Invasion and Extinction Dynamics of Mating Types Under Facultative Sexual Reproduction

Peter Czuppon*[,†,1,2] and George W. A. Constable[‡,1]

*Center for Interdisciplinary Research in Biology, CNRS, Collège de France, PSL Research University, Paris, 75231 France, [†]Institute of Ecology and Environmental Sciences of Paris, Sorbonne Université, UPEC, CNRS, IRD, INRA, Paris, 75252 France and [‡]Department of Mathematical Sciences, The University of Bath, BA2 7AY, United Kingdom

ORCID IDs: 0000-0003-1462-7237 (P.C.); 0000-0001-9791-9571 (G.W.C.)

ABSTRACT In sexually reproducing isogamous species, syngamy between gametes is generally not indiscriminate, but rather restricted to occurring between complementary self-incompatible mating types. A longstanding question regards the evolutionary pressures that control the number of mating types observed in natural populations, which ranges from two to many thousands. Here, we describe a population genetic null model of this reproductive system, and derive expressions for the stationary probability distribution of the number of mating types, the establishment probability of a newly arising mating type, and the mean time to extinction of a resident type. Our results yield that the average rate of sexual reproduction in a population correlates positively with the expected number of mating types observed. We further show that the low number of mating types predicted in the rare-sex regime is primarily driven by low invasion probabilities of new mating type alleles, with established resident alleles being very stable over long evolutionary periods. Moreover, our model naturally exhibits varying selection strength dependent on the number of resident mating types. This results in higher extinction and lower invasion rates for an increasing number of residents.

KEYWORDS mating types; isogamy; facultative sex; self-incompatibility; balancing selection; frequency-dependent selection; finite populations

I N isogamous species, the gamete size differentiation that defines the sexes in anisogamous species is absent (Lehtonen *et al.* 2016). Despite their morphological similarity, the gametes of isogamous species are typically not interchangeable, but rather fall into one of a number of genetically determined self-incompatible gamete classes, termed mating types. Syngamy can only occur between gametes of distinct and complementary mating types.

The evolutionary explanation for this self-incompatibility (SI), which limits the number of potential mates available to an isogamous organism, has been the subject of debate, with a number of competing hypotheses proposed (Billiard *et al.* 2011). These include theories that such SI alleles limit inbreeding depression (Charlesworth and Charlesworth 1979), increase encounter rates between gamete pairs

(Hoekstra 1982; Hadjivasiliou *et al.* 2015; Hadjivasiliou and Pomiankowski 2019), allow for ploidy level detection and the instigation of the zygote developmental program (Haag 2007; Perrin 2012), or manage cytoplasmic conflict by promoting uniparental inheritance of organelles (UPI) (Hurst and Hamilton 1992; Hadjivasiliou *et al.* 2012).

Empirical observations of the number of mating types vary between species, ranging from two to many thousands (Kothe 1996). For example, intermediate numbers of mating types are reported in ciliates of the *Tetrahymena*-species [$3 - 9$ different mating types (Doerder *et al.* 1995; Phadke and Zufall 2009)] and slime molds [$2 - 13$ mating types (Bloomfield *et al.* 2010; Clark and Haskins 2010)]. Larger numbers can be found in fungal populations; in *Coprinellus disseminatus* the global population is estimated to contain 123 different mating types (James *et al.* 2006), while *Schizophyllum commune* has a staggering $23,328$ distinct types (Kothe 1999). This naturally leads to the question which type of evolutionary pressures govern this diversity in mating type number? This has been the subject of much debate, motivated in part by a discrepancy between these empirical observations and simple evolutionary reasoning.

From a theoretical standpoint, one might naïvely expect to see a very large number of mating types within any given species, due to the "rare sex advantage" of novel types (Iwasa and Sasaki 1987). Since mating types are self-incompatible, rare types have more opportunities for mating, and, thus, each type experiences negative-frequency dependent selection. Therefore, each novel mating type produced by mutation should establish in the population, and the number of mating types should consistently grow. However, this prediction stands in stark contrast to what is observed in the natural world. Although isogamous species with hundreds, or even thousands, of mating types are possible (Kothe 1996), examples of such species are very rare; the vast majority have very few (typically two) mating types (Hadjivasiliou 2014). Explaining this discrepancy between theory and empirical observation has been the focus of much work, and multiple theories have been proposed (Billiard *et al.* 2011).

One prominent hypothesis is that UPI drives the evolution of two mating types (Hadjivasiliou *et al.* 2012), with larger mating type numbers becoming less stable with the increased complexity of coordinating an organelle donor-receiver program (Hurst and Hamilton 1992). While recent modeling work has shown that, when the frequency-dependent effects of UPI are accounted for, an invading donor-receiver program does not reduce the expected number of mating types (Hadjivasiliou *et al.* 2013), perceived empirical support for the theory comes from the fact that many species with more than two mating types have developed mechanisms to ensure homoplasmy without such a program. For instance, in *Paramecium bursaria*, with up to eight mating types (Phadke and Zufall 2009), and *S. commune*, sexual reproduction is achieved following the exchange of nuclei between cells without cytoplasmic mixing (Birky 1995). However, this in turn leads to an opposing question: if it really is only a donor-receiver program that limits the number of mating types, why do ciliates and Agaricomycetes, with these alternative methods for ensuring homoplasmy, still feature species with very few types [as low as two and three respectively, see James (2015)]?

A further hypothesis (Hoekstra 1982; Hoekstra *et al.* 1991), with renewed attention (Hadjivasiliou and Pomiankowski 2016), suggests that it is cell–cell signaling between gametes that limits the number of mating types. Here coevolution between two resident types has been shown to potentially limit the evolutionary success of a third mutant mating type. If derived from a resident type, this new mutant must essentially cross a fitness valley before it can develop encounter rates with the residents comparable to their current pairwise encounter rate, limiting its invasion potential (Hadjivasiliou and Pomiankowski 2016).

In each of these hypotheses, a biologically plausible emergent mechanism is sought that generates a selective advantage to a pair of mating types, thus limiting their number to two. A notable exception can be found in Iwasa and Sasaki (1987). Here, it was demonstrated that, under certain dynamics for the mating type encounter rate, the advantage to rare mating types could be suppressed. In the limit of infinitely long-lived gametes (that can always survive until a suitable partner becomes available), selection for mating type numbers greater than two could be eliminated entirely. It was verbally suggested that, under such a scenario, genetic drift would purge new mating types, limiting their number. This led to a bimodal prediction for the number of types; populations would either have two (given a particular set of encounter rate dynamics and immortal gametes), or have infinitely many otherwise (with each at infinitely low frequency).

While the conditions required for two mating types in Iwasa and Sasaki (1987) were stringent (and, indeed, the possibility of intermediate numbers of mating types impossible) it was notable for suggesting that genetic drift may have a key role to play in determining mating type number. In a similar vein, recent work has emphasized the relevance of finite population size null models (*i.e.*, models in which all mating types are phenotypically similar) for addressing the distribution of mating type numbers observed in nature (Constable and Kokko 2018; Czuppon and Rogers 2019). These studies stress that, even in the absence of species-specific biological processes, the number of mating types in any real finite population cannot be infinite. Instead, the expected number of types will arise from balance between mutations (which introduce new mating types) and extinctions (which decrease the number of types), leading to a number of mating types well below the studied population size.

As an explanation for the low number of mating types often observed in isogamous species, this mutation-extinction balance hypothesis may seem at first improbable, particularly in the light of a number of classic population genetic studies of SI alleles in plants (Wright 1939, 1960, 1964; Ewens 1964; Nagylaki 1975; Yokoyama and Nei 1979; Yokoyama and Hetherington 1982), see also Clark and Kao (1994) for a review. From a modeling perspective, the dynamics of SI alleles in gametophytic species, those where SI is determined at the haploid pollen stage (Bod'ová *et al.* 2018), closely resemble those of SI mating type alleles in isogamous species. [A comparison with the dynamics of SI alleles in sporophytic species, in which SI of the haploid pollen is determined by the diploid parent, can be more complicated due to the complex dominance relationships that regulate SI in some of these species (Thompson and Taylor 1966; Prigoda *et al.* 2005; Billiard *et al.* 2007)]. In a seminal paper, Wright (1939) predicted that, for the model plant *Oenothera organensis*, a population of ∼500 individuals could sustain ∼13 SI alleles (see also Crosby 1966). As the number of predicted SI alleles would rise with increasing effective population size (which intuitively reduces genetic drift and hence extinction rates), many more SI mating type alleles might naïvely be expected in isogamous species, for which effective population sizes can be of the order of 10^6 (Baranova *et al.* 2015).

While, in the above context, the plausibility of the mutation-extinction balance hypothesis might seem doubtful, an important biological feature of isogamous species with less prominence in plants is the potential to reproduce asexually. In many isogamous species, long periods of asexual reproduction are punctuated by rare bouts of facultative sex. For instance, in the single-celled green alga, *Chlamydomonas reinhardtii* [with two mating types (Goodenough *et al.* 2007)], recent genomic estimates have placed the rate of sexual reproduction to be once in every 770 asexual generations (Hasan and Ness 2018). Sex in yeast, which also typically have two mating types (Butler 2007), appears to be rarer still, with estimates of once in every 1000–3000 asexual generations reported in some species (Tsai *et al.* 2008).

In Hadjivasiliou *et al.* (2016), a computational model was used to show that long periods of asexual reproduction would lead to substantial drift in mating types frequencies. Using a population genetic model, Constable and Kokko (2018) demonstrated further that, as sex becomes increasingly rare, the relative strength of genetic drift over selection for novel mating types is amplified. This leads to a lower expected number of mating types in mutation-extinction balance [a similar observation has been made of Solanaceae gametophytic SI alleles (Vallejo-Marín and Uyenoyama 2008)]. This was true even in the absence of any species-specific selective mechanisms. A survey of available empirical data further supported the view that the rate of facultative sex is a key predictor of the number of mating types in isogamous species.

Analytic bounds on the number of expected mating types under a mutation-extinction balance were calculated in Constable and Kokko (2018), using the stationary distribution of a simple Moran-type model of mating type dynamics. In Czuppon and Rogers (2019), an approximation of the expected number of mating types was calculated under the assumption that sex is obligate. Muirhead and Wakeley (2009) used a similar model to calculate the stationary distribution of the frequency spectrum of mating type alleles; however, estimates on the number of expected mating types were not explicitly calculated. In addition, facultative sex was not modeled (although the mathematical analysis employed can account for this factor).

While such estimates of the number of mating type alleles in the stationary distribution are informative, they obscure the dynamic processes that drive and maintain this equilibrium: the establishment and extinction of alleles. These quantities are important for two reasons.

First, they are both independent of the arrival rate of new SI alleles. Null models of mating type dynamics often equate the arrival rate of new SI alleles with an effective mutation rate, and, in doing so, take a course grained approach to the genetic, molecular, and physiological mechanisms underlying mating type determination. However, this effective mutation rate is a confounding parameter, as it is difficult to estimate empirically. [A similar problem was encountered by Wright, who conceded that there could be argument about what a "reasonable" rate should be (Wright 1939)]. In contrast, the effective mutation rate affects neither the establishment probability of a mutant mating type nor the expected extinction time of a resident type. These quantities thus provide mutation rate independent measures of the evolutionary dynamics of mating types that can be used to test the feasibility of the mutation-extinction hypothesis.

Second, whereas the stationary distribution of mating type number can be observed only over evolutionary time periods, the establishment probability of new mating types and the mean extinction time of mating types can be observed over significantly shorter time periods. These quantities are thus important for providing empirically testable insight into the evolution of mating type number.

Extinction rates have previously been studied in related systems featuring negative frequency-dependent selection, which results in high levels of polymorphism. In Takahata (1990) a one-dimensional stochastic diffusion was used to identify the dynamics of the polymorphic major histocompatibility complex (MHC) system with a time rescaled neutral coalescent. This allowed for estimating the extinction rates of MHC alleles. These results were verified numerically in Takahata and Nei (1990) and later in Slatkin and Muirhead (1999). The same theoretical approach was adapted to gametophytic SI in plants, obtaining results showing that the time to the most recent common ancestor of a sex allele might even exceed the speciation time (Vekemans and Slatkin 1994). In the specific context of mating type alleles, such results have thus far been absent. However, recently, Czuppon and Rogers (2019) calculated the establishment probability of newly arising mating type for populations in which sex is obligate. We will rely on this quantity to obtain our estimate on the mean extinction time of resident mating types.

In this paper, we complement this literature by exploring scenarios of SI under facultative sex. We begin by calculating an analytic expression for the stationary distribution of the number of mating types, extending the results of Constable and Kokko (2018) (where only bounds of the mode of this distribution were calculated). We then provide an expression for the establishment probability of a novel mating type in facultatively sexual populations, generalizing the results of Czuppon and Rogers (2019). Finally, combining these expressions, we calculate the mean time to extinction of a resident mating type allele in a novel way. To be more precise, instead of relying solely on the one-dimensional dynamics of a focal mating type (in mating type frequency space) we make use of the dynamics on the number of mating types (in mating type number space). We conclude by discussing the biological implications of these results.

Model

We consider a population in which the self-incompatible mating type of an individual is determined by one of an infinite set of potential alleles at a single locus. Such single locus determination systems are found, for example, in the social

amoebae *Dictyostelium discoideum* [with three mating types (Bloomfield *et al.* 2010)] and *Didymium squamulosum* [with up to 12 mating types (Clark and Haskins 2010)]. The evolutionary dynamics are determined by a Moran process; generations are overlapping and the population is assumed to be at an ecological equilibrium at which the number of individuals, N, is constant over time. We denote the number of individuals carrying mating type i by n_i, such that $\sum_{i=1}^{\infty} n_i = N$.

Asexual reproduction happens with probability $c \in [0, 1]$. In this case, a randomly chosen individual produces a clone that replaces one of the individuals in the population, also chosen uniformly at random.

During sexual reproduction, which occurs with probability $1 - c$, two randomly chosen individuals mate if they express different mating types. In the case of a successful mating, an offspring individual is produced that replaces one of the other individuals in the population. The mating type of the offspring is chosen at random from among the two parental alleles.

Additionally, we consider the emergence of novel mating types through mutations which occur at rate m. We implement mutations according to the infinitely many alleles model (Kimura and Crow 1964), where the mutated individual expresses a completely new mating type not previously present in the population. For simplicity, we consider the case in which mutation events are decoupled from reproduction. Mutants possess the same characteristics as any other mating type in the population; that is, they are self-incompatible and mate with nonself types at the same *per capita* rate as the resident types.

Mathematically implementing the model described above, the probability per unit time for a type i to increase by one, and a type j to decrease by one, through a birth–death event is given by

$$\mathcal{T}_{ij} = \underbrace{\left(c \frac{n_i}{N} + \frac{(1-c)}{2} \frac{n_i}{N} \frac{\sum_{k \neq i} n_k}{N} \right)}_{\text{reproduction}} \underbrace{\left(\frac{n_j}{N} \right)}_{\text{death}}, \quad \text{for } n_i > 0.$$

(1)

The reproduction term is split into an asexual component (the first term) and a sexual component (the second term). We note that the sum in the sexual reproduction term goes over all non-i mating types present in the population, generating a reproductive advantage to rare mating types. The probability per unit time that a novel mating type i is generated from an ancestral mating type j is given by

$$\mathcal{T}_{ij} \propto m \frac{n_j}{N}, \quad \text{for } n_i = 0,$$

(2)

which can only occur when type i is not already present in the population. As described in Equation S6 in the Supplemental Information, the full expression for the probability transition rate given in Equation 2 features an additional

normalization constant, included to ensure that the mutation rate is independent of the population composition (*i.e.*, that $\sum_{i=1}^{\infty} \mathcal{T}_{ij}\big|_{n_i=0} = m$).

The deterministic limit ($N \to \infty$) of this model is described by a system of ordinary differential equations, where the dynamics (neglecting mutations) of a the ith mating type, with frequency $x_i = \lim_{N \to \infty} \frac{n_i}{N}$, is given by

$$\frac{dx_i}{dt} = \frac{(1-c)}{2} x_i \sum_{j \neq i} x_j (x_j - x_i) .$$

(3)

This system has been studied in Iwasa and Sasaki (1987) (see their Mating kinetics I). The dynamical system possesses an internal stable fixed point where all mating types are present at equal frequencies. Analogous finite population size models lead to stochastic differential equations, and have been analyzed in Czuppon and Rogers (2019). Note that, in that paper, mutation was implemented in a distinct manner, with the assumption that mutation occurs *with* reproductive events. While that implementation is biologically more reasonable, the choice leads to only minor quantitative differences in dynamics.

Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article. The code and simulation outputs can be found on github: https://github.com/gwaconstable/InvExtDynMatTypes. Supplemental material available at Figshare: https://doi.org/10.25386/genetics.8099852.

Results

In this section, we mathematically analyze the null model just presented. We begin by characterizing the long-time equilibrium behavior of the model, which is described by its stationary distribution. This allows us to answer the key question of how many mating types, M, are predicted by the null model as a function of the population size N, mutation rate m, and, importantly, relative rate of asexual reproduction, c. However, while this results in an important benchmark for the expected number of mating types in real populations, it provides little insight into the dynamics of mating type number. To this end, we further calculate the probability that a novel mutant mating type establishes in the population and the expected extinction time of a resident mating type allele. These quantities provide a deeper insight into the ongoing dynamics of mating type number in real populations.

The stationary distribution of individuals carrying each mating type allele

We denote by P_n^{st} the stationary distribution of the number of individuals of each mating type. Constable and Kokko (2018) showed that an analytic solution for P_n^{st} is accessible as the probability transition rates $\mathcal{T}_{ij}$ in Equations 1 and 2 can be decomposed into the product of a birth function $b(n_i)$ and

death function $d(n_j)$ that each depend only on the number of each mating type reproducing and dying respectively;

$$T_{ij} = b(n_i)d(n_j) , \qquad (4)$$

where

$$b(n_i) = c\frac{n_i}{N} + \frac{(1-c)}{2}\frac{n_i}{N}\frac{N-n_i}{N} , \quad \text{if} \quad n_i \geq 1 ,$$

$$b(n_i) \propto m, \qquad \qquad \text{if} \quad n_i = 0, \qquad (5)$$

$$d(n_j) = \frac{n_j}{N}, \qquad \qquad \text{for all } n_j.$$

Under this decomposition, the stationary distribution of mating type alleles takes the exact form

$$P_{\boldsymbol{n}}^{\text{st}} = \frac{1}{\mathcal{N}} \prod_{i=1}^{M-1} \prod_{k=0}^{n_i^{\downarrow}-1} \frac{b(k)d(\Phi_i - k)}{b(\Phi_i - (k+1))d(k+1)} \qquad (6)$$

where $\boldsymbol{n}^{\downarrow}$ is the vector $\boldsymbol{n}$ reordered with its largest entries first, $\mathcal{N}$ is a normalization constant that enforces $\sum_{\boldsymbol{n}} P_{\boldsymbol{n}}^{\text{st}} = 1$, and Φ_i is defined as

$$\Phi_i = N - \sum_{j=1}^{i-1} n_j^{\downarrow}. \qquad (7)$$

The full derivation of Equation 6 is given in Constable and Kokko (2018). In terms of birth–death processes, it is interesting to note that Equation 6 is the product of standard stationary distributions obtained for a two-allele (single variable) population (Karlin and Taylor 1975), moderated by a series of effective population sizes. A similar observation has been made in a related model of multi-allelic selection (Muirhead and Wakeley 2009).

The stationary distribution of the number of mating type alleles

The stationary distribution of populations with M mating types is defined as $\mathcal{P}_M^{\text{st}}$. This gives the probability of finding a population with a given set of parameters (N, c, and m) in a state with M mating types at long times. $\mathcal{P}_M^{\text{st}}$ is related to the distribution $P_{\boldsymbol{n}}^{\text{st}}$ through the following summation;

$$\mathcal{P}_M^{\text{st}} = \sum_{\boldsymbol{n} \in S^{(M)}} P_{\boldsymbol{n}}^{\text{st}}, \qquad (8)$$

where $S^{(M)}$ is the set of all vectors $\boldsymbol{n}$ that represent a population with M present mating types (that is, all vectors $\boldsymbol{n}$ with M nonzero elements).

While Equation 8 can be expressed neatly, evaluating this quantity is problematic as it involves summations of Equation 6 over sets of the infinite vector $\boldsymbol{n}$. To make analytic progress, we note that, when the population size N is large and the per-generation mutation rate $m_g = mN$ is small, at intermediate times the population resides in a quasi-stationary distribution

in the region of a fixed point of the dynamics in the deterministic limit. More precisely, when the population is comprised of M mating types, this fixed point is given by $n_i \approx N/M$ for each of the i present mating types, and $n_i = 0$ otherwise. Following an extinction or mutation event (which changes the number of mating types in the population), the population quickly relaxes to a new quasi-stationary distribution in the region of an alternate fixed point with $M\pm 1$ mating types, i.e., $n_i \approx N/(M\pm 1)$. Within the limit of large N and small m_g then, $P_{\boldsymbol{n}}^{\text{st}}$ can be approximated by a superposition of these quasi-stationary distributions.

Briefly, the quasi-stationary distribution of the population in the region of a fixed point can be calculated by conducting a diffusion approximation on the underlying Moran model and linearizing the resultant advection–diffusion equation about a deterministic fixed point (the van Kampen approximation). The full calculation is conducted in the Supplemental Information, where we show that these quasi-stationary distributions are Gaussian (see Equations S18–S22).

Each of the quasi-stationary distributions can now be renormalized, using Equation 6 to "pin" the height of each quasi-stationary distribution to the height of Equation 6 in the region of the deterministic fixed point [see Hufton *et al.* (2016) and Vasconcelos *et al.* (2017) for similar approaches]. The full calculation is detailed in Section S3 in the Supplemental Information. Substituting the resulting approximation for $P_{\boldsymbol{n}}^{\text{st}}$ into Equation 8, and taking the limit of large N, we find

$$\mathcal{P}_M^{\text{st}} \approx \frac{(2\pi)^{\frac{M}{2}-1}}{\mathcal{M}} \left(\frac{2m}{1+c}\right)^{M-1} \frac{M^{M-\frac{1}{2}}}{M!} \left[\frac{\theta^{M-1}}{\theta-M^{-1}}\right]^{1/2} \times \qquad (9)$$

$$N^{(M-1)/2} \left[(\theta-M^{-1})\left(\frac{\theta}{\theta-M^{-1}}\right)^{M\theta}\right]^N ,$$

where $\mathcal{M}$ is a normalization constant, such that $\sum_{M=1}^{\infty} \mathcal{P}_M^{\text{st}} = 1$ and

$$\theta = \frac{1+c}{1-c}. \qquad (10)$$

Note that when sexual reproduction is obligate, ($c = 0$) and $\theta = 1$, while when sex is facultative and rare, θ becomes large.

Comparing our approximate expression for the distribution of the number of mating types, Equation 9, against the results of stochastic simulation of the population, we find excellent agreement (see Figure 1). We note that, in a facultatively sexual population, where the role of genetic drift is amplified by high clonal reproduction rates, c, extinctions can even lower the number of mating types to one (see Supplemental Information, Section S8). In this scenario, no further sexual reproduction events can take place, and the population reproduces purely asexually until a new mating type is generated by mutation.
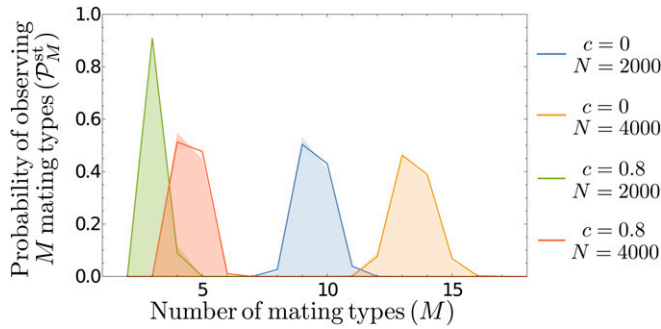
Figure 1 Stationary distribution of mating types, $\mathcal{P}_M^{st}$. For any given rate of asexual reproduction, c, larger populations (larger N) contain more mating types on average. Meanwhile, increasing the rate of asexual to sexual reproduction (increasing c) decreases the expected number of mating types. Analytic results (solid lines) are obtained by evaluating Equation 9. Simulation results (corresponding shaded regions) are obtained using the Gillespie algorithm (Gillespie 1976), averaged over 5×10^6 generations, sampled every 10^2 generations, and following a relaxation period of 10^3 generations from initial conditions with one more mating type than the mode number predicted analytically by the model (see Equation 9). The mutation rate is $m = 10^{-8}$ in all plots. The analytic description can be seen to capture the simulated behavior of the model with very high accuracy.

The mode number of mating types

Since the distribution $\mathcal{P}_M^{st}$ is unimodal, obtaining an estimate for its mode is straightforward numerically. We define by r_M the ratio of probabilities of having M and $M-1$ mating types;

$$r_M = \frac{\mathcal{P}_M^{st}}{\mathcal{P}_{M-1}^{st}}. \tag{11}$$

For any given set of parameters, this ratio is independent of the normalization constant $\mathcal{M}$. Substituting Equation 9 into Equation 11, we obtain a simplified analytic expression for r_M (see Supplemental Information, Equation S42). Since the distribution $\mathcal{P}_M^{st}$ is unimodal, finding its mode is equivalent to finding the value of M for which $\mathcal{P}_M^{st}$ starts to decrease. The approximate mode of $\mathcal{P}_M^{st}$, which we denote M_o, can then be obtained as the solution to the equation

$$r_{M_o} = 1. \tag{12}$$

The function $r_M - 1$ has a single root, and, thus, solving $r_{M_o} = 1$ for M_o is numerically straightforward.

In Figure 2, we plot the mode number of mating types as a function of the per-generation mutation rate, m_g, and the population size, N, for four different rates of asexual to sexual reproduction. Here, we see that facultative sex has a strong influence over the number of mating types, with even a ratio of $9:1$ asexual to sexual divisions (Figure 2b) reducing the expected number of mating types by an order of magnitude compared to the obligately sexual scenario (Figure 2a).

The establishment probability of a new mating type allele

Having approximated the mode of the number of mating types, we now proceed to study the evolutionary dynamics
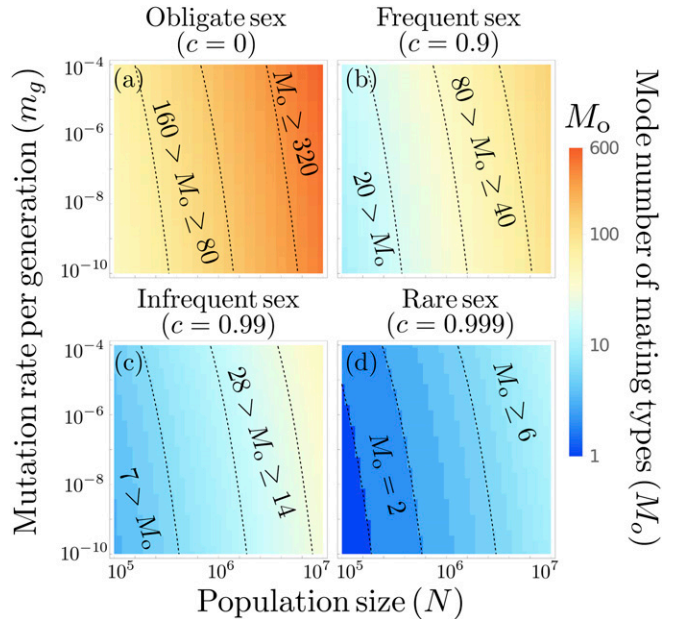

Figure 2 Analytic results on the mode number of mating types, M_o, as a function of the population size, N, and the per-generation mutation rate, $m_g = Nm$, under differing rates of asexual reproduction, c. As the rate of sexual to asexual reproduction is decreased (a–d) so too does the expected number of mating types. When sex is obligate (a), of the order of hundreds of mating types are expected. When sex is facultative and very rare, occurring approximately once to every 1000 asexual reproduction events (d) far fewer mating types are observed. Results are obtained by numerically solving Equation 12.

of the number of mating types. We begin by computing the probability of a successful establishment, Q_M^{Est}, of a newly arising mutant in a population of M resident mating types. We define Q_M^{Est} as the probability that this novel mating type allele (initially at a frequency $1/N$) reaches the new stationary frequency of mating types $(1/(M+1))$ before any of the resident mating types goes extinct.

For $c < 1$, the deterministic equilibrium is an internal fixed point. This is crucial for our approximation of the establishment probability, since we will identify it by the survival probability of a corresponding branching process (see Section S5 in the Supplemental Information). Furthermore, we assume that $N \gg M$, i.e., the population size is sufficiently large that the resident mating types are not quickly lost through genetic drift. Under these assumptions, we can extend the computation from Czuppon and Rogers (2019) for obligate sexual reproduction, and find

$$Q_M^{Est} = \left(\frac{1-c}{1+c}\right)\frac{1}{M}. \tag{13}$$

In the case of obligate sex ($c = 0$), Q_M^{Est} reduces to $1/M$, recovering the result found in Czuppon and Rogers (2019). For facultative sex ($0 < c < 1$), we see that the establishment probability is reduced. Since asexual reproduction increases the time to reach the stationary state in the deterministic system (i.e., the selection strength for even mating type ratios

is reduced, see also Equation 3), newly arising mutants spend more time at low frequencies where they are susceptible to purging by genetic drift. This is reflected by Q_M^{Est} decreasing with c (see Figure 3). Our definition of the establishment probability includes that none of the resident mating types become extinct during the invasion process. This impedes a straightforward comparison with the case of obligate asexual reproduction, $c = 1$ (*i.e.*, a neutral Moran model). The corresponding establishment probability might naïvely be assumed to be $1/(M + 1)$; however, this value ignores the survival of all resident mating types, and, therefore, overestimates the actual establishment probability.

Mean time until the extinction of a mating type allele

Our final goal is to calculate the mean extinction time of a mating type allele. We assume that, initially, the population is close to its equilibrium (*i.e.*, all mating types are approximately at equal frequencies). We then model the arrival of mating types (through mutation events) and their extinction (due to genetic drift) as a birth–death process on the number of mating types. Let β_M be the probability per unit time that a population with M resident mating types gains a new mating type, and δ_M be the probability per unit time that a mating type goes extinct.

Although we are interested in calculating the time between extinction events, we first turn our attention to the stationary probability distribution of this effective "birth–death" process in the number of mating types; $\mathcal{P}_M^{\text{st}}$. Solving the equations for the stationary distribution of the effective birth–death process, we obtain the well-known result (see for instance, Allen 2011, Chapter 6)

$$\mathcal{P}_M^{\text{st}} = \frac{\beta_{M-1}}{\delta_M}\mathcal{P}_{M-1}^{\text{st}}. \qquad (14)$$

Rearranging for δ_M, we find

$$\delta_M = \beta_{M-1}\frac{\mathcal{P}_{M-1}^{\text{st}}}{\mathcal{P}_M^{\text{st}}} = \frac{\beta_{M-1}}{r_M}, \qquad (15)$$

where we have used the definition of r_M given in Equation 11. The mean time between extinction events is the inverse of this effective death rate, *i.e.*,

$$T_M^{\text{Ext}} = \frac{1}{N}\frac{1}{\delta_M} = \frac{r_M}{N\beta_{M-1}}, \qquad (16)$$

where the factor N accounts for the fact that we are measuring the time until an extinction event in time units of generations. Further, we note that, since T_M^{Ext} is determined using $\mathcal{P}_M^{\text{st}}$ (which itself accounts for covariances in allelic frequencies in the stationary state), our expression for T_M^{Ext} captures the effect of allele frequency fluctuations in all M types.

As we have already analytically calculated r_M (see Equation 11), all we now need to evaluate the mean time until the extinction of a mating type allele is an expression for β_M (see Equation 16). We assume that the effective "probability birth
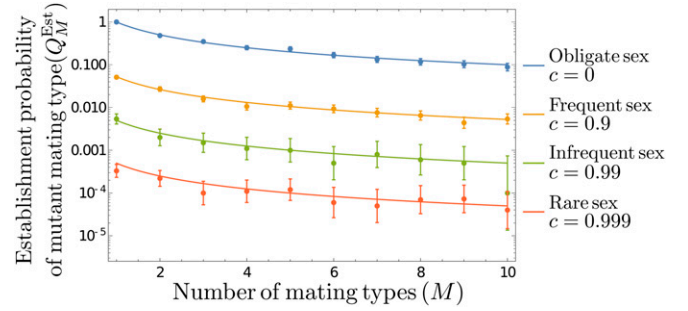

Figure 3 Establishment probability of a novel mating type as a function of M for varying rates of asexual reproduction, c. Analytic results are obtained by evaluating Equation 13. Simulation results are obtained from Gillespie simulations; in each plot $N = 10^5$, and results are averaged over 10^3 runs ($c = 0$), 10^4 runs ($c = 0.9$ and $c = 0.99$), and 10^5 runs ($c = 0.999$). Error bars are ± 2 times the SD of the sample binomial distribution.

rate," β_M (the probability per unit time that the number of mating types in the system increases) is given by the product of the mutation rate and the establishment probability of a mutant allele (see Equation 13):

$$\beta_M \approx mQ_M^{\text{Est}} = m\left(\frac{1-c}{1+c}\right)\frac{1}{M}. \qquad (17)$$

This approximation relies on the frequencies of the resident types being close to their deterministic equilibrium (an assumption made in the derivation of Q_M^{Est}). In general, this will be true if the state containing M types is sufficiently stable and if mutations are occurring sufficiently infrequently that the population has time to relax to this quasi-stationary state. If these conditions are not met, it is possible that a new mutant mating type may arise in a population with a highly uneven mating type distribution, where Q_M^{Est} will no longer provide an accurate estimate of the establishment probability. We discuss these points further below.

Inserting this effective birth rate into Equation 16, and substituting r_M from Equation 11 (see Equation S42), we can express the mean extinction time as

$$T_M^{\text{Ext}} = \sqrt{2\pi}\,\frac{2(M-1)}{1+c}\theta^{\frac{3}{2}}\left(\frac{M}{M-1}\right)^{M-\frac{3}{2}}\left(\frac{\theta-\frac{1}{M}}{\theta-\frac{1}{M-1}}\right)^{-\frac{1}{2}} \qquad (18)$$

$$\times\frac{1}{N^{\frac{3}{2}}}\left[\theta^\theta\left(\frac{\theta-\frac{1}{M}}{\theta-\frac{1}{M-1}}\right)^{1-M\theta}\left(\theta-\frac{1}{M-1}\right)^{-\theta}\right]^N.$$

We first note that, as we would expect, Equation 18 is independent of the mutation rate of novel mating type alleles (the mean time until the extinction of a mating type allele should not depend on the time until a new mating type allele arrives). Further, we can see that, when the population size, N, is large, the mean time to extinction is dominated by the term raised to the power of N in Equation 18 [explaining the linear growth of $\log(T_M^{\text{Ext}})$ for large N in Figure 4]. Although
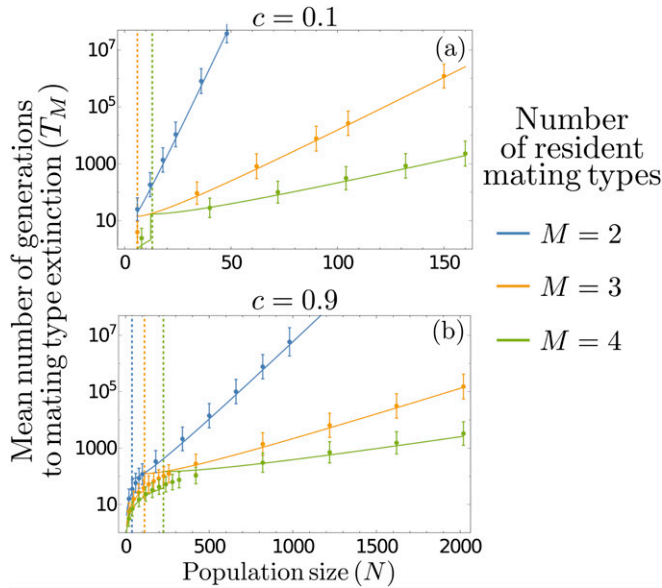
Figure 4 Mean extinction time of the first mating type allele, T_M, as a function of N for populations with initially $M = 2$, $M = 3$, and $M = 4$ mating types. Analytic results (see Equation 20) are plotted as solid lines. Simulation results are averaged over 10^3 Gillespie simulations, with error bars indicating the SD of results. Vertical dashed lines indicate the transition between the approximation of Equation 18 (right of dashed line) and the neutral limit of Equation 19 (left of dashed line), defined by Equation 20. As the rate of asexual reproduction is increased [from $c = 0.1$ in (a) to $c = 0.9$ in (b)], the mean extinction time drops rapidly (note differing scales on x-axis).

Equation 18 is lengthy, it is useful as it allows us to quickly evaluate the mean extinction time for arbitrary parameter values. We can then rapidly explore parameter regimes that would be prohibitively time consuming to simulate (see Figure 5).

The expression derived in Equation 18 becomes increasingly accurate with increasing population size, N, decreasing rates of asexual reproduction, c, and smaller numbers of initial mating types, M. The approximation can break down, however, when either N is small, or c or M are large. In this latter range, as the population dynamics become increasingly dominated by genetic drift, various assumptions involved in the derivation of Equation 18 can become invalid. Most importantly, the Gaussian approximation for the quasi-stationary distribution of a focal mating type breaks down. The true distribution becomes increasingly flat, while the Gaussian approximation does not respect the boundary conditions requiring non-negative number of mating types n_i. Since this results in erroneous approximations of $\mathcal{P}_M^{st}$, (the crucial ingredient calculating Equation 14) when the number of mating types is highly unstable (*i.e.*, when $M \gg M_o$), we do not expect our analytically derived mean extinction time to provide reasonable results in this regime. This reasoning is assessed in more detail and validated in the Supplemental Information (see Section S6).

We now seek a more quantitative measure of when we expect Equation 18 to remain valid. We first note that,
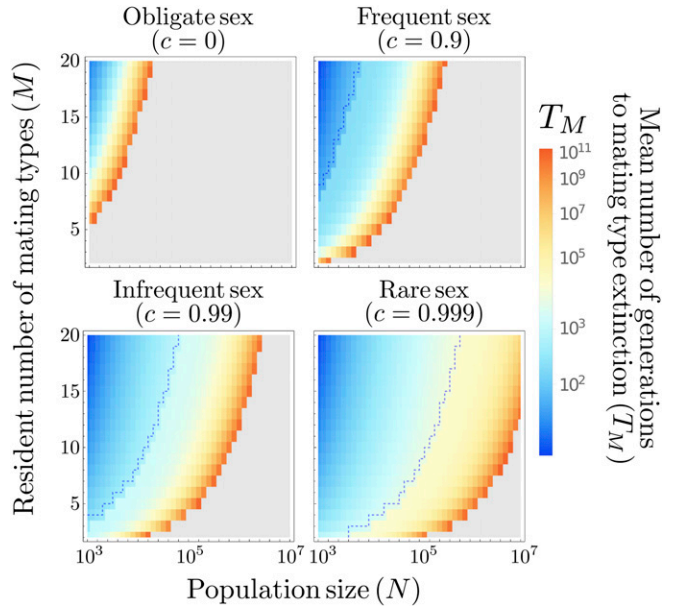

Figure 5 Analytic predictions for the mean extinction time of a resident mating type allele as a function of N and M (see Equation 20). The blue dashed line indicates the parameter regime in which extinction rates become approximately neutral (Equation 19). Extinction times in the gray shaded region exceed 10^{11} generations. As these time are approximately longer than the evolutionary history of fungi, they have been omitted for clarity.

naturally, T_M^{Ext} should increase monotonically with population size, and decrease monotonically with the rate of asexual reproduction and the number of resident mating types. However, on evaluating Equation 18, we find that these expectations are violated as the relative strength of genetic drift increases (*i.e.*, as the initial number of mating types, N/M, becomes small), or, conversely, as the strength of selection for equal numbers of mating types decreases (*i.e.*, c becomes large). In these regimes, the dynamics of the mating type alleles approach neutrality, and extinction time is better approximated by standard results on the neutral multi-allelic Moran model Baxter *et al.* (2007) (see also Equation S51 in the Supplemental Information);

$$T_M^{Neutral} = -N \sum_{s=1}^{M-1} (-1)^{s-1} \binom{M}{s} \frac{s}{M} \log\left(\frac{s}{M}\right). \quad (19)$$

Combining T_M^{Ext} and $T_M^{Neutral}$, the mean time to extinction of a mating type allele can then be approximated by

$$T_M = \begin{cases} T_M^{Ext} & \text{while } \frac{\partial T_M^{Ext}}{\partial N} > 0, \frac{\partial T_M^{Ext}}{\partial c} < 0, \frac{\partial T_M^{Ext}}{\partial M} < 0, \\ T_M^{Neutral} & \text{otherwise.} \end{cases}$$

$$(20)$$

In Figure 4, we can see that this captures the results of simulations very well.

In Figure 5, we plot predicted extinction times for regions of parameter space that are prohibitively time consuming to

investigate numerically. The gray regions that dominate the parameter space for low asexual reproduction rates depict extinction rates $>10^{11}$ generations. To set this in context, we compare this to the evolutionary history of fungi. Fungi first evolved $\sim$1.5 billion years ago (Wang *et al.* 1999). Assuming $\sim$100 generations a year, a conservative estimate considering the doubling time of yeast, which is $\sim$90 min, this leads to an order of magnitude guess of 10^{11} generations for the entire evolutionary history of fungi. Hence, in the context of obligate sex (see Figure 5 top-left panel), mating types alleles are expected to remain fixed in a population while their number remains below six. This condition is severely relaxed in populations in which sex is facultative (see, for example, Figure 5, bottom-right panel). Even though extinction times remain high for realistic effective population sizes, some turnover of mating types might be expected over long time periods, *e.g.*, the evolutionary history of fungi. Moreover, this observation indicates that mating types loss is most likely be driven by extreme population bottlenecks decreasing the effective population size N.

Discussion

The evolutionary mechanisms that drive the number of mating type alleles observed across species has been the topic of numerous theoretical studies. Our analysis in the context of haploid SI adds new results and perspectives to the evolutionary dynamics of this type of balancing selection.

A specificity of mating type SI, as studied here, when compared to gametophytic SI, as prevalent in plants, is the possibility to reproduce asexually. Our findings support and extend the previous result that switching between an asexual and a sexual life cycle significantly reduces the number of mating types in a population (Constable and Kokko 2018). Empirically, available data appears to support the view that more frequent sex is correlated with more mating types [see Table 1, and again Constable and Kokko (2018)]. However, we are hampered from making any strong empirical claims in this area as a result of a paucity of empirical estimates of rates of sex in natural populations. In most species, such estimates are absent; however, with new methodologies and understanding for estimating the rate of sex arising increasingly frequently (Hartfield *et al.* 2018; Nieuwenhuis *et al.* 2018; Ennos and Hu 2019), it is our hope that this gap in the literature will soon be filled.

In Table 1, we test our model quantitatively against four species where estimates for the rate of sex are available. We compare the number of mating types observed empirically with the mode number predicted theoretically. Since estimates for the mutation rate of new mating type alleles and the effective population size parameters are difficult to obtain, we considered a range of values. We find that, while facultative sex explains much of the variation in mating type number, there are quantitative disagreements. In particular, for a range of parameters (particularly large effective population size), the number of mating types is overestimated in

Saccharomyces cerevisiae, *C. reinhardtii* and Tetrahymena, while it is underestimated in *S. commune*.

Given the simplicity of the model that we have proposed, these quantitative disagreements are not unexpected. An interesting open problem is to explore other null models that are capable of explaining these discrepancies. Essentially this task translates to identifying additional mechanisms that decrease selection for more mating types (where their number is overestimated) or increase this selection strength (where their number is underestimated). There exist a number of biologically reasonable potential candidates.

In Equation 1, we have assumed mass action encounter rate dynamics between gametes, leading to a linear relationship between a mating type's frequency and its probability of finding a sexual partner [*i.e.*, $(1 - x_i)$]. This implementation neglects active mate search. In reality, in species such as *C. reinhardtii* or the diatom *Ditylum brightwellii*, which have developed active methods to increase the encounter rate between complementary mating types (Waite and Harrison 1992; Snell and Goodenough 2009), this term might be more accurately described by a decreasing concave up function (see, for example, Ashby and Gupta 2014). This would qualitatively recapitulate the results of Iwasa and Sasaki (1987) (Mating Kinetics 3 and 4), and lead to a decrease in the predicted number of mating types.

Further, our model assumes that a mating type is determined by a single locus. Thus, our model is restricted to bipolar mating type systems, as opposed to tetrapolar systems, where the mating type is determined by alleles at two loci (Nieuwenhuis *et al.* 2013). We expect that additional loci may lead to larger numbers of mating types maintained in the population (as mating type allele combinations can be regenerated through recombination, the extinction rate of mating types will be reduced). Indeed, this would be consistent with the larger number of mating types empirically observed in the tetrapolar *S. commune*, as well as the general observation that tetrapolarity is associated with an increase in mating type diversity (Nieuwenhuis *et al.* 2013). While it is interesting to note that the number of alleles predicted by our model capture the number of *A* and *B* SI complexes in *S. commune* separately (see Table 1), theoretically addressing the mutation-extinction hypothesis for tetrapolar systems in a systematic way, along with the effects of encounter rate dynamics, will be interesting areas for future investigations.

Thus far, we have focused on modeling considerations that may well strengthen the mutation-extinction balance hypothesis. Of course, non-neutral differences between mating types must be considered as well, and a pluralistic view that incorporates the remaining hypotheses for observed mating type numbers may be necessary. For instance, we have not accounted for the fact that newly arising mutants are unlikely to be fully compatible with (or equally as fit as) resident mating types, a biologically important consideration (Power 1976; Hadjivasiliou and Pomiankowski 2016; Krumbeck *et al.* 2019) that would lower the number of types relative to our idealized estimate. In addition, we have not

Table 1 A comparison of empirical values for the number of mating types in various isogamous species with the predictions of the null model

Species	Estimated c	Empirical M	Null model M_o	Q_M^{Est}	T_M (generations)
Saccharomyces cerevisiae[a]	$1 - 1/2000$[b]	2[c]	$9 \geq M_o \geq 1$	$Q_2^{Est} = 1.25 \times 10^{-4}$	$10^{274} \geq T_2 \geq 10^6$
Chlamydomonas reinhardtii	$1 - 1/770$[d]	2[e]	$14 \geq M_o \geq 1$	$Q_2^{Est} = 3.25 \times 10^{-4}$	$10^{707} \geq T_2 \geq 10^9$
Tetrahymena[f]	$1 - 1/100$[g]	$3 - 9$[h]	$38 \geq M_o \geq 4$	$Q_3^{Est} = 1.68 \times 10^{-3}$	$10^{1822} \geq T_3 \geq 10^{20}$
				$Q_9^{Est} = 5.58 \times 10^{-4}$	$10^{153} \geq T_9 \geq 10^4$
Schizophyllum commune	0[i]	$23,328$[j]	$510 \geq M_o \geq 48$	$Q_{23,328}^{Est} = 4.29 \times 10^{-5}$	$41 \geq T_{23,328} \geq 0.5$[k]
				$Q_{288}^{Est} = 3.47 \times 10^{-3}$	$10^{26} \geq T_{288} \geq 58$[k]
				$Q_{81}^{Est} = 1.23 \times 10^{-2}$	$10^{337} \geq T_{81} \geq 10^4$

We assume an effective population size of between $10^7 \geq N \geq 10^5$ in our calculations of M and T_M (note that Q_M^{Est} is independent of N). In our calculation for M, we further assume a per-generation mutation rate, $m_g = mN$, of $10^{-6} \geq m_g \geq 10^{-8}$. For references to empirical data, see the list at the end of the manuscript. For a visualization of the full theoretical distributions of M, $\mathcal{P}_M^{st}$, for each parameter set, see Section S8 in the Supplemental Information.

[a] Considering only strains of *S. cerevisiae* that do not feature mating type switching, for which our model is appropriate.

[b] As addressed within the discussion, our model is only strictly appropriate for populations in which the mating type does not switch between generations. Mating type switching is prevalent among many *Saccharomyces*; however, most populations feature nonswitching strains (Nieuwenhuis *et al.* 2018). We assume *S. cerevisiae* has a similar ecology to *S. paradoxus*, where sex has been estimated to occur between once every 1000 and 3000 generations (Tsai *et al.* 2008) (indeed, estimates for the rates of outcrossing in both species are comparable).

[c] The mating type of Saccharomyces species is determined by one of two alleles at the MAT locus, termed α and a (Butler 2007). We note that often these variants are referred to as idiomorphs (rather than alleles) because of large differences in their size and genetic composition (Butler *et al.* 2004).

[d] The rate of sex in *C. reinhardtii* has been estimated to once in every 770 generations (Hasan and Ness 2018).

[e] *C. reinhardtii* has two mating types, denoted mt^+ and mt^-, each determined by one of two alleles at the MT locus. Similar mating type determination is found among other isogamous algae (Hamaji *et al.* 2013; Sekimoto 2017).

[f] Considering only *Tetrahymena* species with synclonal inheritance, for which our model is appropriate.

[g] We assume that synclonal *Tetrahymena* species [such as *T. americanis*, see (h)] feature similar rates of sex to *T. thermophila*. Our estimate on the rate of sex is based on a minimal sexually immature period of $60 - 100$ fissions in *T. thermophila* (Doerder *et al.* 1995). We note, however, that the frequency of cellular conjugation events observed in other ciliates indicates that lower rates of sex may be appropriate (Lucchesi and Santangelo 2004).

[h] Tetrahymena species *T. americanis*, *T. hegewischi*, *T. hyperangularis*, and *T. pigmentosa* have mating type numbers nine, eight, four, and three respectively (Phadke and Zufall 2009). We take care to focus only on these synclonal species, defined as those in which mating type is inherited deterministically based on parental genotype. In caryonidal species, such as *T. thermophila*, the mating type of progeny is determined stochastically (Phadke and Zufall 2009), and, thus, a comparison of our model with these species is not appropriate.

[i] Molecular analysis suggests that *S. commune* features some of the highest rates of sex within the fungal kingdom (Nieuwenhuis and James 2016). In addition, most of the Agaricomycotina (the class to which *S. commune* belongs) are known to be obligately sexual (Nieuwenhuis and Aanen 2012). We therefore assume a rate of asexual reproduction of $c = 0$ for *S. commune*. We note, however, that the life-cycle of *S. commune* features more complicated dynamics than accounted for by our model; it is multicellular and can exhibit vegetative growth as a haploid mycelium.

[j] *S. commune* is Tetrapolar, with mating type determined by two complementary pathways, controlled by unlinked genetic complexes A (with 288 variants) and B (with 81 variants). Each of these regions consists of two weakly recombining loci, leading to a total of four mating type loci. These loci, denoted Aa, Ab, Ba, and Bb, have 9, 32, 9, and 9 alleles, respectively (Stankis *et al.* 1992). Full compatibility between two mating types is achieved when both the A and B complexes of the mates are of different specificities; semi-compatibility occurs when the only one of these complexes is of different specificity (Raudaskoski and Kothe 2010).

[k] Lower-bound extinction time is effectively neutral (see Equation 20).

considered how forms of homothalism (*i.e.*, the evolution of self-compatibility) may affect our results. For instance most yeasts have evolved the ability to switch mating types between sexual generations (Nieuwenhuis and Immler 2016; Nieuwenhuis *et al.* 2018), while, among ciliates, some species exhibit probabilistic mating type expression (Paixão *et al.* 2011). Indeed, as demonstrated in Hadjivasiliou *et al.* (2016), the rapid loss of mating type diversity when sex is rare and local population sizes are low (see Figure 5) can drive selection for mating type switching in order to maintain the capacity for sexual reproduction. Such factors make a comparison of our model with these species problematic. Further mechanisms that may limit the number of mating types that are not included in our model, including UPI and costly mate search, are reviewed in Billiard *et al.* (2011).

Our model for haploid SI is conceptually similar to extensively studied gametophytic SI systems observed in plants. Yet, we take a novel mathematical approach in estimating the number of mating types supported by a finite population. Instead of using the extinction boundaries of the one-dimensional diffusion approximation of a focal mating type as in most previous studies (Wright 1939, 1960, 1964; Ewens 1964; Yokoyama and Nei 1979; Yokoyama and Hetherington 1982; Vallejo-Marín and Uyenoyama 2008), we study a birth–death process on the total number of mating types. Utilizing just the local description of the stationary distribution around an interior stable fixed point, we circumvent the problem of the diffusion approximation for systems with a stable deterministic behavior being inaccurate at the boundaries (Assaf and Meerson 2017). This leads to robust predictions for the stationary probability distribution $\mathcal{P}_M^{st}$ (Figure 1), and, importantly, the mean extinction time of a mating type allele, T_M (Figure 4).

In calculating this mean extinction time, we deviate from the approaches taken before that rely on the one-dimensional dynamics of a focal mating type (Takahata 1990; Vekemans and Slatkin 1994). This previous method compares the one-dimensional diffusion to a time-rescaled neutral coalescent, thus obtaining estimates for the diversification rate (*i.e.*, the establishment rate of a novel mating type). This comparison is conducted by assuming constant selective strength. In our model, this would be equivalent to assuming that the term

$\sum x_i x_j$ is constant. However, with a varying number of present mating types (and indeed, also with fluctuations in the mating type frequencies themselves), this value is not constant in time. Hence, a comparison with a time-rescaled neutral coalescent does not seem appropriate since the selection strength would need to be reassessed after each coalescence event that is associated with a loss of a mating type. This variation in selection strength has been empirically observed by analyzing the genome of various *Coprinus cinereus* (mushroom fungus) populations worldwide, which shows that selective forces in the mating type dynamics are dependent on the number of distinct lineages (May *et al.* 1999); see also Richman (2000) for other examples in the context of balancing selection. Our technique, explicitly computing the mean extinction time by the effective birth–death process on the number of mating types, avoids this problem. Since the varying selective strength enters in both the establishment probability and the stationary distribution, our computed extinction time accounts for the fluctuating selective pressure.

These explicit estimates of establishment probabilities and typical extinction times yield new insight into the dynamics that drive the low number of mating types predicted by our model in the rare sex regime. One might initially suspect that these numbers are the result of a high turnover of mating type alleles, *i.e.*, frequent extinction and invasion events. However, we find that, in fact, they are a result of very low invasion probabilities for novel mating types, combined with rapidly decreasing extinction times as a function of resident mating type diversity. It is worth mentioning that, although our approximations for the extinction time break down when the resident state becomes highly unstable (*i.e.*, when $M \gg M_o$), these represent states that would be very short-lived in nature, and thus are not relevant from a biological perspective.

Finally, in Table 1, we link our theoretical observation of small numbers of mating type alleles yielding very large extinction times (see Figure 4) to some species examples. Indeed, the large extinction times found by the quantitative study are in line with previous empirical observations of long terminal branches in allelic genealogies under negative-frequency dependent selection, *e.g.*, in fungi (May *et al.* 1999) and Solanaceae (Uyenoyama 1997). Similar to previous simulation studies (Slatkin and Muirhead 1999; Gervais *et al.* 2011), we find that the number of resident mating types strongly influences the diversification rate; the larger the resident number, the lower the diversification rate. This slowdown is ubiquitously observed in natural systems under balancing selection, such as SI in fungi (May *et al.* 1999) or Solanaceae (Uyenoyama 1997) and MHC-systems (Solberg *et al.* 2008). We find that long terminal branches in SI systems, corresponding to old mating type allele ages [with some being older than the corresponding species age (Richman 2000)] emerge naturally under balancing selection for two reasons: (i) the stability of the internal equilibrium can lead to extremely large extinction times of alleles, thus enabling such long branches; and (ii) the lower establishment rate due to a large number of

resident SI alleles slows down the diversification rate, and, hence, decreases the number of newly established mating types.

In conclusion, we analyzed the evolutionary dynamics of self-incompatible mating types in facultatively sexual isogamous species. Our results refine previous estimates on the number of mating types maintained in a finite population as well as on the establishment probability of a newly arising mating type allele. Furthermore, we have used these results to compute the mean extinction time of a focal mating type via an effective birth–death process describing the number of mating types. This estimate naturally incorporates variation in selection strength due to varying numbers of resident mating types, a fact that previous studies have failed to incorporate. We are therefore able to qualitatively explain the empirically and numerically observed slowdown of allelic diversification rates in populations under balancing selection. The here presented methodology is theoretically extendable to other systems exhibiting negative-frequency dependent selection.

Literature Cited

Allen, L. J., 2011   *An Introduction to Stochastic Processes with Applications to Biology*, Ed. 2. Pearson Education, Hoboken, NJ.

Ashby, B., and S. Gupta, 2014   Parasitic castration promotes coevolutionary cycling but also imposes a cost on sex. Evolution 68: 2234–2244.

Assaf, M., and B. Meerson, 2017   WKB theory of large deviations in stochastic populations. J. Phys. A Math. Theor. 50: 263001. https://doi.org/10.1088/1751-8121/aa669a

Baranova, M. A., M. D. Logacheva, A. A. Penin, V. B. Seplyarskiy, Y. Y. Safonova *et al.*, 2015   Extraordinary genetic diversity in a wood decay mushroom. Mol. Biol. Evol. 32: 2775–2783. https://doi.org/10.1093/molbev/msv153

Baxter, G. J., R. A. Blythe, and A. J. McKane, 2007   Exact solution of the multi-allelic diffusion model. Math. Biosci. 209: 124–170. https://doi.org/10.1016/j.mbs.2007.01.001

Billiard, S., V. Castric, and X. Vekemans, 2007   A general model to explore complex dominance patterns in plant sporophytic

self-incompatibility systems. Genetics 175: 1351–1369. https://doi.org/10.1534/genetics.105.055095

Billiard, S., M. López-Villavicencio, B. Devier, M. E. Hood, C. Fairhead *et al.*, 2011   Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol. Rev. Camb. Philos. Soc. 86: 421–442. https://doi.org/10.1111/j.1469-185X.2010.00153.x

Birky, C. W., 1995   Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc. Natl. Acad. Sci. USA 92: 11331–11338. https://doi.org/10.1073/pnas.92.25.11331

Bloomfield, G., J. Skelton, A. Ivens, Y. Tanaka, and R. Kay, 2010   Sex determination in the social amoeba *Dictyostelium discoideum*. Science 330: 1533–1536. https://doi.org/10.1126/science.1197423

Bod'ová, K., T. Priklopil, D. Field, N. Barton, and M. Pickup, 2018   Evolutionary pathways for the generation of new self-incompatibility haplotypes in a nonself-recognition system. Genetics 209: 861–883.

Butler, G., 2007   The evolution of MAT: the ascomycetes, pp. 3–18 in *Sex in Fungi*, edited by J. Heitman, J. W. Kronstad, J. W. Taylor, and L. A. Cassleton. American Society for Microbiology, Washington, DC.

Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin *et al.*, 2004   Evolution of the mat locus and its ho endonuclease in yeast species. Proc. Natl. Acad. Sci. USA 101: 1632–1637. https://doi.org/10.1073/pnas.0304170101

Charlesworth, D., and B. Charlesworth, 1979   The evolution and breakdown of S-allele systems. Heredity 43: 41–55. https://doi.org/10.1038/hdy.1979.58

Clark, A. G., and T.-H. Kao, 1994   Self-incompatibility: theoretical concepts and evolution, pp. 220–242 in *Genetic Control of Self-Incompatibility and Reproductive Development in Flowering Plants*, edited by E. G. Williams, A. E. Clarke, and R. B. Knox. Springer, Dordrecht, The Netherlands. https://doi.org/10.1007/978-94-017-1669-7_11

Clark, J., and E. F. Haskins, 2010   Reproductive systems in the myxomycetes: a review. Mycosphere 1: 337–353.

Constable, G. W. A., and H. Kokko, 2018   The rate of facultative sex governs the number of expected mating types in isogamous species. Nat. Ecol. Evol. 2: 1168–1175. https://doi.org/10.1038/s41559-018-0580-9

Crosby, J. L., 1966   Self-incompatibility alleles in the population of *Oenothera organensis*. Evolution 20: 567–579. https://doi.org/10.1111/j.1558-5646.1966.tb03387.x

Czuppon, P., and D. Rogers, 2019   Evolution of mating types in finite populations: the precarious advantage of being rare. J. Evol. Biol. (in press). https://doi.org/10.1111/jeb.13528

Doerder, F. P., M. A. Gates, F. P. Eberhardt, and M. Arslanyolu, 1995   High frequency of sex and equal frequencies of mating types in natural populations of the ciliate *Tetrahymena thermophila*. Proc. Natl. Acad. Sci. USA 92: 8715–8718. https://doi.org/10.1073/pnas.92.19.8715

Ennos, R. A., and X. S. Hu, 2019   Estimating the number of sexual events per generation in a facultatively sexual haploid population. Heredity (Edinb) 122: 729–741. https://doi.org/10.1038/s41437-018-0171-1

Ewens, W. J., 1964   On the problem of self-sterility alleles. Genetics 50: 1433–1438.

Gervais, C. E., V. Castric, S. Ressayre, and A. Billiard, 2011   Origin and diversification dynamics of self-incompatibility haplotypes. Genetics 188: 625–636. https://doi.org/10.1534/genetics.111.127399

Gillespie, D. T., 1976   A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22: 403–434. https://doi.org/10.1016/0021-9991(76)90041-3

Goodenough, U., H. Lin, and J. H. Lee, 2007   Sex determination in chlamydomonas. Semin. Cell Dev. Biol. 18: 350–361. https://doi.org/10.1016/j.semcdb.2007.02.006

Haag, E., 2007   Why two sexes? Sex determination in multicellular organisms and protistan mating types. Semin. Cell Dev. Biol. 18: 348–349. https://doi.org/10.1016/j.semcdb.2007.05.009

Hadjivasiliou, Z., 2014   *Theoretical studies on the role and evolution of mating types and two sexes*. Ph.D. Thesis, University College London (University of London), London.

Hadjivasiliou, Z., and A. Pomiankowski, 2016   Gamete signalling underlies the evolution of mating types and their number. Philos. Trans. R. Soc. Lond. B Biol. Sci. 371: 20150531. https://doi.org/10.1098/rstb.2015.0531

Hadjivasiliou, Z., and A. Pomiankowski, 2019   Evolution of the mating type locus with suppressed recombination. bioRxiv. https://doi.org/10.1101/518654

Hadjivasiliou, Z., A. Pomiankowski, R. Seymour, and N. Lane, 2012   Selection for mitonuclear co-adaptation could favour the evolution of two sexes. Proc. Biol. Sci. 279: 1865–1872. https://doi.org/10.1098/rspb.2011.1871

Hadjivasiliou, Z., N. Lane, R. Seymour, and A. Pomiankowski, 2013   Dynamics of mitochondrial inheritance in the evolution of binary mating types and two sexes. Proc. Biol. Sci. 280: 20131920. https://doi.org/10.1098/rspb.2013.1920

Hadjivasiliou, Z., Y. Iwasa, and A. Pomiankowski, 2015   Cell-cell signalling in sexual chemotaxis: a basis for gametic differentiation, mating types and sexes. J. R. Soc. Interface 12: 20150342. https://doi.org/10.1098/rsif.2015.0342

Hadjivasiliou, Z., A. Pomiankowski, and B. Kuijper, 2016   The evolution of mating type switching. Evolution 70: 1569–1581. https://doi.org/10.1111/evo.12959

Hamaji, T., P. Ferris, I. Nishii, Y. Nishimura, and H. Nozaki, 2013   Distribution of the sex-determining gene mid and molecular correspondence of mating types within the isogamous genus gonium (volvocales, chlorophyta). PLoS One 8: e64385. https://doi.org/10.1371/journal.pone.0064385

Hartfield, M., S. I. Wright, and A. F. Agrawal, 2018   Coalescence and linkage disequilibrium in facultatively sexual diploids. Genetics 210: 683–701. https://doi.org/10.1534/genetics.118.301244

Hasan, A. R., and R. W. Ness, 2018   The genomic landscape of recombination rate variation in *Chlamydomonas reinhardtii* reveals effects of linked selection. bioRxiv. https://doi.org/doi:10.1101/340992

Hoekstra, R. F., 1982   On the asymmetry of sex: evolution of mating types in isogamous populations. J. Theor. Biol. 98: 427–451. https://doi.org/10.1016/0022-5193(82)90129-1

Hoekstra, R. F., Y. Iwasa, and F. Weissing, 1991   The origin of isogamous sexual differentiation, pp. 155–181 in *Game Equilibrium Models*, edited by I. Selten. Springer, Berlin. https://doi.org/10.1007/978-3-662-02674-8_6

Hufton, P. G., Y. T. Lin, T. Galla, and A. J. McKane, 2016   Intrinsic noise in systems with switching environments. Phys. Rev. E 93: 052119. https://doi.org/10.1103/PhysRevE.93.052119

Hurst, L. D., and W. D. Hamilton, 1992   Cytoplasmic fusion and the nature of sexes. Proc. Biol. Sci. 247: 189–194. https://doi.org/10.1098/rspb.1992.0027

Iwasa, Y., and A. Sasaki, 1987   Evolution of the number of sexes. Evolution 41: 49–65. https://doi.org/10.1111/j.1558-5646.1987.tb05770.x

James, T. Y., 2015   Why mushrooms have evolved to be so promiscuous: insights from evolutionary and ecological patterns. Fungal Biol. Rev. 29: 167–178. https://doi.org/10.1016/j.fbr.2015.10.002

James, T. Y., P. Srivilai, U. Kües, and R. Vilgalys, 2006   Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of

mating-type-specific pheromone receptor function. Genetics 172: 1877–1891. https://doi.org/10.1534/genetics.105.051128

Karlin, S., and H. M. Taylor, 1975   *A First Course in Stochastic Processes*, Ed. 2. Academic Press, San Deigo, CA.

Kimura, M., and J. F. Crow, 1964   The number of alleles that can be maintained in a finite population. Genetics 49: 725–738.

Kothe, E., 1996   Tetrapolar fungal mating types: sexes by the thousands. FEMS Microbiol. Rev. 18: 65–87. https://doi.org/10.1111/j.1574-6976.1996.tb00227.x

Kothe, E., 1999   Mating types and pheromone recognition in the homobasidiomycete *Schizophyllum commune*. Fungal Genet. Biol. 27: 146–152. https://doi.org/10.1006/fgbi.1999.1129

Krumbeck, Y., G. W. A. Constable, and T. Rogers, 2019   Fitness differences suppress the number of mating types in evolving isogamous species. *arXiv* arXiv:1906.07117v1

Lehtonen, J., H. Kokko, and G. A. Parker, 2016   What do isogamous organisms teach us about sex and the two sexes? Philos. Trans. R. Soc. Lond. B Biol. Sci. 371: 20150532. https://doi.org/10.1098/rstb.2015.0532

Lucchesi, P., and G. Santangelo, 2004   How often does conjugation in ciliates occur? Clues from a seven-year study on marine sandy shores. Aquat. Microb. Ecol. 36: 195–200. https://doi.org/10.3354/ame036195

May, G., F. Shaw, H. Badrane, and X. Vekemans, 1999   The signature of balancing selection: fungal mating compatibility gene evolution. Proc. Natl. Acad. Sci. USA 96: 9172–9177. https://doi.org/10.1073/pnas.96.16.9172

Muirhead, C., and J. Wakeley, 2009   Modeling multiallelic selection using a Moran model. Genetics 182: 1141–1157. https://doi.org/10.1534/genetics.108.089474

Nagylaki, T., 1975   The deterministic behavior of self-incompatibility alleles. Genetics 79: 545–550.

Nieuwenhuis, B., and S. Immler, 2016   The evolution of mating-type switching for reproductive assurance. BioEssays 38: 1141–1149. https://doi.org/10.1002/bies.201600139

Nieuwenhuis, B. P., and T. Y. James, 2016   The frequency of sex in fungi. Philos. Trans. R. Soc. Lond. B Biol. Sci. 371: pii: 20150540. https://doi.org/10.1098/rstb.2015.0540

Nieuwenhuis, B. P. S., and D. K. Aanen, 2012   Sexual selection in fungi. J. Evol. Biol. 25: 2397–2411. https://doi.org/10.1111/jeb.12017

Nieuwenhuis, B. P. S., S. Billiard, S. Vuilleumier, E. Petit, M. E. Hood *et al.*, 2013   Evolution of uni- and bifactorial sexual compatibility systems in fungi. Heredity 111: 445–455. https://doi.org/10.1038/hdy.2013.67

Nieuwenhuis, B. P. S., S. Tusso, P. Bjerling, J. Stångberg, J. B. W. Wolf *et al.*, 2018   Repeated evolution of self-compatibility for reproductive assurance. Nat. Commun. 9: 1639. https://doi.org/10.1038/s41467-018-04054-6

Paixão, T., S. S. Phadke, R. B. Azevedo, and R. A. Zufall, 2011   Sex ratio evolution under probabilistic sex determination. Evolution 65: 2050–2060. https://doi.org/10.1111/j.1558-5646.2011.01266.x

Perrin, N., 2012   What uses are mating types? The "developmental switch" model. Evolution 66: 947–956. https://doi.org/10.1111/j.1558-5646.2011.01562.x

Phadke, S. S., and R. A. Zufall, 2009   Rapid diversification of mating systems in ciliates. Biol. J. Linn. Soc. Lond. 98: 187–197. https://doi.org/10.1111/j.1095-8312.2009.01250.x

Power, H. W., 1976   On forces of selection in the evolution of mating types. Am. Nat. 110: 937–944. https://doi.org/10.1086/283118

Prigoda, N. L., A. Nassuth, and B. Mable, 2005   Phenotypic and genotypic expression of self-incompatibility haplotypes in *Arabidopsis lyrata* suggests unique origin of alleles in different

dominance classes. Mol. Biol. Evol. 22: 1609–1620. https://doi.org/10.1093/molbev/msi153

Raudaskoski, M., and E. Kothe, 2010   Basidiomycete mating type genes and pheromone signaling. Eukaryot. Cell 9: 847–859. https://doi.org/10.1128/EC.00319-09

Richman, A., 2000   Evolution of balanced genetic polymorphism. Mol. Ecol. 9: 1953–1963. https://doi.org/10.1046/j.1365-294X.2000.01125.x

Sekimoto, H., 2017   Sexual reproduction and sex determination in green algae. J. Plant Res. 130: 423–431. https://doi.org/10.1007/s10265-017-0908-6

Slatkin, M., and C. A. Muirhead, 1999   Overdominant alleles in a population of variable size. Genetics 152: 775–781.

Snell, W. J., and U. W. Goodenough, 2009   Flagellar adhesion, flagellar-generated signaling, and gamete fusion during mating, pp. 369–394 in *The Chlamydomonas Sourcebook*, Vol. 3, edited by E. H. Harris, D. B. Stern, and G. B. Witman. Elsevier, Amsterdam. https://doi.org/10.1016/B978-0-12-370873-1.00049-6

Solberg, O. D., S. J. Mack, A. K. Lancaster, R. M. Single, Y. Tsai *et al.*, 2008   Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum. Immunol. 69: 443–464. https://doi.org/10.1016/j.humimm.2008.05.001

Stankis, M. M., C. A. Specht, H. Yang, L. Giasson, R. C. Ullrich *et al.*, 1992   The A alpha mating locus of Schizophyllum commune encodes two dissimilar multiallelic homeodomain proteins. Proc. Natl. Acad. Sci. USA 89: 7169–7173. https://doi.org/10.1073/pnas.89.15.7169

Takahata, N., 1990   A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA 87: 2419–2423. https://doi.org/10.1073/pnas.87.7.2419

Takahata, N., and M. Nei, 1990   Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124: 967–978.

Thompson, K. F., and J. P. Taylor, 1966   Non-linear dominance relationships between S alleles. Heredity 21: 345–362. https://doi.org/10.1038/hdy.1966.36

Tsai, I. J., D. Bensasson, A. Burt, and V. Koufopanou, 2008   Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. Proc. Natl. Acad. Sci. USA 105: 4957–4962. https://doi.org/10.1073/pnas.0707314105

Uyenoyama, M. K., 1997   Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. Genetics 147: 1389–1400.

Vallejo-Marín, M., and M. K. Uyenoyama, 2008   On the evolutionary modification of self-incompatibility: implications of partial clonality for allelic diversity and genealogical structure, pp. 53–71 in *Self-Incompatibility in Flowering Plants*. Springer, Berlin, Heidelberg.

Vasconcelos, V. V., F. P. Santos, F. C. Santos, and J. M. Pacheco, 2017   Stochastic dynamics through hierarchically embedded Markov chains. Phys. Rev. Lett. 118: 058301. https://doi.org/10.1103/PhysRevLett.118.058301

Vekemans, X., and M. Slatkin, 1994   Gene and allelic genealogies at a gametophytic self-incompatibility locus. Genetics 137: 1157–1165.

Waite, A., and P. J. Harrison, 1992   Role of sinking and ascent during sexual reproduction in the marine diatom *Ditylum brightwellii*. Mar. Ecol. Prog. Ser. 87: 113–122. https://doi.org/10.3354/meps087113

Wang, D. Y.-C., S. Kumar, and S. B. Hedges, 1999   Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc. R. Soc. Lond. B Biol. Sci. 266: 163–171. https://doi.org/10.1098/rspb.1999.0617

Wright, S., 1939   The distribution of self-sterility alleles in populations. Genetics 24: 538–552.

Wright, S., 1960   On the number of self-incompatibility alleles maintained in equilibrium by a given mutation rate in a population of a given size: a re-examination. Biometrics 16: 61–85. https://doi.org/10.2307/2527956

Wright, S., 1964   The distribution of self-incompatibility alleles in populations. Evolution 18: 609–619. https://doi.org/10.1111/j.1558-5646.1964.tb01675.x

Yokoyama, S., and M. Nei, 1979   Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants. Genetics 91: 609–626.

Yokoyama, S., and L. E. Hetherington, 1982   The expected number of self-incompatibility alleles in finite plant populations. Heredity 48: 299–303. https://doi.org/10.1038/hdy.1982.35

Communicating editor: A. Agrawal