The K_a/K_s and π_a/π_s Ratios under Different Models of Gametophytic and Sporophytic Selection

Ling-Ling Li[1,2], Yu Xiao[1,2], Xi Wang[1,2], Zi-Han He[1,2], Yan-Wen Lv[1,2], and Xin-Sheng Hu[1,2,*]

[1]College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China

[2]Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, South China Agricultural University, Guangzhou 510642, China

*Corresponding author: E-mail: xinsheng@scau.edu.cn.

Abstract

Alternation of generations in plant life cycle provides a biological basis for natural selection occurring in either the gametophyte or the sporophyte phase or in both. Divergent biphasic selection could yield distinct evolutionary rates for phase-specific or pleiotropic genes. Here, we analyze models that deal with antagonistic and synergistic selection between alternative generations in terms of the ratio of nonsynonymous to synonymous divergence (K_a/K_s). Effects of biphasic selection are opposite under antagonistic selection but cumulative under synergistic selection for pleiotropic genes. Under the additive and comparable strengths of biphasic allelic selection, the absolute K_a/K_s for the gametophyte gene is equal to in outcrossing but smaller than, in a mixed mating system, that for the sporophyte gene under antagonistic selection. The same pattern is predicted for K_a/K_s under synergistic selection. Selfing reduces efficacy of gametophytic selection. Other processes, including pollen and seed flow and genetic drift, reduce selection efficacy. The polymorphism (π_a) at a nonsynonymous site is affected by the joint effects of selfing with gametophytic or sporophytic selection. Likewise, the ratio of nonsynonymous to synonymous polymorphism (π_a/π_s) is also affected by the same joint effects. Gene flow and genetic drift have opposite effects on π_a or π_a/π_s in interacting with gametophytic and sporophytic selection. We discuss implications of this theory for detecting natural selection in terms of K_a/K_s and for interpreting the evolutionary divergence among gametophyte-specific, sporophyte-specific, and pleiotropic genes.

Key words: mating system, evolutionary rate, alternation of generations, positive selection, purifying selection.

Significance

This study develops a new theory of how the ratio of nonsynonymous to synonymous divergence (K_a/K_s) changes under phase variation of selection during the plant life cycle. Effects of selfing on K_a/K_s are emphasized.

Introduction

K_a (or K_s) is the average number of nonsynonymous (or synonymous) nucleotide differences between protein-coding gene sequences per nonsynonymous (or synonymous) site. The relative rate of synonymous to nonsynonymous divergence per nucleotide site, K_a/K_s, is often used to detect natural selection occurring in a protein-coding gene at the evolutionary time scale. Neutrality or selection (positive or purifying) is signaled when K_a/K_s is equal to or unequal to 1 (>1 or <1), respectively. Given a gene sequence, selection may exhibit "spatial" variation across nonsynonymous sites in both the type (positive and purifying) and strength (K_a/K_s values) of selection along the sequence. Such spatial variation of selection could partly be cancelled out if the analysis is based on individual genes, such as the branch model from phylogenetic analysis with maximum likelihood (PAML) package in which each branch is hypothesized to have the same K_a/K_s at any nonsynonymous site, and could

yield purifying selection at the gene level (Yang 2006). Here, we demonstrate another type of variation of selection that occurs between gametophyte and sporophyte phases during the plant life cycle. This phase variation in terms of K_a/K_s could influence detection of natural selection as well.

Phase variation in K_a/K_s could arise from multiple processes. One is the potentially divergent selection between gametophyte and sporophyte phases during the life cycle. Haldane (1932) and Haldane and Jayakar (1963) thought that antagonistic selection could occur between two phases where one allele is favorable in the gametophyte phase but deleterious in the sporophyte phase, or vice versa. Theoretical studies show that antagonistic selection helps to maintain polymorphisms (Damgaard et al. 1994; Immler et al. 2012; Otto et al. 2015; Peters and Weis 2018). An alternative type is synergistic selection between gametophyte and sporophyte phases, which purges deleterious alleles but enhances positive selection of advantageous alleles (Walsh and Charlesworth 1992; Charlesworth and Charlesworth 1992; Damgaard 2000). Gametophytic selection can increase sporophyte fitness (Mulcahy and Mulcahy 1987; Winsor et al. 1987; Beaudry et al. 2020). Both types of selection can cause phase variation in K_a/K_s.

Evidence supports different types of selection in gametophytic and sporophytic phases. Phase-specific selection could likely occur in genes that are only expressed in the gametophyte or the sporophyte phase (Page and Grossniklaus 2002; Honys and Twell 2003; Rutley and Twell 2015; Liu et al. 2015; Zhang et al. 2021). Synergistic selection can be observed for genes that are co-expressed in both phases (Tanksley et al. 1981; Borg et al. 2009; Arunkumar et al. 2013; Frank and Scanlon 2015), whereas antagonistic selection is likely to occur in genes that are differentially expressed in both phases (Gossmann et al. 2016; Beaudry et al. 2020). These empirical observations provide a biological basis for modeling and testing phase variation of K_a/K_s.

The second process is the joint effects of selfing and biphasic selection, which could alter K_a/K_s in a complicated way. Selfing enhances the purging of deleterious alleles via reducing the frequency of heterozygotes that mask deleterious alleles in the sporophyte phase. Selfing affects the mutation load in metapopulations (Roze and Rousset 2004) and, in the case of gametophytic selection, will have a strong effect on allele frequencies in gametes (Hu et al. 2019). Previous theory indicates that selfing/inbreeding favors the gametophyte generation, while outcrossing favors the sporophyte generation (Otto and Marks 1996). This can change the relative selection strengths between the two phases in various plant species and causes phase variation of K_a/K_s throughout the life cycle.

The third process is the joint effects of gene flow and biphasic selection (Hu et al. 2019). Pollen flow directly influences gametophytic selection through its effects on the

frequencies of deleterious alleles but indirectly affects sporophytic selection through the mating system. Seed flow directly interacts with sporophytic selection through introducing both homozygotes and heterozygotes to recipient populations. Both pollen and seed flow could bring about migration loads (Wright 1977; Hu and Li 2003; Lopez et al. 2008) and reduce selection efficacy. In addition, because K_a/K_s is related to the fixation of alternative alleles in genetically distinct populations (Yang 2006; Kryazhimskiy and Plotkin 2008), it is of significance to examine how gene flow impedes the speciation process. Gene flow via haploid pollen and diploid seeds as vectors could also be responsible for the observed phase variation of K_a/K_s during the life cycle.

Other processes, such as the interaction of mating system with genetic drift (Glemin 2007; Glemin and Muyle 2014) or with mutation, could alter K_a/K_s (Li et al. 2023). Genetic drift enhances fixation of deleterious alleles and hence reduces selection efficacy (Kimura 1962), which makes K_a/K_s for deleterious alleles approach 1.0. Selfing species are hypothesized to have a greater K_a/K_s ratio than outcrossing species. However, this ratio often refers to that measured in the sporophyte phase rather than in the gametophyte phase. Thus, it is of significance to study how different models of biphasic selection change K_a/K_s under multiple processes in both theory and practice. For comparison, we also consider π_a/π_s where π_a and π_s are the average heterozygosities at the nonsynonymous and synonymous sites, respectively. Conceptually, these two ratios reflect different time scales, with the K_a/K_s ratio measuring the long-term divergence of fixed alleles and the π_a/π_s ratio measuring either short- or long-term polymorphisms of coexisting alleles. Empirical studies showed that these two ratios differ in characterizing molecular evolution and speciation (Wang et al. 2021; Li et al. 2023). It is interested to compare the similarity and differences between the two ratios under different models of selection.

Previous theories relevant to this topic include deterministic or stochastic selection in haploid and diploid phases in a random mating system, with emphasis on the conditions for maintaining polymorphisms (Scudo 1967; Ewing 1977), or fixation/loss of an allele or equilibrium (Hartl 1976). The theory also includes the haplodiploid selection model that has the same population genetic consequences as the X-linked genes in sexual selection model (Lester and Selander 1979; Avery 1984; Hall and Goodisman 2012), the haploid competition controlled by parents (Otto et al 2015), and the sex-specific selection where sexually antagonistic or synergistic selection takes place (Connallon and Calsbeek 2020; Lund-Hansen et al. 2021). Theory for maintaining biphasic cycles, including antagonistic selection, is also developed through analyzing the interplay between genetic and ecological effects (Rescan et al. 2016) or the unequal haploid and diploid fitness (Scott and Rescan

2016). Here, we develop new models of selection in gametophyte and sporophyte phases of plant life cycle, with emphasis on molecular evolution measured by K_a/K_s and π_a or π_a/π_s. We begin by describing the model of antagonistic selection and then the model of synergistic selection. A mainland–island model of population structure is employed to look at the effects of pollen and seed flow on these two measures. We finally discuss potential implications of the new theory.

The Theory

Methodology

The model deals with a hermaphrodite plant species in a conventional mainland–island model. The mainland population is sufficiently large in size and stable in genetic composition. Only unidirectional gene flow from the mainland to the island is considered. Figure 1 shows the sequential events of life cycle modeled. We consider weak selection in both haploid gametophyte and diploid sporophyte phases and the same order of migration rate and drift effects. The terms with the second or higher order of these parameters are neglected for mathematic tractability. Effects of mutation on K_a/K_s and π_a/π_s at a nucleotide site are neglected in this study. The variables and parameters used in the model are summarized in supplementary table S1, Supplementary Material online.

Consider a biallelic site in autosomes, with an ancestral allele A and a mutant allele a. Two selection schemes are analyzed in the gametophyte and sporophyte phases (table 1). In the antagonistic selection scheme, allele A is deleterious in the gametophyte phase but favorable in the sporophyte phase, while allele a is favorable in the gametophyte phase but deleterious in the sporophyte phase. This selection scheme facilitates maintenance of polymorphisms (Haldane 1932; Haldane and Jayakar 1963), but genetic drift could yield fixation of the mutant allele. In the synergistic selection scheme, the mutant allele is deleterious in both gametophytic and sporophytic phases but can be fixed in a small population (Kimura 1962).

Based on the life cycle modeled (fig. 1), we derive systematic change of allele frequency, which is caused by gene flow and biphasic selection. Following Wright's idea (Wright 1969), we derive the changes in allele frequencies due to selfing and outcrossing separately and then combine them to calculate genotypic frequencies before seed flow. The diffusion model is then applied to derive the fixation probability of the mutant allele and the density distribution of its frequency after genetic drift (Kimura 1962). We focus on K_a/K_s and π_a (or π_a/π_s in the case with polymorphisms at a neutral site) in the island population under multiple scenarios. Based on the general analytical formulae, numerical examples are used to illustrate the patterns of K_a/K_s and π_a or π_a/π_s. Two Mathematica notebooks are provided

for calculating K_a/K_s and π_a or π_a/π_s for numerical analyses (Wolfram 1996).

Antagonistic Selection

General Case

Let p_{AA}, p_{Aa}, and p_{aa} be the genotypic frequencies in current adults in the island population. The per-generation change of gene frequency is derived below according to the life cycle (fig. 1). For the outcrossing part, with a probability of $1 - \alpha$, the fitness of gametes A and a is set as $w_A = 1 - s_h$ and $w_a = 1$ in both pollen and ovules, respectively, where s_h is the selection coefficient. The average fitness in pollen is calculated as $w_P = w_A p_A^* + w_a p_a^*$, where p_A^* and p_a^* are the allele frequencies after pollen flow. The average fitness in ovules is $w_O = w_A p_A + w_a p_a$ where p_A and p_a are the allele frequencies in adults because ovules do not migrate. Genotypic frequencies in seeds are then calculated according to the random combination between pollen and ovules.

For the selfing part, with a probability of α, gametophytic selection only occurs in heterozygotes. The mean fitness in pollen or ovules is $1 - \frac{1}{2}s_h$. Genotypic frequencies are calculated in seeds produced by selfing. Following Wright's (1969) idea, the overall genotypic frequencies in seeds are obtained by combining the selfing and outcrossing parts.

Genotypic frequencies after seed flow are then calculated. For antagonistic selection (table 1), the fitness of three genotypes is set as $W_{AA} = 1$, $W_{Aa} = 1 - hs_d$, and $W_{aa} = 1 - s_d$ where s_d is the selection coefficient and h is the degree of dominance. Allele a is completely masked by allele A in fitness when $h = 0$. Strong selection against allele a occurs when $h = 1$ (complete dominance). The additive selection model occurs in the sporophyte phase when $h = 1/2$. The average fitness in the sporophyte phase is $\overline{W} = W_{AA} p_{AA}^{**} + W_{Aa} p_{Aa}^{**} + W_{aa} p_{aa}^{**}$, where p_{ij}^{**} is the frequency of genotype ij after seed flow. The per-generation systematic changes in gene and genotypic frequencies are derived in detail in supplementary appendix S1, Supplementary Material online in the Supplementary Material.

For a mixed mating system, an equilibrium relationship for a single locus between the inbreeding coefficient F and the selfing rate α is approximated by $F = \frac{\alpha}{2 - \alpha}$ (Haldane 1924; Wright 1969). According to equation (S1.22) of supplementary appendix S1, Supplementary Material online, the per-generation systematic change in allele frequency for the mutant allele is expressed as

$$M_{\Delta p_a} = p_a^{***} - p_a$$

$$= \tilde{m}(Q_a - p_a) + s_h p_A p_a \left(1 - \frac{1}{2}\alpha(1 + F)\right)$$

$$- s_d p_A p_a \left(p_a + \frac{1}{2}\alpha p_A(1 + F) + h(p_A - p_a)\left(1 - \frac{1}{2}\alpha(1 + F)\right)\right),$$
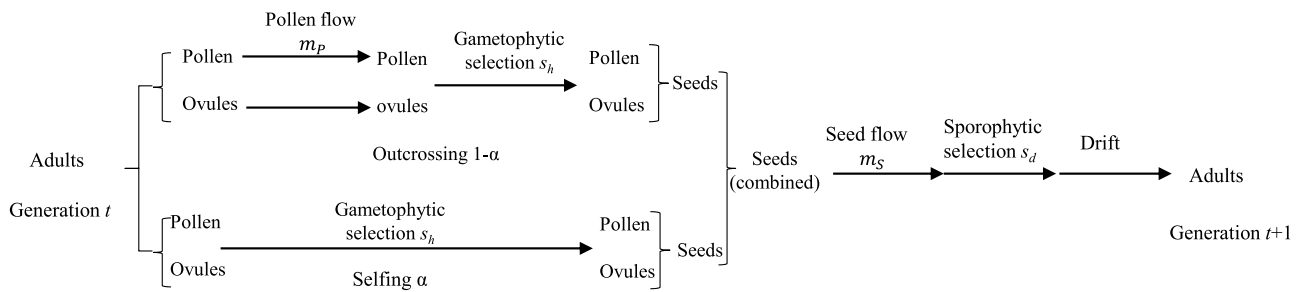
$$(1)$$

Fig. 1.—The life cycle used in model. The sequential events are indicated in plant life cycle, including generation of pollen and ovules, pollen flow, selection in the gametophyte phase, mixed mating, seed flow, selection in the sporophyte phase, and genetic drift. The arrow lines show occurrence of sequential events.

Table 1

Two Selection Schemes in Gametophyte and Sporophyte Phases

Selection Scheme	Gametophytic Selection		Sporophytic Selection		
	A	a	AA	Aa	aa
Antagonistic selection	$1 - s_h$	1	1	$1 - hs_d$	$1 - s_d$
Synergistic selection	1	$1 - s_h$	1	$1 - hs_d$	$1 - s_d$

where p_a^{***} is the allele frequency after sporophytic selection, $\tilde{m}$ stands for a composite migration rate, and $\tilde{m} = m_S + \frac{1-\alpha}{2}m_P$, in which m_P and m_S are the migration rates of pollen and seeds from the mainland population, respectively. Q_a ($Q_A + Q_a = 1$) is the frequency of allele a in migrants from the mainland population. The per-generation systematic change in equation (1) is caused by gene flow, gametophytic selection, and sporophytic selection. The per-generation systematic change for allele frequency p_A, $M_{\Delta p_A}$, is equal to the negative $M_{\Delta p_a}$, that is, $M_{\Delta p_A} = -M_{\Delta p_a}$.

We now incorporate the genetic drift effects into the change of allele frequencies (fig. 1). Let N_e be the effective size of the island population. The genetic drift process generates variation of allele frequency but does not change the mean allele frequency. The variance for the per-generation change of gene frequency Δp_a is given by

$$V_{\Delta p_a} = \frac{p_a^{***}(1 - p_a^{***})}{2N_e}$$
$$\approx \frac{p_a(1 - p_a)}{2N_e}. \tag{2}$$

Similarly, the variance for the per-generation change of allele frequency, Δp_A, is equal to $V_{\Delta p_a}$ according to the binomial distribution of allele frequency, $V_{\Delta p_A} = V_{\Delta p_a}$.

From Caballero and Hill (1992), the effective population size in a mixed mating system is re-expressed as $N_e = \frac{N}{1+F} = (1 - \frac{1}{2}\alpha)N$ under our assumption of a single locus without background selection, where N is the actual population

size. Selfing reduces the effective population size by a maximum rate of 50%.

According to Kimura (1962), the fixation probability of the mutant allele, $\varphi(p_0)$, with an initial frequency p_0 in a population of effective size N_e is calculated as

$$\varphi(p_0) = \frac{\int_0^{p_0} G(p_a)dp_a}{\int_0^1 G(p_a)dp_a}, \tag{3}$$

where

$$G(p_a) = exp\left(-2\int \frac{M_{\Delta p_a}}{V_{\Delta p_a}}dp_a\right). \tag{4}$$

Bessho and Otto (2017) provide an alternative formula for calculating the fixation probability for a selected allele with haploid–diploid life cycles under random mating only. In their case, the fixation probability is calculated by weighting the fixation probabilities in haploid and diploid phases with their fractions, which is different from the preceding calculation.

Substitution of equation (1) and equation (2) into equation (4) yields

$$G(p_a) = (1 - p_a)^{-2(2-\alpha)N\tilde{m}Q_A} p_a^{-2(2-\alpha)N\tilde{m}Q_a}.$$
$$exp(2(1-\alpha)(1-2h)Ns_d p_a^2 + 2N((\alpha + 2(1-\alpha)h)s_d - 2(1-\alpha)s_h)p_a).$$
$$\tag{5}$$

For antagonistic selection, the mutant allele (a) could attain a high frequency or approach fixation, depending upon the relative effects of selection, migration, and genetic drift.

If the mutant allele is neutral ($s_h = s_d = 0$), that is, in the case of a synonymous site, its fixation probability with an initial frequency of p_0 is

$$\varphi_0(p_0) = \frac{\int_0^{p_0} p_a^{-2(2-\alpha)N\tilde{m}Q_a}(1 - p_a)^{-2(2-\alpha)N\tilde{m}Q_A}dp_a}{\int_0^1 p_a^{-2(2-\alpha)N\tilde{m}Q_a}(1 - p_a)^{-2(2-\alpha)N\tilde{m}Q_A}dp_a}. \tag{6}$$

When the island population is completely isolated from the mainland population ($\tilde{m} = 0$), gene frequency in the island population is operated by the genetic drift only. The fixation probability is $\varphi_0(p_a) = \frac{\int_0^{p_0} 1 \, dp_a}{\int_0^1 1 \, dp_a} = p_0$, which equals the initial allele frequency, for example, $\varphi_0\left(\frac{1}{2N}\right) = \frac{1}{2N}$ (Kimura 1962).

Let u be the mutation rate from allele A to a at the nonsynonymous site, without backward mutation. The substitution rate with allele a at the nonsynonymous site, measured by K_a, is calculated as

$$K_a = 2N_e u * \varphi(p_0), \tag{7}$$

where $2N_e u$ is the expected number of nonsynonymous mutations (mutant allele a) entering the population per generation and $\varphi(p_0)$ is the fixation probability of the mutant allele with initial frequency p_0 (Bustamante 2005; Kryazhimskiy and Plotkin 2008). K_a measures the rate of nonsynonymous substitution between gene sequences per nonsynonymous site.

Let the substitution rate per synonymous site (the neutral allele), K_s, be calculated as

$$K_s = 2N_e \mu * \varphi_0(p_0). \tag{8}$$

The $2N_e \mu$ is the expected number of synonymous (neutral) mutations entering the population per generation, and $\varphi_0(p_0)$ is the fixation probability of a neutral mutant with initial frequency p_0. For a completely isolated population ($\tilde{m} = 0$), K_s is equal to the mutation rate, $K_s = 2N_e \mu \times \frac{1}{2N_e} = \mu$. If both migration and drift processes are involved, K_s is calculated by substituting equation (6) into equation (8).

Denote a ratio of the evolutionary rate at the nonsynonymous site relative to that at the neutral site by $\frac{K_a}{K_s}$, which is calculated by

$$\frac{K_a}{K_s} = \frac{\varphi(p_0)}{\varphi_0(p_0)} \cdot \frac{u}{\mu}. \tag{9}$$

Unequal mutation rates ($u \neq \mu$) also influence K_a/K_s, which is not focused on here.

We now consider polymorphisms at nonsynonymous and synonymous sites (π_a or π_a/π_s). Under the joint effects of migration, selection, and genetic drift, the density distribution of allele frequency at equilibrium is calculated as $\phi(p_a) = \frac{C}{V_{\Delta p_a}} \exp\left(2\int \frac{M_{\Delta p_a}}{V_{\Delta p_a}} dp_a\right) = \frac{C}{V_{\Delta p_a} G(p_a)}$ (Kimura 1962), where C is the constant that satisfies $\int_0^1 \phi(p_a) dp_a = 1$ (Wright 1969). Substitution of equations (1) and (2) into $\phi(p_a)$ yields the following expression:

$$\phi(p_a) = C(2-\alpha)N \cdot$$
$$exp(-2N(1-\alpha)(1-2h)s_d p_a^2 - 2N((\alpha + 2h(1-\alpha))s_d$$
$$- 2(1-\alpha)s_h)p_a) \cdot p_a^{2N(2-\alpha)\tilde{m}Q_a - 1}(1-p_a)^{2N(2-\alpha)\tilde{m}Q_A - 1}. \tag{10}$$

The expected allele frequency, $\bar{p}_a$, is numerically calculated by $\bar{p}_a = \int_0^1 p_a \phi(p_a) dp_a$, and $\bar{p}_A$ is equal to $1 - \bar{p}_a$.

Under gametophytic and/or sporophytic selection, when both alleles may coexist in the island population, the average heterozygosity per nonsynonymous site, π_a, is calculated as

$$\pi_a = \int_0^1 2p_a(1-p_a)\phi(p_a) dp_a = 2(\bar{p}_A \bar{p}_a - V_p), \tag{11}$$

where V_p is the variance of allele frequency at equilibrium, which is induced by the genetic drift process.

At the neutral site ($s_d = s_h = 0$), the density distribution of allele frequency of equation (10) is simplified as

$$\phi_0(p_a) = \frac{\Gamma(2N(2-\alpha)\tilde{m})}{\Gamma(2N(2-\alpha)\tilde{m}Q_A)\Gamma(2N(2-\alpha)\tilde{m}Q_a)} \cdot$$
$$p_a^{2N(2-\alpha)\tilde{m}Q_a - 1}(1-p_a)^{2N(2-\alpha)\tilde{m}Q_A - 1}. \tag{12}$$

The genetic diversity at the synonymous site is maintained by migration and genetic drift. The average heterozygosity per synonymous site, π_s, is calculated by

$$\pi_s = \int_0^1 2p_a(1-p_a)\phi_0(p_a) dp_a$$
$$= 2Q_A Q_a \cdot \left(1 - \frac{1}{1 + 2(2-\alpha)N\tilde{m}}\right), \tag{13}$$

where the proportion in parentheses is the reduction due to genetic drift effects. When only genetic drift process operates on the neutral site, polymorphisms could be transiently maintained before any allele is lost or fixed (e.g., $\pi_s = 0$).

In the following parts of this section, we consider specific cases, which can be applied to interpret the evolutionary divergence among genes expressed in gametophyte phase, sporophyte phase, and two phases. Additive and nonadditive selection models in the sporophyte phase are separately addressed.

Additive Selection (h = 1/2)

Consider a specific case where the island population is completely isolated from the mainland population ($\tilde{m} = 0$). Let mutation rates be equal at synonymous and nonsynonymous sites ($u = \mu$). Following Kimura (1962), we assume that the nonsynonymous site is evolved by selection and genetic drift processes. When only the gametophytic selection takes place ($s_d = 0$, $s_h \neq 0$), the mutant allele (a) is favorable and is expected to be fixed although drift could slightly reduce the fixation probability (Kimura 1962). This case could occur for the genes that

are expressed only in the gametophyte phase. The relative evolutionary rate is approximated by

$$\frac{K_a}{K_s} = \frac{1 - \exp(-4N(1-\alpha)s_h p_0)}{1 - \exp(-4N(1-\alpha)s_h)} \cdot \frac{1}{p_0}. \quad (14)$$

The initial frequency of allele a at both the neutral and selective sites is set as p_0 in deriving (14). It can be shown that K_a/K_s increases as s_h increases. Let $\epsilon = 4N(1-\alpha)s_h$. The partial differential of $\frac{K_a}{K_s}$ with respect to s_h is

$$\frac{\partial\left(\frac{K_a}{K_s}\right)}{\partial s_h}$$
$$= \frac{4N(1-\alpha)}{p_0(1-e^{-\epsilon})^2} e^{-\epsilon(1+p_0)} (p_0 e^{\epsilon} - p_0 - e^{\epsilon p_0} + 1), \quad (15)$$

which is greater than 0 from the results in supplementary appendix S2, Supplementary Material online. Similarly, a large population enhances gametophytic selection $\left(\partial\left(\frac{K_a}{K_s}\right)/\partial N > 0\right)$ and facilitates the fixation of allele a. However, selfing reduces the selection efficacy $\left(\partial\left(\frac{K_a}{K_s}\right)/\partial\alpha < 0\right)$ and hence reduces the substitution rate with allele a.

Figure 2 shows two numerical examples under gametophytic selection only ($N = 30$, $s_d = 0$, and $s_h = 0.03$ or 0.06), with the initial frequency of the mutant allele $p_0 = \frac{1}{2N}$. Selfing reduces the efficacy of gametophytic selection. Because the mutant allele approaches fixation despite the impeding effects from selfing, the level of polymorphism at this site decreases to 0 ($\pi_a = 0$). Under the sole genetic drift process, the neutral site is polymorphic for finite time before the mutant allele is fixed or lost ($\pi_s = 0$), and selfing accelerates this process because it reduces the effective population size.

When only sporophytic selection takes place ($s_d \neq 0$, $s_h = 0$), the mutant allele a is deleterious and is under purifying selection. In this situation, the relative evolutionary rate is approximated by

$$\frac{K_a}{K_s} = \frac{\exp(2Ns_d p_0) - 1}{\exp(2Ns_d) - 1} \cdot \frac{1}{p_0}. \quad (16)$$

The initial frequency of allele a at the neutral site is set as p_0 in deriving equation (16). A smaller population size and/or weaker selection pressure can impede the efficacy of purifying selection against the mutant allele, $\partial\left(\frac{K_a}{K_s}\right)/\partial N < 0$ and $\partial\left(\frac{K_a}{K_s}\right)/\partial s_d < 0$. If the initial allele frequency is $p_0 = \frac{1}{2N}$, equation (16) reduces to the conventional result obtained by Kimura (1962). Note that N is the actual population size; equation (16) can also be re-expressed in terms of
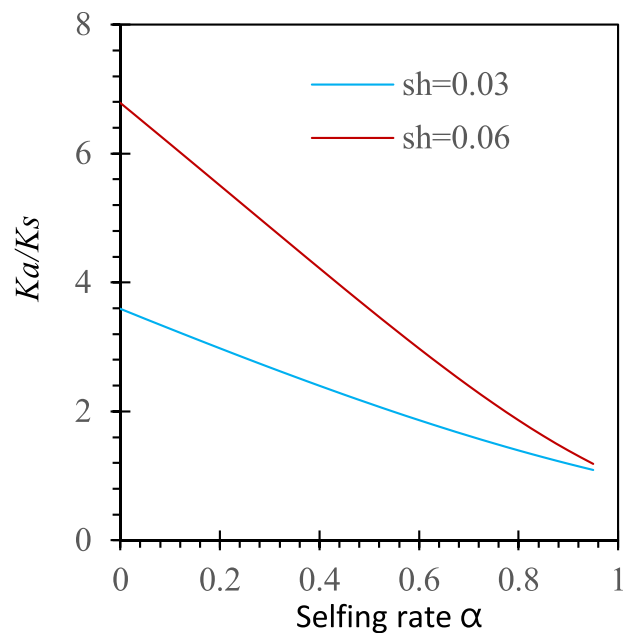

FIG. 2.—Examples of gametophytic selection changing K_a/K_s under different selfing rates. Results are derived from equation (14) with the selection coefficients $s_h = 0.03$ or 0.06 and $s_d = 0$, the degree of dominance $h = 0.5$, the population size $N = 30$, the initial frequencies $p_0 = 1/2N$, and the migration rates of seeds and pollen $m_S = m_P = 0$.

the effective population size through the relationship of $N = 2N_e/(2-\alpha)$, that is

$$\frac{K_a}{K_s} = \frac{4N_e s_d}{(2-\alpha)\left(\exp\left(\frac{4N_e s_d}{2-\alpha}\right) - 1\right)}. \quad (17)$$

Selfing increases the evolutionary rate K_a/K_s $\left(\partial\left(\frac{K_a}{K_s}\right)/\partial\alpha > 0\right)$ despite purifying selection against the mutant allele ($K_a/K_s < 1$). This is realized through reducing effective population size rather than reducing the efficacy of gametophytic selection.

To compare the relative evolutionary rates between gametophyte- and sporophyte-specific genes, we assume comparable selection strength in two phases ($s_d = 2s_h$). In the outcrossing system ($\alpha = 0$), the two phase-specific genes have equal absolute K_a/K_s values from equations (14) and (16), given the same initial allele frequency of the mutant allele. However, in the mixed or selfing system ($0 < \alpha \leq 1$), which leads to $s_d = 2s_h > 2(1-\alpha)s_h$, the absolute K_a/K_s in equation (14) is smaller than the absolute K_a/K_s in equation (16), given their same initial allele frequencies. Gametophyte-specific genes are expected to have a lower absolute evolutionary rate than sporophyte-specific genes in the mixed mating system.

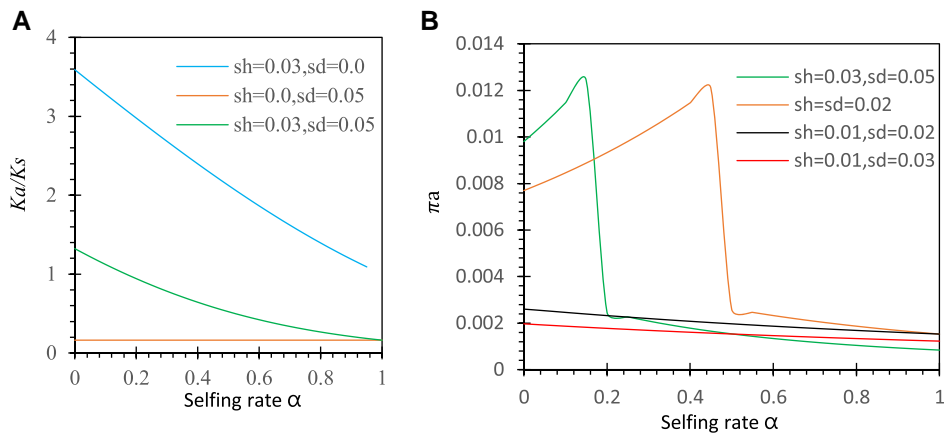When both gametophytic selection and sporophytic selection take place, this likely occurs for those genes that

Fig. 3.—Examples of the patterns of K_a/K_s and π_a under antagonistic selection. Results are derived from a Mathematica notebook. In (A), three sets of selection parameter settings are the gametophytic selection ($s_d = 0$ and $s_h = 0.03$), the sporophyte selection ($s_d = 0.05$ and $s_h = 0.0$), and biphasic selection ($s_d = 0.05$ and $s_h = 0.03$). In (B), four biphasic selection settings are examined: $s_d = 0.05$ and $s_h = 0.03$; $s_d = s_h = 0.02$; $s_d = 0.02$, $s_h = 0.01$; and $s_d = 0.03$ and $s_h = 0.01$. Other parameters are the population size $N = 30$, the degree of dominance $h = 0.5$, and the migration rates of seeds and pollen $m_S = m_P = 0$.

are co-expressed in two phases. The evolutionary rate for the mutant allele is calculated by

$$\frac{K_a}{K_s} = \frac{1 - exp(2N(s_d - 2(1-\alpha)s_h)p_0)}{1 - exp(2N(s_d - 2(1-\alpha)s_h))} \cdot \frac{1}{p_0}. \quad (18)$$

An increase in sporophytic selection coefficient (s_d) reduces K_a/K_s, whereas an increase in gametophytic selection coefficient (s_h) increases K_a/K_s. Equation (18) indicates that biphasic selection is completely offset in outcrossing ($\alpha = 0$) under comparable strengths of allelic selection ($s_d = 2s_h$), which yields $K_a/K_s = 1$ (note $G(p_a) = 1$ under these conditions). Selfing ($\alpha \neq 0$) modifies the extent of offset by reducing the efficacy of gametophytic selection.

From equation (18), we obtain the following relationship:

$$\frac{\partial\left(\frac{K_a}{K_s}\right)}{\partial s_h} = -2(1-\alpha)\frac{\partial\left(\frac{K_a}{K_s}\right)}{\partial s_d}. \quad (19)$$

Under comparable allelic selection between the two phases ($s_d = 2s_h$), the responses of the evolutionary rate K_a/K_s to the change in gametophytic and sporophytic selection are the same in the outcrossing system ($\alpha = 0$), that is, $\partial\left(\frac{K_a}{K_s}\right)/\partial s_h = \left|\partial\left(\frac{K_a}{K_s}\right)/\partial\left(\frac{s_d}{2}\right)\right|$. However, the mixed mating system ($0 < \alpha < 1$) can lead to the gametophytic selection to be less efficient than the sporophyte selection, that is, $\partial\left(\frac{K_a}{K_s}\right)/\partial s_h < \left|\partial\left(\frac{K_a}{K_s}\right)/\partial\left(\frac{s_d}{2}\right)\right|$. Complete selfing ($\alpha = 1$) removes the gametophytic selection but enhances purifying selection against the mutant allele in the sporophyte phase.

Figure 3A shows the pattern of K_a/K_s for the mutant allele under the antagonistic selection compared with those under gametophytic and sporophyte selection alone. The results indicate that the opposite effects from antagonistic selection can partially offset with the given parameter settings ($s_d = 0.05$ and $s_h = 0.03$). Purifying selection in the sporophyte phase dominates and produces small K_a/K_s under the condition of $s_d > 2(1-\alpha)s_h$. Positive selection in the gametophytic phase dominates and produces large K_a/K_s under the condition of $s_d < 2(1-\alpha)s_h$. A complete offset is expected under the condition of $s_d = 2(1-\alpha)s_h$, which leads to $K_a/K_s = 1$ (neutral case).

Figure 3B shows the pattern of π_a as the selfing rate increases from 0 to 1. The π_a value increases with the selfing rate under the condition of $s_d - 2(1-\alpha)s_h < 0$ but decreases with the selfing rate under the condition of $s_d - 2(1-\alpha)s_h > 0$. The turning point is located at the selfing rate $\alpha = 1 - s_d/2s_h$, where completely balancing selection occurs.

Nonadditive Selection ($h \neq 1/2$)

Under nonadditive selection in the sporophyte phase, we consider two specific cases ($h = 0$ or 1) to illustrate the effects of dominance selection on evolutionary rate. Figure 4A shows a comparison of the results under $h = 1$ versus $h = 0$ with the same initial allele frequency $p_0 = 1/2N$ and without gametophytic selection ($s_h = 0$). Purifying selection against allele a is enhanced under complete dominance ($h = 1$) but weakened when allele a is completely masked by allele A in heterozygotes ($h = 0$). This results in a smaller K_a/K_s under $h = 1$ than under $h = 0$. However, K_a/K_s values tend to be the same in two cases as the selfing rate increases from 0 to 1. The π_a value exhibits the pattern similar to that of K_a/K_s as the selfing
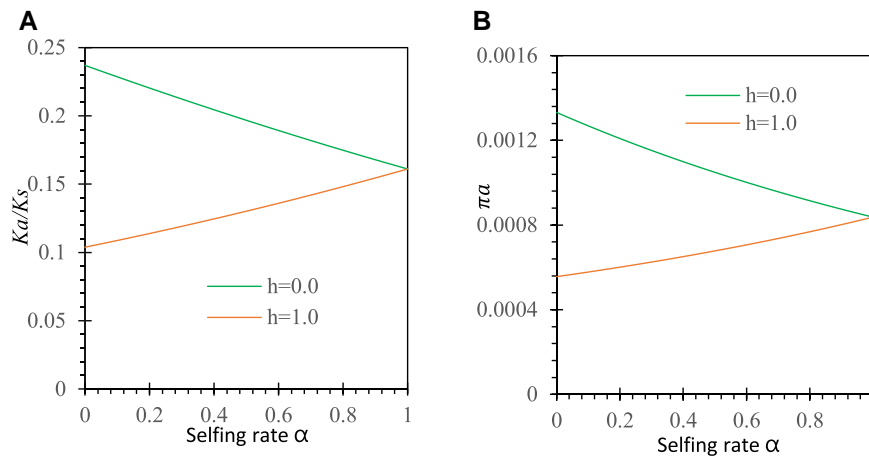
Fig. 4.—Examples of dominance effects on K_a/K_s and π_a under antagonistic selection. Results are derived from equations (3), (5), (8), (9), and (11): (A) K_a/K_s ratio under $h = 1$ and 0 and (B) π_a under $h = 1$ and $h = 0$. Other parameters are the selection coefficients $s_d = 0.05$ and $s_h = 0$, the population size $N = 30$, the initial allele frequencies $p_0 = 1/2N$, and the migration rates of seeds and pollen $m_S = m_P = 0$.

rate increases from 0 to 1 (fig. 4B). Complete dominance of the mutant allele over the ancestral allele ($h = 1$) has larger effects of purifying selection on the mutant allele in the selfing and mixed mating systems.

Synergistic Selection

In the synergistic selection scheme, allele a is deleterious in both gametophyte and sporophyte phases but could be maintained under migration or has a small probability of fixation under large genetic drift effects (Kimura 1962). The fitness of gametes and zygotes is set in table 1. Under the mainland–island model, the per-generation systematic change of gene frequency, $M_{\Delta p_a}$, can be directly obtained by replacing positive s_h in equation (1) with negative s_h. Although biphasic selection is in the same direction, selfing can regulate the impacts of gametophytic selection against the mutant allele (Hu and Li 2003; Hu 2015).

The variance for the per-generation change of allele frequency due to genetic drift, $V_{\Delta p_a}$, remains the same as equation (2). The fixation probability of allele a, $\varphi(p_0)$, is calculated using equation (3) except that $G(p_a)$ is calculated using a negative s_h in equation (5). K_a/K_s for the mutant allele is numerically calculated using equation (9) under equal mutation rates between synonymous and nonsynonymous sites. The distribution of allele frequency at equilibrium, $\phi(p_a)$, is calculated by replacing the negative s_h in equation (10) with positive s_h. The average heterozygosity at the nonsynonymous site, π_a, is calculated using equation (11). The average heterozygosity at the synonymous site, π_s, is given by equation (13) when polymorphisms are maintained by migration and drift processes. Using these equations, we numerically evaluate π_a or π_a/π_s in different cases.

Additive Selection (h = 1/2)

Consider a specific case where the island population is completely isolated from the mainland population. Under the additive selection model ($h = 1/2$) and equal mutation rates at synonymous and nonsynonymous sites ($u = \mu$), K_a/K_s is derived as

$$\frac{K_a}{K_s} = \frac{1 - exp(2N(s_d + 2(1 - \alpha))s_h)p_0)}{1 - exp(2N(s_d + 2(1 - \alpha))s_h))} \frac{1}{p_0}. \quad (20)$$

Unlike the case of antagonistic selection, equation (20) indicates the cumulative effects of biphasic selection on K_a/K_s. When the initial frequency is $p_0 = \frac{1}{2N}$, equation (20) is approximated by $\frac{K_a}{K_s} = \frac{2N(s_d + 2(1-\alpha)s_h)}{e^{2N(s_d + 2(1-\alpha)s_h)} - 1}$, indicating that selfing reduces the contribution of gametophytic selection to K_a/K_s because of the negative selfing (α) effects in the term of gametophytic selection.

Similarly, under comparable selection strength in two phases ($s_d = 2s_h$), the gametophyte- and sporophyte-specific genes have equal K_a/K_s values in the outcrossing system ($\alpha = 0$), given the same initial allele frequency of the mutant allele. However, in the mixed or selfing system ($0 < \alpha \leq 1$), which leads to $s_d = 2s_h > 2(1 - \alpha)s_h$, K_a/K_s for gametophyte-specific genes is smaller than that for sporophyte-specific genes, given their same initial allele frequencies.

For genes undergoing gametophytic selection only, K_a/K_s increases as the selfing rate increases (supplementary fig. S1A, Supplementary Material online). Selfing impedes gametophytic selection against the mutant allele. For genes undergoing sporophytic selection only, K_a/K_s for the mutant allele has the same expression as equation (15) or (16) in antagonistic selection and increases as the selfing rate increases in terms of effective population
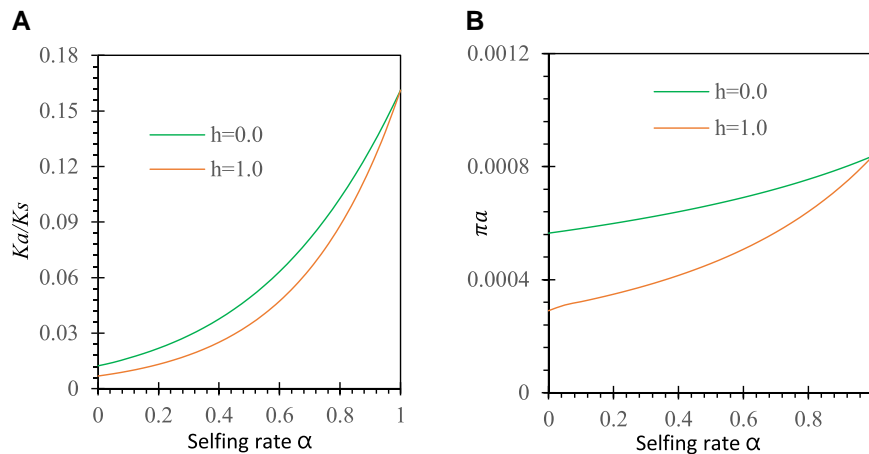
Fig. 5.—Examples of dominance effects on K_a/K_s and π_a under synergistic selection. Results are derived from a Mathematica notebook. The parameters except for the degree of dominance ($h = 0$ and 1) are the gametophytic selection ($s_h = 0.03$), sporophytic selection ($s_d = 0.05$), the population size $N = 30$, the initial allele frequencies $p_0 = 1/2N$, and the migration rates of seeds and pollen $m_S = m_P = 0$.

size. However, K_a/K_s is not sensitive to the selfing rate in terms of actual population size (supplementary fig. S1A, Supplementary Material online). For genes undergoing biphasic selection, cumulative effects reinforce purifying selection against the mutant allele (supplementary fig. S1A, Supplementary Material online).

The polymorphism at the nonsynonymous site, π_a, exhibits the patterns similar to those of K_a/K_s as the selfing rate increases from 0 (outcrossing) to 1 (selfing) (supplementary fig. S1B, Supplementary Material online). Biphasic selection further reduces polymorphic level, compared with those under single-phase selection.

For genes undergoing biphasic selection, comparable strengths of selection ($s_d = 2s_h$) lead to equal contribution from each phase to the evolutionary rate in outcrossing systems ($\alpha = 0$). However, selfing reduces the proportion of gametophytic selection in a mixed mating system ($\alpha \neq 0$). Similarly, we obtain the following relationship:

$$\partial\left(\frac{K_a}{K_s}\right)\bigg/\partial s_h = (1-\alpha)\partial\left(\frac{K_a}{K_s}\right)\bigg/\partial\left(\frac{s_d}{2}\right). \quad (21)$$

Selfing reduces sensitivity of K_a/K_s to the change of gametophytic selection, resulting in a lower efficiency of gametophytic selection than sporophytic selection.

From equation (20), the response of K_a/K_s to the change of the selfing rate is

$$\partial\left(\frac{K_a}{K_s}\right)\bigg/\partial\alpha = \frac{4Ns_h e^{x(1+p_0)}}{(e^{-x}-1)^2 p_0}(p_0 e^{-x} - p_0 - e^{-xp_0} + 1), \quad (22)$$

where $x = 2N(s_d + 2(1-\alpha)s_h)$. From supplementary appendix S2, Supplementary Material online, this response

is positive (>0), indicating that selfing increases K_a/K_s through reducing the efficacy of gametophytic selection.

Nonadditive Selection ($h \neq 1/2$)

For genes undergoing sporophyte selection only ($s_h = 0$, $s_d \neq 0$), the same patterns as those in the case of antagonistic selection ($s_h = 0$, $s_d \neq 0$) are expected for the change of K_a/K_s or π_a with the selfing rate. This is because $G(p_a)$ and $\phi(p_a)$ functions are the same between the two cases. Figure 5A shows that selfing enhances K_a/K_s under biphasic selection. As expected, the complete underdominance of the ancestral allele over the mutant allele ($h = 1$) gives great selection pressure against the mutant allele. When the ancestral allele completely masks the mutant allele in heterozygotes ($h = 0$), selection pressure is alleviated, which produces K_a/K_s slightly higher than that under $h = 1$.

Figure 5B shows that selfing impedes gametophytic selection but increases π_a under biphasic selection. Polymorphism is enhanced when the mutant allele is completely masked by the ancestral allele ($h = 0$). As expected, when the mutant allele is underdominant ($h = 1$), polymorphism is reduced.

Compared with the results from sporophyte selection only (fig. 4), the presence of gametophytic selection adds additional selection against the mutant allele (fig. 5). Both K_a/K_s and π_a decrease in a mixed mating system ($0 < \alpha < 1$).

Pollen and Seed Flow

Gene flow with higher polymorphisms in migrants counteracts directional selection in the recipient population (Wright 1969). This impedes fixation of mutant or ancestral alleles but may increase polymorphisms (π_a or π_a/π_s). To examine effects of pollen flow, we consider a predominantly
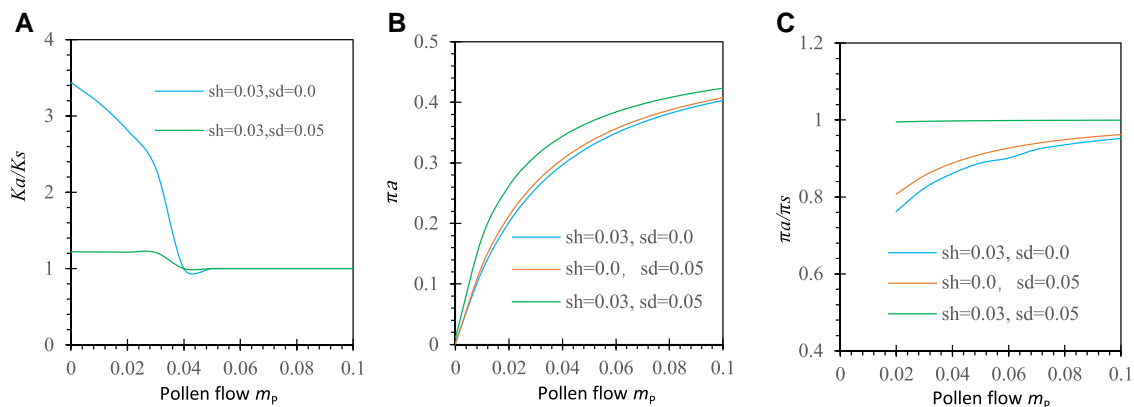
FIG. 6.—Examples of the effects of pollen flow on K_a/K_s and π_a or π_a/π_s under antagonistic selection in a predominantly outcrossing system ($\alpha = 5\%$). Results are derived from a Mathematica notebook: (A) K_a/K_s; (B) π_a; and (C) π_a/π_s. The parameters except for selection coefficients shown in figures are the population size $N = 30$, the initial allele frequencies $p_0 = 1/2N$, the degree of dominance $h = 0.5$, the migration rate of seeds $m_S = 0$, and the migrant allele frequencies $Q_A = Q_a = 0.5$.

outcrossing species that allows pollination from alien pollen. For instance, let the migrant allele frequency $Q_a = 0.5$ in the mainland population, with the maximum polymorphism. Under antagonistic selection, fixation of either the mutant allele or the ancestral allele is very small under constant rate of pollen or seed flow. Figure 6A shows that, when the mutant allele is under gametophytic ($s_h = 0.03$, $s_d = 0$, and $\alpha = 5\%$) or biphasic selection ($s_h = 0.03$, $s_d = 0.05$, and $\alpha = 5\%$), the evolution rate (K_a/K_s) decreases to 1 as the migration rate of pollen increases from 0 to 0.1. When the mutant allele is deleterious ($s_h = 0$, $s_d = 0.05$, and $\alpha = 5\%$), K_a/K_s (<1) also decreases as the migration rate of pollen increases (note that $G(p_a)$ is not integrable using Mathematica notebooks when m_P is greater than 0.04 with the parameter settings). Figure 6B shows that π_a increases as the migration rate of pollen increases from 0 to 0.1. Likewise, π_a/π_s also increases as the migration rate of pollen increases (fig. 6C). Biphasic selection enhances π_a/π_s compared with the case of single-phase selection. Although seed flow is not restricted by the type of mating system, it generates the patterns similar to those generated by pollen flow in changing K_a/K_s and π_a or π_a/π_s (supplementary fig. S2, Supplementary Material online).

Under synergistic selection, gene flow reduces K_a/K_s but increases π_a or π_a/π_s. In the predominantly outcrossing system (e.g., $\alpha = 5\%$), K_a/K_s decreases as m_P increases from 0 to 0.03 under both single-phase selection and biphasic selection (supplementary fig. S3, Supplementary Material online). Polymorphisms in terms of π_a and π_a/π_s increase as the migration rate of pollen increases from 0 to 0.1 (supplementary fig. S3, Supplementary Material online). Unlike the case of antagonistic selection, cumulative effects are present under biphasic selection (supplementary fig. S3, Supplementary Material online). Similarly, seed flow generates the patterns similar to those generated by pollen flow

in changing K_a/K_s, π_a, and π_a/π_s (supplementary fig. S4, Supplementary Material online). The difference between seed and pollen flow is due to larger effects of seed flow, given their same migration rates.

Genetic Drift

To evaluate the effects of genetic drift, we fix all parameters except for changing population size (N) in the isolated island population ($m_P = m_S = 0$). As expected, a large population improves selection efficiency but still generates different patterns of evolutionary rates between the two selection schemes. Under antagonistic selection, K_a/K_s substantially increases as the population size increases for genes undergoing gametophytic selection only (fig. 7A). This is because positive selection for the mutant allele is enhanced in larger populations. For genes undergoing sporophytic selection only, larger populations slightly reduce K_a/K_s because the power of purging deleterious alleles increases (fig. 7A). For genes undergoing biphasic selection and additive selection model ($h = 1/2$) in the sporophyte phase, K_a/K_s could increase as the population increases under the condition of $s_d - 2(1 - \alpha)s_h < 0$ (fig. 7A) or decrease under the condition of $s_d - 2(1 - \alpha)s_h > 0$ (data not provided here). Opposite effects between gametophytic and sporophyte selection mostly offset in reducing K_a/K_s when the population size is large.

Both gametophytic and sporophyte selection alone can substantially reduce π_a as the genetic drift effects become small (fig. 7B). The antagonistic selection between two phases facilitates maintenance of higher polymorphisms, compared with the case of single-phase selection (fig. 7B).

Under synergistic selection, the mutant allele is deleterious in both phases. A large population size facilitates purging of deleterious alleles in both phases, resulting in
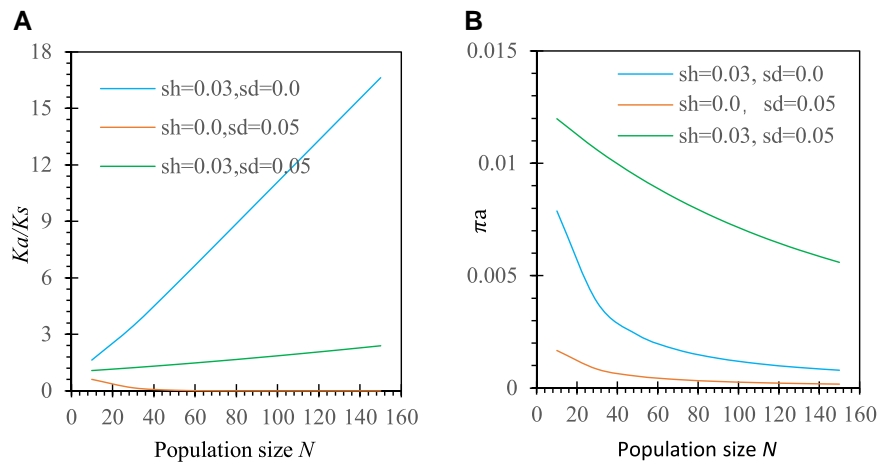
Fig. 7.—Examples of genetic drift effects K_a/K_s and π_a under antagonistic selection in a predominantly outcrossing system ($\alpha = 5\%$): (A) K_a/K_s and (B) π_a. Results are derived from a Mathematica notebook. Parameters except for the selection coefficient and population size shown in figures are the initial allele frequencies $p_0 = 1/2N$, the degree of dominance $h = 0.5$, and the migration rates of pollen and seeds $m_P = m_S = 0$.

decreasing patterns of K_a/K_s and π_a as the genetic drift effects become small (supplementary fig. S5, Supplementary Material online). Unlike the case of antagonistic selection, cumulative effects of purging deleterious alleles are present for genes under biphasic selection, which further reduces K_a/K_s and π_a (supplementary fig. S5, Supplementary Material online).

Discussion

The ratio of nonsynonymous to synonymous divergence at a nucleotide site, K_a/K_s, is widely applied to detecting natural selection with molecular sequence data. To estimate K_a/K_s, we may use pairwise or multiple homologous (mostly orthologous but less frequently paralogous) protein-coding gene sequences. With the pairwise sequences, we count the numbers of synonymous and nonsynonymous sites in each sequence and their differences at synonymous and nonsynonymous sites. We then calculate the proportions of synonymous and nonsynonymous differences. The evolutionary rate per nucleotide site is estimated according to a specific model of molecular evolution, such as the one-parameter model (Jukes and Cantor 1969). With multiple gene sequences, the maximum likelihood estimates of K_a/K_s at a nucleotide site can be obtained using phylogeny-based method (Yang 2006). For instance, the CODEML program from PAML (Yang 2007) or MEGA package (Kumar et al. 2018) can be applied to estimating this ratio. Although it is well known that K_a/K_s varies across sites along a gene sequence, such a "spatial" variation across sites with positive and purifying selection partly offsets and makes it difficult to detect selection at the gene level. Here, we demonstrate that phase variation of selection during the life cycle may also make it difficult to detect

selection for genes co-expressed across phases. Selfing can regulate K_a/K_s through its interaction with biphasic selection. In addition, we show that π_a or π_a/π_s also exhibits different patterns under one-phasic or biphasic selection. All these results aid in characterizing molecular evolution under phase variation of selection and diverse systems of mating.

Specific Conclusions

Several specific conclusions can be drawn below:

1. Opposite effects occur in antagonistic selection between gametophytic and sporophytic phases, leading to compensation in estimating K_a/K_s for mutant alleles under biphasic selection. However, cumulative effects on K_a/K_s occur for mutant alleles under synergistic biphasic selection. Under the additive selection model ($h = 1/2$) and comparable strength of biphasic allelic selection ($s_d = 2s_h$), gametophytic selection is less effective than sporophytic selection in changing K_a/K_s in the mixed or selfing mating system ($0 < \alpha \leq 1$) but equally effective in the outcrossing system ($\alpha = 0$).

2. Selfing oppositely interacts with gametophytic and sporophytic selection in shaping K_a/K_s and π_a or π_a/π_s in either antagonistic or synergistic selection. Under the additive selection model, selfing can increase or reduce K_a/K_s, depending upon the relative selection strength of the mutant allele in two phases. Selfing reduces efficacy of gametophytic selection. Under the nonadditive selection model ($h \neq 1/2$), partial dominance ($h > 1/2$) strengthens the purging of deleterious alleles in the sporophyte phase, while partial recessivity ($h < 1/2$) facilitates masking deleterious alleles in

heterozygotes and weakens the purging of deleterious alleles.

3. Gene flow reduces K_a/K_s of the mutant allele in both antagonistic and synergistic selection. Gene flow reduces selection efficacy but may increase π_a or π_a/π_s. Seed flow has larger effects than pollen flow on K_a/K_s and π_a or π_a/π_s, given their same migration rates.

4. Genetic drift has opposite effects on K_a/K_s for genes expressed in one or two phases under antagonistic selection, depending on the relative selection coefficients and the type of mating system. However, genetic drift has the same directional effects on K_a/K_s for genes expressed in two phases under synergistic selection.

Model Assumption and Comparison

Previous studies have examined the properties of K_a/K_s and π_a/π_s by emphasizing the effects of selfing on the effective population size and genetic hitchhiking (Glemin 2007; Glemin and Muyle 2014). They showed that the reduction in effective population size can relax selection against deleterious alleles (Glemin et al. 2019; Li et al. 2023). They confirmed that difference between the two ratios is pronounced between selfers and outcrossers at equilibrium. Here, we also study the effects of selfing on the effective population size but exclude genetic hitchhiking or background selection effects (Maynard-Smith and Haig 1974; Charlesworth et al. 1993). Besides, we consider phase variation of selection in a life cycle and separation of gene flow via pollen and seeds. The multiple-site–based method was used by Glemin (2007), where three probabilities of nonsynonymous mutations (neutral, deleterious, and advantageous) were jointly considered according to the infinite-site mutation model. The present analysis is individual site based, assuming that the nonsynonymous site is under weak selection (Welch et al. 2008) while the synonymous site is neutral.

The site model is appropriate for interpreting the results derived from individual site analyses, such as the site model presented by Yang (2007). Two alleles are considered at a single site because more than two alleles at a nucleotide site are infrequent in natural populations. Fixation of a mutant allele is governed by selection, drift, and migration processes. When the mutation effects are included, a more complicated fixation probability is needed to derive, such as the theory of mutational influx equilibrium (Wright 1938; Sawyer and Hartl 1992) and the nonequilibrium theory of site-frequency spectrum (Evans et al. 2007). The present single-site model differs from the infinite-site model on which site-frequency spectrum theory is based. Polymorphisms can be maintained at synonymous sites under mutational influx ($\pi_s \neq 0$). Also, inclusion of mutational influx in a single-site model could likely yield a comparable π_a/π_s ratio to that derived from site-frequency spectrum

theory in an isolated population ($\tilde{m} = 0$; Sawyer and Hartl 1992). This needs further clarifications. If the continuous selection effects of mutants are considered, such as a gamma distribution (Piganeau and Eyre-Walker 2003), an appropriate extension is required (Kryazhimskiy and Plotkin 2008).

The theory assumes that sites are independent from each other. When multiple sites are considered simultaneously, linkage disequilibria (LD) between sites could be generated by selfing through reducing recombination rate (Glemin 2007) or by seed and pollen flow. Although analytical fixation probability for a mutant allele is unavailable under LD, it is speculated that fixation probability of a mutant allele at a nonsynonymous site could be influenced by its linked selective sites (Maynard-Smith and Haig 1974; Charlesworth et al. 1993). Positive (or negative) LD between a mutant allele and its linked background adaptive alleles facilitate (or impede) its fixation. Similarly, fixation of a mutant neutral allele could be facilitated by its closely linked adaptive allele or impeded by its linked deleterious alleles. K_a/K_s and π_a or π_a/π_s are also influenced by LD. Thus, a caution is needed in interpreting the observed evolutionary rate at nucleotide sites. The present theory confines to the single site–based analysis.

In a relevant theory of antagonistic selection, Peters and Weis (2018) also showed that selfing is critical in spreading and maintaining pollen-expressed genes related to competitiveness. Selfing impedes the fixation of the alleles conferring greater pollen competitiveness. Here, besides the pollen selection (competition), allele selection in ovules is considered because ovule selection could not be excluded in reality (Hu and Li 2003). One restrictive assumption is that only the same fitness is allowed for an allele between pollen and ovules. This could be violated for those genes that are differentially expressed in male and female gametophytes. For gametophyte specific genes, the present model only confines to those genes comparably expressed across male and female gametophytes.

Concerning the effects of gene flow on K_a/K_s and π_a/π_s, the unidirectional migration from mainland to island population reduces selection efficacy but aids in maintaining polymorphisms. The precondition of this function is the presence of a higher level of polymorphisms in mainland population than in recipient population. The impacts of pollen flow are counteracted by mating system through the discounting of alien pollen, different from seed flow. The same conclusion about effects of gene flow on K_a/K_s is also reported in a model of subdivided populations (Glemin 2007), except that gene flow is not separated into pollen and seed flow.

Our model only considers weak selection and the same order of gene flow and genetic drift effects. If selection is strong, the advantageous allele would be rapidly fixed in one phasic selection alone or in biphasic synergistic

selection. Strong selection facilitates fixation of alternative alleles in different species and yields a large K_a/K_s ratio (>1) in one species. However, even in the presence of strong selection, antagonistic selection contributes to compensating K_a/K_s and keeping it to approach 1. If migration is high, selection efficacy could be substantially reduced if the allele frequencies are different between migrants and the recipient population (Hu et al. 2003).

Implications

Several implications can be derived from the present theory. The first implication concerns the detection of natural selection through comparing K_a/K_s estimates with $K_a/K_s = 1$ (neutrality) for the genes expressed in two phases. Although antagonistic selection enhances maintenance of allelic polymorphism (Haldane 1932; Damgaard et al. 1994; Peters and Weis 2018), the opposite effects of biphasic selection weaken the power of detecting selection. This is analogous to the opposite effects of positive and purifying selection across sites along a sequence, which weakens the power of detecting selection at the gene level (Yang 2006). The phase variation of selection in plant life cycles is neglected in our current analysis in terms of K_a/K_s. In addition, selfing reduces the efficacy of gametophytic selection or reduces the cumulated effects of gametophytic and sporophytic selection in the nonadditive ($h \neq 1/2$) synergistic scenario. This also weakens the statistical test of natural selection in terms of K_a/K_s.

The second implication concerns the sampling strategy for studying molecular evolution. Both evolutionary rate and selection strength in terms of K_a/K_s could be different between gametophytic and sporophytic phases (Szovenyi et al. 2013; Immler 2019). Samples collected from the sporophyte phase (e.g., tree leaves) or diploid-expressed genes contain the compound effects from both gametophytic and sporophytic selection, which is difficult to separate. Samples collected from the gametophyte phase (e.g., pollen grains or ovules) or haploid-expressed genes can be used to assess the strength of gametophytic selection (Szovenyi et al. 2013). Separation of gametophytic from sporophytic selection remains a challenge when studying genes with pleiotropic effects in both phases.

The third implication is the theoretical application to interpreting empirical findings of distinct evolutionary rates for genes expressed in alternative generations. The significance of gametophytic selection is previously appreciated (Ottaviano and Mulcahy 1989; Charlesworth and Charlesworth 1992) but recently re-emphasized (Immler and Otto 2018; Beaudry et al. 2020). Gametophytic selection could potentially affect the evolution in predominantly haploid organisms (Immler 2019). For instance, pollen-specific genes had stronger purifying or positive selection than sporophytic-specific genes in highly outcrossing

species *Capsella grandiflora* (Arunkumar et al. 2013), but a reverse pattern was observed in *Macrocystis pyrifera* (Molano et al. 2022). The present theory predicts that the evolutionary rate of gametophyte-specific genes is shaped by the selfing rate and exhibits different patterns in selfers and ourcrossers. Empirical evidence already supports this prediction. For instance, Szovenyi et al. (2013) inferred that selection for haploid-specific genes was not more efficient in the haploid stage than for the diploid-specific genes in the diploid stage in the selfing species *Arabidopsis thaliana* (Page and Grossniklaus 2002) and the haploid selfing moss *Funaria hygrometrica* (Szovenyi et al. 2014). This is also evident in a separate study in *A. thaliana*, *Oryza sativa*, *Glycine max*, and in *Arabis alpina* (Gossmann et al. 2016; Gutierrez-Valencia et al. 2021). More empirical studies are needed to verify this prediction.

For genes co-expressed in haploid and diploid phases, empirical results show that they evolve more slowly than one phase-specific gene (Park and Choi 2010; Szovenyi et al. 2013). When a gene has substantially differential expression in two phases, it is likely related to different selection pressures between phases. Based on our theory, antagonistic selection is potentially involved because synergistic selection yields a faster overall evolutionary rate for the biphase-expressed genes than for one phase-specific genes. Other types of selection schemes cannot be excluded.

Finally, the theory helps to infer the mode of selection (antagonistic vs. synergistic) by combining the patterns of K_a/K_s and π_a or π_a/π_s and the type of mating system. For genes with pleiotropic effects in haploid and diploid phases, there are substantial differences in synergistic selection but small differences in antagonistic selection between selfing and outcrossing species in the K_a/K_s and π_a or π_a/π_s ratios. For genes expressed in the sporophytic phase, similar patterns of K_a/K_s and π_a or π_a/π_s are expected between selfing and outcrossing species. For genes expressed in the gametophytic phase only, substantial differences are expected between selfing and outcrossing species in K_a/K_s and π_a or π_a/π_s. These predictions need verifications in future studies.

The relative strength of gametic versus zygotic selection plays an important role in shaping K_a/K_s for genes with pleiotropic effects. Practical observations of high K_a/K_s ratios in species with mixed mating systems could imply strong positive selection in the sporophytic phase (Wang et al. 2021; Li et al. 2023). This likely occurs in most plants that have a much longer sporophyte phase than gametophyte phase, such as forest trees. Long-lived species could have more chances of undergoing stronger selection pressure in the sporophyte phase than that in the gametophyte phase, in contrast to short-lived species. Under this situation, sporophytic selection is likely stronger than gametophytic selection (Hu et al. 2019), and gametophytic

selection could be negligible. Otto and Marks (1996) showed that, in theory, selfing or inbreeding favors the gametophyte phase, while outcrossing favors the sporophyte phase. If K_a/K_s is low in selfing/inbreeding species, purifying gametophytic selection could be strong, and vice versa. If K_a/K_s is high in predominantly outcrossing species, gametophytic selection could also be strong, and vice versa. These predictions await appropriate data collection.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Data Availability

The Supplementary Material includes supplementary appendices S1 and S2, table S1, and figures S1–S5, Supplementary Material online. Two Mathematica notebooks for numerical calculations are provided in separate files.

Literature Cited

Arunkumar R, Josephs EB, Williamson RJ, Wright SI. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. Mol Biol Evol. 30:2475–2486.

Avery PJ. 1984. The population genetics of haplo-diploids and X-linked genes. Genet Res. 44:321–341.

Beaudry FEG, Rifkin JL, Barrett SCH, Wright SI. 2020. Evolutionary genomics of plant gametophytic selection. Plant Commun. 1(6): 100115.

Bessho K, Otto SP. 2017. Fixation probability in a haploid–diploid population. Genetics 205:421–440.

Borg M, Brownfield L, Twell D. 2009. Male gametophyte development: a molecular perspective. J Exp Bot. 60:1465–1478.

Bustamante CD. 2005. Population genetics of molecular evolution. In: Nielsen R, editor. Statistical methods in molecular evolution. Statistics for biology and health. New York, NY: Springer. p. 63–99.

Caballero A, Hill WG. 1992. Effects of partial inbreeding on fixation rates and variation of mutant genes. Genetics 131:493–507.

Charlesworth D, Charlesworth B. 1992. The effects of selection in gametophytic stage on mutational load. Evolution 46:703–720.

Charlesworth D, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Connallon T, Calsbeek CR. 2010. Fitness consequences of sex-specific selection. Evolution 64(6):1671–1682.

Damgaard C. 2000. Fixation of advantageous alleles in partially self-fertilizing populations: the effect of different selection models. Genetics 154:813–821.

Damgaard C, Guldbrandtsen B, Christiansen FB. 1994. Male gametophytic selection against a deleterious allele in a mixed mating model. Hereditas 120:13–18.

Evans SN, Shvests Y, Slatkin M. 2007. Non-equilibrium theory of the allele frequency spectrum. Theor Popul Biol. 71:109–119.

Ewing EP. 1977. Selection at the haploid and diploid phases: cyclical variation. Genetics 87:195–207.

Frank MH, Scanlon MJ. 2015. Transcriptomic evidence for the evolution of shoot meristem function in sporophyte-dominant land plants through concerted selection of ancestral gametophytic and sporophytic genetic programs. Mol Biol Evol. 32:355–367.

Glemin S. 2007. Mating system and the efficacy of selection at the molecular level. Genetics 177:905–916.

Glemin S, Francois CM, Galtier N. 2019. Genome evolution in outcrossing vs. selfing vs. asexual species. Methods Mol Biol. 1910:331–369.

Glemin S, Muyle A. 2014. Mating systems and selection efficacy: a test using chloroplastic sequence data in angiosperms. J Evol Biol. 27: 1386–1399.

Gossmann TI, Saleh D, Schmid MW, Spence MA, Schmid KJ. 2016. Transcriptomes of plant gametophytes have a higher proportion of rapidly evolving and young genes than sporophytes. Mol Biol Evol. 33(7):1669–1678.

Gutierrez-Valencia J, et al. 2021. Genomic signatures of sexual selection on pollen-expressed genes in *Arabis alpina*. Mol Biol Evol. 39:msab349.

Haldane JBS. 1924. A mathematical theory of natural and artificial selection. Part II. Proc Camb Philos Soc Biol Sci. 1:158–163.

Haldane JBS. 1932. The causes of evolution. New York: Harper.

Haldane JBS, Jayakar SD. 1963. Polymorphism due to selection of varying direction. J Genet. 58:237–242.

Hall DW, Goodisman MAD. 2012. The effects of kin selection on rates of molecular evolution in social insects. Evolution 66:2080–2093.

Hartl DL. 1976. Stochastic selection in both haplophase and diplophase. J Math Biol. 3:263–269.

Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol. 132(2):640–652.

Hu SX. 2015. Mating system as a barrier to gene flow. Evolution 69(5): 1158–1177.

Hu XS, et al. 2019. Mating system shifts a species range. Evolution 73(2):158–174.

Hu XS, Li BL. 2003. On migration load of seeds and pollen grains in a local population. Heredity (Edinb). 90:162–168.

Hu XS, Zeng W, Li BL. 2003. Impacts of one-way gene flow on genetic variance components in a natural population. Silvae Genet. 52(1): 18–24.

Immler S. 2019. Haploid selection in "diploid" organisms. Annu Rev Ecol Evol Syst. 50:219–236.

Immler S, Arnqvist G, Otto SP. 2012. Ploidally antagonistic selection maintains stable genetic polymorphism. Evolution 66(1):55–65.

Immler S, Otto SP. 2018. The evolutionary consequences of selection at the haploid gametic stage. Am Nat. 192:241–249.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro Hn, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–123.

Kimura M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713–719.

Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. PLoS Genet. 4(12):e1000304.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 35:1547–1549.

Lester LJ, Selander RK. 1979. Population genetics of haplodiploid insects. Genetics 92(4):1329–1345.

Li LL, et al. 2023. On the theories of plant mating system and molecular evolution and their applications. Sci Sin Vit. 51:50–63.

Liu Y, Joly V, Dorion S, Rivoal J, Matton DP. 2015. The plant ovule secretome: a different view toward pollen-pistil interactions. J Proteome Res. 14(11):4763–4775.

Lopez S, Rousset F, Shaw FH, Shaw RG, Ronce O. 2008. Migration load in plants: role of pollen and seed dispersal in heterogeneous landscapes. J Evol Biol. 21:294–309.

Lund-Hansen KK, Olito C, Morrow EH, Abbott JK. 2021. Sexually antagonistic coevolution between the sex chromosomes of *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 118(8): e2003359118.

Maynard-Smith J, Haig D. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23:22–35.

Molano G, Diesel J, Montecinos GJ, Alberto F, Nuzhdin SV. 2022. Sporophyte stage genes exhibit stronger selection than gametophyte stage genes in haplodiplontic giant kelp. Front Mar Sci. 8: 774076.

Mulcahy DL, Mulcahy GB. 1987. The effects of pollen competition. Am Sci. 75:44–50.

Ottaviano E, Mulcahy DL. 1989. Genetics of angiosperm pollen. Adv Genet. 26:1–64.

Otto SP, Marks JC. 1996. Mating systems and the evolutionary transition between haploidy and diploidy. Biol J Linn Soc. 57:197–218.

Otto SP, Scott MF, Immler S. 2015. Evolution of haploid selection in predominantly diploid organisms. Proc Natl Acad Sci U S A. 112: 15952–15957.

Page DR, Grossniklaus U. 2002. The art and design of genetic screens: *Arabidopsis thaliana*. Nat Rev Genet. 3(2):124.

Park SG, Choi SS. 2010. Expression breadth and expression abundance behave differently in correlations with evolutionary rates. BMC Evol Biol. 10:241.

Peters MAE, Weis AE. 2018. Selection for pollen competitive ability in mixed-mating systems. Evolution 72(11):2513–2536.

Piganeau G, Eyre-Walker A. 2003. Estimating the distribution of fitness effects from DNA sequence data: implications for molecular clock. Proc Natl Acad Sci U S A. 100(18):10335–10340.

Rescan M, Lenormand T, Roze D. 2016. Interactions between genetic and ecological effects on the evolution of life cycles. Am Nat. 187: 19–34.

Roze D, Rousset F. 2004. Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. Genetics 167:1001–1015.

Rutley N, Twell D. 2015. A decade of pollen transcriptomics. Plant Reprod. 28(2):73–89.

Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. Genetics 132:1161–1176.

Scott MF, Rescan M. 2016. Evolution of haploid–diploid life cycles when haploid and diploid fitnesses are not equal. Evolution 71: 215–226.

Scudo FM. 1967. Selection on both haplo and diplophase. Genetics 56:693–704.

Szovenyi P, et al. 2013. Selection is no more efficient in haploid than in diploid life stages of an angiosperm and a moss. Mol Biol Evol. 30(8):1929–1939.

Szovenyi P, et al. 2014. Efficient purging of deleterious mutations in plants with haploid selfing. Genome Biol Evol. 6(5): 1238–1252.

Tanksley SD, Zamir D, Rick CM. 1981. Evidence for extensive overlap of sporophytic and gametophtic gene expression in Lycopersicon esculentum. Science 213:453–455.

Walsh N, Charlesworth D. 1992. Evolutionary interpretations of differences in pollen tube growth rates. Q Rev Biol. 67:19–37.

Wang XJ, et al. 2021. The genomic selfing syndrome accompanies the evolutionary breakdown of heterostyly. Mol Biol Evol. 38: 168–180.

Welch JJ, Eyre-Walker A, Waxman D. 2008. Divergence and polymorphism under the nearly neutral theory of molecular evolution. J Mol Evol. 67:418–426.

Winsor J, Davis L, Stephenson A. 1987. The relationship between pollen load and fruit maturation and the effect of pollen load on offspring vigor in *Cucurbita pepo*. Am Nat. 129:643–656.

Wolfram S. 1996. The Mathematica book. Cambridge: Cambridge University Press.

Wright S. 1938. The distribution of gene frequencies under irreversible mutation. Proc Natl Acad Sci U S A. 24:353–359.

Wright S. 1969. Evolution and the genetics of populations. The theory of gene frequencies, vol. 2. Chicago: University Chicago Press.

Wright S. 1977. Evolution and the genetics of populations. Vol. 3, experimental results and evolutionary deductions. Chicago: The University of Chicago Press.

Yang Z. 2006. Computational molecular evolution. Oxford, UK: Oxford University Press.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Zhang JX, et al. 2021. Differential gene expression patterns during gametophyte development provide insights into sex differentiation in the dioicous kelp *Saccharina japonica*. BMC Plant Biol. 21(1): 335.

Associate editor: David Enard