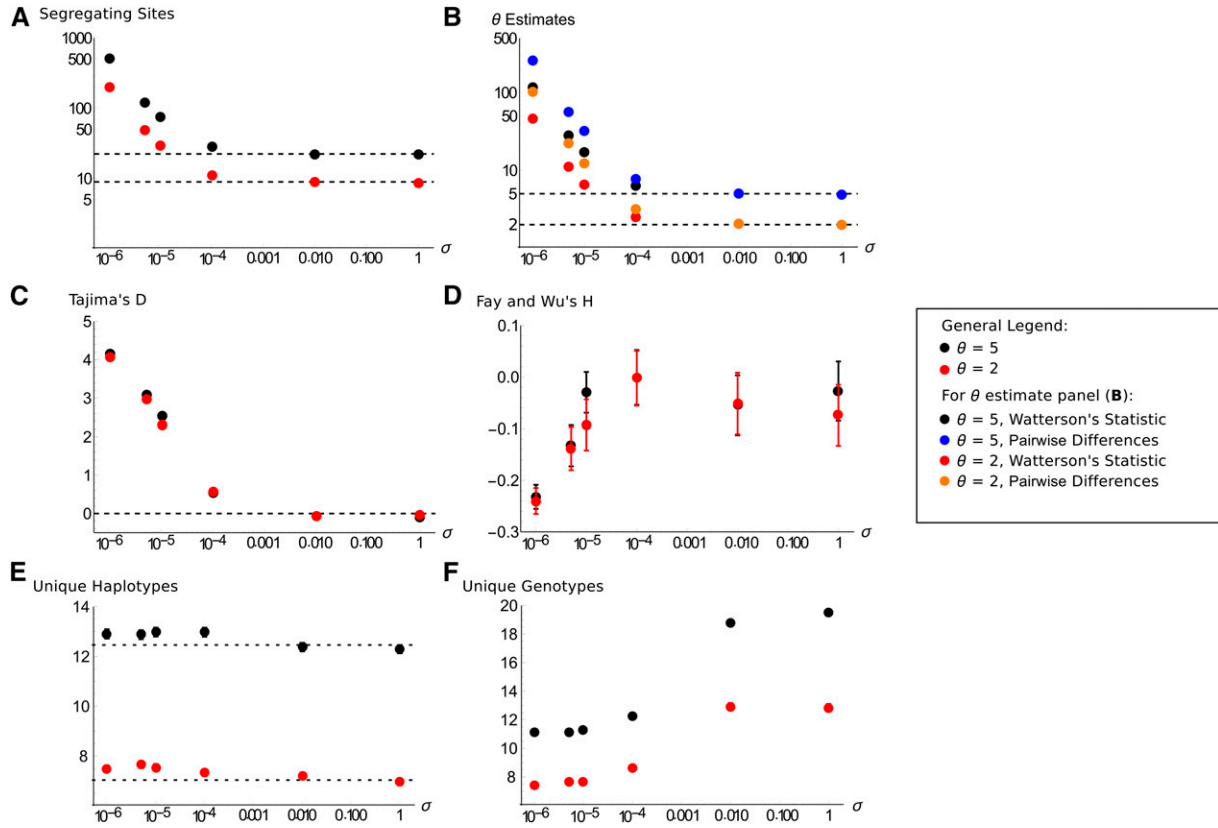


Figure 3 How classic genomic summary statistics are affected by infrequent sex in a nonsubdivided population. All graphs are as a function of σ ; parameters are $N = 10,000$, $\theta = 5$ (black points by default) or $\theta = 2$ (red points by default), based on 25 paired samples, averaged over 1000 coalescent simulations. Error bars are 95% confidence intervals; if they cannot be seen, they lie within the points. (A) Mean number of segregating sites in a sample; the horizontal bar shows the expectation for $\sigma = 1$ (obligate sex). (B) Estimates of θ using Watterson's θ (black or red points for $\theta = 5$ or 2) or the mean pairwise differences between pairs of samples (blue or orange points for $\theta = 5$ or 2). (C) Tajima's D . The horizontal line at zero is the unbiased expectation for obligate sex. (D) Fay and Wu's normalized H statistic. (E) The mean number of unique haplotypes present: the obligate sex expectation (dashed line) is calculated using Equation 3.85 of Ewens (2004). (F) The mean number of unique genotypes present.

investigated (Milligan 1996; Nordborg and Donnelly 1997; Nordborg 1997, 2000; Nordborg and Krone 2002); the key result is that the genealogy is equivalent to a Kingman coalescent, but with the population size scaled by the selfing rate $N_e = N(2 - S)/S$ so that the expected coalescent time is reduced. We can implement similar methods to determine how inbreeding and asexuality interact to affect coalescent times.

The transition matrix T with selfing and asexuality is written as (see File S1A for further information)

$$T = \begin{bmatrix} 1 - \frac{1}{N} & \frac{1}{2N} & \frac{1}{2N} \\ \sigma(1-S)\left(1 - \frac{1}{N}\right) & 1 - \sigma\left(1 - \frac{1}{2N}\right) + \frac{\sigma S}{2}\left(1 - \frac{1}{N}\right) & \frac{\sigma(1-S)}{2N} + \frac{\sigma S}{2} \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

This equation demonstrates that self-fertilization only affects within-individual configurations. If a bout of selfing arises, then the samples coalesce with probability $1/2$; otherwise, they will remain paired within the same individual. Using the ma-

trix-inversion method of Slatkin (1991) and rescaling to the coalescent timescale ($\tau = 2Nt$), the mean coalescent times are obtained as

$$E[\tau_b] \approx \left(1 - \frac{S}{2}\right) + \frac{1}{\Omega} \quad (9)$$

$$E[\tau_w] \approx 1 - S + \frac{2}{\Omega}$$

This equation makes explicit how selfing reduces the mean coalescent times. Ceplitis (2003, Equation 5) also derived equivalent equations in the limit of low rates of sex. However, there appears to be a typographical error in his term for coalescence for two samples taken from different individuals and different chromosome arms. Similarly, Bengtsson (2003) derived accurate terms for the eigenvalues of G in the limit of weak sex, but plots of the effect of selfing appear to be inaccurate (compare his Figure 2 with our Figure 2C). These differences are discussed further in File S1A.

Selfing and asexual reproduction have opposing effects on the mean coalescent times and hence differently affect $E[\tau_b]$ and $E[\tau_w]$. With selfing, $E[\tau_b] > E[\tau_w]$ when rates of sex

are high, but the reverse is true when rates of sex are low (Figure 2, B and C). With obligate sexual reproduction, complete selfing reduces the between-individual mean coalescent time by half, while the within-individual mean coalescent time is instantaneous (Figure 2B). For $\Omega = 2$, both mean coalescent times equal 1 with complete selfing (Figure 2C).

Note that because the main focus of this study is on the effects of sexual relative to asexual reproduction on diversity patterns, we did not implement self-fertilization in the facultative sexual coalescent simulation.

Gene conversion

Mitotic gene conversion can strongly affect the genetic architecture of organisms and can be especially important in facultative sexual organisms, as shown recently in the genome sequence of bdelloid rotifers (Flot *et al.* 2013). If a gene conversion event arises in which one allele sample replaces another during reproduction (irrespective of reproduction type), only one of the two samples is effectively passed onto its offspring.

We assume that unbiased gene conversion occurs at rate γ . The transition matrix becomes

$$\mathbf{T} = \begin{bmatrix} 1 - \frac{1}{N} & \frac{1 - \gamma}{2N} & \frac{1 + \gamma}{2N} \\ \sigma(1 - S) \left(1 - \frac{1}{N}\right) & (1 - \gamma)(1 - \sigma) + \frac{\sigma(1 - S)(1 - \gamma)}{2N} + \frac{\sigma S(1 - \gamma)}{2} & \frac{\sigma(1 - S)(1 + \gamma)}{2N} + \frac{\sigma S(1 + \gamma)}{2} + (1 - \sigma)\gamma \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

We can solve this matrix as usual; we further set $S = 0$ to simplify the analysis (solutions with selfing are provided in File S1A). Furthermore, we make the substitutions $\Omega = 2N\sigma$ and $\Gamma = 2N\gamma$ and rescale to the coalescent timescale [so that $\Gamma, \Omega = \mathcal{O}(1)$]. The mean coalescent times then can be written as

$$\begin{aligned} E[\tau_b] &\approx \left(\frac{1 + \phi}{2 + \phi}\right) + \frac{1}{\Gamma(2 + \phi)} \\ E[\tau_w] &\approx \left(\frac{\phi}{2 + \phi}\right) + \frac{2}{\Gamma(2 + \phi)} \end{aligned} \quad (11)$$

where $\phi = \Omega/\Gamma$. Assuming that ϕ is not too small, each result in Equation 11 is dominated by the first term on the right-hand side. If sex is rare relative to gene conversion ($\phi \rightarrow 0$), the leading terms go to 1 and 0, respectively, for $E[\tau_b]$ and $E[\tau_w]$. For $E[\tau_w]$, we must then consider the second term, which goes to $1/\Gamma$ as the first term goes to 0. This shows that low sex with comparatively high gene conversion behaves similarly to selfing in that $E[\tau_b]$ is half the standard value, and $E[\tau_w]$ is much smaller than $E[\tau_b]$. The dependence on the ratio ϕ means that a reduction in expected coalescent time with gene conversion can occur at much larger rates

of sex relative to population size ($\sigma \gg 1/N$) than the increase in mean coalescent time that we discussed earlier, although the absolute rate of sex must be low ($\sigma \ll 1$).

The reason that gene conversion is important when sex is low can be understood by considering coalescence from two samples taken from separate individuals. In the absence of gene conversion, two samples are put into the same individual, on average, two separate times (requiring at least one bout of sex) before coalescing (*e.g.*, the first time the two samples descended from different homologous chromosomes, but the second time they descended from the same chromosome). When the rate of sex is low, each time the two samples are put into the same individual, they persist together in that genotype for many asexual generations. If gene conversion is high relative to the rate of sex, the samples are likely to coalesce via gene conversion before the genotype is broken apart by sex. Thus, two samples need only be put into the same individual once before coalescing, so the coalescent time is half what it would be otherwise. If two samples start in the same individual, then the time to coalescence is simply the waiting time until gene conversion $1/\Gamma$. This waiting time can be much lower than that for two

alleles in separate individuals to enter the same ancestral individual (equal to half a time unit on the coalescent scale) if $\Gamma \gg 1$.

Note that when gene conversion is rare relative to sex ($\Gamma \rightarrow 0$), Equation 11 collapses to the results assuming no gene conversion (Equation 6). The effect of gene conversion on mean coalescent time is shown in Figure 4 (A and B), demonstrating how mean coalescent time is greatly reduced when gene conversion is strong compared to sex.

Gene conversion rates are generally weak, with Flot *et al.* (2013) estimating gene conversion rates of 10^{-5} to 10^{-6} per site in bdelloid rotifers. However, these rates are feasibly the same order (or higher) as $1/N_e$ for most species, which is the same rate at which infrequent sex starts affecting expected coalescent times. Thus, the strong impact of gene conversion when sex is rare can explain why little within-individual divergence has been reported in facultative and obligate sexual genomes despite very low (or zero) rates of sex (Crease and Lynch 1991; Schön *et al.* 1998; Normark 1999; Schön and Martens 2003; Flot *et al.* 2013).

As with selfing, gene conversion reduces the mean coalescent times by increasing the probability that two samples within an individual will coalesce. However, there are

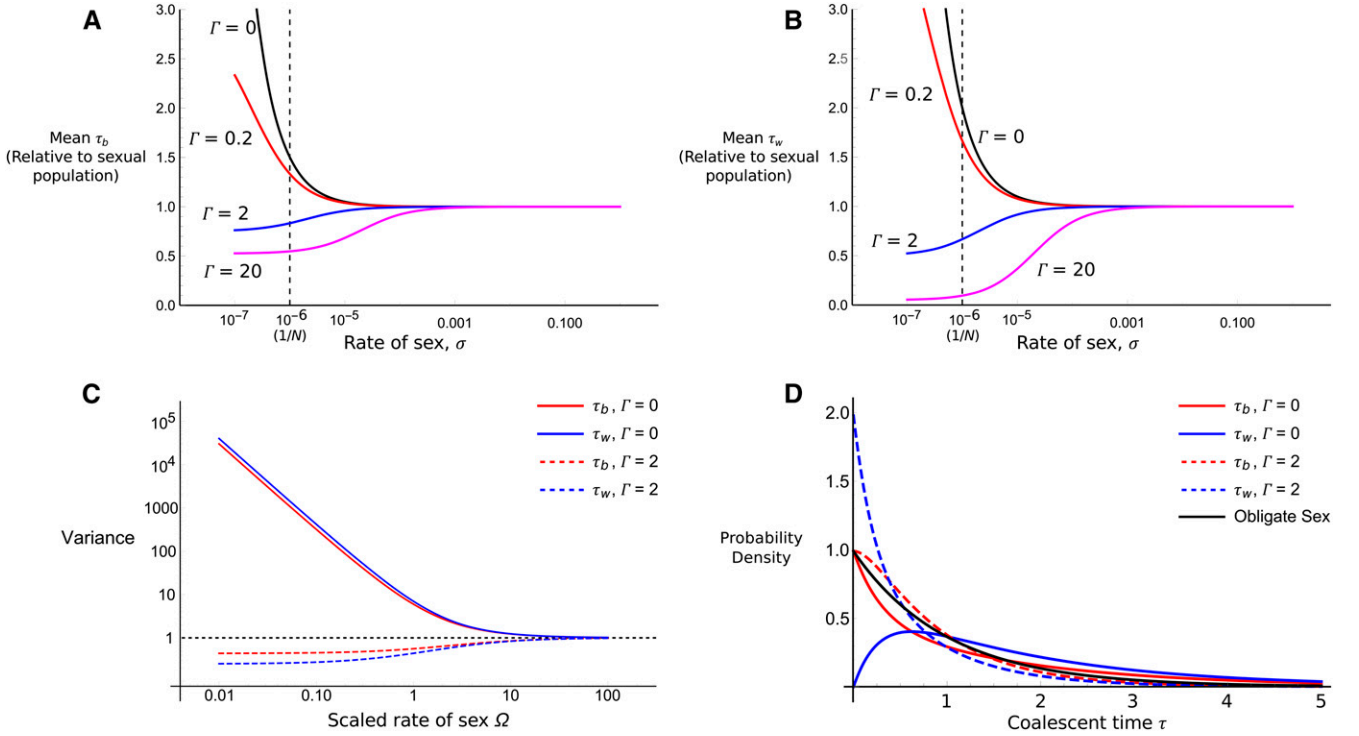


Figure 4 (A and B) Expected coalescent time as a function of the rate of sexual reproduction. The population is of size $N = 10^6$ for two samples taken from separate individuals (A) or the same individual (B). This value of N was used so that $1/N$ is on the same order as gene conversion rates obtained from empirical data (see text for details). Gene conversion rates of $\Gamma = 0$ (black), 0.2 (red), 2 (blue), and 20 (magenta) are shown. (C) The variance of the coalescent process as a function of the scaled rate of sex ($\Omega = 2N\sigma$), scaled to that for the fully sexual coalescent. Red lines are for τ_b ; blue for τ_w ; solid lines for $\Gamma = 0$; dashed line for $\Gamma = 2$. Horizontal dashed line is the variation with complete sex. (D) Probability density of the partial sex coalescent (assuming that $\Omega = 2$ for red and blue lines and $\sigma = 1$ for the black line, *i.e.*, obligate sex) as a function of coalescent time units. Line colors are as in C, with the addition of the obligate sexual distribution (black solid line).

important distinctions between the two processes and their interaction with asexual reproduction. Selfing occurs as a reproductive alternative to asexual reproduction, and the effects of selfing and low sex on mean coalescent time are additive, as shown in Equation 9. In contrast, gene conversion interacts with asexual reproduction rather than being an alternative to it. Gene conversion is only a significant force when asexual reproduction is sufficiently common that two samples within an individual are kept together for many generations; this is why ϕ , the ratio of sex to gene conversion, is important. Nonetheless, both selfing and gene conversion reduce coalescence times and increase homozygosity. Distinguishing between them may be possible through examination of multilocus models because the two processes affect linkage disequilibrium differently.

Variance and probability distribution with and without gene conversion: Asexuality and gene conversion alter not only the mean coalescent time but also other aspects of the probability distribution. Previously, Bengtsson (2003) derived the distribution for the case with partial sex and selfing; here we present a derivation to extend these results to consider gene conversion. We achieve this by following the method of Herbots (1997) and set up a series of linear equa-

tions to find the Laplace transform, and therefore the probability distribution, of the coalescent process. The method is outlined in File S2A.

If we denote the random variable of the between- and within-individual expected coalescent times as τ_b and τ_w , respectively, the variances of these times are

$$\begin{aligned} \text{var}[\tau_b] &= \frac{3 + 2\Omega + (\Gamma + \Omega)^2}{(2\Gamma + \Omega)^2} \\ \text{var}[\tau_w] &= \frac{4 + \Omega(2 + 2\Gamma + \Omega)}{(2\Gamma + \Omega)^2} \end{aligned} \quad (12)$$

Furthermore, the probability density functions of each time $f(\tau_b)$ and $f(\tau_w)$ can be written as a mixture of two exponential distributions

$$\begin{aligned} f(\tau_b) &= (A_b)\lambda_1 \exp(-\lambda_1 \tau_b) + (B_b)\lambda_2 \exp(-\lambda_2 \tau_b) \\ f(\tau_w) &= (A_w)\lambda_1 \exp(-\lambda_1 \tau_w) + (B_w)\lambda_2 \exp(-\lambda_2 \tau_w) \end{aligned} \quad (13)$$

where $\lambda_{1,2} = -(1/2)[-2 - \Gamma - \Omega \pm \sqrt{(\Gamma - 2)^2 + 2\Gamma\Omega + \Omega^2}]$ are the rate parameters of the exponential functions (where 1 and 2 refer to the plus and minus variants of the second term), and the A and B terms are scaling constants, *i.e.*,

$$\begin{aligned}
A_b &= -\frac{\Omega + 2\Gamma - \lambda_1}{\lambda_1(\lambda_1 - \lambda_2)} \\
B_b &= \frac{\Omega + 2\Gamma - \lambda_2}{\lambda_2(\lambda_1 - \lambda_2)} \\
A_w &= -\frac{\Omega + \Gamma(2 - \lambda_1)}{\lambda_1(\lambda_1 - \lambda_2)} \\
B_w &= \frac{\Omega + \Gamma(2 - \lambda_2)}{\lambda_2(\lambda_1 - \lambda_2)}
\end{aligned} \tag{14}$$

Figure 4C plots the variance (Equation 12) with and without gene conversion. With no gene conversion, the variance greatly increases as the rate of sex decreases, reflecting the greatly increased and sporadic coalescent times that result with infrequent sex. However, with gene conversion present at rate $\Gamma = 2$, the variance is lower than for the obligate sex coalescent. Figure 4D plots the corresponding probability distributions. Without gene conversion, the probability density distributions for τ_b and τ_w are qualitatively similar to those for two-allele samples from either different demes or the same deme in a subdivided population model (Herbots 1997). $f(\tau_b)$ has a mode at zero, whereas $f(\tau_w)$ has a nonzero mode. This result is indicative of how paired samples have to be split by sex before they are able to coalesce, just as samples from different demes must migrate into the same deme before coalescence can occur. Hence, coalescence at early times is not possible. If gene conversion is present, the distribution of τ_b is weakly affected by the distribution of τ_w and becomes strongly skewed with a mode at zero. Both the variance and probability distributions with gene conversion provide further evidence that low rates of gene conversion greatly reduce the coalescent times.

Effect of gene conversion on summary statistics: In multi-sample simulations, we found that relatively low rates of gene conversion ($\Gamma = 2$) generally lead to the opposite behavior of the summary statistics shown in Figure 3 owing to increased homozygosity. That is, most estimators decrease with lower sex, except for Fay and Wu's H . The number of unique haplotypes also decreases for low rates of sex, as in cases without gene conversion (File S2B). Therefore, the simultaneous reduction in both haplotype and sequence diversity with decreased rates of sex, in contrast to observing allelic sequence divergence, can be used to determine the presence of gene conversion (or a similar effect such as selfing) in facultative sexual organisms.

Sequence-dependent gene conversion

The preceding section assumes that gene conversion is possible, but if enough differences accumulate between alleles, then gene conversion may no longer operate. Here we consider the time needed for two sequences to achieve sufficiently high divergence to halt gene conversion. Using a different approach, we reach the same finding as Walsh (1987). He considered duplicated gene sequences diverging at a certain

rate, with possible gene conversion homogenizing the two sequences. If each gene acquires k_c mutations, further conversion is prevented, and each sample begins to act independently.

Assume that gene conversion occurs at rate Γ and that $\theta_d = 2N\mu_d$ is the scaled mutation rate for a gene that can lead to sequence divergence. If k_c such mutations occur, then no further gene conversion is possible. We further assume that sex is sufficiently rare (*i.e.*, $\Omega \ll 1$) that it does not affect this process. Conditional on an event occurring, the probability that it is a conversion is $\Gamma/(\Gamma + 2\theta_d)$, and the probability that it is a mutation is $\theta_d/(\Gamma + 2\theta_d)$ (note that the factor of 2 exists because mutation can occur on either of the two gene copies). For complete divergence to occur, k_c mutations must arise before gene conversion homogenizes both sequences. Hence, following from a conversion event, the probability p_c that sufficient divergence occurs before another gene conversion event is $p_c = [2\theta_d/(\Gamma + 2\theta_d)]^{k_c}$. Thus, p_c is equivalent to Equation 6 of Walsh (1987), which was obtained by a different approach.

Given that $p_c > 0$, the critical level of divergence will occur eventually, but we can ask how long we expect this process to take. Let n denote the number of conversion events that happen before divergence eventually occurs. Thus, n follows a geometric distribution with expectation $E(n) = (1 - p_c)/p_c$ or $[1 + (\Gamma/2\theta_d)]^{k_c} - 1$ (Walsh 1987, Equation 7). For example, if $\Gamma/\theta_d = 10$ and $k_c = 10$, then there are expected to be $\sim 6 \times 10^7$ gene conversion events before sufficient divergence is reached. Given that both Γ/θ_d and k_c are likely to be greater than 10, $E(n)$ is likely to be even larger than in this example. Hence, situations where divergence stops gene conversion are highly unlikely to occur.

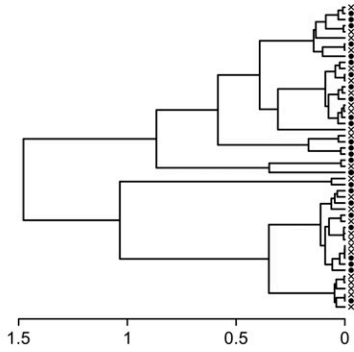
Coalescent times for an island model

Here we show how a simple demographic model, the island model, also can be implemented into the facultative sexual analysis. The population consists of d demes with N_d individuals in each, so the total population size is $N_T = d \times N_d$. In each generation, an individual can migrate to another deme with probability $m = \mathcal{O}(1/N_d) \ll 1$; this assumption ensures that no more than one sample migrates per generation. We also can write this migration rate as $M = 2N_T m$ so that processes can be considered in continuous time. To keep solutions tractable, we do not consider selfing in this analysis.

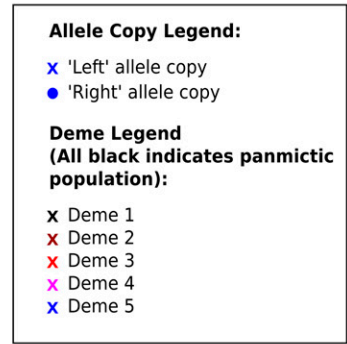
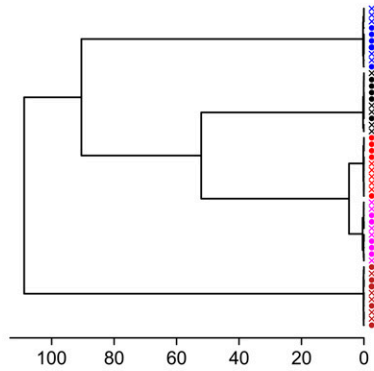
In this model, two alleles can be found in three noncoalesced states. Pairs of samples can be taken either from different demes, from different individuals within a deme, or from the same individual. File S2A shows how the transition matrix \mathbf{G} is formed and solved to obtain mean coalescent times for different demes t_d , within demes t_b , and within individuals t_w . Conveniently, in the continuous timescale $[\Omega, \Gamma, M = \mathcal{O}(1)]$, $E[\tau_b]$ and $E[\tau_w]$ are exactly as in Equation 11 with $N = N_d d$, and $E[\tau_d]$ is the same as $E[\tau_b]$ but with an additional $(d - 1)/2M$ term to denote isolation by distance (Slatkin 1991) (see also File S1B). Hence, the effects of population structure, as well as the joint effect of gene conversion and

Without Gene Conversion

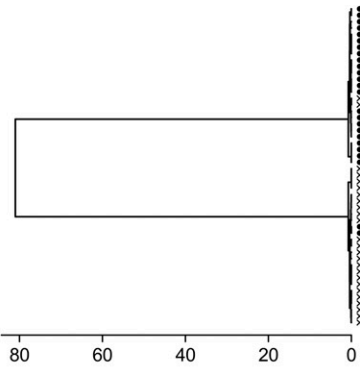
A Kingman Coalescent ($\sigma = 1$)



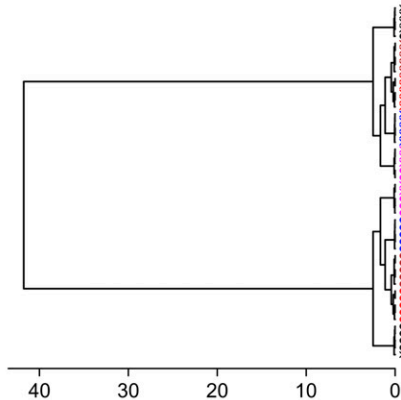
B Five Demes ($\sigma = 1, M = 0.01$)



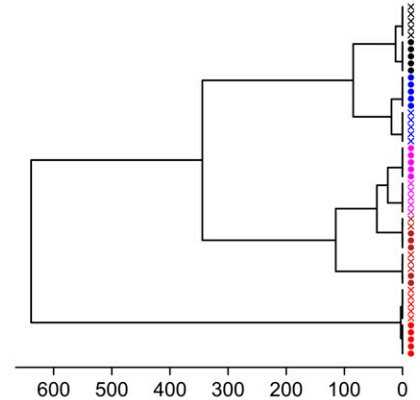
C Low sex Coalescent ($\sigma = 0.01/N_T$)



D Low sex, five demes, high migration ($\sigma = 0.01/N_T, M = 1$)

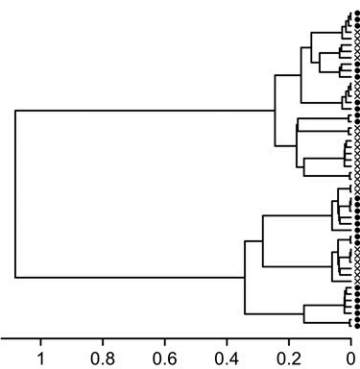


E Low sex, five demes, low migration ($\sigma = 0.01/N_T, M = 0.01$)

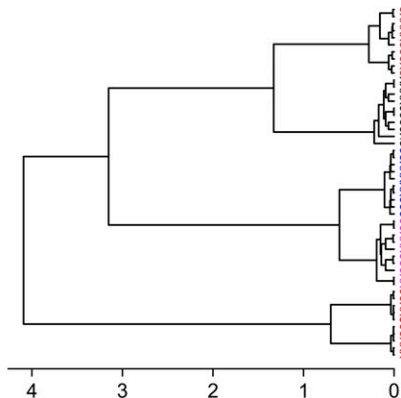


With Gene Conversion $\Gamma = 2$

F Low sex Coalescent ($\sigma = 0.01/N_T$)



G Low sex, five demes, high migration ($\sigma = 0.01/N_T, M = 1$)



H Low sex, five demes, low migration ($\sigma = 0.01/N_T, M = 0.01$)

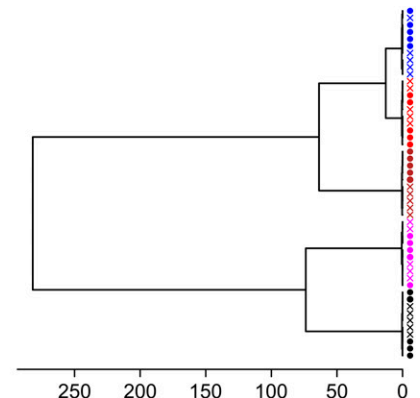


Figure 5 Examples of typical genealogies produced by facultative sex coalescent simulations. For all simulations, $N_T = 10,000$, and 25 paired samples were simulated. Scale bars indicate time in past in units of $2N_T$ generations. Subdivision by either allelic copy or deme is shown in the legend. (A) Traditional (Kingman) coalescent in a nonsubdivided population with obligate sex. (B) Obligate sexual population in an island model, with $d = 5$ and $M = 2N_T m = 0.01$. (C and F) Nonsubdivided population with low sex $\Omega = 2N_T \sigma = 0.02$. (D and G) Low-sex population ($\Omega = 2N_T \sigma = 0.02$) in an island model with a relatively high migration rate ($d = 5, M = 2N_T m = 1$). (E and H) Low-sex population ($\Omega = 2N_T \sigma = 0.02$) in an island model with a relatively low migration rate ($d = 5, M = 2N_T m = 0.01$). No gene conversion is present for A–E, but it is present at a rate $\Gamma = 2$ in F–H.

low sex, are simply additive. As in a nonsubdivided population, sex has to be on the order of the inverse of the total population size $\Omega = \mathcal{O}(1)$ to have a significant impact on the expected coalescent time. If gene conversion is present with rate $\Gamma = \mathcal{O}(1)$, then mean coalescent times are greatly reduced. Hence, even a small rate of gene conversion reduces within-individual diversity that otherwise would be caused by low rates of sex. Further information is available in [File S1B](#).

For multisample genealogies without gene conversion, several distinct types of tree topologies can occur depending on the relative rates of sex and migration. A standard genealogy with obligate sex and high migration is shown in [Figure 5A](#). It is well known ([Wakeley 2000](#); [Wakeley and Aliacar 2001](#)) that with obligate sex and limited migration, samples separate themselves into subclades principally by their geographic location ([Figure 5B](#)), meaning that most of the variation is among demes; the same result occurs with facultative sex, provided that sex is not too rare. However, if sex is low, then there tends to be two deep lineages, present within all demes, representing the long expected coalescent times between left/right haplotypes, assuming that gene conversion is rare relative to sex ([Figure 5C](#)). That is, most of the variation is within individuals created by excess heterozygosity. When migration and sex are both low but migration is high relative to sex [*i.e.*, $\Omega \ll M \sim \mathcal{O}(1)$], then principally two major subclades arise based on the chromosome arm of the sample. Within these major clades, five further clades are apparent owing to the geographic location of the samples ([Figure 5D](#)). A large fraction of the genetic variance is within individuals, but there is also a considerable amount of variation among populations. Finally, if migration rates are of the same order or lower than the rate of sex, the opposite pattern is seen: there are five clades owing to geographic separation, with pairs of subclades then forming as a result of the lack of genetic segregation ([Figure 5E](#)). The frequency spectra for each of these five examples are shown in [File S2C](#) (a–e). With low sex, sequences present in half the allelic samples are very common, representing ongoing divergence at the two different allele copies. Furthermore, once isolation by distance is additionally prevalent, polymorphisms tend to arise in multiples of five (because $d = 5$ in our example), reflecting segregation both by deme and by allelic copy.

With even small rates of gene conversion [$\Gamma = \mathcal{O}(1)$], spatial structure remains, but little genetic partitioning is evident, leading to greater mixing of the two chromosome arms ([Figure 5, F–H](#)). Such trees give a misleading appearance of either a higher overall rate of sex or spatially heterogeneous rates of sex. [File S2C](#) (f and h) shows the relevant site-frequency spectra: low-frequency variants are much more likely to be produced as in the obligate sex case, highlighting how even weak gene conversion creates spurious genomic signatures of frequent sex.

Diversity measurements and parameter estimation: In the absence of gene conversion and with low rates of sex, it is possible to use the migration model results with $\Gamma = 0$ to

derive diversity-based estimators for the mutation, migration, and sex rates. [File S2A](#) outlines how to derive the following scaled estimators for the mutation rate $\theta = 4N_T\mu$, sex rate $\Omega = 2N_T\sigma$, and migration rate $M_e = 2N_Tm/(d-1)$ (we use different notation to differentiate M_e from the parameter $M = 2N_Tm$ introduced earlier):

$$\begin{aligned}\hat{\theta} &= 2\bar{\pi}_b - \bar{\pi}_w \\ \hat{\Omega} &= \frac{2\bar{\pi}_b - \bar{\pi}_w}{(\bar{\pi}_w - \bar{\pi}_b)} = \frac{\hat{\theta}}{(\bar{\pi}_w - \bar{\pi}_b)} \\ \hat{M}_e &= \frac{2\bar{\pi}_b - \bar{\pi}_w}{2(\bar{\pi}_d - \bar{\pi}_b)} = \frac{\hat{\theta}}{2(\bar{\pi}_d - \bar{\pi}_b)}\end{aligned}\quad (15)$$

where $\bar{\pi}_d$ is the average between-deme pairwise diversity, $\bar{\pi}_b$ is the average within-deme diversity, and $\bar{\pi}_w$ is the average within-individual diversity. We also investigate what sampling strategies should be used to most accurately estimate these parameters. Briefly, paired within-individual samples should be used as much as possible, and if the means of $\hat{\Omega}$ and \hat{M}_e are taken from multiple chromosomes, the ratio of means should be taken instead of the mean of ratios. However, these estimators assume that within-individual coalescent times are the longest owing to low rates of sex. If gene conversion occurs on the same order or greater than the rate of sex, within-individual diversity would be removed, so these estimators perform badly. Hence, such estimators would work only in the case where $\Gamma \ll \Omega \approx \mathcal{O}(1)$, which may not apply to many partially asexual species that have large population sizes.

Heterogeneity in rates of sex

Results so far have assumed that the rate of sex remains constant over genealogic time. However, this assumption may not be realistic for a broad range of taxa. Facultative sexual organisms are known to adjust their rate of sexual reproduction depending on certain biotic and abiotic conditions that can vary in time and space. We describe here how to extend the preceding derivations to account for either temporal or spatial heterogeneity. Unless otherwise stated, we do not consider gene conversion in these derivations. For brevity, we only outline key results, with further explanation available in [File S2A](#).

Temporally heterogeneous sexual reproduction: Consider a case in which there is a single population, but the rates of sex vary over time. At time t in the past, the rate of sex is σ_t , and the noncoalescent matrix for that generation \mathbf{G}_t is given by Equation 10, replacing σ with σ_t . The distributions of coalescent times (in generations) for within- and between-individual samples are given by

$$\begin{aligned}P_b(t) &= (1, 0) \left(\prod_{t'=0}^{t-1} \mathbf{G}_{t'} \right) \begin{pmatrix} c_{b,t} \\ c_{w,t} \end{pmatrix} \\ P_w(t) &= (0, 1) \left(\prod_{t'=0}^{t-1} \mathbf{G}_{t'} \right) \begin{pmatrix} c_{b,t} \\ c_{w,t} \end{pmatrix}\end{aligned}\quad (16)$$

where

$$c_{b,t} = \frac{1 + \gamma}{2N}$$

$$c_{w,t} = \frac{\sigma_t}{2N} + \gamma \left[1 - \sigma_t \left(1 - \frac{1}{2N} \right) \right] + S \left[\frac{(N-1)(1+\gamma)\sigma_t}{2N} \right] \quad (17)$$

are the probabilities that two samples that are between individuals or within individuals at time $t - 1$ coalesce at time t .

Numerical evaluation of various scenarios involving temporal heterogeneity indicates that the distribution of coalescent times can be well approximated using the arithmetic average rate of sex and otherwise ignoring temporal heterogeneity. This works well provided that the fluctuations in rates of sex are fast relative to the average coalescent time so that lineages can be expected to experience “average” rates of sex over their coalescent history. When very long periods ($\sim N$ generations) of low sex are punctuated by very brief periods of high rates of sex, use of the average rate of sex can be misleading. For example, we consider a species that reproduces exclusively asexually most of the time, but once every g

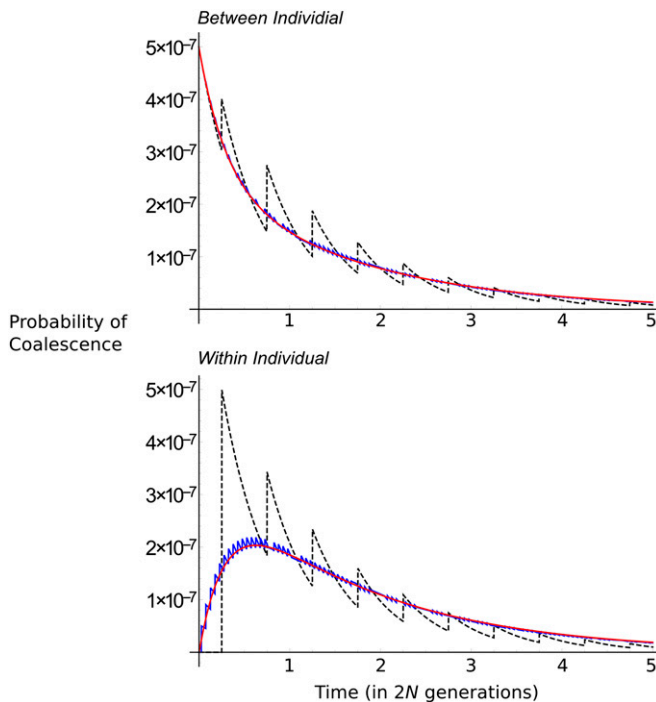


Figure 6 Distribution of coalescent times with temporal heterogeneity in rate of sex, assuming no self-fertilization or gene conversion. The rate of sex is assumed to be zero most of the time, but once every g generations, a fraction σ^* of offspring is produced sexually, giving an average rate of sex of $\bar{\sigma} = \sigma^*/g$. Here $\bar{\sigma} = 1/N$ ($N = 10^6$) for all cases, but this is achieved via very high heterogeneity ($\sigma^* = 1$ and $g = N$; black dashed line), moderate heterogeneity ($\sigma^* = 0.1$ and $g = 0.1N$; blue line), or no heterogeneity ($\sigma^* = 1/N$ and $g = 1$; red line). The first sexual episode occurs $g/2$ generations in the past. The upper and lower panels show the distributions of coalescent times for between- and within-individual samples, respectively.

generations a fraction σ^* of offspring is produced sexually, giving an average rate of sex of $\bar{\sigma} = \sigma^*/g$. As illustrated in Figure 6, when $\sigma^* = 1/10$ and $g = N/10$ (i.e., $\bar{\sigma} = 1/N$), the distributions closely follow those expected in a population with a constant rate of sex of $\sigma = 1/N$. However, the distributions are notably different when $\sigma^* = 1$ and $g = N$ (i.e., $\bar{\sigma} = 1/N$). During the periodic episodes of sex, within-individual samples are converted to between-individual samples, allowing for their subsequent coalescence; this creates periodic spikes in the distribution of coalescent times.

Spatially heterogeneous sexual reproduction: We next consider the case in which there is an island model, but rates of sex occur at a lower rate σ_L in some demes and at a higher rate σ_H in others (and these rates stay constant within each deme over time). There are d_L low-sex demes and d_H high-sex demes, with $d_L + d_H = d_T$. Although we have assumed only two rates of sex, the same approach can be extended to consider an arbitrary number of rates. Given the variation in sex rates considered, we will work with unscaled parameters of σ .

When calculating mean pairwise coalescent times, there are now seven states to consider. For between-deme comparisons, two samples can be taken from either two different low-sex demes, two different high-sex demes, or one from a low-sex deme and the other from a high-sex deme. In addition, for within-deme samples (either from the same or separate individuals), the two samples can be from either a high-sex deme or a low-sex deme.

One can form the 7×7 \mathbf{G} matrix, but in this case, deriving each term is complicated, and the overall solutions are cumbersome. Hence, the full derivation is saved for [File S1D](#). Simpler equations can be obtained if it is assumed that migration $m \ll 1$ and $\sigma_L \ll \sigma_H \ll 1$. These simpler terms are described in [File S2A](#). In particular, elegant terms can be obtained for the difference in mean coalescent times between high- and low-sex demes

$$\Delta t_w = E[t_{w,L}] - E[t_{w,H}] \approx \frac{d_T - 1}{2d_{Hm}}$$

$$\Delta t_b = E[t_{b,L}] - E[t_{b,H}] \approx \frac{2 + d_T(d_T^2 - 3)}{2d_T m \{d_T[1 + d_T(1 + 2mN_d)] - 2\}} \quad (18)$$

These equations illustrate two points regarding the differences in expected coalescent times between samples from low-sex demes compared with those from high-sex demes. The differences increase as migration rates decrease, and they are also independent of the rates of sex. This is so because we have assumed that $\sigma_L \ll \sigma_H \ll 1$. Consequently, alleles from different left/right haplotypes cannot or are extremely unlikely to coalesce within the low-sex demes. Such alleles must first move to high-sex demes before coalescence can occur. The difference shown in Equation 18 represents the waiting time for alleles in low-sex demes to migrate to high-sex

demes before coalescence can occur. Equation 18 reflects the true differences as long as σ_L is low enough compared to the migration rate (File S1D).

We can ask how average coalescent times correspond to those expected in a metapopulation without spatial heterogeneity in sex rates but with the same average rate of sex [*i.e.*, using $\sigma = (d_L/d_T)\sigma_L + (d_H/d_T)\sigma_H$]. File S2D highlights how using the weighted mean in the homogeneous result underestimates true mean coalescent times. This is so because using an average rate of sex does not account for how sequences taken from low-sex demes contribute disproportionately to the overall coalescent time.

File S2A shows how we apply diversity-based estimators for the rates of sex to simulation data from a spatially heterogeneous case in the absence of gene conversion. If migration is not too high ($2mN_T \ll 1$), then application of these estimators to individual demes can give a decent estimation of the rate of sex within them. Estimates are less accurate if migration is higher ($2mN_T \approx 1$), instead predicting a rate of sex intermediate between the two true values.

Discussion

We have outlined here a comprehensive theoretical treatment to consider coalescent processes with arbitrary rates of sex. After reapplying structured coalescent theory to previous models (Bengtsson 2003; Cepitis 2003), we show how to calculate expected coalescent times for pairs of samples taken either between or within individuals (Equation 6). We show how this structured coalescent can be further extended to take other cases into account, such as self-fertilization (Equation 9), gene conversion (Equation 11), and migration according to an island model. From the migration case, we further derived statistical estimators for the rates of mutation, migration, and sex (Equation 17).

Our multisample simulation results highlight that with very low rates of sexual reproduction, strong departures from the standard neutral model expectations for summary statistics such as Tajima's D and Fay and Wu's H are expected, reflecting the deep genealogical divergence between alleles within individuals. Thus, species with very low rates of sex can exhibit diversity patterns resembling long-term balancing selection and/or population subdivision. This highlights the importance of analysis of within vs. between individual patterns of diversity to help untangle the relative role of demographic history and selection from partial sex on the shape of genealogies.

To more accurately account for demographic effects, we found two solutions. First, newly derived estimators for the mutation, sex, and migration rates based on pairwise diversity (Equation 17) can be accurate if good sampling practice is followed (File S2A). However, these simple estimators can produce misleading results under a wide variety of scenarios that cannot be accounted for because of insufficient degrees of freedom in the equations. These include the presence of gene conversion rates of similar magnitude to sex and spatial

or temporal heterogeneity in sex. We therefore suspect that the best manner to account for complex demography would be to use coalescent simulations in an inference framework, such as approximate Bayesian computation (ABC) (Sunnåker *et al.* (2013)), to predict how summary statistics change under different regimes. Our new simulation package could be used in an ABC-type inference while varying the rate of sex to simultaneously infer sex and demography.

One major result is that once sex becomes on the same order as gene conversion, the latter becomes a powerful force in reducing within-individual diversity and therefore removing genetic signatures of rare sex. Beyond the challenges this presents for estimating rates of sexual reproduction, this result also has important evolutionary consequences because the absence of permanent heterozygosity can influence not only the structuring of neutral diversity but also the exposure of both deleterious and advantageous recessive mutations to selection. Although elevated heterozygosity has been observed in some partial and obligate asexual organisms (Tucker *et al.* 2013; Hollister *et al.* 2015), there is increasing evidence that a major source of high within-individual diversity may be hybrid origins of asexual lineages rather than ongoing accumulation of diversity owing to new mutations (Tucker *et al.* 2013). Under biologically realistic rates of gene conversion, this initially high heterozygosity may be rapidly eroded by gene conversion, decreasing neutral variation within and between individuals (Tucker *et al.* 2013). Furthermore, these potentially high rates of gene conversion make it unlikely that two allelic copies develop sufficient divergence before conversion resets their independent evolution (see also Walsh 1987).

These analyses provide a more thorough theoretical basis for which the genealogies of samples from facultative sexual organisms can be reconstructed to provide the null distribution of neutral diversity for such populations. We have outlined methodology so that it is clear how the results presented here can be extended to cover even more complex scenarios; nevertheless, there are several apparent routes for extending this work. Although we have shown how to implement demography via an island model, it would be desirable for future models to consider a broader array of demographic scenarios. Real-world populations go through a bewildering array of changes, including extinction, recolonization, population expansion, and bottlenecking (Whitlock and McCauley 1999; Veeramah and Hammer 2014). Advanced coalescent models for obligate sexual organisms have been developed to determine how complex demographic effects affect genealogies (Rosenberg and Nordborg 2002; Hudson 2002); it will certainly be worthwhile to implement similar demographic scenarios into facultative sex coalescent simulations given the known impact of spatial effects on clonal species distribution (Arnaud-Haond *et al.* 2007).

The other major extension that will prove important is to implement recombination. In this case, the outputs will no longer be a genealogy, but rather a recombination graph (Griffiths 1981, 1991), although it would still be possible to

create genealogies for unrecombined genomic segments. Recombination graphs can provide greater power for detecting demographic effects with fewer samples, although it can be more computationally demanding to run relevant analyses (Schiffels and Durbin 2014). Adding recombination will prove important for facultative sexual coalescents because while sex has to be very rare [$\mathcal{O}(1/N)$] to affect the nonrecombining genealogies presented here, differences in recombination graphs might become apparent with higher rates of sex. This is so because it is expected that recombination correlates linearly with sex rate. A similar result was derived for partially selfing organisms by Nordborg (2000). Genomic sequences of facultative sexual organisms can reveal evidence of rare recombination events, providing proof of cryptic sex [examples were shown by Grimsley *et al.* (2010) for the marine algae *Ostreococcus* spp. and Signorovitch *et al.* (2005) for Placozoa]. In addition, ancestral graphs would be able to separate out areas of the genome affected by gene conversion, leading to much more refined measurements of how this force affects genetic architecture. In particular, gene conversion events in the absence of sex will lead to much higher rates of gene conversion relative to crossing over. This will have a strong effect on short- vs. long-range linkage disequilibrium along chromosomes, which may enable accurate estimators of the rate of sex even in the presence of significant rates of gene conversion. Recombination graphs of facultative sexual organisms thus can greatly increase the power by which the effects of infrequent sex and demography can be disentangled.

Acknowledgments

We thank two anonymous reviewers and John Wakeley for comments on the manuscript. M.H. is supported by a Marie Curie International Outgoing Fellowship (grant MC-IOF-622936, project SEXSEL). This work also was supported by Discovery Grants (to A.F.A. and S.I.W.) from the Natural Sciences and Engineering Research Council of Canada.

Literature Cited

- Ankarklev, J., N. M. B. Brancucci, I. Goldowitz, P.-Y. Mantel, and M. Marti, 2014 Sex: how malaria parasites get turned on. *Curr. Biol.* 24: R368–R370.
- Arnaud-Haond, S., C. M. Duarte, F. Alberto, and E. A. Serrão, 2007 Standardizing methods to address clonality in population studies. *Mol. Ecol.* 16: 5115–5139.
- Balloux, F., L. Lehmann, and T. de Meeùs, 2003 The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635–1644.
- Becks, L., and A. F. Agrawal, 2010 Higher rates of sex evolve in spatially heterogeneous environments. *Nature* 468: 89–92.
- Bengtsson, B. O., 2003 Genetic variation in organisms with sexual and asexual reproduction. *J. Evol. Biol.* 16: 189–199.
- Brookfield, J. F. Y., 1992 DNA fingerprinting in clonal organisms. *Mol. Ecol.* 1: 21–26.
- Burt, A., D. A. Carter, G. L. Koenig, T. J. White, and J. W. Taylor, 1996 Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA* 93: 770–773.
- Butlin, R., 2002 The costs and benefits of sex: new insights from old asexual lineages. *Nat. Rev. Genet.* 3: 311–317.
- Ceplitis, A., 2003 Coalescence times and the Meselson effect in asexual eukaryotes. *Genet. Res.* 82: 183–190.
- Chang, H.-H., E. L. Moss, D. J. Park, D. Ndiaye, S. Mboup *et al.*, 2013 Malaria life cycle intensifies both natural selection and random genetic drift. *Proc. Natl. Acad. Sci. USA* 110: 20129–20134.
- Crease, T. J., and M. Lynch, 1991 Ribosomal DNA variation in *Daphnia pulex*. *Mol. Biol. Evol.* 8: 620–640.
- de Meeùs, T., and F. Balloux, 2005 *F*-statistics of clonal diploids structured in numerous demes. *Mol. Ecol.* 14: 2695–2702.
- D'Souza, T. G., and N. K. Michiels, 2008 Correlations between sex rate estimates and fitness across predominantly parthenogenetic flatworm populations. *J. Evol. Biol.* 21: 276–286.
- Ewens, W. J., 2004 *Mathematical Population Genetics: 1. Theoretical Introduction* (Interdisciplinary Applied Mathematics Series, Vol. 27), Ed. 2. Springer, New York.
- Felsenstein, J., 2004 *Inferring Phylogenies*, Vol. 2. Sinauer Associates, Sunderland, MA.
- Flot, J.-F., B. Hespels, X. Li, B. Noel, I. Arkhipova *et al.*, 2013 Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* 500: 453–457.
- Goddard, M. R., H. C. J. Godfray, and A. Burt, 2005 Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* 434: 636–640.
- Griffiths, R. C., 1981 Neutral two-locus multiple allele models with recombination. *Theor. Popul. Biol.* 19: 169–186.
- Griffiths, R. C., 1991 The two-locus ancestral graph. *Lect. Notes Monogr. Ser.* 18: 100–117.
- Grimsley, N., B. Pequin, C. Bachy, H. Moreau, and G. Piganeau, 2010 Cryptic sex in the smallest eukaryotic marine green algae. *Mol. Biol. Evol.* 27: 47–54.
- Grishkan, I., A. B. Korol, E. Nevo, and S. P. Wasser, 2003 Ecological stress and sex evolution in soil microfungi. *Proc. Biol. Sci.* 270: 13–18.
- Hadany, L., and S. P. Otto, 2007 The evolution of condition-dependent sex in the face of high costs. *Genetics* 176: 1713–1727.
- Halkett, F., J.-C. Simon, and F. Balloux, 2005 Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol. Evol.* 20: 194–201.
- Hand, M. L., and A. M. G. Koltunow, 2014 The genetic control of apomixis: asexual seed formation. *Genetics* 197: 441–450.
- Herbots, H. M., 1997 The structured coalescent, pp. 231–255 in *Progress in Population Genetics and Human Evolution* (IMA Volumes in Mathematics and Its Applications, Vol. 87), edited by P. Donnelly, and S. Tavaré. Springer, New York.
- Hojsgaard, D., and E. Hörandl, 2015 A little bit of sex matters for genome evolution in asexual plants. *Front. Plant Sci.* 6: 82.
- Hollister, J. D., S. Greiner, W. Wang, J. Wang, Y. Zhang *et al.*, 2015 Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Mol. Biol. Evol.* 32: 896–905.
- Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18: 337–338.
- King, K. C., L. F. Delph, J. Jokela, and C. M. Lively, 2009 The geographic mosaic of sex and the Red Queen. *Curr. Biol.* 19: 1438–1441.
- Kingman, J. F. C., 1982 On the genealogy of large populations. *J. Appl. Probab.* 19: 27–43.
- Levin, T. C., and N. King, 2013 Evidence for sex and recombination in the choanoflagellate *Salpingoeca rosetta*. *Curr. Biol.* 23: 2176–2180.
- Mark Welch, D. B., and M. Meselson, 2000 Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* 288: 1211–1215.

- Mark Welch, D. B., J. L. Mark Welch, and M. Meselson, 2008 Evidence for degenerate tetraploidy in bdelloid rotifers. *Proc. Natl. Acad. Sci. USA* 105: 5145–5149.
- Milligan, B. G., 1996 Estimating long-term mating systems using DNA sequences. *Genetics* 142: 619–627.
- Möhle, M., 1998 A Convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Probab.* 30: 493–512.
- Morran, L. T., M. D. Parmenter, and P. C. Phillips, 2009 Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* 462: 350–352.
- Nordborg, M., 1997 Structured coalescent processes on different time scales. *Genetics* 146: 1501–1514.
- Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154: 923–929.
- Nordborg, M., and P. Donnelly, 1997 The coalescent process with selfing. *Genetics* 146: 1185–1195.
- Nordborg, M., and S. M. Krone, 2002 Separation of time scales and convergence to the coalescent in structured populations, pp. 194–232 in *Modern Developments in Theoretical Population Genetics*, edited by M. Slatkin, and M. Veuille. Oxford University Press, Oxford, UK.
- Normark, B. B., 1999 Evolution in a putatively ancient asexual aphid lineage: recombination and rapid karyotype change. *Evolution* 53: 1458–1469.
- R Development Core Team, 2014 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Redfield, R. J., 1988 Evolution of bacterial transformation: Is sex with dead cells ever better than no sex at all? *Genetics* 119: 213–221.
- Rosenberg, N. A., and M. Nordborg, 2002 Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3: 380–390.
- Schaefer, I., K. Domes, M. Heethoff, K. Schneider, I. Schön *et al.*, 2006 No evidence for the “Meselson effect” in parthenogenetic oribatid mites (Oribatida, Acari). *J. Evol. Biol.* 19: 184–193.
- Schiffels, S., and R. Durbin, 2014 Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46: 919–925.
- Schön, I., R. K. Butlin, H. I. Griffiths, and K. Martens, 1998 Slow molecular evolution in an ancient asexual ostracod. *Proc. Biol. Sci.* 265: 235–242.
- Schön, I., and K. Martens, 2003 No slave to sex. *Proc. Biol. Sci.* 270: 827–833.
- Signorovitch, A. Y., S. L. Dellaporta, and L. W. Buss, 2005 Molecular signatures for sex in the Placozoa. *Proc. Natl. Acad. Sci. USA* 102: 15518–15522.
- Slatkin, M., 1991 Inbreeding coefficients and coalescence times. *Genet. Res.* 58: 167–175.
- Snell, T. W., J. Kubanek, W. Carter, A. B. Payne, J. Kim *et al.*, 2006 A protein signal triggers sexual reproduction in *Brachionus plicatilis* (rotifera). *Mar. Biol.* 149: 763–773.
- Sunnåker, M., A. G. Busetto, E. Numminen, J. Corander, M. Foll *et al.*, 2013 Approximate Bayesian computation. *PLOS Comput. Biol.* 9: e1002803.
- Tucker, A. E., M. S. Ackerman, B. D. Eads, S. Xu, and M. Lynch, 2013 Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proc. Natl. Acad. Sci. USA* 110: 15740–15745.
- Veeramah, K. R., and M. F. Hammer, 2014 The impact of whole-genome sequencing on the reconstruction of human population history. *Nat. Rev. Genet.* 15: 149–162.
- Wakeley, J., 2000 The effects of subdivision on the genetic divergence of populations and species. *Evolution* 54: 1092–1101.
- Wakeley, J., 2009 *Coalescent Theory: An Introduction*, Vol. 1. Roberts & Company Publishers, Greenwood Village, CO.
- Wakeley, J., and N. Aliacar, 2001 Gene genealogies in a metapopulation. *Genetics* 159: 893–905.
- Walsh, J. B., 1987 Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics* 117: 543–557.
- Whitlock, M. C., and D. E. McCauley, 1999 Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. *Heredity* 82: 117–125.
- Yonezawa, K., T. Ishii, and T. Nagamine, 2004 The effective size of mixed sexually and asexually reproducing populations. *Genetics* 166: 1529–1539.
- Yoshida, K., H. A. Burbano, J. Krause, M. Thines, D. Weigel *et al.*, 2014 Mining herbaria for plant pathogen genomes: back to the future. *PLoS Pathog.* 10: e1004028.

Communicating editor: J. Wakeley

GENETICS

Supporting Information

www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178004/-/DC1

Coalescent Times and Patterns of Genetic Diversity in Species with Facultative Sex: Effects of Gene Conversion, Population Structure, and Heterogeneity

Matthew Hartfield, Stephen I. Wright, and Aneil F. Agrawal

File S1 Mathematica notebook of derivations. (.nb, 2,806 KB)

Available for download as a .nb file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178004/-/DC1/FileS1.nb

A Further information on model derivations

This section provides further information on how the transition probabilities for individual models are derived.

Basic Model

The logic behind each entry for the transition matrix \mathbf{T} (Equation 1 in the main text) is as follows.

- T_{11} : Gene 1 in one individual picks any gene copy; the probability of this is 1. Gene 2 has to pick a copy in any other of the $1 - 1/N$ parents. Since this parent is distinct and produces separate offspring, the effect of reproductive mode averages out.
- T_{12} : Gene 1 in one individual ‘picks’ any gene copy. Gene 2 has to pick the alternate gene copy in the same parent, with probability $1/(2N)$. As before, due to production of separate offspring (in the current generation), the effect of reproductive mode averages out.
- T_{13} : Gene 2 has to pick the same gene copy in the same parent as gene 1, with probability $1/(2N)$.
- T_{21} : Gene 2 has to pick the a gene copy in a different parent ($p = 1 - 1/N$). Reproduction must be sexual (with probability σ) in order for both copies to be placed in the same offspring.
- T_{22} : Either reproduction is asexual with probability $1 - \sigma$, in which case the state is maintained. Otherwise reproduction is sexual, with the second allele

copy choosing the different chromosome arm in the same individual as the first, with net probability $\sigma/2N$.

- T_{23} : The parent reproduces sexually, and transmits the same chromosome arm in the population to the same offspring; this occurs with $p = \sigma/(2N)$. Coalescence from this state is impossible with asexual reproduction.

Including Selfing

The discordance matrix (\mathbf{G}) with selfing is already well-known (NORDBORG and DONNELLY 1997). Selfing only affects the within-individual transition probabilities: if two samples are taken from the same individual, they are descended from different allele samples with probability $1/2$; or the same sample (coalescence) with probability $1/2$ as well. With selfing included, the discordance matrix \mathbf{G} is written as:

$$\begin{aligned} \mathbf{G} &= \sigma \cdot (1 - S) \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} \\ 1 - \frac{1}{N} & \frac{1}{2N} \end{pmatrix} + \sigma \cdot S \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} \\ 0 & \frac{1}{2} \end{pmatrix} + (1 - \sigma) \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} \\ \sigma(1 - S)(1 - \frac{1}{N}) & 1 - \sigma(1 - \frac{1}{2N}) + \frac{\sigma S}{2}(1 - \frac{1}{N}) \end{pmatrix} \end{aligned} \tag{1}$$

Including Gene Conversion

Let the rate of gene conversion be γ , which is unbiased (so both allele copies can be equally affected). Since the outcome of gene conversion is to transmit the same

allele copy to all offspring, it reduces the probability of two different allele copies being descended from the same individual by $1 - \gamma$, and increased coalescence probability by a factor $1 + \gamma$.

Here, the discordance matrix is altered to be as follows:

$$\begin{aligned} \mathbf{G} &= \sigma \cdot (1 - S) \begin{pmatrix} 1 - \frac{1}{N} & \frac{1-\gamma}{2N} \\ 1 - \frac{1}{N} & \frac{1-\gamma}{2N} \end{pmatrix} + \sigma \cdot S \begin{pmatrix} 1 - \frac{1}{N} & \frac{1-\gamma}{2N} \\ 0 & \frac{1-\gamma}{2} \end{pmatrix} + (1 - \sigma) \begin{pmatrix} 1 - \frac{1}{N} & \frac{1-\gamma}{2N} \\ 0 & 1 - \gamma \end{pmatrix} \\ &= \begin{pmatrix} 1 - \frac{1}{N} & \frac{1-\gamma}{2N} \\ \sigma(1 - S)(1 - \frac{1}{N}) & (1 - \gamma)(1 - \sigma) + \frac{\sigma(1-S)(1-\gamma)}{2N} + \frac{\sigma S(1-\gamma)}{2} \end{pmatrix} \end{aligned} \tag{2}$$

Variance and probability distribution with and without gene conversion

Here we describe how to determine the probability distribution and variance of the coalescent processes using Laplace transformations (HERBOTS 1997). If we denote the random variable of the between- and within-individual coalescent times as τ_b , τ_w respectively, then their Laplace transformations are the solution to the equations:

$$\begin{aligned} E[e^{-s\tau_b}] &= E[e^{-sX_b}] \left(\frac{1}{2} + \frac{1}{2} E[e^{-s\tau_w}] \right) \\ E[e^{-s\tau_w}] &= E[e^{-sX_w}] \left(\frac{\Gamma}{\Gamma + \Omega} + \frac{\Omega}{\Gamma + \Omega} E[e^{-s\tau_w}] \right) \end{aligned} \tag{3}$$

X_b and X_w are random variables of the waiting times for between- and within-individual samples before they change state. Standard coalescent theory (WAKE-

LEY 2009) states that these variables are exponentially distributed with rates 2 and $(\Omega + \Gamma)$ respectively. The Laplace transform of exponential distributions with rates λ is $\lambda/(\lambda + s)$. Equation 3 shows that once there is a state change, between-individual samples either coalesce or enter the same individual, each with probability 1/2. Alternatively, if within-individual samples change state, they coalesce with probability $\Gamma/(\Gamma + \Omega)$, or split into separate individuals with probability $\Omega/(\Gamma + \Omega)$. Further information on these calculations are available in Section A of File S1. Solving this system of equations yields the following terms for τ_b, τ_w :

$$\begin{aligned} E[e^{-s\tau_b}] &= \frac{s + 2\Gamma + \Omega}{\Omega + 2\Gamma + s(2 + \Gamma + \Omega) + s^2} \\ E[e^{-s\tau_w}] &= \frac{\Omega + \Gamma(2 + s)}{\Omega + 2\Gamma + s(2 + \Gamma + \Omega) + s^2} \end{aligned} \tag{4}$$

Due to similarity between moment-generating functions and Laplace transformations, we can find the variance of τ_b, τ_w by finding the first and second derivatives of Equations 4 at $s = 0$ to obtain (OTTO and DAY 2007):

$$\begin{aligned} Var[\tau_b] &= \frac{3 + 2\Omega + (\Gamma + \Omega)^2}{(2\Gamma + \Omega)^2} \\ Var[\tau_w] &= \frac{(4 + 2\Omega(1 + \Gamma) + \Omega^2)}{(2\Gamma + \Omega)^2} \end{aligned} \tag{5}$$

To determine the probability distribution, we split Equation 4 into partial fractions of the following form:

$$E[e^{-s\tau_i}] = \left(\frac{A_i}{-\lambda_1}\right) \left(\frac{-\lambda_1}{s - \lambda_1}\right) + \left(\frac{B_i}{-\lambda_2}\right) \left(\frac{-\lambda_2}{s - \lambda_2}\right) \quad (6)$$

Here, λ_1, λ_2 are the roots of the denominator of the Laplace transforms in Equation 4, equal to $-(1/2)(-2 - \Gamma - \Omega \pm \sqrt{(\Gamma - 2)^2 + 2\Gamma\Omega + \Omega^2})$. By comparing Equation 6 to known Laplace transformations, it demonstrates that τ_b, τ_w are sums of exponential distributions with rate parameters λ_1, λ_2 respectively. Some algebra reveals the form of A and B as given by Equation 13 in the main text.

Island Model

First, consider two samples taken from different demes. The only possible change in state is migration of one of the two samples into the same deme as the other, with probability $2m/(d - 1)$. Coalescence of the samples, or migration of one sample into the same individual is unlikely, of $O(1/N^2)$.

For two samples taken from the same deme but different individuals, they can either become two samples originating from different demes if one of the two migrates to any other deme. This event arises with probability $2m + O(m^2)$. Alternatively, the samples can either be descended from the same individual (and form an within-individual pair), or coalesce within the deme with probability $1/2N_d$. Thus the status quo is maintained with probability $1 - 1/N_d - 2m$.

When two samples are taken from the same individual, several outcomes are possible.

- One of the two samples can migrate into a different deme (with probability

$2m$), but only if the two samples are first split by sex (with probability σ).

The net probability of state change is $2\sigma m$.

- Alternatively, the two samples can split by sex (σ), enter different individuals ($1 - 1/N_d$), and neither migrate ($1 - 2m$). Net probability of state change to two samples from the same deme but in different individuals is $\sigma(1 - 1/N_d)(1 - 2m)$.
- The two samples coalesce with probability $\sigma/(2N_d)$.
- Therefore, the status quo is maintained with probability of 1 minus the sum of the aforementioned probabilities.

Hence, the non-identity matrix \mathbf{G} is:

$$\mathbf{G} = \begin{pmatrix} 1 - \frac{2m}{d-1} & \frac{2m}{d-1} & 0 \\ 2m & 1 - 2m - \frac{1}{N_d} & \frac{1}{2N_d} \\ 2\sigma m & \sigma \left(1 - \frac{1}{N_d}\right) (1 - 2m) & 1 - \sigma \left(1 - \frac{1}{N_d}\right) (1 - 2m) - 2\sigma m - \frac{\sigma}{2m} \end{pmatrix} \quad (7)$$

Solving to obtain mean coalescent times for different demes t_d , within-deme t_b , and within-individual t_w :

$$\begin{aligned} E[t_d] &= \left(2N_d d + \frac{d-1}{2m}\right) + \left(\frac{1-\sigma}{\sigma}\right) \left(\frac{N_d}{2m + N_d}\right) \\ E[t_b] &= 2N_d d + \left(\frac{1-\sigma}{\sigma}\right) \left(\frac{N_d}{2m + N_d}\right) \\ E[t_w] &= 2N_d d + \left(\frac{1-\sigma}{\sigma}\right) \left(\frac{2N_d}{2m + N_d}\right) \end{aligned} \quad (8)$$

Adding gene conversion to the rate matrix \mathbf{G} is straightforward, since it is only the within-individual transition terms that are affected. That is, the last term of the second column is altered to $(1 - \gamma)/(2N_d)$, and the coalescence probability within-deme is altered to $(\sigma(1 + \gamma))/(2N_d) - (1 - \sigma)\gamma$.

Diversity measurements and parameter estimation in the absence of gene conversion

Estimates of coalescent times can be used to determine the average pairwise diversity between samples. In the infinite-sites model (HUDSON 1990), the total number of new mutations accumulating along a branch is Poisson distributed with mean μT , for μ the per-site mutation rate and T the branch length. Therefore, the expected number of pairwise differences between two samples is $2\mu T$.

If we use the approximate terms for the average coalescent time between samples in an island model (Equation 8), mean pairwise diversity estimates are obtained:

$$\begin{aligned}\bar{\pi}_d &= \theta \left(1 + \frac{d-1}{4mN_d d} + \left(\frac{1}{2N_d d} \right) \left(\frac{1-\sigma}{\sigma} \right) \right) \\ \bar{\pi}_b &= \theta \left(1 + \left(\frac{1}{2N_d d} \right) \left(\frac{1-\sigma}{\sigma} \right) \right) \\ \bar{\pi}_w &= \theta \left(1 + \left(\frac{1}{N_d d} \right) \left(\frac{1-\sigma}{\sigma} \right) \right)\end{aligned}\tag{9}$$

Here, $\theta = 4N_T\mu = 4N_d d\mu$ is a standard compound parameter for the scaled population mutation rate. We can solve the system of Equations (9) to obtain estimators for the mutation rate $\hat{\theta}$, the population-level rate of sex $\hat{\Omega} = 2N_d d\sigma$,

and the scaled migration rate $\widehat{M}_e = (2N_d dm)/(d - 1)$:

$$\begin{aligned}\widehat{\theta} &= 2\bar{\pi}_b - \bar{\pi}_w \\ \widehat{\Omega} &= \frac{2\bar{\pi}_b - \bar{\pi}_w}{(\bar{\pi}_w - \bar{\pi}_b)} = \frac{\widehat{\theta}}{(\bar{\pi}_w - \bar{\pi}_b)} \\ \widehat{M}_e &= \frac{2\bar{\pi}_b - \bar{\pi}_w}{2(\bar{\pi}_d - \bar{\pi}_b)} = \frac{\widehat{\theta}}{2(\bar{\pi}_d - \bar{\pi}_b)}\end{aligned}\tag{10}$$

Note that the estimator $\widehat{\Omega}$ further assumes that the net population size $N_D d \gg \widehat{\theta}$. Unfortunately, it is not possible to obtain an estimator for gene conversion, since there are not enough degrees of freedom, and because effects of sex and gene conversion are connected when both are rare. CEPLITIS (2003) derived equivalent expressions for $\widehat{\Omega}$, however they were not tested by simulations.

The above estimators are based on pairwise diversity. However, it is also possible to obtain estimators based on the expected homozygosity (HUDSON 1990; NORDBORG and DONNELLY 1997). We show in Section B how these alternative estimators of $\widehat{\theta}$ and $\widehat{\Omega}$ can be derived, but are very poor estimators of key population parameters.

Sampling and estimating mutation and sex rates in a non-subdivided population

Using the coalescent simulation, we tested the accuracy of the estimators for the scaled mutation rate, $\widehat{\theta}$, and rate of sex $\widehat{\Omega}$, based on average pairwise diversity (Equation 10). We analysed output from simulations with 25 paired samples, but also investigated the output from 30 single samples and 10 paired samples, as well as 10 single samples and 20 paired samples. Hence one can investigate what

sampling regime is best to obtain the most accuracy in estimating mutation and sex rates.

Without gene conversion. Results are outlined in Table A. We see that, broadly, $\hat{\theta}$ accurately matches the true θ for all the parameter range, except when some single samples are present and the rate of sex is very low ($N\sigma \ll 1$). However, measures of Ω fare less well, if we simply take the mean of estimates across simulations. For $\theta = 5$, estimates are only of the same order of magnitude if 25 paired samples are taken: even then, the mean estimates usually overestimate the true Ω value by a factor of two. If $\theta = 2$, the estimator $\hat{\Omega}$ fares better for all sampling schemes, but still overestimates the true Ω value.

We next tested whether more accurate estimates of Ω can be obtained by not taking the mean of values, but instead first calculating the mean of the numerator and denominators in Equation 10, then take the ratio of means. This approach was advocated by WEIR and COCKERHAM (1984) to improve estimates of population subdivision using F -statistics. Confidence intervals were calculated for such statistics using 1000 bootstrap samples. For high rates of sex ($N\sigma \gg 1$) estimates are inaccurate and have a high absolute value. This results is reflective of the fact that for $N\sigma \gg 1$ it is hard to discern between obligate and facultative sex, meaning the denominator will be close to zero since there is little difference for between- and within-individual variation measures. Hence very noisy estimates are produced. However, for low rates of sex ($N\sigma = O(1)$ or lower), $\hat{\Omega}$ with ratio of means matches up extremely well with the true value, with estimates being most accurate if only paired samples are used.

We also tested estimators for an island model scenario. In these simulations, the total population size was fixed at 10,000 and $\theta = 5$; as well as varying rates of sex,

Table A: Estimates of θ and Ω from coalescent simulations for a non-subdivided population. Values in parentheses are 95% confidence intervals.

$N = 10,000$, True $\theta = 5$				
Sampling Regime	True $2N\sigma = \Omega$	Mean $\hat{\theta}$	Mean $\hat{\Omega}$	$\hat{\Omega}$, Ratio of Means
30 single samples, 10 paired samples	20000	5.08 (± 0.197)	-0.213 (± 5.58)	-172 (-1197, 1551)
	200	5.05 (± 0.199)	4.65 (± 13.3)	-354 (-2584, 2398)
	2	4.99 (± 0.250)	0.430 (± 3.41)	1.99 (1.77, 2.26)
	0.2	4.83 (± 0.516)	0.420 (± 0.334)	0.195 (0.172, 0.218)
	0.1	4.91 (± 0.842)	0.370 (± 0.185)	0.0962 (0.0785, 0.115)
	0.02	2.64 (± 2.76)	0.0501 (± 0.0743)	0.0108 (0.00211, 0.0240)
10 single samples, 20 paired samples	20000	5.00 (± 0.183)	6.77 (± 9.56)	-194 (-2355, 1383)
	200	4.85 (± 0.170)	0.0502 (± 7.88)	112 (-755, 1087)
	2	5.16 (± 0.260)	4.17 (± 3.84)	2.14 (1.91, 2.41)
	0.2	5.10 (± 0.450)	0.608 (± 0.169)	0.205 (0.183, 0.229)
	0.1	4.75 (± 0.616)	0.274 (± 0.160)	0.0942 (0.0815, 0.109)
	0.02	4.27 (± 1.39)	0.224 (± 0.251)	0.0174 (0.0123, 0.0237)
25 paired samples	20000	4.87 (± 0.179)	-15.2 (± 14.3)	705 (-3322, 2394)
	200	5.02 (± 0.174)	-3.91 (± 12.0)	105 (-396, 717)
	2	5.05 (± 0.213)	2.25 (± 1.80)	2.01 (1.84, 2.21)
	0.2	4.84 (± 0.436)	0.489 (± 0.183)	0.183 (0.164, 0.203)
	0.1	4.85 (± 0.566)	0.357 (± 0.0723)	0.0965 (0.0844, 0.111)
	0.02	5.30 (± 1.54)	0.0929 (± 0.0496)	0.0213 (0.0159, 0.0289)
$N = 10,000$, True $\theta = 2$				
Sampling Regime	True $N\sigma = \Omega$	Mean $\hat{\theta}$	Mean $\hat{\Omega}$	$\hat{\Omega}$, Ratio of Means
30 single samples, 10 paired samples	20000	2.01 (± 0.0869)	1.40 (± 7.49)	539 (-1081, 2072)
	200	2.02 (± 0.0910)	2.02 (± 8.77)	156 (-1501, 1697)
	2	2.04 (± 0.110)	-1.19 (± 2.44)	2.09 (1.85, 2.38)
	0.2	1.94 (± 0.215)	3.41 (± 4.28)	0.190 (0.166, 0.218)
	0.1	1.91 (± 0.279)	0.180 (± 0.200)	0.0959 (0.0806, 0.113)
	0.02	1.08 (± 0.670)	-0.218 (± 0.546)	0.0108 (0.00457, 0.0178)
10 single samples, 20 paired samples	20000	2.02 (± 0.0887)	-35.1 (± 43.3)	125 (-1048, 1245)
	200	1.94 (± 0.0753)	0.162 (± 17.9)	135 (-827, 1417)
	2	1.97 (± 0.112)	3.00 (± 2.62)	2.18 (1.94, 2.48)
	0.2	2.16 (± 0.252)	0.680 (± 0.919)	0.224 (0.194, 0.256)
	0.1	1.67 (± 0.168)	0.417 (± 0.169)	0.0842 (0.0751, 0.0953)
	0.02	1.51 (± 0.632)	-0.109 (± 0.289)	0.0155 (0.0101, 0.0226)
25 paired samples	20000	1.99 (± 0.0856)	3.23 (± 7.49)	-215 (-2291, 1950)
	200	2.00 (± 0.0788)	-1.85 (± 4.89)	175 (-2038, 2354)
	2	2.03 (± 0.0924)	4.72 (± 2.72)	1.93 (1.78, 2.12)
	0.2	1.94 (± 0.173)	0.787 (± 0.323)	0.190 (0.169, 0.214)
	0.1	2.19 (± 0.284)	0.381 (± 0.237)	0.109 (0.0950, 0.126)
	0.02	1.88 (± 0.427)	0.0857 (± 0.0296)	0.0191 (0.0149, 0.0241)

individuals were split across either 2 or 5 equally-sized demes. The net migration rate $2N_T m$ is fixed at 1 or 0.01. For simplicity, we only simulated 25 paired samples. If the ratio of means are taken to estimate Ω and $M_e = 2N_T m / (d - 1)$, then these estimates seem to match up very well with true values, with the known exception of estimating Ω when sex is common (Table B).

Table B: Estimates of θ , Ω , and M_e from coalescent simulations in an island model. All simulations use 25 paired samples. Values in parentheses are 95% confidence intervals.

$N_T = 10,000, d = 2, \text{ True } \theta = 5, 2N_T m = M = 1$					
True $2N\sigma = \Omega$	Mean $\hat{\theta}$	Mean $\hat{\Omega}$	Mean \hat{M}_e	$\hat{\Omega}$, Ratio of Means	\hat{M}_e , Ratio of Means
20000	4.90 (± 0.180)	-5.32 (± 10.1)	1.45 (± 3.18)	-551 (-2397, 3231)	0.904 (0.838, 0.980)
200	5.07 (± 0.191)	4.17 (± 5.33)	2.63 (± 6.84)	151 (-1581, 1714)	1.05 (0.965, 1.14)
2	4.89 (± 0.224)	1.93 (± 1.58)	0.00603 (± 3.48)	1.94 (1.75, 2.18)	0.957 (0.875, 1.05)
0.2	5.69 (± 0.575)	0.598 (± 0.177)	4.10 (± 2.27)	0.224 (0.198, 0.254)	1.18 (1.03, 1.34)
0.1	4.69 (± 0.526)	0.489 (± 0.468)	2.57 (± 1.58)	0.0870 (0.0771, 0.0994)	0.939 (0.773, 1.16)
0.02	5.27 (± 1.13)	0.0655 (± 0.0122)	3.29 (± 1.10)	0.0209 (0.0167, 0.0257)	0.783 (0.572, 1.08)
$N_T = 10,000, d = 5, \text{ True } \theta = 5, 2N_T m = M = 1 (M_e = 0.25)$					
True $2N\sigma = \Omega$	Mean $\hat{\theta}$	Mean $\hat{\Omega}$	Mean \hat{M}_e	$\hat{\Omega}$, Ratio of Means	\hat{M}_e , Ratio of Means
20000	4.94 (± 0.217)	-2.07 (± 2.72)	0.322 (± 0.0167)	-707 (-1725, 1658)	0.241 (0.229, 0.253)
200	4.87 (± 0.232)	2.93 (± 4.13)	0.345 (± 0.0239)	536 (-1130, 1855)	0.246 (0.233, 0.261)
2	5.00 (± 0.295)	2.28 (± 2.14)	0.370 (± 0.0304)	1.95 (1.72, 2.24)	0.249 (0.233, 0.266)
0.2	5.11 (± 0.433)	0.427 (± 0.850)	0.404 (± 0.0366)	0.198 (0.179, 0.221)	0.253 (0.232, 0.275)
0.1	5.28 (± 0.547)	0.228 (± 0.0785)	0.427 (± 0.0379)	0.104 (0.0921, 0.117)	0.280 (0.254, 0.306)
0.02	5.43 (± 1.07)	0.0487 (± 0.00706)	0.401 (± 0.0454)	0.0220 (0.0179, 0.0264)	0.251 (0.206, 0.304)
$N_T = 10,000, d = 2, \text{ True } \theta = 5, 2N_T m = M = 0.01$					
True $2N\sigma = \Omega$	Mean $\hat{\theta}$	Mean $\hat{\Omega}$	Mean \hat{M}_e	$\hat{\Omega}$, Ratio of Means	\hat{M}_e , Ratio of Means
20000	4.71 (± 0.689)	-3.81 (± 14.3)	0.0507 (± 0.0173)	226 (-904, 540)	0.00968 (0.00831, 0.0114)
200	4.88 (± 0.928)	-11.8 (± 10.1)	0.0464 (± 0.0109)	33.1 (-212, 314)	0.0101 (0.00831, 0.0124)
2	4.52 (± 0.897)	1.44 (± 1.78)	0.0571 (± 0.0236)	1.77 (1.23, 2.67)	0.00951 (0.00757, 0.0117)
0.2	6.33 (± 1.57)	0.0914 (± 0.0772)	0.0549 (± 0.0171)	0.257 (0.191, 0.339)	0.0132 (0.0102, 0.0167)
0.1	4.77 (± 1.55)	-0.0863 (± 0.226)	0.0528 (± 0.0421)	0.0978 (0.0705, 0.135)	0.00877 (0.00610, 0.0120)
0.02	6.63 (± 2.41)	0.110 (± 0.278)	0.0845 (± 0.0436)	0.0262 (0.0172, 0.0362)	0.0130 (0.00872, 0.0183)
$N_T = 10,000, d = 5, \text{ True } \theta = 5, 2N_T m = M = 0.01 (M_e = 0.0025)$					
True $2N\sigma = \Omega$	Mean $\hat{\theta}$	Mean $\hat{\Omega}$	Mean \hat{M}_e	$\hat{\Omega}$, Ratio of Means	\hat{M}_e , Ratio of Means
20000	5.55 (± 1.81)	-7.26 (± 16.1)	0.00329 (± 0.000898)	-26.5 (-203, 166)	0.00270 (0.00189, 0.00355)
200	4.55 (± 1.76)	5.87 (± 5.65)	0.00303 (± 0.00111)	-9.67 (-96.0, 59.2)	0.00225 (0.00142, 0.00322)
2	4.40 (± 2.58)	0.640 (± 0.128)	0.00267 (± 0.00112)	1.87 (0.629, 7.80)	0.00223 (0.00110, 0.00386)
0.2	5.18 (± 2.26)	0.0199 (± 0.0544)	0.00345 (± 0.00147)	0.202 (0.119, 0.311)	0.00253 (0.00151, 0.00369)
0.1	3.70 (± 1.39)	0.0256 (± 0.0506)	0.00327 (± 0.00141)	0.0745 (0.0484, 0.107)	0.00184 (0.00124, 0.00255)
0.02	5.21 (± 3.51)	0.0287 (± 0.0297)	0.00346 (± 0.00171)	0.0217 (0.00996, 0.0407)	0.00269 (0.00125, 0.00472)

Including gene conversion. The estimators presented in Equation 10 are based on the assumption of no gene conversion. In reality, gene conversion is always present but there are not enough degrees of freedom from the two-sample model to independently estimate its rate. We hence examined how well these estimators (derived assuming no gene conversion) perform when applied to simulation data in which gene conversion is present (Table C). Consistent with the results of Equation 11 in the main text, we find that gene conversion does not have much effect on the estimates if $\gamma \ll \sigma$. However, if γ becomes large relative to σ then Ω can be badly underestimated. Hence, even if only a low rate of gene conversion is present in facultatively-sexual species, this will create an excess of homozygosity and low genotype diversity, and will cause these estimators to perform badly.

Table C: Estimates of θ and Ω from a non-subdivided population, $N = 10,000$, true $\theta = 5$, $\Omega = 0.2$ ($\sigma = 0.1/N$) and variable rates of gene conversion. All simulations use 25 paired samples. Values in parentheses are 95% confidence intervals.

Rate of Gene Conversion γ	Mean $\hat{\theta}$	Mean $\hat{\Omega}$	$\hat{\Omega}$, Ratio of Means
$0.01 \times \sigma$	5.01 (± 0.426)	0.617 (± 0.0957)	0.196 (0.178, 0.217)
$0.1 \times \sigma$	5.00 (± 0.450)	-1.42 (± 3.57)	0.247 (0.221, 0.278)
$1 \times \sigma$	4.98 (± 0.292)	0.0914 (± 5.47)	0.779 (0.709, 0.865)
$10 \times \sigma$	5.13 (± 0.219)	0.460 (± 3.90)	-4.00 (-4.26, -3.79)
$50 \times \sigma$	4.92 (± 0.182)	-2.38 (± 0.262)	-2.25 (-2.26, -2.23)
$100 \times \sigma$	5.00 (± 0.203)	-2.05 (± 0.226)	-2.12 (-2.12, -2.11)

Approximations with spatially heterogeneous rates of sex

This section outlines approximations for coalescent times if assuming that migration $m \ll 1$ and $\sigma_L \ll \sigma_H \ll 1$. Here, the between-deme mean coalescent times for two high-sex demes ($t_{D,HH}$), a high-sex and low-sex deme ($t_{D,HL}$), and two low-sex demes ($t_{D,LL}$) are obtained as:

$$\begin{aligned} E[t_{D,HH}] &\approx \frac{D_1}{f_H} + \frac{1}{f_H \sigma_H} \\ E[t_{D,HL}] &\approx E[t_{D,HH}] + \frac{D_2}{f_H} \\ E[t_{D,LL}] &\approx E[t_{D,HH}] + \frac{2D_2}{f_H} \end{aligned} \tag{11}$$

where f_H is the fraction of high-sex demes d_H/d_T , and D_1, D_2 are compound parameters:

$$\begin{aligned} D_1 &= \frac{(d_T(1 + 4mN_d) - 1)((d_T - 1)(d_H + d_T(1 + d_H)) + 2d_H d_T^2 m N_d)}{2md_T(d_T(1 + d_T(1 + 2mN_d)) - 2)} \\ D_2 &= \frac{(d_T - 1)^2}{2d_T m(d_T(1 + d_T(1 + 2mN_d)) - 2)} \approx \frac{1}{2d_T m(1 + 2mN_d)} \end{aligned} \tag{12}$$

The approximation for D_2 arises as d_T becomes large. Similarly, the approximate mean coalescent times between-individuals in a high- or low-sex deme ($t_{b,H}$ and $t_{b,L}$ respectively), and within-individual times ($t_{w,H}$ and $t_{w,L}$) can be built up as:

$$\begin{aligned}
E[t_{w,H}] &\approx 2 \left(d_T N_d \frac{D_3}{f_H} + \frac{1}{f_H \sigma_H} \right) \\
E[t_{w,L}] &\approx E[t_{w,H}] + \frac{2}{f_H} \left(\frac{d_T - 1}{2d_T m} \right) \\
E[t_{b,H}] &\approx 2d_T N_d \frac{D_3}{f_H} + \frac{1}{f_H \sigma_H} \\
E[t_{b,L}] &\approx E[t_{b,H}] + \frac{D_4}{f_H}
\end{aligned} \tag{13}$$

with D_3, D_4 being further compound parameters:

$$\begin{aligned}
D_3 &= \frac{(1 - d_T + f_H(1 - d_T^2(1 + 2mN_d)))}{2 - d_T(1 + d_T + 2d_T m N_d)} \approx f_H \\
D_4 &= \frac{2 + d_T(d_T^2 - 3)}{2d_T m(d_T(1 + d_T(1 + 2mN_d)) - 2)} \approx \frac{d_T(1 + 2mN_d) - 1}{2d_T m(1 + 2mN_d)^2}
\end{aligned} \tag{14}$$

Again, approximations are for large d_T . Furthermore, due to the recursive nature of Equation 13, elegant terms can also be obtained for the difference in mean coalescent times between high-sex and low-sex demes (given as Equation 16 in the main text):

$$\begin{aligned}
\Delta t_w &= E[t_{w,L}] - E[t_{w,H}] \approx \frac{d_T - 1}{2d_H m} \\
\Delta t_b &= E[t_{b,L}] - E[t_{b,H}] \approx \frac{2 + d_T(d_T^2 - 3)}{2d_T m(d_T(1 + d_T(1 + 2mN_d)) - 2)}
\end{aligned} \tag{15}$$

Another simple case we can consider is where there exists a single deme of each type ($d_L = d_H = 1$), and where both σ_L and σ_H are small but of the same order. Here, the difference in mean coalescent times for samples from low-sex

demes compared to those from high-sex demes becomes:

$$\begin{aligned}\Delta t_w &\approx \frac{2(1 + 2mN_d)(\sigma_H - \sigma_L)}{\sigma_H\sigma_L + 4m^2N_d(\sigma_H + \sigma_L) + 2m(\sigma_H + \sigma_L + 2N_d\sigma_H\sigma_L)} \\ \Delta t_b &\approx \frac{\sigma_H - \sigma_L}{\sigma_H\sigma_L + 4m^2N_d(\sigma_H + \sigma_L) + 2m(\sigma_H + \sigma_L + 2N_d\sigma_H\sigma_L)}\end{aligned}\tag{16}$$

In this case, we can see that the difference in expected coalescent times grows with the difference in sex rates, and decreases with migration.

Comparison with non-heterogeneous results. We can ask how average coalescent times correspond to those expected in a metapopulation without spatial heterogeneity in sex rates, but with the same average rate of sex (i.e., using $\sigma = (d_L/d_T)\sigma_L + (d_H/d_T)\sigma_H$). Figures D(a) and D(c) demonstrate that if the rate of migration and low-sex rates are not too low ($2N_Tm = 1$ and $N_T\sigma_L \approx 1$ respectively), then the average spatially heterogeneous within-individual coalescent time can be well-approximated by using a mean of sex rates. However, as both the migration rate and low-sex rates decrease, then using non-heterogeneous results with the mean rate of sex greatly underestimates the true mean coalescent time. Plotting the ratio of the coalescent times for low- and high-sex demes respectively demonstrate that coalescent times in the low-sex case increases disproportionately with reduced σ_L and m , compared to those in high-sex demes (Figures D(b), D(d)). Similar results arise for between-deme cases, as well as those for $\sigma_H = 1$ (Section D of File S1).

Simulation Results with spatially heterogeneous rates of sex

We ran coalescent simulations over a four-deme island model, either with or without heterogeneity. With heterogeneity, two of the demes had a high rate of sex, while the other two had a low rate. In a homogeneous scenario ran for comparison, the single rate of sex is that used in the heterogeneous case, averaged over demes. Note that none of these simulations consider gene conversion, which could invalidate these results if present at a sufficiently high frequency.

Table D shows that, if migration is relatively high ($2N_Tm = 1$) and both rates of sex are small ($\sigma = O(1/N_T)$), estimates of sex Ω are close to the average over demes, albeit mean values tend to be underestimates. With lower migration rates though ($2N_Tm = 0.01$), Ω estimates are substantially lower than the average over demes, and are negative on average if σ_L is very low ($0.01/N_T$). Furthermore, if the high-sex demes exhibit obligate sex ($\sigma_H = 1$), then measures of Ω predict that the population only has a low rate of sex, even though the population-wide average tends towards obligate sex. All these results are again reflective of the analytical solutions (Figure D); with spatial heterogeneity (especially with low migration rates), low-sex demes contribute disproportionately to the coalescence time.

However, if one were to instead estimate Ω separately for just the high-sex and low-sex demes, then in cases of low migration, the estimates do seem to accurately measure the true σ values used in these demes, with the additional caveat that obligate sex demes have large, noisy Ω estimates (Table E). With higher rates of migration, the two separate estimates of Ω are less accurate and appear to converge towards each other, reflecting higher rates of gene flow between distinct areas. Such partitioning demonstrates that, with low rates of migration, accurate

estimates of sex in each area are possible if there are enough samples in each deme.

Table D: Estimates of θ , Ω , and M_e from coalescent simulations with spatially heterogeneous rates of sex. All simulations used 25 paired samples with $N = 10,000$ over four demes.

Two high-sex, two low-sex demes			
Parameters Used	Mean $\hat{\theta}$	Mean $\hat{\Omega}$, Ratio of Means	Mean \widehat{M}_e , Ratio of Means
$M_e = 1/3, \sigma_L = 0.1/N, \sigma_H = 1/N$	4.80 (± 0.286)	0.810 (0.735, 0.895)	0.310 (0.289, 0.334)
$M_e = 1/3, \sigma_L = 0.01/N, \sigma_H = 1/N$	5.26 (± 0.349)	0.785 (0.712, 0.870)	0.335 (0.313, 0.361)
$M_e = 0.01/3, \sigma_L = 0.1/N, \sigma_H = 1/N$	4.12 (± 1.84)	0.321 (0.182, 0.519)	0.00278 (0.00162, 0.00419)
$M_e = 0.01/3, \sigma_L = 0.01/N, \sigma_H = 1/N$	-1.36 (± 4.02)	-0.0175 (-0.0583, 0.04259)	-0.000902 (-0.00298, 0.00220)
$M_e = 1/3, \sigma_L = 0.1/N, \sigma_H = 1$	4.79 (± 0.266)	2.43 (2.15, 2.80)	0.318 (0.297, 0.339)
$M_e = 0.01/3, \sigma_L = 0.1/N, \sigma_H = 1$	2.05 (± 0.489)	-265 (-1236, 1106)	0.00134 (0.00103, 0.00170)
$M_e = 1/3, \sigma_L = 0.01/N, \sigma_H = 1$	5.09 (± 0.268)	2.36 (2.09, 2.68)	0.322 (0.301, 0.343)
$M_e = 0.01/3, \sigma_L = 0.01/N, \sigma_H = 1$	-4.08 (± 1.16)	-0.0523 (-0.0654, -0.0361)	-0.00273 (-0.00350, -0.00190)
All demes have same rate of sex, average rates of heterogeneous case			
Parameters Used	Mean $\hat{\theta}$	Mean $\hat{\Omega}$, Ratio of Means	Mean \widehat{M}_e , Ratio of Means
$M_e = 1/3, \sigma_L = 0.1/N, \sigma_H = 1/N$	4.88 (± 0.265)	1.02 (0.935, 1.13)	0.321 (0.299, 0.342)
$M_e = 1/3, \sigma_L = 0.01/N, \sigma_H = 1/N$	4.73 (± 0.284)	0.957 (0.865, 1.06)	0.316 (0.295, 0.339)
$M_e = 0.01/3, \sigma_L = 0.1/N, \sigma_H = 1/N$	5.27 (± 1.92)	1.17 (0.664, 2.26)	0.00349 (0.00236, 0.00486)
$M_e = 0.01/3, \sigma_L = 0.01/N, \sigma_H = 1/N$	4.39 (± 2.11)	1.03 (0.481, 2.27)	0.00292 (0.00176, 0.00459)
$M_e = 1/3, \sigma_L = 0.1/N, \sigma_H = 1$	5.01 (± 0.229)	-606 (-2618, 1464)	0.336 (0.317, 0.357)
$M_e = 0.01/3, \sigma_L = 0.1/N, \sigma_H = 1$	5.10 (± 1.78)	-28.4 (-164, 195)	0.00327 (0.00230, 0.00449)
$M_e = 1/3, \sigma_L = 0.01/N, \sigma_H = 1$	4.97 (± 0.213)	434 (-1819, 2350)	0.325 (0.307, 0.344)
$M_e = 0.01/3, \sigma_L = 0.01/N, \sigma_H = 1$	5.46 (± 1.68)	-37.0 (-391, 238)	0.00365 (0.00262, 0.00487)

Table E: Estimates of Ω from the coalescent simulations in Table D, but if estimating over only the high-sex or low-sex demes respectively.

Parameters Used	Mean $\hat{\Omega}$ from high-sex demes	Mean $\hat{\Omega}$ from low-sex demes
$M = 1/3, \sigma_L = 0.1/N, \sigma_H = 1/N$	1.27 (1.10, 1.46)	0.560 (0.495, 0.640)
$M = 1/3, \sigma_L = 0.01/N, \sigma_H = 1/N$	1.37 (1.19, 1.56)	0.515 (0.454, 0.590)
$M = 0.01/3, \sigma_L = 0.1/N, \sigma_H = 1/N$	2.15 (0.862, 9.43)	0.278 (0.00241, 0.00650)
$M = 0.01/3, \sigma_L = 0.01/N, \sigma_H = 1/N$	5.58 (-64.4, 30.3)	0.0255 (0.0113, 0.0422)
$M = 1/3, \sigma_L = 0.1/N, \sigma_H = 1$	137 (-1191, 1112)	1.07 (0.931, 1.23)
$M = 0.01/3, \sigma_L = 0.1/N, \sigma_H = 1$	-265 (-859, 681)	0.211 (0.111, 0.344)
$M = 1/3, \sigma_L = 0.01/N, \sigma_H = 1$	99.5 (-587, 1100)	1.10 (0.946, 1.28)
$M = 0.01/3, \sigma_L = 0.01/N, \sigma_H = 1$	6.08 (3.64, 18.5)	0.0209 (0.0106, 0.0326)

B Additional Results

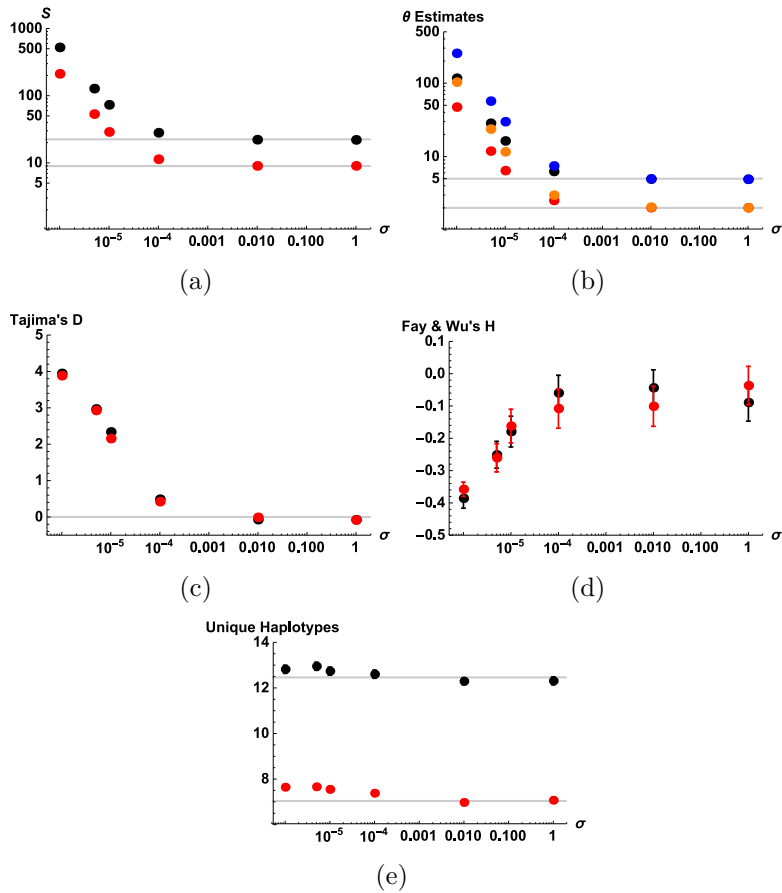


Figure A: How classic genomic summary statistics are affected by infrequent sex in a non-subdivided population, if based on *50 single samples*. All graphs are as a function of σ ; parameters are $N = 10,000$, $\theta = 5$ (black points by default) or $\theta = 2$ (red points by default), averaged over 1000 coalescent simulations. Error bars are 95% confidence intervals; if they cannot be seen, they lie within points. (a) Mean number of segregating sites in a sample; the horizontal bar shows the expectation for $\sigma = 1$ (obligate sex). (b) Estimates of θ using Watterson's θ (black or red points for $\theta = 5$ or 2), or the mean pairwise differences between pairs of samples (blue or orange points for $\theta = 5$ or 2). (c) Tajima's D. The horizontal line at zero is the (unbiased) expectation for obligate sex. (d) Fay and Wu's normalised H statistic. (e) The mean number of unique haplotypes present: the obligate-sex expectation is calculated using Equation 3.85 of EWENS (2004). Since only single samples were simulated, it was not possible to measure mean genotype diversity.

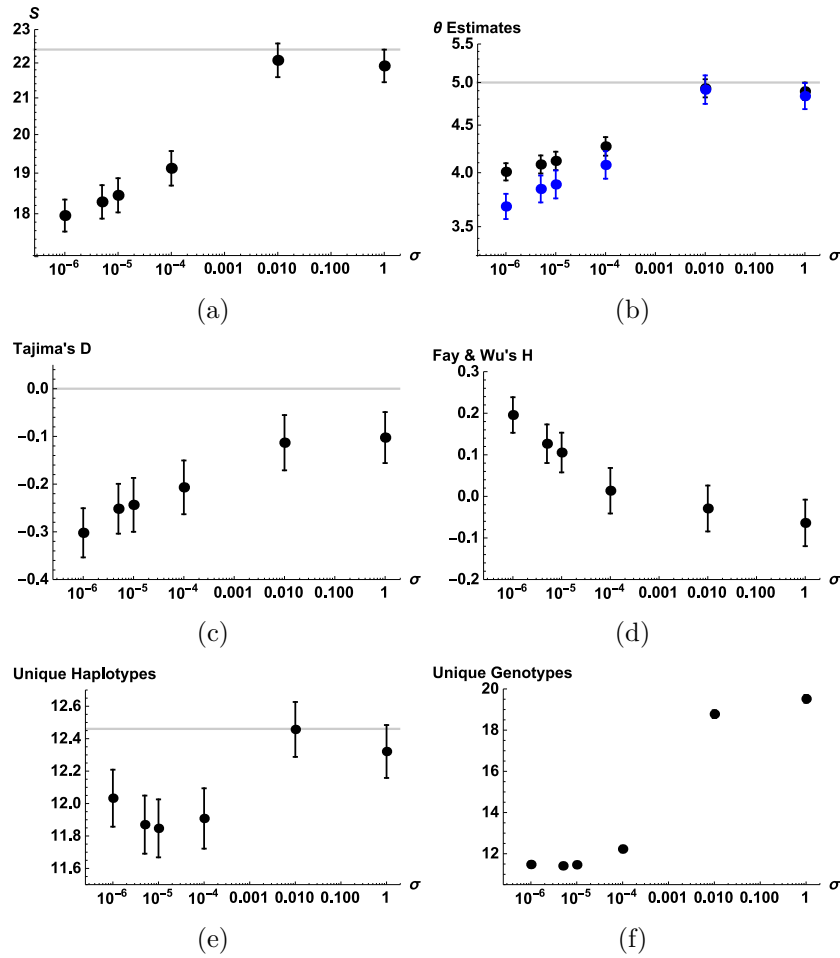


Figure B: How classic genomic summary statistics are affected by infrequent sex in a non-subdivided population, with gene conversion present at rate $\gamma = 1/N$ (or $\Gamma = 2$). All graphs are as a function of σ ; parameters are $N = 10,000$, $\theta = 5$ based on 25 paired samples, averaged over 1000 coalescent simulations. Similar results are in File S1 (Section A) for 50 individual samples. Error bars are 95% confidence intervals. (a) Mean number of segregating sites in a sample; the horizontal bar shows the expectation for $\sigma = 1$ (obligate sex). (b) Estimates of θ using Watterson's θ (black points), or the mean pairwise differences between pairs of samples (blue points). (c) Tajima's D. The horizontal line at zero is the (unbiased) expectation for obligate sex. (d) Fay and Wu's normalised H statistic. (e) The mean number of unique haplotypes present: the oblige-sex expectation is calculated using Equation 3.85 of EWENS (2004). (f) The mean number of unique genotypes present.

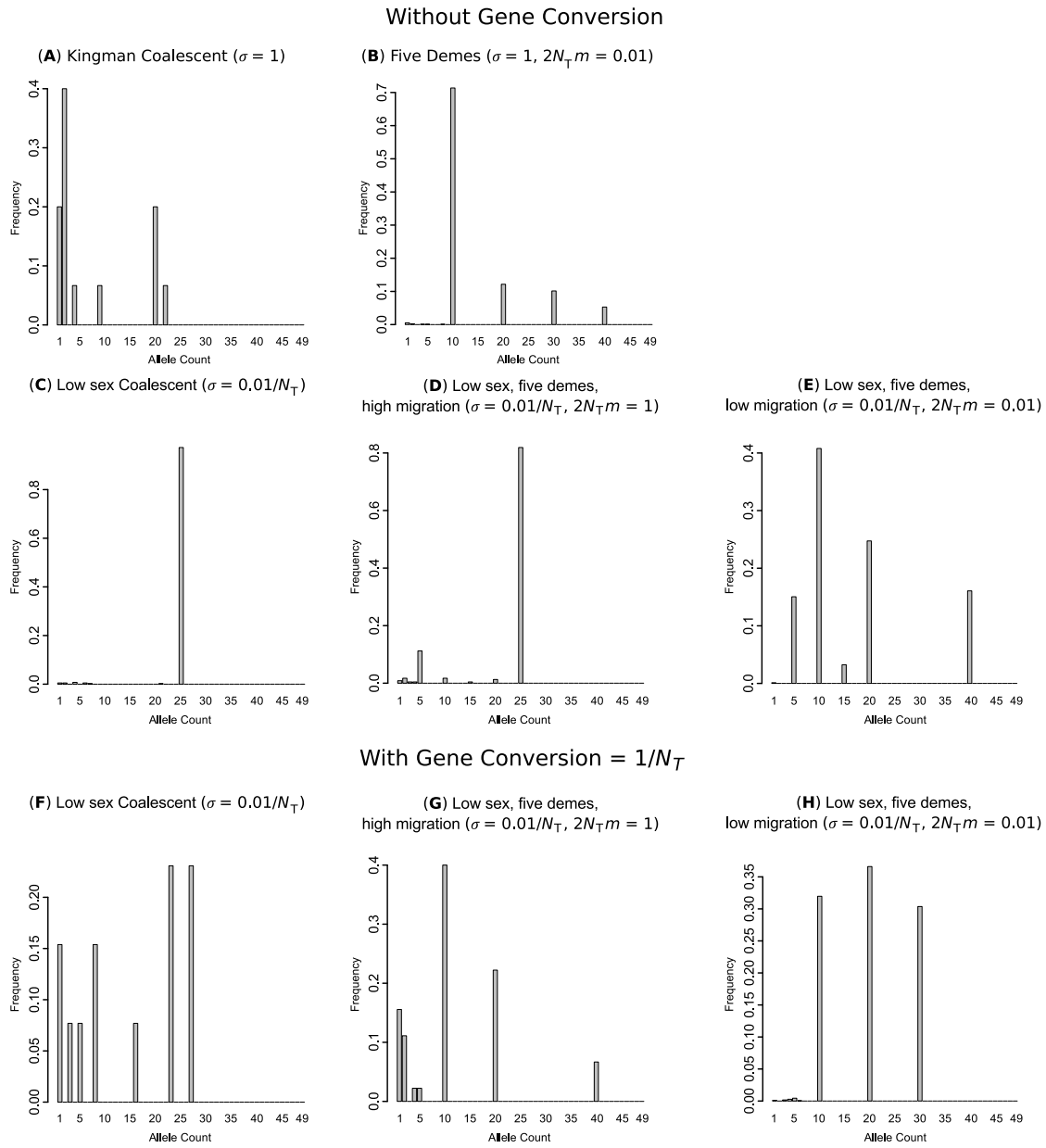


Figure C: The equivalent site frequency spectra (SFS) produced by the same simulations that created the genealogies in Figure 5 in the main text. Parameter values are the same.

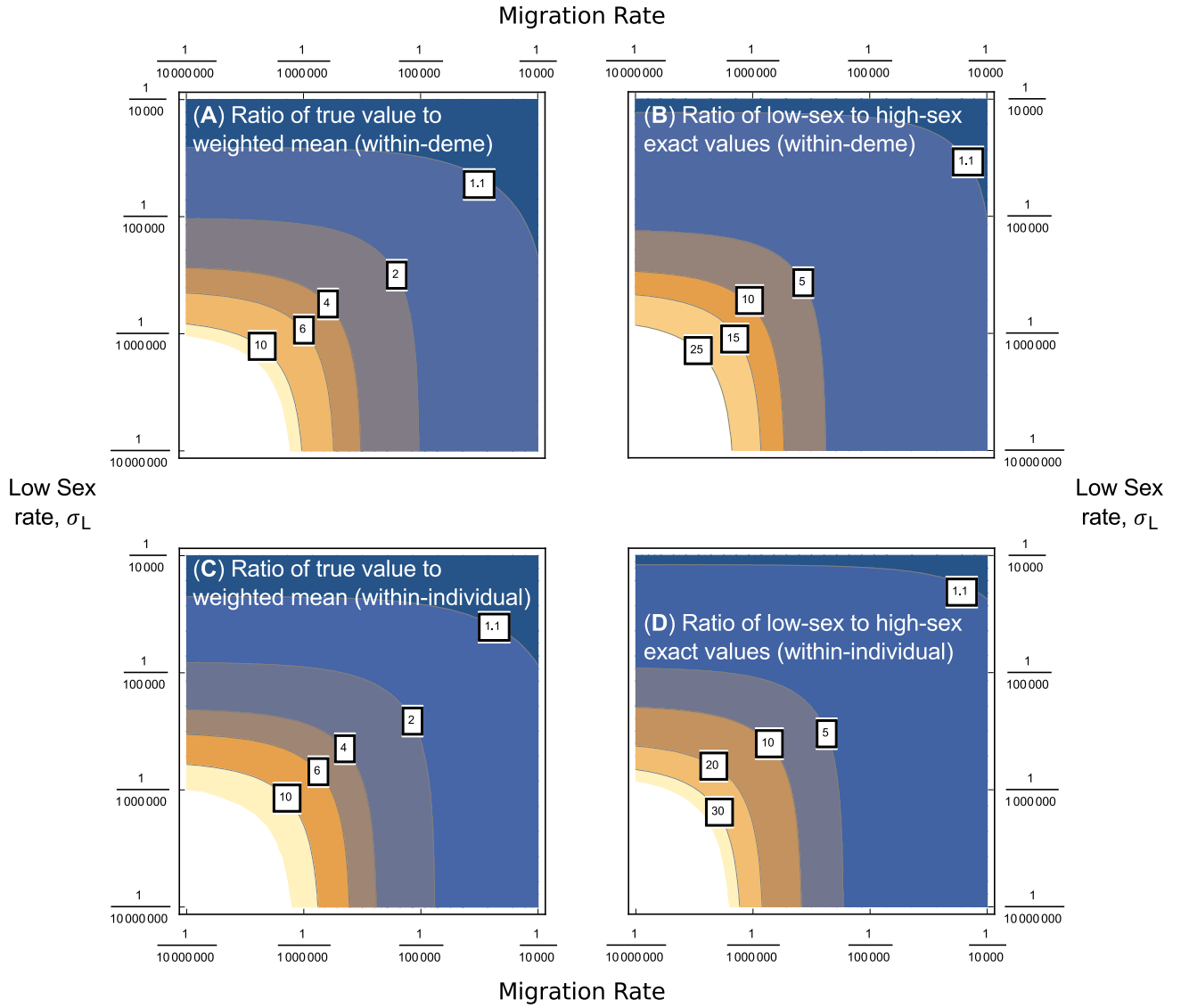


Figure D: Contour plots of the ratio of either the exact within-individual mean coalescent times in a spatially heterogeneous population to those predicted by a weighted mean of sex rates ((a) and (c)), or the ratio of the exact mean coalescent times in low-sex demes compared to high-sex demes ((b) and (d)). Plots are presented as functions of the migration rate m and the low rate of sex σ_L . Some contour labels are added for clarity. (a) and (b) are for the between-individual case, with (c) and (d) are the within-individual case. Other parameters are $N = 10,000$, $\sigma_H = 1/N$, $d_T = 4$ with $d_H = d_L = 2$.

(a) Estimators based on homozygosity

Estimators for mutation and sex rates can also be derived based on the average pairwise homozygosity in the population. That is, if two samples were picked at random, what is the probability that they are identical by descent? The basic idea is that, over a pair of branches of time t , the average number of polymorphisms is Poisson distributed with mean $2\mu t$. Hence the probability that no mutations arise between them is $e^{-2\mu t}$. This term can be included into to the probability distribution of coalescence times, to calculate the mean homozygosity. A classic result for the Kingman coalescent is that the homozygosity probability is $1/(1 + \theta)$ (HUDSON 1990).

Accordingly, one can calculate the expected between- and within-individual homozygosity in a facultative sexual population, using the unscaled mean coalescent times (see section A of File S1). For between-individual homozygosity H_b , we integrate $e^{-2\mu t}$ over an exponential distribution with mean t_b :

$$\begin{aligned} E(H_b) = \widetilde{H}_b &= \int_{t=0}^{t=\infty} \frac{e^{-\frac{t}{2N(1-\frac{1}{2N})+\frac{1}{\sigma}}}}{2N\left(1-\frac{1}{2N}\right)+\frac{1}{\sigma}} e^{-2\mu t} dt \\ &= \frac{2N\Omega}{2N(\theta + \Omega(1 + \theta)) - \theta\Omega} \end{aligned} \quad (17)$$

with $\theta = 4N\mu$ and $\Omega = 2N\sigma$ as before. In the limit of large N , Equation 17 simplifies to:

$$\widetilde{H}_b \approx \frac{\Omega}{\theta + \Omega(1 + \theta)} \quad (18)$$

Similarly, one can derive the expected within-individual homozygosity \widetilde{H}_w as

follows:

$$\begin{aligned}
\widetilde{H}_w &= \int_{t=0}^{t=\infty} \frac{e^{-\frac{t}{2N(1-\frac{1}{N})+\frac{2}{\sigma}}} e^{-2\mu t}}{2N\left(1-\frac{1}{N}\right)+\frac{2}{\sigma}} dt \\
&= \frac{N\Omega}{N(2\theta + \Omega(1 + \theta)) - \theta\Omega} \\
&\approx \frac{\Omega}{2\theta + \Omega(1 + \theta)}
\end{aligned} \tag{19}$$

Solving Equation 18 and 19 leads to estimates of Ω and θ based on homozygosity:

$$\begin{aligned}
\widetilde{\theta}_H &= \frac{2}{\widetilde{H}_b} - \frac{1}{\widetilde{H}_w} - 1 \\
\widetilde{\Omega}_H &= \frac{2\widetilde{H}_w - \widetilde{H}_b(1 + \widetilde{H}_w)}{\widetilde{H}_b - \widetilde{H}_w}
\end{aligned} \tag{20}$$

Table F shows estimates of θ and Ω using Equation 20. Irrespective of the sampling scheme, estimates are extremely poor, rarely matching up with the true values, and are often inadmissible (less than zero), especially for Ω . Unlike those based on polymorphism, measurements become worse if the ratio of means are used. As coalescent times are extended with low rates of sex, the probability of homozygosity is greatly reduced compared to obligate sex cases. As a consequence, the denominators of Equation 20 can be very low, consequently causing massively inflated estimates of θ and Ω . For this reason, homozygosity-based estimators are not recommended at all for genealogies with low rates of sex.

Table F: Estimates of θ and Ω from coalescent simulations for a non-subdivided population, using homozygosity-based estimators. Values in parentheses are 95% confidence intervals.

$N = 10,000, \text{ True } \theta = 5$					
Sampling Regime	True $2N\sigma = \Omega$	Mean $\hat{\theta}_H$	Mean $\hat{\Omega}_H$	$\hat{\theta}_H$, Ratio of Means	$\hat{\Omega}_H$, Ratio of Means
30 single samples, 10 paired samples	20000	5.92 (0.315)	-2.27 (1.58)	3.85 (3.38, 4.32)	284 (-605, 469)
	200	5.72 (0.299)	0.131 (1.73)	3.97 (3.56, 4.44)	-28.2 (-273, 163)
	2	4.57 (0.337)	-0.854 (0.487)	31.3 (26.3, 38.2)	-2.29 (-2.34, -2.24)
	0.2	6.52 (0.530)	-0.757 (0.439)	608 (384, 1146)	-2.02 (-2.03, -2.01)
	0.1	4.91 (0.404)	-1.22 (0.431)	921 (533, 2101)	-2.01 (-2.02, -2.00)
	0.02	1.73 (0.250)	-0.996 (0.00621)	8538 (3281, 35963)	-2.00 (-2.00, -2.00)
10 single samples, 20 paired samples	20000	5.08 (0.348)	-0.442 (1.62)	3.80 (3.49, 4.10)	32.9 (-152, 253)
	200	4.39 (0.346)	-1.62 (2.08)	4.04 (3.66, 4.44)	-37.7 (-469, 210)
	2	0.341 (0.443)	0.310 (2.43)	24.6 (20.4, 30.3)	-2.35 (-2.41, -2.30)
	0.2	-0.152 (0.527)	-1.05 (0.180)	201 (141, 303)	-2.04 (-2.05, -2.02)
	0.1	-0.979 (0.523)	-0.854 (0.261)	323 (163, 805)	-2.02 (-2.03, -2.02)
	0.02	-5.14 (0.593)	-0.992 (0.0146)	6169 (2489, 32478)	-2.00 (-2.00, -2.00)
25 paired samples	20000	4.60 (0.334)	-0.937 (1.36)	3.59 (3.28, 3.94)	273 (-1061, 755)
	200	4.32 (0.370)	-0.642 (1.29)	4.12 (3.78, 4.46)	-23.33 (-78.0, -14.5)
	2	-1.67 (0.536)	-1.44 (0.636)	25.8 (21.2, 31.9)	-2.32 (-2.37, -2.27)
	0.2	-4.49 (0.555)	-1.03 (0.0717)	232 (142, 425)	-2.03 (-2.05, -2.02)
	0.1	-4.39 (0.583)	-1.03 (0.0610)	417 (266, 818)	-2.02 (-2.03, -2.01)
	0.02	-5.64 (0.302)	-0.997 (0.00327)	4594 (2179, 30691)	-2.00 (-2.00, -2.00)
$N = 10,000, \text{ True } \theta = 2$					
Sampling Regime	True $2N\sigma = \Omega$	Mean $\hat{\theta}_H$	Mean $\hat{\Omega}_H$	$\hat{\theta}_H$, Ratio of Means	$\hat{\Omega}_H$, Ratio of Means
30 single samples, 10 paired samples	20000	1.92 (0.167)	0.133 (1.80)	1.33 (1.20, 1.47)	-161 (-920, 600)
	200	2.08 (0.180)	-2.09 (1.72)	1.48 (1.34, 1.63)	-68.3 (-653, 524)
	2	0.599 (0.214)	-1.54 (0.889)	5.89 (5.17, 6.81)	-2.66 (-2.76, -2.58)
	0.2	-0.274 (0.194)	-0.996 (0.259)	50.7 (37.6, 75.1)	-2.06 (-2.07, -2.05)
	0.1	0.375 (0.229)	-1.06 (0.177)	123 (84.8, 191)	-2.03 (-2.04, -2.02)
	0.02	-0.857 (0.237)	-0.988 (0.0275)	505 (236, 1981)	-2.01 (-2.01, -2.00)
10 single samples, 20 paired samples	20000	1.91 (0.186)	-1.77 (3.67)	1.53 (1.41, 1.65)	-42.5 (-327, 172)
	200	1.91 (0.178)	-1.21 (3.90)	1.58 (1.46, 1.71)	-46.7 (-449, 253)
	2	2.31 (0.378)	-1.75 (1.80)	6.04 (5.26, 7.03)	-2.59 (-2.67, -2.52)
	0.2	-4.75 (0.410)	-0.859 (0.266)	57.3 (42.5, 82.4)	-2.06 (-2.07, -2.04)
	0.1	-3.60 (0.445)	-0.968 (0.0564)	103 (64.7, 190)	-2.03 (-2.04, -2.02)
	0.02	-4.85 (0.433)	-0.981 (0.0335)	600 (253, 3325)	-2.00 (-2.01, -2.00)
25 paired samples	20000	2.05 (0.166)	0.402 (1.62)	1.35 (1.26, 1.45)	64.1 (-709, 751)
	200	2.10 (0.176)	-0.514 (1.69)	1.53 (1.42, 1.65)	-51.5 (-373, 190)
	2	-3.91 (0.494)	-0.696 (0.874)	6.82 (6.03, 7.76)	-2.56 (-2.63, -2.50)
	0.2	-6.10 (0.548)	-0.844 (0.263)	42.9 (32.7, 58.5)	-2.06 (-2.08, -2.05)
	0.1	-5.41 (0.560)	-0.984 (0.0905)	97.7 (64.7, 158)	-2.03 (-2.04, -2.02)
	0.02	-5.60 (0.449)	-0.971 (0.0478)	707 (343, 2604)	-2.01 (-2.01, -2.00)

Literature Cited

- CEPLITIS, A., 2003 Coalescence times and the Meselson effect in asexual eukaryotes. *Genet. Res.* **82**: 183–190.
- EWENS, W. J., 2004 *Mathematical population genetics : 1. Theoretical introduction*, volume 27 of *Interdisciplinary applied mathematics*. Springer, New York, second edition.
- HERBOTS, H. M., 1997 The structured coalescent. In P. Donnelly and S. Tavaré, editors, *Progress in Population Genetics and Human Evolution*, volume 87 of *IMA Volumes in Mathematics and its Applications*. Springer, New York, 231–255.
- HUDSON, R. R., 1990 Gene Genealogies and the Coalescent Process. In D. J. Futuyma and J. Antonovics, editors, *Oxford Surveys in Evolutionary Biology*, volume 7. Oxford Univ. Press, Oxford, 1–42.
- NORDBORG, M., and P. DONNELLY, 1997 The coalescent process with selfing. *Genetics* **146**: 1185–1195.
- OTTO, S. P., and T. DAY, 2007 *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press, Princeton.
- WAKELEY, J., 2009 *Coalescent theory: an introduction*, volume 1. Roberts & Company Publishers, Greenwood Village, Colorado.
- WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating F -Statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

FileS3 Simulation code, written in R. (.zip, 194 KB)

Available for download as a .zip file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178004/-/DC1/FileS3.zip