

# Population Genetics of Reproductive Genes in Haplodiploid Species

Amy L. Dapper <sup>1,\*</sup>, Garrett P. Slater<sup>2</sup>, Katherynne Shores<sup>1</sup>, and Brock A. Harpur <sup>2</sup>

<sup>1</sup>Department of Biological Sciences, Mississippi State University, MS 39762, USA

<sup>2</sup>Department of Entomology, Purdue University, West Lafayette, IN 47907, USA

\*Corresponding author: E-mail: dapper@biology.msstate.edu.

Accepted: 03 May 2022

## Abstract

Many animal species are haplodiploid: their fertilized eggs develop into diploid females and their unfertilized eggs develop into haploid males. The unique genetic features of haplodiploidy raise the prospect that these systems can be used to disentangle the population genetic consequences of haploid and diploid selection. To this end, sex-specific reproductive genes are of particular interest because, while they are shared within the same genome, they consistently experience selection in different ploidal environments. However, other features of these genes, including sex-specific expression and putative involvement in postcopulatory sexual selection, are potentially confounding factors because they may also impact the efficacy of selection asymmetrically between the sexes. Thus, to properly interpret evolutionary genomic patterns, it is necessary to generate a null expectation for the relative amount of polymorphism and divergence we expect to observe among sex-specific genes in haplodiploid species, given differences in ploidal environment, sex-limited expression, and their potential role in sexual selection. Here, we derive the theoretical expectation for the rate of evolution of sex-specific genes in haplodiploid species, under the assumption that they experience the same selective environment as genes expressed in both sexes. We find that the null expectation is that reproductive genes evolve more rapidly than constitutively expressed genes in haplodiploid genomes. However, despite the aforementioned differences, the null expectation does not differ between male- and female-specific reproductive genes, when assuming additivity. Our theoretical results provide an important baseline expectation that should be used in molecular evolution studies comparing rates of evolution among classes of genes in haplodiploid species.

**Key words:** polymorphism, divergence, haplodiploid, sexual selection, sex-specific expression.

## Significance

The unique genetic features of haplodiploidy (haploid males and diploid females) raise the prospect that these systems can be used to disentangle the population genetic consequences of haploid and diploid selection. However, confounding factors, such as sex-specific expression, must be considered in molecular evolution studies comparing rates of evolution among classes of genes in haplodiploid species. To this end, we derive the null expectation for the rate of evolution of sex-specific genes in haplodiploid species, under the assumption that they experience the same selective environment as genes expressed in both sexes. Our results provide essential theoretical context for future and on-going genomic studies of haplodiploid species.

## Introduction

The efficacy of selection, and as a consequence, the evolutionary rate is expected to differ dramatically between haploid and diploid states (Crow and Kimura 1965; Hartl 1972;

Kondrashov and Crow 1991). This occurs primarily because dominance interactions can fully or partially mask the fitness effects of an allele in diploid genomes, while alleles expressed in a haploid state are directly exposed to selection.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

As a result, the selection is more effective at fixing beneficial alleles and removing deleterious ones in haploids (Crow and Kimura 1965; Hartl 1972; Kondrashov and Crow 1991). Haploid genomes are also predicted to have a lower mutation rate, simply by carrying one less copy of the genome (Otto and Gerstein 2008), but see Sharp et al. (2018), further impacting their expected rate of evolution. The prevalence of diploidy, in contrast with the advantages, is attributed to several purported advantages over haploidy: diploidy can increase genetic diversity through recombination and random segregation of alleles, it can mask the effects of recessive deleterious alleles, and having a second copy of a locus allows for the repair of de novo mutations with a template (Crow and Kimura 1965; Kondrashov and Crow 1991; Roze 2009).

There has been renewed interest in identifying and quantifying the consequences of alternation between haploidy and diploidy on patterns of genetic diversity, genomic architecture, phenotypic diversity, and ultimately evolution (Joseph and Kirkpatrick 2004; Lenormand and Dutheil 2005; Szövényi et al. 2013; Otto et al. 2015; Immler 2019; Raices et al. 2019). To this end, there are several advantages to using haplodiploid, or arrhenotokous species, over predominantly diploid, or diplontic, species. In diplontic species, technical limitations make it difficult to obtain empirical evidence for the impact of haploid selection (Joseph and Kirkpatrick 2004; Immler 2019), which stem from the challenge of identifying genes that exclusively experience haploid selection. For example, mRNA products produced by haploid sperm genomes may be shared via cytoplasmic bridges, making them effectively diploid (Erickson 1973; Braun et al. 1989). Additionally, the fraction of genes that experience haploid selection—via gametic expression, genomic imprinting, X-inactivation, or sex-linkage—may be small. In mice, best estimates indicate that somewhere between 0.1% and 3.3% of genes are exposed to haploid selection (Joseph and Kirkpatrick 2004).

In all haplodiploid species, haploid male gametes do not undergo recombination and any allele expressed by a male is directly exposed to natural selection. Given that most male hymenoptera are free-living organisms, it is likely that a significant fraction of a genome is haploid-expressed (as we demonstrate in our companion paper; Slater et al. 2022). Furthermore, because male-specific expression correlates directly with haploid expression, it is technically easier to identify genes that exclusively experience haploid selection. Take the honey bee as an example. Haploid males take 24 days to develop from eggs, have an adult lifespan between 20 and 40 days, and their sperm survives inside a queen's spermatheca for her entire lifespan of approximately 3 years (Page and Peng 2001). This means that many genes in honey bees likely experience some degree of selection in the haploid state and this is likely true of most hymenoptera.

However, there are two confounding factors that must be taken into account when studying the consequences

of haploid selection in haplodiploid species. First, autosomal genes that exclusively experience haploid selection are also male-specific, meaning that they are not expressed and, thus are shielded from selection in a large fraction of the population. Thus, the null expectation is that beneficial alleles are less likely to reach fixation and deleterious alleles will take longer to be purged from the population in this class of protein-coding genes (Dapper and Wade 2016, 2020). Additionally, male hymenoptera serve primarily a sexual role and the putative function of many of these genes is in postcopulatory sexual selection. For genes that predominantly function in sperm competition or cryptic female choice, the effective strength of selection is determined by the harmonic mean number of mates per female (Dapper and Wade 2016, 2020). This is the case because such genes do not experience selection in females that mate with only a single male. Hymenoptera species vary dramatically in their mating systems, and thus the opportunity for postcopulatory sexual selection, across the phylogeny (Boomsma et al. 2005).

In order to quantify the impact of haploid selection in arrhenotokous species, it is necessary to take into account the sex-limited expression and potential role in sexual selection. Here, we derive the null expectation for levels of polymorphism and divergence among sex-specific genes in arrhenotokous species in comparison with their constitutively expressed counterparts. We find that, despite differences in ploidy, levels of polymorphism and divergence are expected to be elevated equally in male- and female-specific genes. This occurs because the efficacy of haploid selection in males is exactly offset by sex-limited expression. We also provide an explicit framework for predicting how differences in mating system among arrhenotokous species will impact the rate of evolution of genes involved predominantly in sperm competition and cryptic female choice. Our theoretical results provide an important baseline for the interpretation of molecular evolution studies aimed at quantifying the role of haploid selection in arrhenotokous species (see Slater et al. 2022).

### Sex-Specific Expression Counteracts Haploid Selection in Male-Limited Genes

First consider a constitutively expressed locus in an arrhenotokous species with two alleles,  $A$  and  $a$ . In haploid males, there are two possible genotypes at this locus,  $A$  or  $a$ , with the respective relative fitnesses ( $w$ ) of 1 and  $1 - s$  when the selection coefficient,  $s$ , describes the selective disadvantage of the  $a$  allele relative to the  $A$  allele. In diploid females, there are three possible genotypes at this locus,  $AA$ ,  $Aa$ , or  $aa$ , with respective relative fitnesses of 1,  $1 - s/2$ , and  $1 - s$ , assuming dosage compensation with respect to fitness. Due to the difference in genetic environments (haploid vs. diploid) between the sexes, the change in allele

frequency in males is expected to lag slightly behind the allele frequency of females, generating an asymmetry. However, for small values of  $s$ , the allele frequencies of the sexes have been shown to converge quickly (Haldane 1926; Nagylaki 1979), allowing us to reasonably assume that the frequency of the  $A$  allele ( $p$ ) at time  $t$  is approximately equal in both sexes [ $p_m(t) \cong p_f(t) = p(t)$ ] (Avery 1984). Thus, when  $s$  is assumed to be small and the  $a$  allele is rare, the change in allele frequency in a constitutively expressed locus can be closely approximated as

$$\Delta p(t) \simeq \frac{1}{3} \Delta p_m(t) + \frac{2}{3} \Delta p_f(t) \simeq \frac{sp(t)(1-p(t))[(1+2h) + 2p(t)(1-2h)]}{3\bar{W}} \quad (1)$$

where  $\bar{W}$  is the average fitness of the population with respect to this locus and  $h$  is the dominance coefficient with respect to fitness in females (see also Avery 1984). When considering the additive case ( $h = 0.5$ ), this equation simplifies to:

$$\Delta p(t) \simeq \frac{2sp(t)q(t)}{3\bar{W}} \quad (2)$$

where  $q(t)$  is the frequency of the  $a$  allele at time  $t$  [ $q(t) = 1 - p(t)$ ] (see also Avery 1984). Assuming that the average fitness of the population is very near 1 (appropriate when the  $a$  allele is very rare and  $s$  is small), the approximation further simplifies to  $\Delta p = 2/3spq$ . Thus, a beneficial allele is expected to rise in frequency  $1/3$  faster than a constitutively expressed allele in a diploid species ( $\Delta p = spq/2$ , Crow and Kimura 1970), but  $1/3$  slower than a constitutively expressed allele in a haploid species ( $\Delta p = spq$ , Crow and Kimura 1970). Likewise, the decline in the frequency of a deleterious allele will be  $1/3$  faster than a constitutively expressed allele in a diploid species, but  $1/3$  slower than a constitutively expressed allele in a haploid species. Note that the relative strength of selection on constitutively expressed alleles in arrhenotokous species ( $s_e = 2/3$ ) represents the harmonic mean of the relative strength of selection in the two genetic environments (haploid:  $s_e = 1$ ; diploid:  $s_e = 1/2$ ).

Now let's consider an autosomal male-specific locus with the same two alleles,  $A$  and  $a$ . In this case, the selective disadvantage of the  $a$  allele is incurred in males ( $w_A = 1$ ,  $w_a = 1 - s$ ) and the allele is neutral with respect to fitness in females ( $w_{AA} = w_{Aa} = w_{aa} = 1$ ,  $\Delta p_f(t) = 0$ ). With the same assumptions outlined above, the change in frequency of the  $A$  allele can be closely approximated as

$$\Delta p_{\text{male-specific}}(t) \simeq \frac{1}{3} \Delta p_m(t) \simeq \frac{sp(t)[1-p(t)]}{3\bar{W}} \quad (3)$$

Alternatively, we can consider an autosomal female-specific locus with the same two alleles,  $A$  and  $a$ . In this case, the selective disadvantage of the  $a$  allele is incurred

in females ( $w_{AA} = 1$ ,  $w_{Aa} = 1 - hs$ , and  $w_{aa} = 1 - s$ ) and the allele is neutral with respect to fitness in males ( $w_A = w_a = 1$ ,  $\Delta p_m(t) = 0$ ). Thus, the change in frequency of the  $A$  can be closely approximated as

$$\Delta p_{\text{female-specific}}(t) \simeq \frac{2}{3} \Delta p_f(t) \simeq \frac{2sp(t)[1-p(t)][(1-h) + p(2h-1)]}{3\bar{W}} \quad (4)$$

Assuming additivity ( $h = 0.5$ ), this equation simplifies to:

$$\Delta p_{\text{female-specific}}(t) \simeq \frac{sp(t)[1-p(t)]}{3\bar{W}} \quad (5)$$

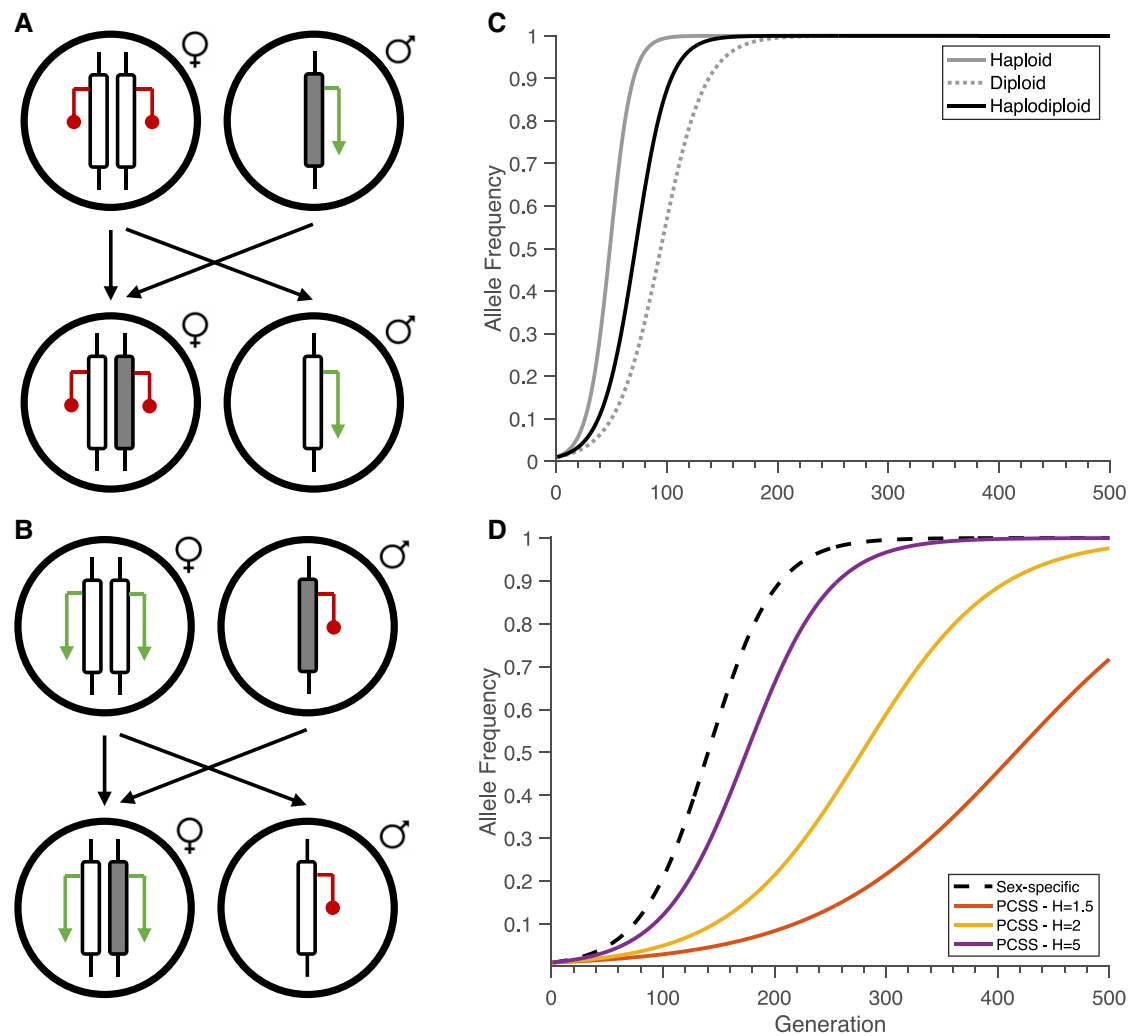
Thus, despite the difference in genetic environments between the sexes, the expectation for the change in frequency of a beneficial or deleterious allele at a male- or female-specific locus is identical and half that expected for a constitutively expressed allele in an arrhenotokous species (see eq. 2, fig. 1).

### Null Expectation for Relative Levels of Polymorphism

Under standard population genetics assumptions, including large population size and independence among sites, we can use the expected equilibrium frequency at mutation-selection balance ( $q^*$ ) to estimate the expected genic diversity within populations ( $\pi_N/\pi_S$ ) (Van Dyken and Wade 2010; Dapper and Wade 2016). In other words, under these conditions, we expect levels of polymorphism and heterozygosity to be approximately equal. Thus, we can compare the expected level of gene diversity in sex-specific proteins by modifying the relative strength of selection appropriately and comparing the expected level of gene diversity ( $q^*$ ) in a population at the mutation-selection balance. If we assume that the mutation rate from  $A$  to  $a$  is equal to  $\mu$ , and both  $\bar{W}$  and  $p$  are near one, the expected level of gene diversity for additive loci in arrhenotokous species is approximately equal to:

$$q^* \simeq \frac{3\mu}{2s} \quad (6)$$

for constitutively expressed alleles. This result is identical to the that of X-linked genes, which are expected to be about a third as polymorphic as their diploid, autosomal counterparts ( $q_{\text{diploid}}^* \simeq 2\mu/s$ , Crow and Kimura 1970) at the mutation-selection equilibrium as a result of haploid selection in males. Likewise, for sex-



**Fig. 1.**—(A) In arrhenotokous species, haploid-specific alleles are only expressed (green arrow) in males. However, as all fertilized embryos develop as diploid females, this paternal allele will not be expressed (red circle), and thus not experience selection, in the next generation. (B) On the other hand, half of the diploid-specific alleles inherited by daughters will not have experienced selection in the parental generation. (C) In arrhenotokous species, autosomal beneficial alleles ( $s = 0.1$ ,  $h = 0.5$ ) are expected to rise to fixation slower (black, solid) than expected in haploid species (gray, solid), but faster than expected for diploid species (gray, dashed). (D) Sex-specific loci (black, dashed) are expected to take longer to near fixation than comparable alleles at loci that are expressed in both sexes. This occurs because the allele is not expressed, and therefore, does not experience selection in one sex. If the alleles primary function is in post-copulatory sexual selection, the expected time to near fixation decreases as the harmonic mean number of mates per female ( $H$ ) increases (solid, colored). The converse is true for the loss of deleterious alleles.

specific loci, the expected level of gene diversity is approximately equal to:

$$q_{\text{sex-specific}}^* \approx \frac{3\mu}{s}. \quad (7)$$

Assuming the mutation rate and selective effect of new mutations are the same for both classes of genes, the expected relative level of gene diversity of sex-specific alleles is twice that expected for constitutively expressed control genes in arrhenotokous species ( $q_{\text{sex-specific}}^*/q^* = 2$ ) (See [Supplementary Files](#)).

While haploid genomes are often assumed to have a lower mutation rate than diploid genomes (Crow and Kimura 1965; Kondrashov and Crow 1991), this complication is not relevant in this case because we are comparing groups of autosomal genes that spend the same amount of time in the genomes of diploid and haploid individuals.

### Null Expectation for Relative Levels of Divergence

Over time, segregating alleles within finite populations invariably either become fixed in the population ( $p = 1$ ) or

completely lost ( $p = 0$ ), excepting cases of strong balancing selection. The probability that any particular alleles becomes fixed in a population, rather than lost, is a function of its starting frequency in the population,  $q_0$ , its selective effect,  $s$ , and the size of the population,  $N$ . The probability of fixation of a nonneutral mutation ( $s \neq 0$ ), is derived from the expected mean,  $E[\Delta p(t)]$ , and variance of the rate of change in allele frequency,  $\text{Var}[\Delta p(t)]$  (see Kimura 1964). The probability of fixation for an autosomal constitutively expressed allele in a haplodiploid population is

$$u(q) \simeq \frac{1 - e^{(-8/3)N_e s p_0}}{1 - e^{(-8/3)N_e s}} \quad (8)$$

(derived in Avery 1984). Assuming that  $p_0 = 2/3N$  (the starting allele frequency of a new mutation in a haplodiploid population), and the effective population size,  $N_e$ , is roughly equivalent to the total population size,  $N$ , this equation simplifies to:

$$u(q) \simeq \frac{(16/9)s}{1 - e^{(-8/3)N_e s}} \quad (9)$$

when  $|s|$  is assumed to be small (Avery 1984). As we show above, the expected change in allele frequency of a sex-specific allele is half that of an otherwise equivalent constitutively expressed allele, such that:

$$E[\Delta p(t)]_{\text{sex-specific}} \simeq \frac{sp(t)q(t)}{3} \quad (10)$$

(from eq. 5). However, because we are considering autosomal loci, the alleles are present in the genomes of both sexes and, as such, the expected variance is not different from an otherwise equivalent constitutively expressed allele in the same genome. Thus, we find that the probability of fixation of a sex-specific allele in an arrhenotokous population is

$$u(q)_{\text{sex-specific}} \simeq \frac{1 - e^{(-4/3)N_e s p_0}}{1 - e^{(-4/3)N_e s}} \quad (11)$$

which similarly simplifies to:

$$u(q)_{\text{sex-specific}} \simeq \frac{(8/9)s}{1 - e^{(-4/3)N_e s}} \quad (12)$$

In the case of neutral mutations ( $s = 0$ ), the probability of fixation,  $u(q)$ , is simply equal to its starting frequency in the population. For new mutations in an arrhenotokous population, the starting allele frequency is  $2/3N$ , such that:

$$u(q_{s=0}) = \frac{2}{3N} \quad (13)$$

Holding population size constant, we can see that the probability of fixation of a new neutral mutation in an arrhenotokous population is expected to be higher by a factor of  $1/3$  than a diploid population ( $1/2N$ ) and lower by a factor of  $1/3$  than a haploid population ( $1/N$ ). Importantly, the probability of fixation of a neutral mutation is the same for sex-specific and constitutively expressed genes, provided they are located on autosomes.

If we assume that mutations that do not change the resulting amino acid sequences (synonymous mutations) do not impact fitness, and are thus neutral, we can use the probability of fixation of a neutral mutation to estimate the expected rate of synonymous substitutions ( $dS$ ). Using a similar approach, we can use the probability of fixation of a nonneutral mutation to estimate the expected rate of nonsynonymous substitutions ( $dN$ )—those that change the amino acid sequence of a protein-coding gene. As follows, the expected ratio of nonsynonymous to synonymous substitutions ( $dN/dS$ ) for constitutively expressed loci in arrhenotokous species is

$$E\left[\frac{dN}{dS}\right] \simeq \frac{3N}{2} \frac{1 - e^{(-16/9)\bar{s}}}{1 - e^{(-8/3)N_e \bar{s}}} \simeq \frac{(8/3)N_e \bar{s}}{1 - e^{(-8/3)N_e \bar{s}}} \quad (14)$$

when  $\bar{s}$ , the average selective effect of a new mutation, is assumed to be small (i.e., most new mutations are mildly deleterious). Likewise, for sex-specific genes, the effective strength of selection is half that of a constitutively expressed gene, giving:

$$E\left[\frac{dN}{dS}\right]_{\text{sex-specific}} \simeq \frac{(4/3)N_e \bar{s}}{1 - e^{(-4/3)N_e \bar{s}}} \quad (15)$$

Thus, we can clearly see that the expected rate of evolution of a protein-coding gene ( $dN/dS$ ) is a function of the product of population size ( $N$ ) and the average selective effect of nonsynonymous mutations ( $\bar{s}$ ). A comparison of equations (14) and (15) also shows us that, when the average selective effect of nonsynonymous substitutions ( $\bar{s}$ ) is weakly negative and the same for constitutively expressed and sex-specific genes, the expected rate of evolution of sex-specific genes is considerably higher (See [Supplementary Files](#)). Conversely, under strong, pervasive directional selection ( $\bar{s} > 0$ ), the rate of evolution will be slower for sex-specific genes.

### Egg-Specific Genes Experience More Effective Selection

The results described above do not extend to the rare case that a female-specific gene is expressed by the gamete, rather than adult, genome. In other words, genes that are

expressed only by the haploid egg genome. In this case, the locus will experience haploid selection in females, giving:

$$\Delta p_{\text{female-specific}}(t) \simeq \frac{2}{3} \Delta p_f(t) \simeq \frac{2sp(t)[1 - p(t)]}{3W}. \quad (16)$$

A comparison with equation (2) shows us that these genes are expected to evolve at the same rate as constitutively expressed genes, despite their limited expression profile, because haploid selection increases the efficacy of selection. A similar effect is not observed in sperm-specific genes because the ploidal environment does not differ from genes expressed in adult males (both are haploid).

### Dominance Generates Sex Differences in Expected Evolutionary Rates

Dominance interactions among alleles at a given locus are only possible in diploid females. As follows, when we relax the assumption of additivity, the null expectation for the evolutionary rate of female-specific genes is altered, while the corresponding expectations for male-specific genes remains the same ( $q^* \simeq 3\mu/s$ ). In the case of complete dominance, selection acts more efficiently on heterozygous females, reducing the expected level of polymorphism by half:

$$q_{\text{female-specific}}^* \simeq \frac{3\mu}{2s}. \quad (17)$$

In contrast, female-specific recessive alleles only experience selection in homozygous genotypes ( $q^2$ ), which occur more rarely than in the diploid case because only 2/3 copies of the alleles are in diploid females in any given generation:

$$q_{\text{female-specific}}^* \simeq \sqrt{\frac{3\mu}{2s}} \quad (18)$$

elevating the expected level of polymorphism among female-specific genes relative to male-specific or constitutively expressed loci.

### Postcopulatory Sexual Selection

If the primary function of these genes is in postcopulatory sexual selection, the expected rate of evolution is expected to be further slowed as a function of the genic variance within a female reproductive tract (Dapper and Wade 2016, 2020). Such that:

$$\Delta p_{\text{sex-specific}}(t) \simeq \frac{sp(t)[1 - p(t)]H - 1}{3W H}. \quad (19)$$

where  $H$  is the harmonic mean number of mates per female (figure 1). Likewise, the expected level of gene diversity in

sex-specific genes will be further elevated by a factor of  $H - 1/H$  (Dapper and Wade 2016), such that:

$$q_{\text{sex-specific}}^* \simeq \frac{3\mu}{s} \left( \frac{H}{H - 1} \right) \quad (20)$$

and the expected rate of evolution of sex-specific genes will be further elevated by a factor of  $H - 1/H$  (Dapper and Wade 2016), such that:

$$E_{\text{sex-specific}} \left[ \frac{dN}{dS} \right] \simeq \frac{(4/3)N_e \bar{s} (H - 1)/H}{1 - e^{-(4/3)N_e \bar{s} (H - 1)/H}}. \quad (21)$$

It is important to note that the sex ratio of the population of reproductive individuals in arrhenotokous populations can deviate dramatically from 50:50 and vary widely between species. In hymenoptera, especially, sex ratios vary considerably among solitary and eusocial species (Trivers and Hare 1976; Boomsma 1991; Boomsma and Grafen 1991). Such deviations are reflected in the expected rate of evolution ( $dN/dS$ ) in two ways: (1) deviations from 50:50 sex ratio lowers the effective population size,  $N_e$  and (2) when considering genes that primarily function in postcopulatory sexual selection, mating system variation impacts the harmonic mean number of mates per female,  $H$ . When considering the relative rate of evolution of constitutive and sex-specific loci within the same genome,  $N_e s$  is common to the comparison, but  $H$  is not.

### Discussion

In arrhenotokous species, one sex develops from fertilized eggs and is diploid (usually females) and the other sex develops from unfertilized eggs and is haploid (usually males). This asymmetry raises the prospect that such haplodiploid systems can be used to empirically quantify and compare the efficacy of selection in different ploidal environments within the same population. One approach is to compare the rate of evolution of male- and female-specific genes, which are consistently exposed to either haploid and diploid selection, respectively. However, such comparisons must also take into account two other factors that are expected to influence the efficacy of selection: (1) sex-specific expression and (2) putative involvement in postcopulatory selection. Here, we derive the null expectation for the rate of evolution of reproductive genes in haplodiploid species, accounting for differences in ploidy, sex-specific expression, and postcopulatory sexual selection. Our null expectation is generated by assuming that, on average, alleles of reproductive genes have the same fitness consequences ( $s$ , selection coefficient) when expressed as alleles of nonreproductive constitutively expressed protein-coding genes. Thus, it provides the correct benchmark with which to compare and interpret the rates of molecular evolution

of reproductive genes. Importantly, these theoretical results make explicit predictions about the expected level of polymorphism and divergence among sex-specific genes that can be directly applied to population genomics studies of haplodiploid species (see Slater et al. 2022, [Supplementary Material](#) online).

We find that the levels of polymorphism and divergence of reproductive genes in haplodiploid species are expected to exceed that of constitutively expressed genes due to relaxed selection, which is consistent with our theoretical predictions for reproductive genes in standard haploid and diploid systems (Dapper and Wade 2016, 2020). This occurs because we are considering sex-specific autosomal genes, which are carried in the genomes of both sexes, but are only expressed in one. Furthermore, the putative function of many of these genes is in postcopulatory sexual selection (i.e., sperm competition, cryptic female choice). The efficacy of postcopulatory sexual selection is determined by the harmonic mean number of mate per female because such genes do not experience selection in females that mate only once or to males that are genetically identical at the loci in question (Dapper and Wade 2016, 2020). This is especially relevant in arrhenotokous species, like hymenopterans, because the males serve primarily sexual roles and there is extreme variation in mating system among species (Boomsma et al. 2005).

The consistent differences in the ploidal environment suggest the possibility that degree to which we expect polymorphism and divergence to be elevated may differ between male- and female-specific genes in haplodiploid species. However, we show that this is not the case. While haploid selection is expected to increase the efficacy of selection at male-specific loci compared with diploid female-specific loci, this is offset by the reduction in the efficacy of selection due to sex-specific expression, which is greater for males in arrhenotokous species. This asymmetry is illustrated by considering the parental environment of a sex-specific autosomal allele. For males, such an allele will always be maternally inherited and thus, will not have experienced selection in the previous generation. For females, such an allele will have a 50% chance of being paternally inherited and thus, a comparatively higher chance of being exposed to selection in the previous generation. While this result cannot be directly extended to species with biphasic life-cycles because each individual experiences a haploid and a diploid life-stage, it is likely that related considerations such as expression breadth and noise may also result in relaxed selection on haploid-specific loci (Szövényi et al. 2013). Such effects may obscure the expected increase in the efficacy of selection due to ploidal environment, as observed in *Arabidopsis* and moss in Szövényi et al. (2013).

In contrast with diploid systems, there are two restrictive cases expected to generate asymmetries in the efficacy of selection on male- and female-specific genes in haplodiploid species: (1) when sex-specific loci are expressed by

the gamete, rather than the adult genomes and (2) when dominance interactions frequently occur between alleles at sex-specific loci. Both conditions alter the efficacy of selection on female-specific genes without affecting the efficacy of selection on male-specific genes. In the first case, egg-specific loci experience an increase in the efficacy of selection due to haploid expression patterns. In the second case, the direction and magnitude nature of the dominance interaction determines the impact on the efficacy of selection. At the extremes, if new mutations tend to be completely dominant, the efficacy of selection on female-specific loci is twice that of their male counterparts—similar to egg-specific loci. In contrast, if new mutations tend to be completely recessive with respect to fitness, they remain hidden from selection in diploid females when rare, decreasing the efficacy of selection on female-specific loci. However, it remains unclear how often either of the conditions are met in haplodiploid species.

Haplodiploid genomes share the same population genetic framework as the X chromosome in diploid species with XY sex-determination systems (Crow and Kimura 1965; Avery 1984). Thus, the null expectations that we derive here also apply to sex-specific loci found on the X chromosome. However, unlike the arrhenotokous case, there are two potential sets of control genes—constitutively expressed loci located on the X chromosome and constitutively expressed loci located on the autosomes. Importantly, the null expectation for levels of polymorphism and divergence observed among sex-specific X-linked genes is greater relative to other X-linked constitutively expressed loci (2-fold elevation) than to constitutively expressed loci on the autosomes (1.33-fold elevation).

Incorporating the correct null hypothesis is necessary to rigorously interpret the results of molecular evolution studies that compare rates of polymorphism and divergence between reproductive and constitutively expressed loci. In order to infer evidence of pervasive and rapid positive selection of reproductive genes, one must observe levels of divergence that exceed this null expectation. Conversely, observations of divergence values among reproductive genes lower than the null expectation may be evidence of stronger negative selection even if the rates of evolution still exceed those observed among constitutively expressed loci. Furthermore, we show that the expectation for the rate of evolution of haploid- and diploid-specific loci is not different when taking into account both the effect of ploidal environment and sex-limited expression. Thus, significant differences in the rate of evolution of these genes can be attributed to differences in the strength of selection acting on these classes of genes.

### Supplementary Material

[Supplementary data](#) are available at *Genome Biology and Evolution* online.

## Acknowledgments

We are grateful to the associate editor and two anonymous reviewers for helpful comments on earlier versions of this manuscript. This material is based on work supported by the National Science Foundation under Grant No. DEB 2143063 (to A.L.D.).

## Data Availability

This manuscript presents theoretical results and does not include new data, including sequences or transcripts. We have generated an interactive Matlab notebook that allows users to easily find the null expectation for the rate of evolution of reproductive genes in arrhenotokous species given a control dataset. The notebook is intended to be made available alongside the published manuscript.

## Literature Cited

- Avery P. 1984. The population genetics of haplo-diploids and x-linked genes. *Genet Res.* 44(3):321–341.
- Boomsma JJ. 1991. Adaptive colony sex ratios in primitively eusocial bees. *Trends Ecol Evol.* 6(3):92–95.
- Boomsma JJ, Baer B, Heinze J. 2005. The evolution of male traits in social insects. *Annu Rev Entomol.* 50:395–420.
- Boomsma JJ, Grafen A. 1991. Colony-level sex ratio selection in the eusocial hymenoptera. *J Evol Biol.* 4(3):383–407.
- Braun RE, Behringer RR, Peschon JJ, Brinster RL, Palmiter RD. 1989. Genetically haploid spermatids are phenotypically diploid. *Nature* 337(6205):373–376.
- Crow JF, Kimura M. 1965. Evolution in sexual and asexual populations. *Am Nat.* 99(909):439–450.
- Crow JF, Kimura M. 1970. An introduction to population genetics theory. New York: Harper and Row.
- Dapper AL, Wade MJ. 2016. The evolution of sperm competition genes: the effect of mating system on levels of genetic variation within and between species. *Evolution* 70(2):502–511.
- Dapper AL, Wade MJ. 2020. Relaxed selection and the rapid evolution of reproductive genes. *Trends Genet.* 36(9):640–649.
- Erickson R. 1973. Haploid gene expression versus meiotic drive: the relevance of intercellular bridges during spermatogenesis. *Nat New Biol.* 243(128):210–212.
- Haldane J. 1926. A mathematical theory of natural and artificial selection. *Math Proc Cambridge Philos Soc.* 23(4):363–372.
- Hartl DL. 1972. A fundamental theorem of natural selection for sex linkage or arrhenotoky. *Am Nat.* 106(950):516–524.
- Immler S. 2019. Haploid selection in “diploid” organisms. *Annu Rev Ecol Evol Syst.* 50(1):219–236.
- Joseph SB, Kirkpatrick M. 2004. Haploid selection in animals. *Trends Ecol Evol.* 19(11):592–597.
- Kimura M. 1964. Diffusion models in population genetics. *J Appl Probab.* 1(2):177–232.
- Kondrashov AS, Crow JF. 1991. Haploidy or diploidy: which is better? *Nature* 351(6324):314–315.
- Lenormand T, Dutheil J. 2005. Recombination difference between sexes: a role for haploid selection. *PLoS Biol.* 3(3):e63.
- Nagylaki T. 1979. Selection in dioecious populations. *Ann Hum Genet.* 43(2):143–150.
- Otto SP, Gerstein AC. 2008. The evolution of haploidy and diploidy. *Curr Biol.* 18(24):R1121–R1124.
- Otto SP, Scott MF, Immler S. 2015. Evolution of haploid selection in predominantly diploid organisms. *Proc Natl Acad Sci U S A.* 112(52):15952–15957.
- Page RE Jr, Peng CY. 2001. Aging and development in social insects with emphasis on the honey bee, *Apis mellifera* L. *Exp Gerontol.* 36(4–6):695–711.
- Raices JB, Otto PA, Vibranovski MD. 2019. Haploid selection drives new gene male germline expression. *Genome Res.* 29(7):1115–1122.
- Roze D. 2009. Diploidy, population structure, and the evolution of recombination. *Am Nat.* 174:S79–S94.
- Sharp NP, Sandell L, James CG, Otto SP. 2018. The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc Natl Acad Sci U S A.* 115(22):E5046–E5055.
- Slater GP, Dapper AL, Harpur BA. 2022. Haploid and sexual selection shape the rate of evolution of genes across the Honey bee (*Apis mellifera* L.) genome. *Genome Biol Evol.* doi:10.1093/gbe/evac063
- Szövényi 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.
- Trivers RL, Hare H. 1976. Haploidploidy and the evolution of the social insect. *Science* 191(4224):249–263.
- Van Dyken JD, Wade MJ. 2010. The genetic signature of conditional expression. *Genetics* 184(2):557–570.

Associate editor: Judith Mank