Approximating the coalescent under facultative sex

Matthew Hartfield

Institute of Evolutionary Biology, The University of Edinburgh, Edinburgh EH9 3FL, United Kingdom.

Email: m.hartfield@ed.ac.uk

Running Head: Approximating facultative sexual genealogies

Key words: Facultative sex, gene conversion, genome analysis, population genetics, coalescent theory.

Abstract

Genome studies of facultative sexual species are providing insight into the evolutionary consequences of mixed reproductive modes. However, it is unclear if the evolutionary history of facultative sexuals' genomes can be captured by standard population genetic models; in particular, whether they can be approximated by Wright–Fisher dynamics while assuming a rescaled effective population size N_e. Here, I determine when the gene genealogies of diploid facultative sexuals, which can either reproduce by sex or parthenogenesis, can be approximated by the standard coalescent, and further derive coalescent–based estimates of N_e. When sex is frequent (occurring at a frequency much greater than $1/N$, for N the actual population size), the underlying genealogy can be approximated by the standard coalescent and N_e is approximately equal to N. When sex is very rare (at frequency much lower than $1/N$), N_e can be defined for two alleles, which is independent of N and is instead influence by the frequencies of sex and mitotic gene conversion. However, the genealogy of more than two alleles does not approximate to the standard coalescent, so a general N_e cannot be defined for the entire population under rare sex. These results clarify the conditions for when the genetic history of facultative sexual populations can be captured by standard models, and sheds light on the best sampling strategies to investigate their evolutionary history.

Introduction

Facultative sex, where individuals can either reproduce sexually or asexually, is pervasive in nature (Neiman and Schwander 2011; Hartfield 2016). By switching reproduction, it is assumed that these organisms can reap the benefits of both modes (e.g., through shuffling genotypes and increasing fecundity, respectively). Genome sequence data from these organisms are currently being extensively studied to determine the evolutionary consequences of mixed reproductive modes. However, research in this area is still limited compared to the extensive body of modelling work that assumes obligate sex. Analyses of genome data from facultative sexuals will be aided if it can be determined when general population genetic models can be used to investigate the evolutionary history of these species.

The underlying determinants of genetic diversity in facultative sexual organisms has been analysed using coalescent theory (Kingman 1982; Wakeley 2009). In the earliest models (Brookfield 1992; Burt *et al.* 1996; Bengtsson 2003; Ceplitis 2003), a proportion of diploid individuals reproduce parthenogenetically, and the remainder reproduce sexually. Parthenogenesis limits the extent that haplotypes segregate among individual lineages. Under rare sex (occurring with frequency at most on the order of $1/N$ for N the population size, hereafter denoted $\mathcal{O}(1/N)$), the mean coalescent time for two alleles taken from the same locus within the same individual is much longer than that for two alleles taken from different individuals. Consequently, the increased within–individual coalescent times leads to "allelic sequence divergence" that increases heterozygosity (Mark Welch and Meselson 2000; Butlin 2002). Hartfield *et al.* (2016) subsequently introduced mitotic gene conversion into the model, which increases the frequency of within–individual coalescence.

Here, the within–individual coalescent time (and resulting diversity) can be lower that that in equivalent obligate sexuals. Given how pervasive gene conversion is during mitotic recombination (LaFave and Sekelsky 2009; Lee *et al.* 2009), then it should be a potent force in removing diversity in facultative sexuals.

Less attention has been given to determining if (and when) genealogies under facultative sex can be captured by the standard coalescent (Kingman 1982). If these approximations are available, they are attractive because the myriad models developed using the standard coalescent can then become applicable to facultative sexuals. These approximations usually arise after specifying an appropriate effective population size N_e. Initially defined by Wright (1931), N_e is defined as the population size needed for some aspect of a model to match a corresponding Wright–Fisher model with the same size. N_e influences many aspects of genetic evolution, including both the rate at which neutral alleles are lost by genetic drift (Fisher 1930; Wright 1931) and new alleles are introduced by mutation (Watterson 1975). Furthermore, for an allele with selection coefficient s, the efficacy of natural selection acting on it is determined by $N_e s$ (Kimura 1971). N_e was historically defined either based on the maximum non–unit eigenvalue of the model's transition matrix (the 'eigenvalue' N_e); the probability that two alleles are identical by descent (the 'inbreeding' N_e); or the variance in allele frequencies (the 'variance' N_e) (Whitlock and Barton 1997; Ewens 2004; Charlesworth and Charlesworth 2010). A more recent definition that has gained interest is the 'coalescent effective population size', which is the N_e needed for alleles to coalesce at the same rate as in the standard coalescent (Whitlock and Barton 1997; Laporte and Charlesworth 2002; Sjödin *et al.* 2005). Coalescent N_e values have the advantage of being relatable to genome data being analysed, as the coalescent reflects how the underlying geneal-

ogy shapes the observed genetic diversity (Sjödin *et al.* 2005). For non–standard coalescent models, N_e can be defined when there is a separation of timescales between an initial demographic shift and coalescence of remaining alleles (Sjödin *et al.* 2005). Coalescent N_e values have been obtained in the case of self–fertilisation (Nordborg and Donnelly 1997; Nordborg and Krone 2002), seed banks (Kaj *et al.* 2001), autotetraploids (Arnold *et al.* 2012), fluctuating population sizes (Sjödin *et al.* 2005), unequal sex–ratios (Wakeley 2009), and for various models of population structure (Wakeley 2004; Sjödin *et al.* 2005; Wakeley 2009). If a coalescent N_e can be defined, and shown to hold even when many alleles have been sampled, then existing tools for genome inference can be applied to genome data from facultative sexuals, after appropriately rescaling N_e. However, sometimes no single N_e can be defined and it instead depends on either what allele samples are available, or the size of specific parameters. For example, under an island model N_e differs depending on the size of the migration rate between demes (Wakeley 2004).

Previous research has elucidated the neutral forces affecting N_e under facultative sex. Orive (1993) determined that the prevalence of multiple asexual stages before the onset of sexual reproduction tended to reduce N_e (similar results were obtained by Berg and Lascoux (2000)). Conversely, Balloux *et al.* (2003) demonstrated that low rates of sex inflated N_e as among different alleles, but decreased N_e as measured over genotypes. Increased variance in asexual and sexual reproductive output can further inflate some measures of N_e (Yonezawa *et al.* 2004). However, it is unclear how mitotic gene conversion affects N_e, nor which of these measurements are applicable to coalescent models.

Hartfield *et al.* (2016) used a separation–of–timescale argument to show how sex and mitotic gene conversion affect coalescent times, if their frequencies were

both $\mathcal{O}(1/N)$. In this case, rare sex and gene conversion would shape diversity both between and within individuals on the same timescale, so the whole population's genetic history cannot be captured by a single N_e. However, this argument only covers a special case of the coalescent process. Here, I extend these separation–of–timescale arguments to show that a coalescent N_e can be defined if sex is very frequent (acting with probability much greater than $1/N$). If sex is very rare (acting with probability much less than $1/N$), it is possible to define a coalescent N_e when comparing two alleles, but it is not possible to obtain a single N_e if there are more than two alleles in the genealogy.

Methods

Using Möhle's theorem to determine coalescent N_e

The standard coalescent assumes that alleles are haploid and are exchangeable (Cannings 1974; Kingman 1982). This assumption implies that it does not matter whether allele samples are taken from the same or different individuals. After appropriate time rescaling (usually by a factor of $\mathcal{O}(N)$, e.g., by $2N$ generations in a Wright–Fisher model (Wakeley 2009)), two alleles coalesce at a rate proportional to $\binom{n}{2}$, for n the number of alleles (Table 1 outlines notation used in this analysis). Hence the time between coalescent events is exponentially distributed with rate $\binom{n}{2}$, which equals 1 for $n = 2$. Under facultative sex, the exchangeability assumption breaks down and instead a 'structured' coalescent is needed (Nordborg 1997). In particular, the genealogical history of two alleles differ if they are sampled from different individuals (hereafter 'unpaired' alleles), or from different

chromosomes within the same individual (hereafter 'paired' alleles). However, over longer timescales, it may still be the case that alleles coalesce at a steady rate. If this is the case then a coalescent N_e can be inferred by rescaling time so that coalescent events occur at the same rate as in the standard coalescent (see Nordborg and Krone (2002); Sjödin *et al.* (2005) for more formal definitions).

Symbol	Usage
N	Actual population size (with $2N$ haplotypes)
N_e	Effective population size
n	Number of sampled alleles in the coalescent process
σ	Proportion of individuals that reproduce sexually
γ	Probability of within–individual coalescence via mitotic gene conversion
F	Inbreeding coefficient
$\mathbb{T}$	Transition matrix over one generation
$\mathbb{A}$	Events in $\mathbb{T}$ that are $\mathcal{O}(1)$
$\mathbb{B}$	Events in $\mathbb{T}$ that are $\mathcal{O}(\xi)$ for $\xi \ll 1$
$\mathbb{P}$	'Short–term' outcomes that act over $\mathcal{O}(1)$ generations
$\mathbb{G}$	'Long–term' outcomes that act over $\mathcal{O}(1/\xi)$ generations
Ω	Scaled rate of sex, $2N\sigma$
Γ	Scaled rate of mitotic gene conversion, $2N\gamma$
λ	Probability of either sex or gene conversion occurring, $\lambda = \sigma + \gamma$
ϕ	Ratio of sex to gene conversion, $\phi = \sigma/\gamma$
Λ	Scaled total probability of an event, $\Lambda = N^2\lambda$
T_b, T_w	Between (within) individual coalescent time for two alleles

Table 1. Glossary of Notation.

Möhle's theorem (Möhle 1998) is often used to separate events over short and long timescales, to determine how time should be rescaled. For example, in a subdivided population with weak migration, the coalescence rate differs if the two alleles reside in the same population, or in different populations (Slatkin 1991). Let $\mathbb{T}$ represent the discrete–time transition matrix of a structured coalescent process over one generation. Further assume that $\mathbb{T}$ can be decomposed into the sum

$\mathbb{T} = \mathbb{A} + \mathbb{B}/N + o(1/N)$, where N is the total population size, and as $N \to \infty$ $\mathbb{A} = \mathbb{T}$, $\mathbb{B} = N(\mathbb{T} - \mathbb{A})$, and $o(1/N)$ are terms that approach zero faster than $1/N$ (Wakeley 2009).

Möhle (1998) proved that, if $\mathbb{T}$ can be written in this way, then there exists an analogous continuous–time rate matrix $\Pi(\tau) = \mathbb{P}e^{-\tau\mathbb{G}}$, where $\mathbb{P} = \lim_{r\to\infty} \mathbb{A}^r$ and $\mathbb{G} = \mathbb{P}\mathbb{B}\mathbb{P}$. $\mathbb{P}$ represents 'short–term' events that occur on timescales much shorter than N generations, which represent an initial adjustment to alleles in the recent past. $\mathbb{G}$ represents 'long–term' events that occur on $\mathcal{O}(N)$ generations. Time can then be rescaled so that coalescence events occur at the same long–term rate as in the standard model; coalescent N_e is subsequently inferred from this rescaling.

The above argument suggests that 'slow' events are determined by the population size N. However, Möhle's theorem can be invoked for any scaling parameter $\xi \ll 1$ by writing $\mathbb{T} = \mathbb{A} + \xi\mathbb{B} + o(\xi)$, and as $\xi \to 0$ $\mathbb{A} = \mathbb{T}$ and $\mathbb{B} = (\mathbb{T} - \mathbb{A})/\xi$. $\mathbb{G}$ then represents events that occur at timescales on $\mathcal{O}(1/\xi)$. This variant of Möhle's theorem was used by Wakeley (2004) to calculate coalescent approximations under different models of population structure, and will also be used in this analyses.

Simulations

Analytical results for N_e will be compared to stochastic simulations written in C, which were based on those used in Agrawal and Hartfield (2016). The simulation tracks a single neutral bi–allelic locus in a facultative sexual population forwards in time. Each generation, a proportion σ of individuals reproduces sexually, with offspring genotypes generated according to Hardy–Weinburg equilibrium frequencies. The remaining fraction $1 - \sigma$ of individuals reproduce as asexual clones.

Mitotic gene conversion acts with probability γ, which converts heterozygotes to homozygotes with equal probability (i.e., gene conversion is unbiased). Using these deterministic expectations, the number of genotypes among N individuals is drawn from a multinomial distribution to implement random drift. Neutral mutations are sequentially introduced, each time from a single copy. The heterozygosity $x(1-x)$ (for x the derived allele frequency) is summed over the neutral allele trajectory, until the mutation either became fixed or lost. 10 million neutral alleles were introduced and their summed heterozygosity values calculated; the mean over all introductions equals N_e/N (Charlesworth *et al.* 1993; Nordborg *et al.* 1996). Confidence intervals were calculated from 1,000 bootstraps.

Results

N_e as measured from two alleles

I consider a diploid population of size N, with individuals reproducing sexually with probability σ and asexually (parthenogenetically) with probability $1-\sigma$. Mitotic gene conversion (hereafter 'gene conversion') acts with probability γ, which causes a within–individual pair of alleles to coalesce. It is twice the probability of gene conversion affecting a single site, as there are two possible donor strands where it can initiate (Hartfield *et al.* 2018). The genealogy is determined by a structured coalescent (Bengtsson 2003; Ceplitis 2003; Hartfield *et al.* 2016), so Möhle's theorem can be applied. I will first look at two–allele results to determine the coalescent N_e as measured from them, then extend work to more than two alleles to subsequently determine if the genealogies can be approximated by the

standard coalescent. I will also relate N_e values to $F-$statistics (Wright 1951) to quantify the inbreeding coefficients under each scenario. The results can be summarised by three phases, which depend on the relative frequencies of sex and gene conversion compared to sexual reproduction. These results are summarised in Figure 1, and are as follows:

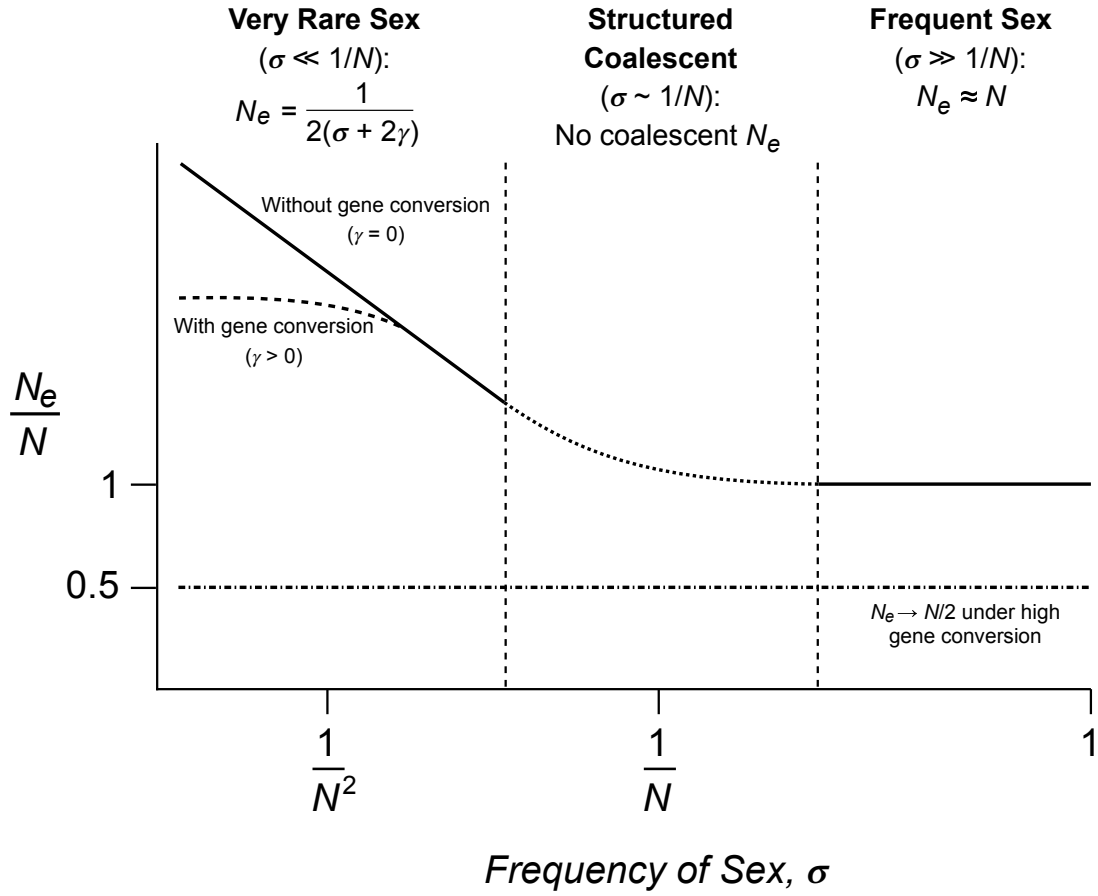

Figure 1. Schematic outline of the coalescent N_e for two alleles under facultative sex. Note that the N_e results given at the top of the figure assume that gene conversion γ acts on the same order as sexual reproduction. If gene conversion is much more frequent, then $N_e \rightarrow N/2$ as shown by the dotted–dashed line.

1. **The 'frequent sex' regime** ($\sigma \gg 1/N$).

(a) *If gene conversion is rare* ($\gamma \ll 1$), then due to the high occurrence of genetic segregation the resulting coalescent process is similar to the standard coalescent, so $N_e \approx N$.

(b) *If gene conversion is high* ($\gamma \to 1$), $N_e \to N/2$ as heterozygosity is removed.

2. **The 'structured coalescent' regime** ($\sigma \sim 1/N$).

   (a) *If gene conversion also acts with probability* $\sim 1/N$, state–transitions (i.e., whether the two alleles lie in the same or different individuals) and coalescent events occur at the same relative frequencies. Hence population diversity cannot be captured by a single N_e.

   (b) *If gene conversion is high* ($\gamma \gg 1/N$), $N_e \to N/2$, similar to the 'frequent sex' regime.

3. **The 'very rare sex' regime** ($\sigma \ll 1/N$).

   (a) *If gene conversion is also very weak* ($\gamma \ll 1/N$) then the coalescent $N_e = \frac{1}{2(\sigma+2\gamma)}$ and is hence independent of N.

   (b) *If gene conversion is much higher than sex* ($\gamma \gg 1/N$) then either no coalescent N_e exists (if $\gamma \sim 1/N$) or $N_e \to N/2$ (if $\gamma \gg 1/N$); simulations suggest that $N_e \to N/2$ in most cases.

The effects of regimes (1)(a) and (2)(a) on genealogies were described in Hartfield *et al.* (2016). Here, I more fully elucidate the role of these regimes on shaping N_e, and introduce results for the 'very low sex' regime (3).

The ‘frequent sex’ regime

The transition matrix for the facultative sex coalescent is (Hartfield *et al.* 2016, Eq. 10 without self–fertilisation):

$$\mathbb{T} = \begin{pmatrix} 1 - \frac{1}{N} & \frac{1-\gamma}{2N} & \frac{1+\gamma}{2N} \\ \sigma(1 - \frac{1}{N}) & (1-\gamma)(1-\sigma) + \frac{\sigma(1-\gamma)}{2N} & \frac{\sigma(1+\gamma)}{2N} + (1-\sigma)\gamma \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

The two alleles transition between three states, as denoted by each row: (1) each allele lies in different individuals; (2) each allele is taken from different chromosomes from the same individual; (3) the two alleles have coalesced. Figure 2 outlines a schematic of these outcomes in the coalescent process. For example, entry 3 of row 1 indicates that two alleles in different individuals coalesce if they are descended from the same allele in the parent, or if there was gene conversion in the parent. If the two alleles are taken from different alleles within an individual, they coalesce (entry 3 of row 2) if there was either sexual reproduction followed by inheritance from the same parent, or gene conversion in the absence of sex. This model assumes that a single individual can act as both parents during sexual reproduction. Hartfield *et al.* (2016) contains further details on how the transition probabilities are formed.

As a first example of how Möhle’s theorem can provide insight into how structured coalescent processes can be approximated, I partition $\mathbb{T}$ with respect to the scaling factor $1/2N$ to determine the fast and slow–rate events. Since I do not make any further assumptions on the frequencies of sex and gene conversion, then

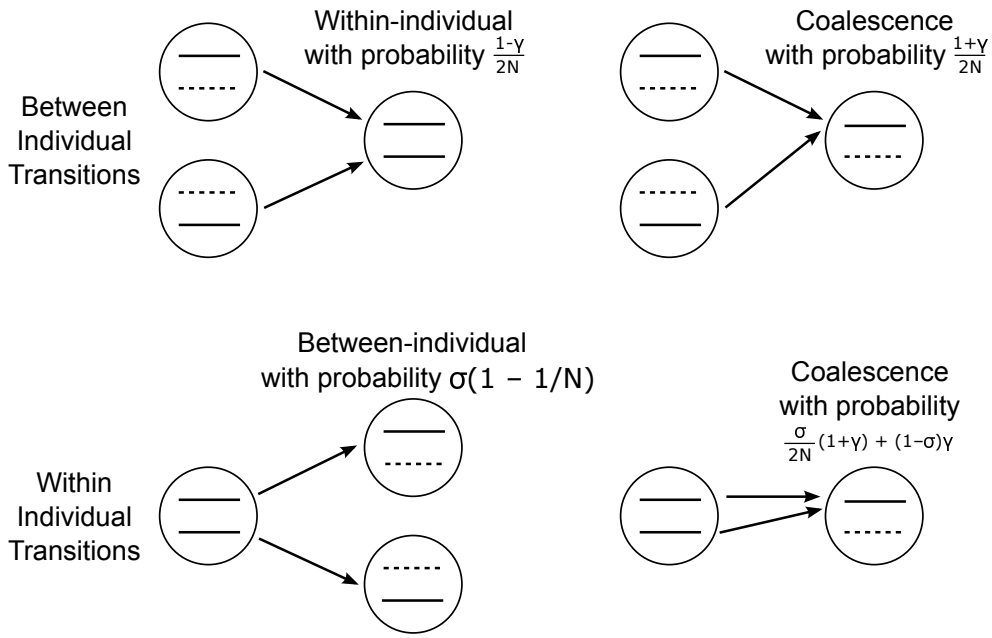

Figure 2. Schematic of the transition probabilities in the facultative sex coalescent. Sampled alleles are shown as solid lines, while dashed lines are homologous alleles that are not sampled. Figure is originally from Hartfield *et al.* (2016) and is reprinted with permission from the Genetics Society of America.

the ensuing result applies when both these events are frequent (more precisely, when σ, $\gamma \gg 1/N$). More details on all matrix calculations are available in Supplementary File S1.

$\mathbb{T}$ can be written as $\mathbb{A} + \mathbb{B}/2N$, with the sub–matrices defined as:

$$\mathbb{A} = \begin{pmatrix} 1 & 0 & 0 \\ \sigma & (1-\gamma)(1-\sigma) & (1-\sigma)\gamma \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbb{B} = \begin{pmatrix} -2 & 1-\gamma & 1+\gamma \\ -2\sigma & (1-\gamma)\sigma & (1+\gamma)\sigma \\ 0 & 0 & 0 \end{pmatrix} \quad (2)$$

Using Möhle's theorem, the short–term matrix $\mathbb{P}$ and long–term matrix $\mathbb{G}$

equal:

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\sigma}{\sigma+\gamma(1-\sigma)} & 0 & \frac{\gamma(1-\sigma)}{\sigma+\gamma(1-\sigma)} \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbb{G} = \begin{pmatrix} -\left(1+\frac{\gamma}{\sigma+\gamma(1-\sigma)}\right) & 0 & 1+\frac{\gamma}{\sigma+\gamma(1-\sigma)} \\ -\left(\frac{\sigma(\gamma(2-\sigma)+\sigma)}{(\sigma+\gamma(1-\sigma))^2}\right) & 0 & \frac{\sigma(\gamma(2-\sigma)+\sigma)}{(\sigma+\gamma(1-\sigma))^2} \\ 0 & 0 & 0 \end{pmatrix} \tag{3}$$

The effective population size N_e is inferred by inspecting $\mathbb{P}$ and $\mathbb{G}$. $\mathbb{P}$ shows that over short timescales (much less than $2N$ generations in the past), alleles will either segregate into different individuals with probability $\frac{\sigma}{\sigma+\gamma(1-\sigma)}$ or coalesce with probability $\frac{\gamma(1-\sigma)}{\sigma+\gamma(1-\sigma)}$. $\mathbb{G}$ implies that, if alleles have not coalesced, they do so over the long term with an increased rate of $\left(1+\frac{\gamma}{\sigma+\gamma(1-\sigma)}\right)$ per $2N$ generations. For a Wright–Fisher population, the coalescent timescale is obtained in the standard model by scaling time by $2N$. The coalescent N_e is obtained by equating $2N_e$ to the value needed for two alleles to coalesce at rate 1. This rescaling is obtained by setting $N_e = N/(1+\frac{\gamma}{\sigma+\gamma(1-\sigma)})$.

This result is a similar form to the general reduction in $N_e = N/(1+F)$ obtained under various forms of inbreeding (Caballero and Hill 1992), with F equal to $\frac{\gamma}{\sigma+\gamma(1-\sigma)}$. If the probability of gene conversion is low relative to the frequency of sex (i.e., $\gamma \ll 1$) then $F \approx \gamma/\sigma \ll 1$, so $N_e \approx N$. F increases with γ up to its maximum value of 1 when $\gamma = 1$. In practice this outcome is unlikely, given the low probability of gene conversion affecting a single site. For example, Sharp and Agrawal (2016) estimated a mitotic gene conversion frequency of $\sim 10^{-6}$ per basepair in *Drosophila melanogaster*.

The 'structured coalescent' regime

The next case of interest is the 'structured coalescent' regime, where $\sigma \sim \mathcal{O}(1/N)$. Here, I will also assume $\gamma \sim \mathcal{O}(1/N)$ and later relax this assumption. I approximate the transition matrix $\mathbb{T}$ (Equation 1) after dropping terms of $\mathcal{O}(1/N^2)$:

$$\mathbb{T} = \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} & \frac{1}{2N} \\ \sigma & 1 - \sigma - \gamma & \gamma \\ 0 & 0 & 1 \end{pmatrix} + \mathcal{O}(1/N^2) \tag{4}$$

I apply Möhle's theorem, again using the scaling factor $1/2N$ to determine the slow–rate matrix. In this case $\mathbb{A}$ is the identity matrix, hence $\mathbb{G}$ equals the slow–rate matrix $\mathbb{B}$:

$$\mathbb{G} = \begin{pmatrix} -2 & 1 & 1 \\ \Omega & -\Omega - \Gamma & \Gamma \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

$\Omega = 2N\sigma$ and $\Gamma = 2N\gamma$ are population–scaled rates of sex and gene conversion (note that γ is scaled by $2N$, as used in Hartfield *et al.* (2016), rather than $4N$ as in Hartfield *et al.* (2018)). A variant of Equation 5 was presented in Hartfield *et al.* (2016, Equation 5) with $\Gamma = 0$. Equation 5 shows that a within–individual pair of alleles will eventually segregate into different individuals via sex, or coalesce via gene conversion. If alleles do segregate into different individuals, then over time they will either end up in the same individual or coalesce. However, since between–

and within–individual processes occur on the same timescale [$\mathcal{O}(N)$ generations], then there is no single N_e that captures diversity for the whole population (see also the conclusions reached by Sjödin *et al.* (2005)). Further information on this structured coalescent case, including the distribution of coalescent times, is given in Hartfield *et al.* (2016).

The 'very low sex' regime

The final case of interest is where the frequency of sex is extremely low relative to the population size ($\sigma \ll 1/N$). Here I will use an alternative version of Möhle's theorem, where the slow–rate matrix is scaled by a parameter other than the population size. Specifically, I will assume both rare sex and gene conversion (i.e., σ, $\gamma \ll \mathcal{O}(1/N)$) and use $\lambda = \sigma + \gamma$ to determine the slow–rate matrix. It is also convenient to use the ratio $\phi = \sigma/\gamma$, which determines whether diploid genotypes experience allelic sequence divergence ($\phi > 1$) or convergence due to gene conversion ($\phi < 1$) (Hartfield *et al.* 2016).

After transforming σ, γ into their new variables and removing terms of $\mathcal{O}(\lambda^2)$, the transition matrix $\mathbb{T}$ can be written as $\mathbb{A} + \lambda\mathbb{B} + o(\lambda^2)$, with the sub–matrices equal to:

$$\mathbb{A} = \begin{pmatrix} 1 - \frac{1}{N} & \frac{1}{2N} & \frac{1}{2N} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbb{B} = \begin{pmatrix} 0 & -\frac{1}{2N(1+\phi)} & \frac{1}{2N(1+\phi)} \\ (1 - \frac{1}{N})(\frac{\phi}{1+\phi}) & \frac{\phi}{2N(1+\phi)} - 1 & (1 + \frac{\phi}{2N})(\frac{1}{1+\phi}) \\ 0 & 0 & 0 \end{pmatrix} \tag{6}$$

Applying Möhle's theorem to obtain $\mathbb{P}$, $\mathbb{G}$:

$$\mathbb{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbb{G} = \begin{pmatrix} 0 & -\frac{2+\phi}{4(1+\phi)} & \frac{2+\phi}{4(1+\phi)} \\ 0 & -\frac{2+\phi}{2(1+\phi)} & \frac{2+\phi}{2(1+\phi)} \\ 0 & 0 & 0 \end{pmatrix} \tag{7}$$

Here, the short–term matrix $\mathbb{P}$ shows that two alleles in different individuals will either segregate into the same individual to become within–individual alleles, or coalesce with equal probability. Two alleles from the same individual will remain as such, so it is the only remaining state. The long–term matrix $\mathbb{G}$ further shows that these alleles will coalesce at rate $\frac{2+\phi}{2(1+\phi)}$ per $1/\lambda$ generations. Hence if discrete time is scaled by $\frac{1}{\lambda}/\frac{2+\phi}{2(1+\phi)}$ then a within–individual allele pair will coalesce at rate 1 per rescaled time unit.

This N_e holds if investigating at a within–individual allele pair. However, if looking at two random allele samples then we also have to consider the between–individual coalescent process. How this change in sampling affects the results can be illustrated if looking at the distribution of coalescent times. Following the rescaling outlined above, the probability that the coalescent time T_b (if they are in different individuals), or T_w (if within the same individual), is less than τ equals:

$$\begin{aligned} P(T_b < \tau) &= \frac{1}{2} + \frac{1}{2}(1 - e^{-\tau}) \\ P(T_w < \tau) &= 1 - e^{-\tau} \end{aligned} \tag{8}$$

Recall that under the standard coalescent, two alleles coalesce according to an exponential distribution with rate 1 (Kingman 1982). So while within–individual

coalescent times follow this distribution, between–individual coalescent times are instead defined by a mixture distribution. The two alleles coalesce according to the standard coalescent half of the time, but coalesce 'instantly' the other half of the time (more specifically, on a timescale much less than $\mathcal{O}(\lambda)$), and exhibit reduced diversity. The alleles that coalesce quickly are likely those taken from the same genetic background in different individuals (i.e., both from the same set of homologous chromosomes). Hence while the mean of T_b is less than T_w, N_e is determined by focussing on the half of distribution that does approximate the standard coalescent. This argument implies that over all individuals, N_e will be reduced by half compared to the time rescaling obtained from Mohle's theorem. Thus in the very low–sex regime, $2N_e = \frac{1}{2\lambda} / \frac{2+\phi}{2(1+\phi)}$. Restoring back the σ, γ terms gives:

$$N_e = \frac{1}{2(\sigma + 2\gamma)} \tag{9}$$

The key result here is that under very low frequencies of sex, two–allele N_e is independent of the actual population size N. It can be possible to re–write Equation 9 using the compound parameters Ω, Γ to show how N_e is related to N:

$$N_e = N\left(\frac{1}{\Omega + 2\Gamma}\right) \tag{10}$$

If $\Gamma = 0$ then $N_e = \frac{N}{\Omega}$ (or $\frac{1}{2\sigma}$); given that $\Omega \ll 1$ in this regime, then rare sex will inflate N_e due to allelic sequence divergence creating extensive within–individual diversity. Further note that $N_e = N$ for $\Omega = 1$ (or $\sigma = \frac{1}{2N}$), indicating an upper limit for σ for which this calculation is valid (otherwise $N_e < N$, inconsistent with results from the high–sex regime). If $\Omega = 0$ then $N_e = \frac{N}{2\Gamma}$ (or $\frac{1}{4\gamma}$). If $\Gamma > 1/2$

then gene conversion will reduce N_e as it removes heterozygosity, similar to self–fertilisation (but gene conversion will only affect part of the chromosome). For smaller gene conversion frequencies ($\Gamma < 1/2$), N_e will still be inflated compared to N due to a lack of segregation, but gene conversion will reduce the extent of allelic divergence.

Effect of very low sex on the inbreeding coefficient, F. Under various modes of regular inbreeding, the effective population size can be written as $N_e = \frac{N}{1+F}$ (Wright 1931; Caballero and Hill 1992). By comparing the N_e results under facultative sex to this general form, the two equate if $F = 2\Gamma + \Omega - 1$ (Figure 3). If $\Omega < 1 - 2\Gamma$ then F is negative, as sex is sufficiently rare to cause some degree of allelic sequence divergence, increasing heterozygosity. F reaches its minimum of -1 when both Ω, Γ are zero. Otherwise, F is positive as gene conversion removes heterozygous sites. $F = 1$ when $\Gamma = 1 - \Omega/2$; at $\Omega = 1$, which is the upper limit at which the approximation is valid, then $\Gamma = 1/2$, subsequently implying an upper limit for gene conversion at which the N_e calculation is valid.

Effects under higher frequencies of gene conversion

Up to now, I assumed that gene conversion occurs at the same approximate frequency as sexual reproduction. If gene conversion is much more frequent than sex (i.e., $\gamma \gg \sigma$), then the resulting N_e depends on whether it occurs very frequently with $\gamma \gg 1/N$, or with $\gamma \sim \mathcal{O}(1/N)$. In the 'frequent sex model', $N_e = N/2$ if the frequency of gene conversion approaches high values. This result can also be derived using Möhle's theorem without making assumptions about the size of γ (see Appendix A). If $\gamma \sim 1/N$ the coalescent follows the 'structured' coalescent where a single N_e cannot be defined, but simulation results (see next section)

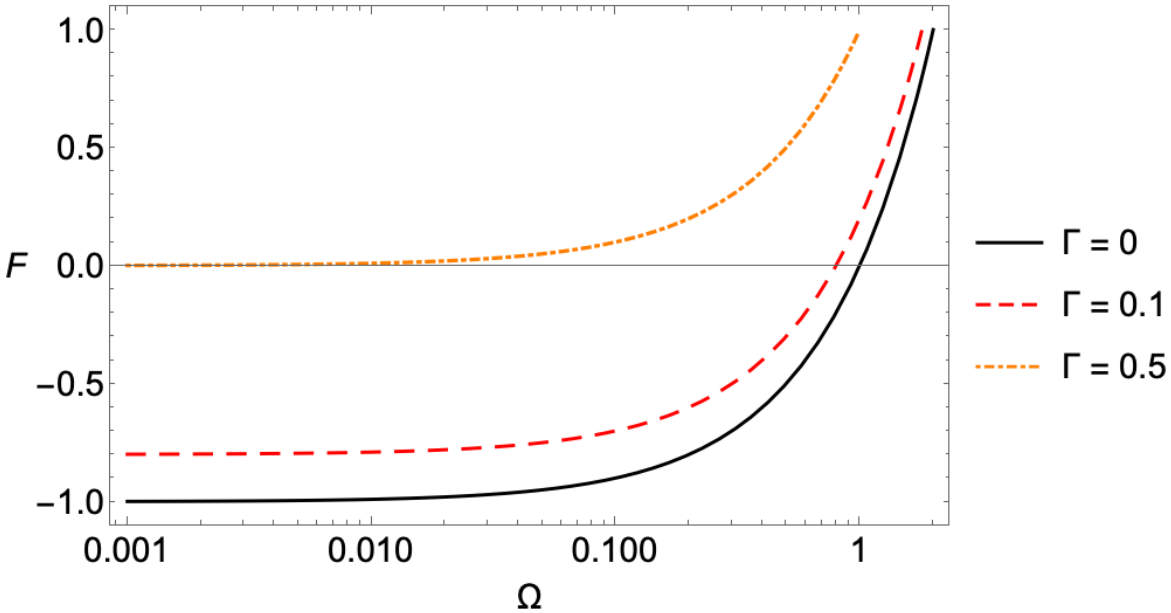

Figure 3. Inbreeding coefficient F under rare sex. Plot of $F = 2\Gamma + \Omega - 1$ as a function of $\Omega = 2N\sigma$. Different lines represent different $\Gamma = 2N\gamma$ values as shown in the legend.

suggest that N_e still tends to $N/2$ in this case.

Simulation Comparisons

Figure 4 plots N_e/N for different values of Ω and Γ, as compared to simulations. For $\Gamma \leq 0.5$ (Figure 4a) the analytical results are generally accurate, but become underestimates as Γ approaches 0.5. The analytical solutions also underestimate N_e if $\Omega \sim 1$, when a coalescent N_e cannot be defined. These results demonstrate how rare sex can substantially inflate N_e. For example, if $\Omega = 0.001$ then N_e is 1,000–fold larger than N in the absence of gene conversion. As Ω exceeds 1 then N_e converges to N.

For very high frequencies of gene conversion ($\Gamma > 0.5$; Figure 4b) N_e is generally

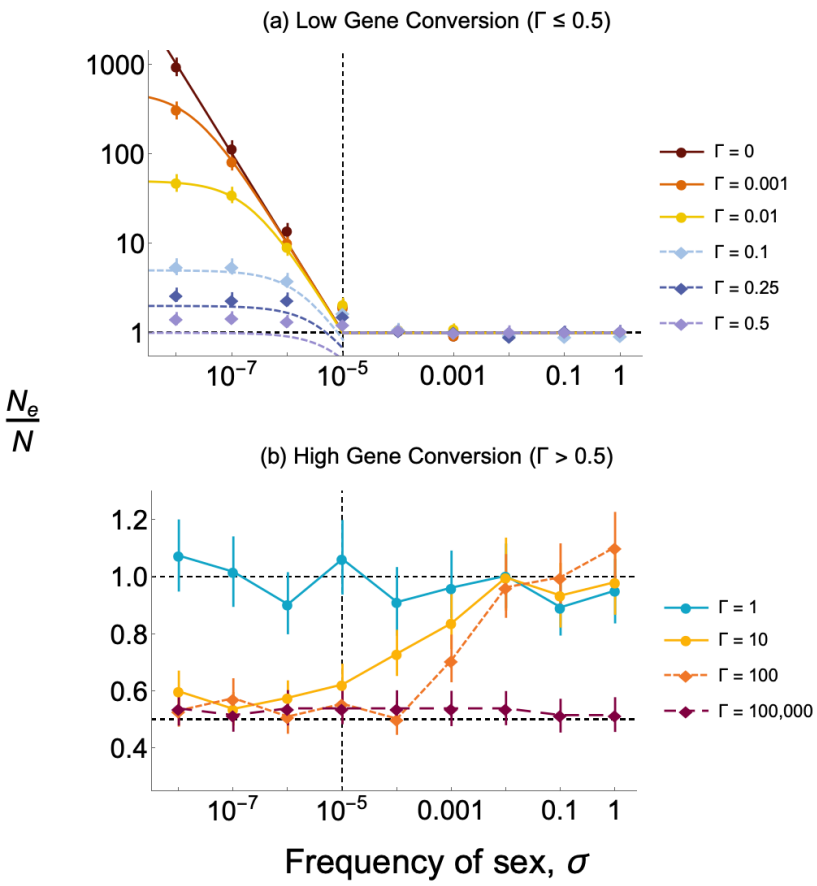

Figure 4. Simulation Comparisons of N_e as measured from two alleles. Plots of N_e/N as a function of the frequency of sex σ. Different colours represent different occurrences of gene conversion, as shown in each subplot legend. Simulations assume a diploid population of size $N = 50{,}000$. Points are simulation results, with bars representing 95% confidence intervals for the mean value (if they cannot be seen, they completely lie within the point). (a) Results for low gene conversion ($\Gamma \leq 0.5$). Solid lines are analytical approximations ($N_e = N$ for $\Omega > 1$, corresponding to $\sigma > 10^{-5}$, as shown by the vertical dashed line; $N_e/N = 1/(\Omega + 2\Gamma)$ for $\Omega < 1$). Horizontal dashed line denotes $N_e/N = 1$. (b) Results for high frequencies of gene conversion ($\Gamma > 0.5$). Horizontal dashed lines show $N_e/N = 0.5$ and 1; vertical dashed line denotes $\sigma = 10^{-5}$. Note the different $y-$axis scales in each subplot. Simulation results are also available in Supplementary File S1.

on the same order as N. As Γ increases then N_e tends to $N/2$ due to gene conversion removing heterozygosity. The amount of gene conversion needed for $N_e = N/2$ depends on the frequency of sex. For example, N_e approaches $N/2$ if $\Gamma = 10$ for $\Omega = 0.001$, but an extremely large value of $\Gamma = 100,000$ is needed in obligate sexual populations.

Extension to multiple alleles

So far, I have shown how Möhle's theorem can be used to derive a coalescent N_e under different frequencies of sex, for the simplest case of two alleles. A related question is whether these scalings apply with multiple alleles, so that the standard coalescent can be used to approximate their genealogy, and the two–allele N_e results are applicable to the entire population. Here, I outline heuristic arguments to demonstrate that this is the case when sex is frequent, but not when sex is very rare.

Multiple alleles with frequent sex

The short–term matrix $\mathbb{P}$ (Equation 3) shows that each pair of alleles from the same individual will quickly segregate out into different individuals, or coalesce (the latter being unlikely in the biologically realistic case of $\gamma \ll \sigma$). Let there be n alleles remaining in different individuals after this readjustment. The transition matrix of the subsequent coalescent process is modelled using three states: (1) n alleles are present in n distinct individuals; (2) n alleles are present in $n-1$ distinct individuals; (3) there is a coalescent event. Because sex is so frequent, I further assume that it is unlikely that n alleles will be present in less than $n-1$ individuals.

The model only considers genealogical history up to the first coalescent event. The transition matrix is the same as for the two–allele case (Equation 1) except that the first row now equals:

$$\mathbb{T}_{\texttt{row 1}} = \begin{pmatrix} 1 - \binom{n}{2}\frac{1}{N} & \binom{n}{2}\frac{1-\gamma}{2N} & \binom{n}{2}\frac{1+\gamma}{2N} \end{pmatrix} \tag{11}$$

Applying Möhle's theorem with the long–term matrix scaled by $1/2N$ gives the same $\mathbb{P}$ (Equation 3), but $\mathbb{G}$ is multiplied by a factor $\binom{n}{2}$. Hence after rescaling time by $2N/(1 + \frac{\gamma}{\sigma+\gamma(1-\sigma)})$, the coalescent rate equals $\binom{n}{2}$ as with the standard coalescent. Hence the $N_e = N/(1 + \frac{\gamma}{\sigma+\gamma(1-\sigma)})$ value obtained for two alleles also holds for n alleles, and demonstrates that this N_e value does approximate the evolutionary history of the whole population.

Multiple alleles with very low sex

To illustrate results under very low frequencies of sex, I outline a transition matrix for the case of two sets of alleles from two individuals (so there are four alleles in total). There are five states representing all different partitions of these alleles among individuals (i.e., two paired alleles; one paired and two unpaired alleles; four unpaired alleles), and possible coalescent events (two individual alleles coalesce; or two paired alleles simultaneously coalesce). As with the previous case, this model only determines the behaviour until the first coalescence event.

The transition matrix under this scenario is outlined in Appendix B; results with more alleles can be derived in a similar fashion if desired. As before, Möhle's theorem is applied with $\lambda = \sigma + \gamma$ determining long–term events. The short–term

matrix $\mathbb{P}$ becomes:

$$\mathbb{P} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{11}{12} & \frac{1}{12} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{12}$$

and $\mathbb{G} = 0$. Hence, there will be at least one coalescent event over short timescales (states 4 and 5 represent single and double coalescent events, respectively). In particular, if there are initially two sets of paired alleles then they will coalesce into a single pair by the end of the initial phase (as given in row 1). Unlike the standard coalescent, these coalescent events do not occur at regular times proportional to $\binom{n}{2}$, but much more frequently. Heuristically, this property arises since individuals (and hence pairs of alleles) coalesce with probability $\mathcal{O}(1/N)$, but the final two alleles coalesce with probability $\mathcal{O}(\lambda) \sim \mathcal{O}(1/N^2)$. This process creates a 'star–like' genealogy, with very short terminal branches, and two long internal branches representing divergence of the two remaining alleles (see Figure 5 of Hartfield *et al.* (2016)). Thus, unlike the frequent–sex case, the very–low sex coalescent process does not approximate the standard coalescent for more than two alleles. Hence, it does not appear that a single N_e value exists for the whole population

These results also shed light on why the two–allele N_e measurement becomes independent of N for very low frequencies of sex. In the recent past, all alleles will coalesce until one within–individual allele pair remains. Subsequent genetic diversity will then accumulate between these alleles; the only process that can

arrest this allelic divergence is either sexual reproduction (which is a prerequisite for coalescence), or gene conversion. The four–allele results further demonstrate that this diverged genotype will be present in all individuals in the population, so while N_e as measured within an individual will be elevated, genotype diversity will be reduced (similar results were found by Balloux *et al.* (2003); Hartfield *et al.* (2016)).

Discussion

Here, I have outlined calculations to determine when the standard coalescent approximates gene genealogies under facultative sex. I first determined coalescent N_e for the simplest case of two alleles. Explicit forms of a pairwise coalescent N_e exist for either very high or very low frequencies of sex (Figures 1, 4). If sex is common ($\sigma \gg 1/N$), then $N_e \approx N$, while for rare sex ($\sigma \ll 1/N$) N_e can be inflated due to allelic sequence divergence, and is independent of the actual population size. Gene conversion can bring N_e back towards N, and reduce it towards $N/2$ if it is very frequent. It is also possible to determine what effects rare sex has on inbreeding coefficients (Figure 3). However, I have also shown that unless sex is common, these N_e values do not hold for more than two alleles, so there does not exist a single N_e for the whole population. Hence, not all standard results for a Wright–Fisher population would carry over to one under facultative sex. For example, although within–individual N_e measurements increases under rare sex with low gene conversion, it does not necessarily follow that selection will be more efficient. This is because between–individual diversity will still be low, as most individuals will share the same set of diverged haplotypes.

The reasons why a single N_e does not exist is exemplified by how the use of Möhle's theorem differs from that in previous studies. In other cases (e.g., population structure or self–fertilisation), there is first rapid coalescence within a group [e.g., within a subpopulation under an island model with weak migration (Wakeley 2004), or within individuals under self–fertilisation (Nordborg and Donnelly 1997; Nordborg and Krone 2002)]. Coalescence then occurs under a rescaled standard coalescent among remaining alleles between groups. Under facultative sex with weak gene conversion, the opposite behaviour arises; there is rapid coalescence of alleles from different individuals, and extended coalescent times for a pair of alleles within individuals. This behaviour means that further rescaling of coalescence rates by 1/2 is needed to obtain the standard coalescent for a random set of two alleles (Equation 9). N_e is independent of N under very rare sex, and instead depends on the relative frequencies of sex and gene conversion. It is feasible that the occurrence of gene conversion is heterogeneous across a genome or individuals (LaFave and Sekelsky 2009; Lee *et al.* 2009), which will lead to varying estimates of N_e.

Given the fact that a general N_e does not exists if sex is rare, then how can genome data from facultative sexuals be best analysed? If sex is frequent then the coalescent process is similar to the standard coalescent, with a slightly adjusted N_e. It has also been previously shown that the degree to which linkage disequilibrium is broken down by meiotic recombination also scales with the frequency of sex, if it is not rare (Hartfield *et al.* 2018). Together, these results suggest that for species with facultative but frequent sex, using models based on the standard coalescent would well approximate their gene genealogies following appropriate rescaling of recombination rates. However, once sex becomes rare and individuals

start exhibiting allelic divergence (as observed in, e.g., the human pathogen *Trypanosoma brucei gambiense* (Weir *et al.* 2016)), then a coalescent N_e can only be defined for two alleles. In these scenarios, it makes most sense to analyse alleles in a pairwise manner, focussing on within–individual allelic divergence to quantify the relative impacts of sexual reproduction and gene conversion. Methods using the distribution of coalescent times from a small number of alleles (e.g., the generating–function method of Lohse *et al.* (2011)) could be particularly useful to determine how historical frequencies of sex and gene conversion vary across the genome. However, these results chiefly apply when sex is sufficiently infrequent so that allele divergence could occur; these results break down if gene conversion was common, or if the facultative sexual species is recently–derived from an obligate sexual ancestor. For more complex models, it may be necessary to use bespoke coalescent models and simulation software that explicitly consider facultative sex (Hartfield *et al.* 2016, 2018).

Note that these calculations only consider the effect of neutral processes on N_e and coalescent histories. These results can change in the presence of selection. For example, background selection reduces local N_e (and hence local diversity) (Charlesworth *et al.* 1993; Hudson 1994); its effects can be amplified in facultative sexuals due to both a lack of both recombination and segregation (Agrawal and Hartfield 2016). These calculations also do not include population or life–stage structure (Orive 1993). However, the effects of population structure can be easily incorporated into the transition matrix (Hartfield *et al.* 2016), and coalescent approximations can also be determined under different cases of the island model (Wakeley 2004). Note that no single N_e value also exists under the island model under obligate sex, so it is unlikely that one exists for structured facultative sexual

populations.

Overall, these calculations clarify how rare sex affects the genealogical history of a population, and when it can be approximated by the standard coalescent. These calculations can also be used in future modelling work of evolution under facultative sex, to both determine how neutral genetic diversity is affected, and how to best analyse and interpret genome data in other scenarios.

Acknowledgments. I would like to thank Maria Orive for inviting me to speak at the 2019 AGA Presidential Symposium "Sex and Asex: The genetics of complex life histories", and Derek Setter, Konrad Lohse and Brian Charlesworth for providing comments on the manuscript.

Funding. This work was supported by a NERC Independent Research Fellowship (NE/R015686/1).

Data availability. Simulation code, and Supplementary File S1, is available from `http://github.com/MattHartfield/FacSexNe`.

Appendix A The 'structured coalescent' with general gene conversion

To investigate the more general 'structured coalescent' case where gene conversion can act at higher frequencies, I linearise $\mathbb{T}$ without assuming that $\gamma \sim \mathcal{O}(1/N)$. In this case $\mathbb{T} = \mathbb{A} + \mathbb{B}/2N$ is equal to:

$$\mathbb{T} = \begin{pmatrix} 1 - \frac{1}{N} & \frac{1-\gamma}{2N} & \frac{1+\gamma}{2N} \\ \sigma & 1 - \sigma - \gamma + \sigma\gamma & \gamma(1-\sigma) \\ 0 & 0 & 1 \end{pmatrix} + \mathcal{O}(1/N^2) \tag{13}$$

so:

$$\mathbb{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\gamma & \gamma \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbb{B} = \begin{pmatrix} -2 & 1-\gamma & 1+\gamma \\ \Omega & -(1-\gamma)\Omega & -\gamma\Omega \\ 0 & 0 & 0 \end{pmatrix} \tag{14}$$

Applying Möhle's theorem gives the following short– and long–term matrices:

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbb{G} = \begin{pmatrix} -2 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{15}$$

Hence if $\gamma \gg 1/N$ then all within–individual alleles quickly coalesce (over the coalescent timescale), and the resulting paired alleles coalesce at twice the rate as in the standard coalescent. It follows that $N_e = N/2$.

Appendix B Very low sex with four alleles

Here, I outline the coalescent process for four alleles. The underlying transition matrix has five states: (1) two sets of paired alleles; (2) one set of paired alleles, and two unpaired alleles; (3) four unpaired alleles; (4) the number of alleles is reduced by one due to a single coalescent event; (5) the number of alleles is reduced by two due to two paired alleles coalescing (i.e., both pairs are descended from the same individual).

The transition probabilities for each state are as follows. Note that each probability is considered up to $\mathcal{O}(1/N^3)$.

From state 1:

- To state 2: occurs due to sex with probability 2σ (note the factor of two due to two sets of paired alleles).
- To state 3: requires multiple sex events of order $\sigma^2 = \mathcal{O}(1/N^4)$.
- To state 4: occurs due to gene conversion within one of the paired alleles, with probability 2γ.
- To state 5: requires the two individuals to coalesce with probability $\frac{1}{N}$.

From state 2:

- To state 1: Requires that (i) the two unpaired alleles to descend from different alleles within the same individual, and (ii) the paired allele descends from a different individual. The probability is $\frac{1}{2N}(1 - \frac{1}{N})$.
- To state 3: Requires the paired allele to segregate by sex, with probability σ.

- To state 4: Requires either (i) the two unpaired alleles to coalesce, and the paired allele to descend from a different individual, with probability $\frac{1}{2N}(1-\frac{1}{N})$; (ii) 1 unpaired allele coalescing with the paired individual with probability $\frac{2}{N}(1-\frac{1}{N})$; (iii) the paired allele experiences gene conversion with probability γ. The total probability is $\frac{1}{2N}(1-\frac{1}{N})+\frac{2}{N}(1-\frac{1}{N})+\gamma$.

- To state 5: Requires all three individuals to be descended from the same parent with probability $1/N^2$.

From state 3:

- To state 1: Requires all four alleles to place themselves in two different individuals, with each allele having a unique descendant. Probability is $\frac{1}{2N}(1-\frac{1}{N})\frac{1}{2N}=\frac{1}{4N^2}+\mathcal{O}(1/N^3)$.

- To state 2: Requires two alleles to descend from the same parent, and other two from different parents. The total probability is $\frac{1}{2N}(1-\frac{1}{N})(1-\frac{2}{N})=\frac{1}{2N}(1-\frac{3}{N})+\mathcal{O}(1/N^3)$.

- To state 4: Has the same probability as the transition to state 2, as $\frac{1}{2N}$ denotes the probability of two alleles descending from the same ancestor.

- To state 5: Requires two individual coalescent events with probability $\frac{1}{4N^2}$.

States 4 and 5 absorbing.

Hence the transition matrix $\mathbb{T} = \mathbb{A} + \lambda\mathbb{B}$ with:

$$\mathbb{A} = \begin{pmatrix} 1-\frac{1}{N} & 0 & 0 & 0 & \frac{1}{N} \\ \frac{1}{2N} & 1-\frac{3}{N} & 0 & \frac{5}{2N} & 0 \\ 0 & \frac{1}{2N} & 1-\frac{1}{N} & \frac{1}{2N} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad \mathbb{B} = \begin{pmatrix} -2 & \frac{2\phi}{1+\phi} & 0 & \frac{2}{1+\phi} & 0 \\ -\frac{1}{2\Lambda} & \frac{2}{\Lambda}-1 & \frac{\phi}{1+\phi} & -\frac{5}{2\Lambda}+\frac{1}{1+\phi} & \frac{1}{\Lambda} \\ \frac{1}{4\Lambda} & -\frac{3}{2\Lambda} & \frac{5}{2\Lambda} & -\frac{3}{2\Lambda} & \frac{1}{4\Lambda} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{16}$$

where $\Lambda = N^2\lambda$. Applying Möhle's theorem gives Equation 12.

References

Agrawal, A. F. and Hartfield, M. 2016. Coalescence with background and balancing selection in systems with bi- and uniparental reproduction: Contrasting partial asexuality and selfing. *Genetics*, 202(1): 313–326.

Arnold, B., Bomblies, K., and Wakeley, J. 2012. Extending coalescent theory to autotetraploids. *Genetics*, 192(1): 195–204.

Balloux, F., Lehmann, L., and de Meeûs, T. 2003. The population genetics of clonal and partially clonal diploids. *Genetics*, 164(4): 1635–1644.

Bengtsson, B. O. 2003. Genetic variation in organisms with sexual and asexual reproduction. *J. Evol. Biol.*, 16(2): 189–199.

Berg, L. and Lascoux, M. 2000. Neutral genetic differentiation in an island model with cyclical parthenogenesis. *J. Evol. Biol.*, 13(3): 488–494.

Brookfield, J. F. Y. 1992. DNA fingerprinting in clonal organisms. *Mol. Ecol.*, 1(1): 21–26.

Burt, A., Carter, D. A., Koenig, G. L., White, T. J., and Taylor, J. W. 1996. Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA*, 93(2): 770–773.

Butlin, R. 2002. The costs and benefits of sex: new insights from old asexual lineages. *Nat. Rev. Genet.*, 3(4): 311–317.

Caballero, A. and Hill, W. G. 1992. Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics*, 131(2): 493–507.

Cannings, C. 1974. The latent roots of certain Markov chains arising in genetics: A new approach, I. Haploid models. *Adv. Appl. Probab.*, 6(2): 260–290.

Ceplitis, A. 2003. Coalescence times and the Meselson effect in asexual eukaryotes. *Genet. Res.*, 82(03): 183–190.

Charlesworth, B. and Charlesworth, D. 2010. *Elements of Evolutionary Genetics*. Roberts & Company Publishers, Greenwood Village, Colo.

Charlesworth, B., Morgan, M. T., and Charlesworth, D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134(4): 1289–1303.

Ewens, W. J. 2004. *Mathematical population genetics : 1. Theoretical introduction*, volume 27 of *Interdisciplinary applied mathematics*. Springer, New York, second edition.

Fisher, R. A. 1930. *The genetical theory of natural selection*. The Clarendon Press, Oxford.

Hartfield, M. 2016. Evolutionary genetic consequences of facultative sex and outcrossing. *J. Evol. Biol.*, 29(1): 5–22.

Hartfield, M., Wright, S. I., and Agrawal, A. F. 2016. Coalescent times and patterns of genetic diversity in species with facultative sex: Effects of gene conversion, population structure, and heterogeneity. *Genetics*, 202(1): 297–312.

Hartfield, M., Wright, S. I., and Agrawal, A. F. 2018. Coalescence and linkage disequilibrium in facultatively sexual diploids. *Genetics*, 210(2): 683–701.

Hudson, R. R. 1994. How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? *Proc. Natl. Acad. Sci. USA*, 91(15): 6815–6818.

Kaj, I., Krone, S. M., and Lascoux, M. 2001. Coalescent theory for seed bank models. *J. Appl. Prob.*, 38(2): 285–300.

Kimura, M. 1971. Theoretical foundation of population genetics at the molecular level. *Theor. Popul. Biol.*, 2(2): 174–208.

Kingman, J. F. C. 1982. On the genealogy of large populations. *J. App. Prob.*, 19: 27–43.

LaFave, M. C. and Sekelsky, J. 2009. Mitotic Recombination: Why? When? How? Where? *PLoS Genet.*, 5(3): e1000411.

Laporte, V. and Charlesworth, B. 2002. Effective population size and population subdivision in demographically structured populations. *Genetics*, 162(1): 501–519.

Lee, P. S., Greenwell, P. W., Dominska, M., Gawel, M., Hamilton, M., and Petes, T. D. 2009. A Fine-Structure Map of Spontaneous Mitotic Crossovers in the Yeast *Saccharomyces cerevisiae*. *PLoS Genet.*, 5(3): e1000410.

Lohse, K., Harrison, R. J., and Barton, N. H. 2011. A general method for calculating likelihoods under the coalescent process. *Genetics*, 189(3): 977–987.

Mark Welch, D. B. and Meselson, M. 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, 288(5469): 1211–1215.

Möhle, M. 1998. A Convergence Theorem for Markov Chains Arising in Population Genetics and the Coalescent with Selfing. *Adv. Appl. Probab.*, 30(2): 493–512.

Neiman, M. and Schwander, T. 2011. Using parthenogenetic lineages to identify advantages of sex. *Evol. Biol.*, 38(2): 115–123.

Nordborg, M. 1997. Structured coalescent processes on different time scales. *Genetics*, 146(4): 1501–1514.

Nordborg, M. and Donnelly, P. 1997. The coalescent process with selfing. *Genetics*, 146(3): 1185–1195.

Nordborg, M. and Krone, S. M. 2002. Separation of time scales and convergence to the coalescent in structured populations. *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot*, pages 194–232.

Nordborg, M., Charlesworth, B., and Charlesworth, D. 1996. The effect of recombination on background selection. *Genet. Res.*, 67: 159–174.

Orive, M. E. 1993. Effective population size in organisms with complex life-histories. *Theor. Popul. Biol.*, 44(3): 316–340.

Sharp, N. P. and Agrawal, A. F. 2016. Low genetic quality alters key dimensions of the mutational spectrum. *PLoS Biol.*, 14(3): e1002419.

Sjödin, P., Kaj, I., Krone, S., Lascoux, M., and Nordborg, M. 2005. On the meaning and existence of an effective population size. *Genetics*, 169(2): 1061–1070.

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

Wakeley, J. 2004. Recent Trends in Population Genetics: More Data! More Math! Simple Models? *J. Hered.*, 95(5): 397–405.

Wakeley, J. 2009. *Coalescent theory: an introduction*, volume 1. Roberts & Company Publishers, Greenwood Village, Colorado.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.*, 7(2): 256–276.

Weir, W., Capewell, P., Foth, B., Clucas, C., Pountain, A., Steketee, P., Veitch, N., Koffi, M., De Meeûs, T., Kaboré, J., Camara, M., Cooper, A., Tait, A., Jamonneau, V., Bucheton, B., Berriman, M., MacLeod, A., and Soldati-Favre, D. 2016. Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *eLife*, 5: e11473.

Whitlock, M. C. and Barton, N. H. 1997. The effective size of a subdivided population. *Genetics*, 146(1): 427–441.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics*, 16(2): 97–159.

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

Yonezawa, K., Ishii, T., and Nagamine, T. 2004. The effective size of mixed sexually and asexually reproducing populations. *Genetics*, 166(3): 1529–1539.