Neutral genetic diversity in mixed mating systems

Marcy K. Uyenoyama

Department of Biology
Box 90338
Duke University
Durham, NC 27708-0338
USA
marcy@duke.edu

Running head: Mixed mating systems
Keywords: effective number, Ewens Sampling Formula, selfing, hermaphroditism, mating system, conservation genetics

Address for correspondence:

Marcy K. Uyenoyama
Department of Biology
Box 90338
Duke University
Durham, NC 27708-0338
USA

Tel: 919-660-7350
Fax: 919-660-7293
e-mail: marcy@duke.edu

ABSTRACT

Systems of reproduction differ with respect to the magnitude of neutral genetic diversity maintained in a population. In particular, the partitioning of reproductives into mating types and regular inbreeding have long been recognized as key determinants of effective population number. Here, the level of neutral genetic diversity is explored across a range of reproductive systems, including full gonochorism, full hermaphroditism, androdioecy, and gynodioecy. For anisogamous organisms in which reproduction is limited by the availability of large gametes and not of small gametes, the relative proportions of gonochores and hermaphrodites affect levels of neutral genetic diversity under androdioecy and gynodioecy in different ways. The influence on the level of genetic diversity of effective number, sex-specific viability differences, and the evolving quantitative trait of the population sex ratio is compared across these systems of reproduction.

1 Introduction

Genomic patterns of neutral variation reflect the evolutionary context that generated them, including selection targeted to linked and unlinked genomic locations (Stephan 2019), population structure (Barton 2008), and the system of reproduction. This analysis uses a ratio of effective numbers as a standard of comparison across systems of reproduction, including full gonochorism, full hermaphroditism, androdioecy, and gynodioecy. Of particular interest is the effect of the evolution of the sex ratio and sex-specific viability on the level of neutral genetic diversity maintained in diploid populations.

We use coalescence-based arguments to study inbreeding in these systems of reproduction. The coalescent with selfing has been explored by numerous authors (Fu 1997; Nordborg and Donnelly 1997; Möhle 1998). As in those studies, we assume that reproduction (selfing and outcrossing) occurs on a much shorter timescale than mutation and coalescence. We also assume that the size of the sample of genes whose ancestry is examined is much smaller than the inverse of the rates of mutation and coalescence. As a consequence, the state of parent-sharing between genes held by distinct individuals resolves virtually instantaneously either to coalescence or to residence of the lineages in separate individuals (see Figure 3), preserving the convergence of the process to the Kingman (1982) coalescent (Möhle 1998).

1.1 Systems of reproduction

We address anisogamous reproductive systems, in which zygotes represent the fusion between a female gamete (large) and a male gamete (small). Under full gonochorism, each mating type specializes in the production of one type of gamete, with mating occuring exclusively between organisms of opposite mating types. Hermaphroditism introduces the possibility of self-fertilization, the production of offspring through fusion of gametes generated by a single individual. In our models, a proportion s of the offspring that survive to reproductive age are uniparental, with both small and large gametes deriving from the same parent, and the complement $(1 - s)$ are biparental, representing the fusion of a large gamete produced by a female or hermaphrodite with a male gamete randomly derived from the population pool of pollen or sperm.

Androdioecious populations comprise 2 mating types: males, which produce only small gametes, together with hermaphrodites (Figure 1). Androdioecy is rare among plants as well as animals. Model systems include the plant *Datisca glomerata* (Wolf *et al.* 2001) and the nematode *Caenorhabditis elegans* (Steward and Phillips 2002). In our models, males contribute to the population pollen or sperm pool σ_A male gametes per male gamete contributed by hermaphrodites.

Gynodioecious taxa, comprising hermaphrodites together with females, appear to derive from multiple independent origins throughout the flowering plants (Rivkin *et al.* 2016). A female contributes σ_G female gametes per female gamete contributed by a hermaphrodite (Figure 2). Only hermaphrodites produce small gametes, which fertilize large gametes pro-

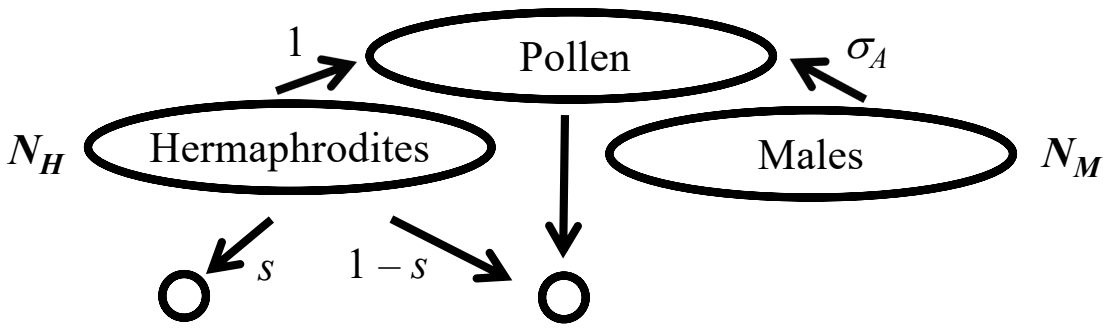

Figure 1: Offspring production under androdioecy. A proportion s of the offspring that survive to reproductive age are uniparental and $(1-s)$ biparental, with the biparental offspring fertilized from the pollen cloud. Each of the N_M males contributes pollen at rate σ_A relative to each of the N_H hermaphrodites.

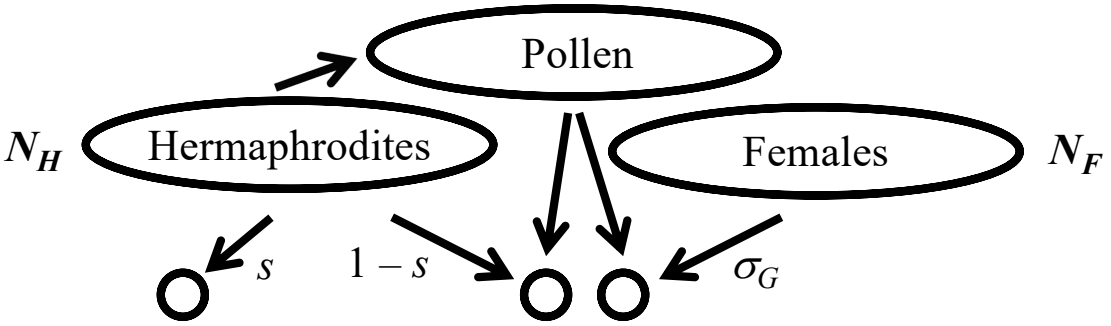

Figure 2: Offspring production under gynodioecy. As under androdioecy (Figure 1), a proportion s of the offspring that survive to reproductive age are uniparental, with the complement biparental. Biparental offspring derive from seeds fertilized by the pollen cloud, to which hermaphrodites alone contribute. Relative to a hermaphrodite, each of the N_F females generates σ_G offspring of reproductive age through seeds, which are fertilized by pollen produced by hermaphrodites.

duced by both females and hermaphrodites. Our models assume that the expression of any positive level of male fertility by hermaphrodites or males is fully sufficient to fertilize the available large gametes. This assumption gives rise to quantitative and qualitative differences in the level of neutral genetic diversity among the reproductive systems studied here.

1.2 Effective numbers

Sewall Wright (*e.g.*, Wright 1939) introduced many definitions of effective number, reflecting a variety of factors influencing genetic diversity. That some have divergent expressions within a given model (Ewens 1982; Caballero and Hill 1992) indicates that the various definitions

address distinct aspects of the evolutionary process.

We use $1/N_P$ to denote the rate of parent-sharing: the probability that a pair of genes randomly sampled from distinct individuals derive from the same parent. In the substantial literature on effective number, N_P has been called "inbreeding effective number" (Crow 1954; Crow and Denniston 1988; Ewens 1982). We use $1/2N_C$ to denote the rate of coalescence: the rate at which lineages ancestral to a pair of genes randomly sampled from distinct diploid individuals coalesce. As coalescence entails descent of the gene pair from the same parent and further from the same gene in that parent, the rates of parent-sharing ($1/N_P$) and coalescence ($1/2N_C$) are clearly closely related; indeed, they differ only by a factor of 2 under some forms of panmixis. Even so, we distinguish between them because they generally have differ conceptually and quantitatively under population structure, including in the reproductive systems considered here. Another notion of effective number corresponds to the total number of reproductives that contribute to the next generation (N). For example, N corresponds to $(N_H + N_M)$ under androdioecy (Figure 1), and to $(N_H + N_F)$ under gynodioecy (Figure 2).

The Ewens Sampling Formula (ESF, Ewens 1972) provides a succinct summary of the pattern of neutral genetic diversity observed in a sample under the infinite-alleles model of mutation, under which each mutation generates a novel allelic class. Under the ESF, the state of a sample comprising n genes corresponds to an allele frequency spectrum (AFS) of the form

$$\boldsymbol{a} = \{a_1, a_2, \ldots, a_n\}, \tag{1}$$

in which a_i denotes the number of allelic classes that appear with multiplicity i in the sample. Sample size corresponds to

$$\sum_i i a_i = n,$$

and the number of distinct alleles in the sample to

$$\sum_i a_i = k.$$

Ewens (1972) derived the probability of any AFS and the probability mass function of the total number of alleles in a sample (K). A key result is that the probability distribution of AFSs conditioned on having observed a certain number of alleles in the sample ($K = k$) is independent of the mutation rate. We distinguish here between the allele frequency spectrum (AFS) and the site frequency spectrum (SFS), which refers to the number of observations of an allele in a biallelic sample. While the SFS distribution can be derived in various ways, a simple and revealing approach involves conditioning the ESF on the observation of $K = 2$ alleles in the sample (Ganapathy and Uyenoyama 2009).

While Ewens's (1972) original paper addressed a panmictic population, Griffiths and Lessard (2005) proved convergence of certain models incorporating population structure to the Kingman (1982) coalescent process under an appropriate rescaling of time. Here, relative effective number (R) corresponds to the key scaling factor in the sense that the forms of population structure studied here can be accommodated by multiplying the mutation

parameter (θ) in the classical ESF by R. Relative effective number (R) corresponds to the large population ratio of two effective numbers: N_C, determined from the rate of coalescence between a pair of genes randomly sampled from distinct individuals, and N, the total number of reproductive individuals. For u the per-generation rate of mutation, the set of parameters governing evolutionary change corresponds to

$$\mathbf{\Phi} = \{\theta, R\}, \tag{2}$$

in which

$$\lim_{\substack{N\to\infty \\ u\to 0}} 4Nu = \theta \tag{3a}$$

$$\lim_{N,N_C\to\infty} \frac{N_C}{N} = R. \tag{3b}$$

The probability of observing AFS $\boldsymbol{a}$ in a random sample of n genes corresponds to the ESF, under the substitution of $R\theta$ for θ:

$$p_n(\boldsymbol{a}) = \frac{n!}{R\theta(R\theta+1)\ldots(R\theta+n-1)} \prod_{i=1}^{n} \left(\frac{R\theta}{i}\right)^{a_i} \frac{1}{a_i!} \tag{4}$$

(Appendix A). For a given value of the scaled mutation rate (θ), neutral genetic diversity increases with R. In the context of the ESF, higher diversity corresponds to greater numbers of alleles observed in samples of a given size and a higher probability that the next-sampled gene represents an allelic class not yet observed in the sample.

Central to the ESF (4) is independence among the sampled lineages. In the presence of inbreeding, lineages of genes sampled from the same organism remain associated by virtue of their shared genealogical history since the most recent outcross. Independence might be ensured by sampling no more than a single lineage per locus per organism. Alternatively, the shared history might be accommodated by regarding sample size (n) as a random variable. At the most recent point at which the ancestral lineages of all sampled genes reside in distinct organisms, n in (4) corresponds to the number of lineages remaining after coalescence between lineages during the generations of inbreeding until the most recent outcross.

Here, we use relative effective number (R) as a standard of comparison of several systems of reproduction, with the total number of reproductives (N) and the scaled rate of mutation (θ) held constant. Both inbreeding and the partitioning of reproductives among mating types tend to reduce R. We examine the implications for neutral genetic diversity of the evolution of the sex ratio to the evolutionarily stable strategy (ESS) in each reproductive system. An ESS corresponds to an "unbeatable" population sex ratio (Hamilton 1967), at which no alleles that modify the relative rate of production of the sexes increase at a geometric rate (see Maynard Smith 1978, Chapter 9). We first address the familiar case of a two-sex model under random mating (full gonochorism), followed by mating systems that allow regular inbreeding.

2 Full gonochorism

We first consider a gonochoristic population with random mating between N_F reproductive females and N_M reproductive males. Let P denote the collective contribution of female reproductives to the pool of autosomal genes transmitted to offspring and $(1-P)$ the collective contribution of male reproductives. Parent-sharing between a random pair of genes sampled from distinct offspring of reproductive age requires either that both genes derive from females or both from males:

$$\frac{1}{N_P} = \frac{P^2}{N_F} + \frac{(1-P)^2}{N_M}.$$

Offspring receive half their autosomal genome from each parent ($P = 1/2$), with the rate of parent-sharing reducing to

$$\frac{1}{N_P} = \frac{1}{2}\left[\frac{1}{2N_F} + \frac{1}{2N_M}\right]. \tag{5}$$

That the effective number of reproductives corresponds to the harmonic mean of twice the number of females and twice the number of males, has been derived under various definitions of effective number (*e.g.*, Wright 1931; Ewens 1982; Caballero 1994). Crow and Kimura (1970, p. 102) appear to have presented the earliest retrospective (coalescent) argument that refers explicitly to the probability of parent-sharing.

In this base case, the effective numbers implied by the rates of parent-sharing and coalescence are in fact equal,

$$N_P = N_C,$$

with relative effective number (3b) corresponding to

$$R = \lim \frac{N_P}{N}.$$

From (5), under the assignment of the arbitrary scaling factor N as the sum of the numbers of female and male reproductives, we obtain

$$R = \frac{N_P}{N_F + N_M} = 4x(1-x), \tag{6}$$

for x the frequency of females among reproductives:

$$x = \frac{N_F}{N_F + N_M}.$$

Clearly,

$$R \leq 1,$$

with equality only if the sex ratio at reproduction is equal to unity ($x = 1/2$).

We now address the effect on R of the evolution of the sex ratio in the population. Fisher's (1958) concept of reproductive value of an individual of a given sex is proportional to the

collective contribution of the sex to the population gene pool: each reproducing female has a reproductive value of

$$\frac{P}{N_F}$$

and each reproducing male

$$\frac{1-P}{N_M}.$$

Let Z denote the number of female reproductives that can be reared per male reproductive. The ESS corresponds to the point at which the returns on investing a unit of reproduction in each sex are equal:

$$\frac{ZP}{N_F} = \frac{1-P}{N_M}$$

(Fisher 1958). This ESS sex ratio among reproductives corresponds to

$$\frac{N_F}{N_M} = \frac{x}{1-x} = \frac{ZP}{1-P}, \tag{7}$$

for x the proportion of females.

In general, Z reflects various forms of sex-specific demands by offspring on parental resources. For simplicity, we here restrict consideration to Z deriving solely from sex-specific viability: Z is the rate at which female zygotes survive to reproductive age relative to male zygotes. Under this interpretation of Z, Fisher (1958) noted that the ESS sex ratio at conception corresponds to

$$\frac{x}{Z(1-x)} = \frac{P}{1-P}.$$

From the perspective of autosomal modifiers expressed in parents, parents are equally related to offspring of the two sexes ($P = 1/2$). The ESS sex ratio at conception is then unity, irrespective of the value of Z.

Substituting the relative frequencies of female and male reproductives at the ESS (7) for $P = 1/2$ into (6), we obtain

$$R = \frac{4Z}{(Z+1)^2}.$$

Relative effective number assumes its maximum value ($R = 1$) only in the absence of sex-specific differences in viability ($Z = 1$). Alternatively, any viability differences ($Z \neq 1$) cause disparities between the sex ratio at conception and at reproductive age, reducing R. For a given scaled mutation parameter (θ) and total number of reproductives (N), the partitioning of reproductives into mating types reduces neutral genetic diversity below the level expected under the original ESF (Ewens 1972).

3 Hermaphroditism

We now address reproductive systems comprising hermaphrodites, which offer the potential for self-fertilization (selfing). In contrast with full gonochorism, the effective numbers inferred

from rates of coalescence (N_C) and parent-sharing (N_P) are in general unequal.

3.1 Regular inbreeding

3.1.1 Relative rate of coalescence

We use s to denote the uniparental fraction: the probability that a random reproductive individual is uniparental. Figure 3 depicts the ancestry of the lineages of the genes held by a random diploid reproductive individual. If the individual is uniparental, the lineages

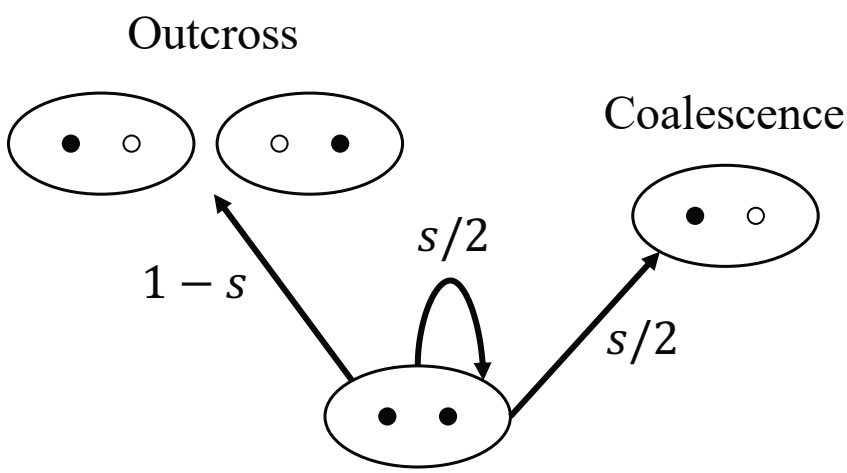

Figure 3: Ancestry of lineages borne by a single individual, for s the probability that a random reproductive individual is uniparental.

either coalesce in the immediately preceding generation with probability c (Coalescence) or remain distinct, with probability $(1-c)$. For clarity, we distinguish between these events even though

$$c = 1 - c = 1/2.$$

Alternatively, with probability $(1-s)$, the individual is biparental, implying that the lineages descend in the immediately preceding generation from lineages held by distinct individuals (Outcross). Of the two end states (Coalescence and Outcross), the process resolves to coalescence with probability

$$\frac{sc}{1-s+sc} \tag{8}$$

and to separation of the lineages into distinct individuals with the complement probability. Accordingly, the large-population limit of the rate of coalescence between a pair of genes sampled from distinct individuals corresponds to

$$\frac{1}{2N_C} = \frac{c}{N_P} + \frac{1-c}{N_P}\,\frac{sc}{1-s+sc} = \frac{1}{2N_P(1-s/2)}. \tag{9}$$

For uniparental fractions on the order of the mutation rate ($s = O(u)$), this rate reduces to $1/2N_P$, the value expected in the absence of regular inbreeding. For biparental fractions on the order of the mutation rate ($1 - s = O(u)$), the rate reduces to $1/N_P$, consistent with the near-cloning of N_P highly homozygous lines.

In the remaining case ($s, (1-s) \gg 1/N_P, u$), resolution to either coalescence of the lineages or to their separation into different carriers occurs virtually instantaneously relative to the rate of parent-sharing and mutation. Möhle (1998) showed that under these conditions, the ancestral process converges to the Kingman coalescent, with the rate of coalescence between lineages held by a single individual (9) increasing with the rate of selfing. Relative effective number (3b) corresponds to

$$R = \lim \frac{N_C}{N} = \lim \frac{N_P(1 - s/2)}{N}. \tag{10}$$

In the special case of full hermaphroditism (absence of gonochores), we assign N as the number of diploid reproductives. Let s^* denote the fraction at conception of zygotes that are uniparental (derived from gametes of a single individual), with the complement $(1 - s^*)$ formed by the fusion of a pair of gametes sampled at random from the local gamete pool. Note that while s^* is defined at conception, the uniparental fraction s refers to the proportion of uniparentals among reproductive individuals. At conception, the proportion of zygotes that are uniparental is

$$s^* + (1 - s^*)/N,$$

reflecting that uniparental offspring can also arise through random union of gametes. Uniparental offspring survive to reproductive age at rate τ relative to biparental offspring. The uniparental fraction corresponds to the probability that a random individual of reproductive age is uniparental:

$$s_H = \frac{[s^* + (1 - s^*)/N]\tau}{[s^* + (1 - s^*)/N]\tau + (1 - s^*)(1 - 1/N)} = \frac{s^*\tau}{s^*\tau + 1 - s^*} + O(u), \tag{11}$$

under the usual large-population assumptions ($1/N = O(u)$). Appendix B shows that under full hermaphroditism, a pair of genes randomly sampled from distinct reproductive individuals derive from the same parent with probability $1/N$, irrespective of the uniparental fraction (s):

$$1/N_P = s^2(1/N)(1) + 2s(1 - s)(2/N)(1/2) + (1 - s)^2(4/N)(1/4) = 1/N. \tag{12}$$

We then obtain from (10)

$$R = 1 - s/2, \tag{13}$$

with

$$N_C = N(1 - s/2),$$

in agreement with Pollak's (1987, p. 354) expression for effective number under partial selfing. Under (12), the transition matrices noted by Nordborg and Donnelly (1997) and Möhle (1998) are consistent with (9) and Figure 3.

3.1.2 ESS sex ratio in dioecious systems

We now address dioecious systems in which N_G gonochores (females or males) and N_H hermaphrodites contribute to the next generation. As in the case of full gonochorism (Section 2), the probability of parent-sharing between a random pair of genes sampled from distinct reproductive individuals depends on the proportion of the gene pool that derives from each mating type:

$$\frac{1}{N_P} = \frac{P^2}{N_G} + \frac{(1-P)^2}{N_H}, \tag{14}$$

for P the proportion of the gene pool contributed by gonochores. Because gonochores necessarily outcross to hermaphrodites,

$$1/2 \geq P, \tag{15}$$

with equality only for cases in which hermaphrodites are also obligated to outcross to gonochores.

The heuristic reproductive value argument presented for full gonochorism would suggest that the ESS sex ratio corresponds to the point of equality between the marginal reproductive values of gonochores and hermaphrodites:

$$\frac{ZP}{N_G} = \frac{1-P}{N_H},$$

for Z the number of gonochore reproductives that can be generated per hermaphrodite. The ESS ratio of gonochores to hermaphrodites implied by this argument is

$$\frac{ZP}{1-P} = \frac{x}{1-x}, \tag{16}$$

for x the proportion of gonochores among reproductives. For each of the mixed mating models addressed here, this expression does in fact agree with the unbeatable sex ratio determined from full dynamic analyses of the evolution of sex ratio modifiers (Uyenoyama and Takebayashi 2017).

At the ESS sex ratio (16), relative effective number (10) corresponds to

$$R = \lim \frac{N_P(1-s/2)}{N} = \frac{Z(1-s/2)}{[ZP+1-P][P+Z(1-P)]}, \tag{17}$$

for s the probability that a random reproductive individual is uniparental. This expression indicates that

$$R \leq (1-s/2),$$

with equality only at $P=0$ (full hermaphroditism) or $Z=1$ (absence of sex-specific viability). In the absence of gonochores, relative effective number R depends on the uniparental fraction (s) alone (13). In their presence, R may depend on all model parameters through their effects on the collective genetic contribution of gonochores (P), in addition to s.

Under the chain rule, the derivative of R with respect to Z is

$$\frac{dR}{dZ} = \frac{\partial R}{\partial Z} + \frac{\partial R}{\partial P}\frac{dP}{dZ} + \frac{\partial R}{\partial s}\frac{ds}{dZ}. \tag{18}$$

From (17), the first term corresponds to

$$\frac{\partial R}{\partial Z} = \frac{(1-Z^2)P(1-P)(1-s/2)}{[P(1-P)(1-Z)^2+Z]^2}, \tag{19}$$

indicating that the partial derivative of R with respect to Z is negative for $Z > 1$ and positive for $Z < 1$. The partial derivative of R with respect to P is non-positive:

$$\frac{\partial R}{\partial P} = \frac{-(1-2P)Z(1-Z)^2(1-s/2)}{[P(1-P)(1-Z)^2+Z]^2} \leq 0, \tag{20}$$

with equality only if

$$1/2 = P$$

(see (15)). Lastly, R declines under higher rates of selfing (s):

$$\frac{\partial R}{\partial s} = \frac{-Z}{P(1-P)(1-Z)^2+Z}. \tag{21}$$

A full examination of the nature of the effect of sex-specific viability Z on R (18) requires determination of the uniparental fraction among reproductives (s) and the contribution to the gene pool of gonochores (P), as well as an assessment of their dependence on Z (dP/dZ, ds/dZ). We now obtain explicit expressions for these components for the models of androdioecy and gynodioecy addressed here (Sections 3.2 and 3.3).

3.2 Androdioecy

Under androdioecy, the uniparental fraction s is independent of the proportion of reproductive males, a reflection of our assumption of a sufficiency of small gametes (sperm or pollen) at any frequency of males or hermaphrodites. Accordingly, the probability that a random individual of reproductive age is uniparental is identical to the expression under full hermaphroditism (11):

$$s_A = s_H. \tag{22}$$

That the uniparental fraction is independent of Z implies

$$\frac{\partial R}{\partial s}\frac{ds}{dZ} = 0$$

in (18).

In an androdioecious population comprising N_H hermaphroditic and N_M male reproductives, males contribute small gametes at rate σ_A relative to hermaphrodites (Figure 1).

Among biparental offspring of reproductive age, the proportion that have a male parent is proportional to σ_A:

$$G_A = \frac{\sigma_A N_M}{\sigma_A N_M + N_H} = \frac{\sigma_A x}{\sigma_A x + 1 - x},$$

for x the proportion of males among reproductives. A gene randomly sampled from an offspring of reproductive age derives from a male parent with probability

$$P_A = (1 - s_A) G_A / 2, \tag{23}$$

and from a hermaphroditic parent with the complement probability. Expression (23) indicates that among biparental reproductives $(1 - s_A)$, a proportion G_A have a male parent, with 1/2 the probability that the paternally-derived gene is sampled.

The ESS sex ratio at reproductive age (16) corresponds to

$$\frac{Z P_A}{1 - P_A} = \frac{x}{1 - x}.$$

Eliminating x from these two expressions for P_A, we obtain the proportion of the gene pool contributed by males at the ESS,

$$P_A = \frac{\sigma_A Z (1 - s_A)/2 - 1}{\sigma_A Z - 1}, \tag{24}$$

which declines with fraction of offspring that are biparental $(1 - s_A)$. Using (24), we obtain an expression for the ESS sex ratio (16) among reproductives:

$$\frac{x}{1 - x} = \frac{\sigma_A Z (1 - s_A)/2 - 1}{\sigma_A (1 + s_A)/2}, \tag{25}$$

in agreement with Equation (7) of Lloyd (1975). It is valid only if

$$\sigma_A Z (1 - s_A)/2 > 1; \tag{26}$$

otherwise, the ESS sex ratio corresponds to full hermaphroditism ($x = 0$, Uyenoyama and Takebayashi 2017).

Expression (24) confirms that males contribute less to the gene pool than hermaphrodites,

$$1/2 > P_A$$

(15), and that $\partial R / \partial P_A$ (20) is negative. As one might expect, the gonochore contribution to the gene pool increases with gonochore viability

$$\frac{dP_A}{dZ} = \frac{\sigma_A (1 + s_A)/2}{(\sigma_A Z - 1)^2} = \frac{(1 - P_A)}{Z(\sigma_A Z - 1)} > 0, \tag{27}$$

in which s_A is independent of Z. We then have

$$\frac{\partial R}{\partial P_A} \frac{dP_A}{dZ} < 0.$$

This relation, together with the negativity of $\partial R/\partial Z$ (19) for $Z > 1$, indicates that relative effective number R declines as Z increases in cases in which males have higher viability than hermaphrodites ($Z > 1$). In this range, R declines from its maximum of $(1 - s_A/2)$, the level expected under full hermaphroditism, as the sex ratio at reproduction departs from the sex ratio at conception in favor of males.

We now address the case in which hermaphrodite viability exceeds male viability ($1 > Z$). Because males are maintained only if (26) holds, we restrict consideration to the range

$$1 > Z > \frac{2}{\sigma_A(1 - s_A)}. \tag{28}$$

For Z near the lower bound, hermaphrodites contribute nearly the entire gene pool ($P_A \approx 0$), implying that R approaches its maximum $(1-s_A/2)$ at both ends of this range. This behavior suggests non-monotonic dependence of R on male viability Z. We obtain an expression for the derivative of R with respect to Z by substituting (19), (20), and (27) into (18). In the range (28), this expression is proportional to a quadratic in Z,

$$\frac{dR}{dZ} \propto Z^2(1 - s_A)\sigma_A^2 + Z(1 - s_A)\sigma_A(\sigma_A - 3) + 4 - \sigma_A(3 + s_A), \tag{29}$$

which is positive at $Z = 1$ and negative at $Z = 2/[\sigma_A(1 - s_A)]$, the limits of the range of validity. From the maximum of $(1 - s_A/2)$ at $Z = 1$, R first declines as Z declines, reaching its minimum value at the single positive root of this quadratic (29). As Z declines further, R increases, again approaching its maximum $(1 - s_A/2)$ as Z approaches its lower bound.

3.3 Gynodioecy

We now address a gynodioecious population comprising N_H hermaphroditic and N_F female reproductives, in which females produce large gametes at rate σ_G relative to hermaphrodites (Figure 2). As before, a proportion s^* of the large gametes produced by a hermaphrodite are self-fertilized, with the complement fertilized from the small gamete pool of the population. To terms larger than the rate of mutation, the probability that a random individual of reproductive age is uniparental is now

$$\begin{aligned} s_G &= \frac{N_H s^* \tau}{\sigma_G N_F + N_H(s^*\tau + 1 - s^*)} \\ &= \frac{(1 - x)s_H}{\tilde{\sigma}_G x + 1 - x}, \end{aligned} \tag{30}$$

for s_H given in (11), x the proportion of females among reproductives, and

$$\widetilde{\sigma}_G = \sigma_G/(s^*\tau + 1 - s^*). \tag{31}$$

Relative to a hermaphrodite, a female generates $\widetilde{\sigma}_G$ offspring of reproductive age through large gametes. The presence of obligately-outcrossing females ($x > 0$) reduces the fraction of offspring that are uniparental:

$$s_G < s_H.$$

This aspect of gynodioecy contrasts with androdioecy, under which the presence of males has no effect on the uniparental fraction (22).

While all uniparental offspring necessarily derive from hermaphrodites, the proportion of biparental offspring that have a female parent is

$$\begin{aligned} G_G &= \frac{\sigma_G N_F}{\sigma_G N_F + N_H(1-s^*)} \\ &= \frac{\widetilde{\sigma}_G x}{\widetilde{\sigma}_G x + (1-x)(1-s_H)}, \end{aligned}$$

for x the proportion of females among reproductives and $\widetilde{\sigma}_G$ given by (31). Using (30), we obtain the probability that a gene randomly sampled from the offspring generation at the point of reproduction derives from a female parent:

$$P_G = (1-s_G)G_G/2 = \frac{\widetilde{\sigma}_G x/2}{\widetilde{\sigma}_G x + 1 - x}. \tag{32}$$

This expression reflects that females contribute half of the autosomal genome of their offspring.

The ESS sex ratio at reproductive age (16) is

$$\frac{x}{1-x} = \frac{ZP_G}{1-P_G},$$

for x the proportion of females among reproductives. Solving (32) at this ESS sex ratio, we obtain the proportion of the gene pool contributed by females at the ESS:

$$P_G = \frac{\widetilde{\sigma}_G Z/2 - 1}{\widetilde{\sigma}_G Z - 1}. \tag{33}$$

The collective contribution of females to the gene pool increases with the relative viability of females (Z), while remaining less than the collective contribution of hermaphrodites (15). The sex ratio among reproductives,

$$\frac{x}{1-x} = \frac{\widetilde{\sigma}_G Z/2 - 1}{\widetilde{\sigma}_G/2}, \tag{34}$$

is valid only if

$$\widetilde{\sigma}_G Z/2 > 1; \tag{35}$$

otherwise, the ESS sex ratio corresponds to full hermaphroditism ($x = 0$, Uyenoyama and Takebayashi 2017).

Lloyd (1975) presented a parameter-dense model of gynodioecy, including multiple stages at which self-pollen may fertilize ovules produced by hermaphrodites ("males" in Lloyd's terminology). Agreement between (34) and Lloyd's Equation (2) requires that his $o = 1/\sigma_G$, $S = 1/Z$, $i = \tau$, and $b = e = 1$ (equal fertilization rates of ovules produced by females and

hermaphrodites). Also required are that either $a = s^*$ with $w = 1$ or $a = 0$ with $w = (1-s^*)$, conditions under which the large bracket in Lloyd's Equation (2) reduces to $(s^*\tau + 1 - s^*)$, the average viability of zygotes produced through ovules generated by hermaphrodites.

The derivative with respect to Z of the collective contribution of females is

$$\frac{dP_G}{dZ} = \frac{\widetilde{\sigma}_G/2}{(\widetilde{\sigma}_G Z - 1)^2} > 0.$$

As $\partial R/\partial P_G$ (20) is negative, the effect of Z on relative effective number through the collective contribution of females is also negative:

$$\begin{aligned}\frac{\partial R}{\partial P_G}\frac{dP_G}{dZ} &= \frac{-(1-2P_G)Z(1-Z)^2(1-s_G/2)\widetilde{\sigma}_G/2}{[P_G(1-P_G)(1-Z)^2+Z]^2(\widetilde{\sigma}_G Z-1)^2} \\ &= \frac{-(1-2P_G)(1-Z)^2(1-s_G/2)(1-P_G)}{[P_G(1-P_G)(1-Z)^2+Z]^2(\widetilde{\sigma}_G Z-1)} < 0.\end{aligned}$$

A novel feature of gynodioecy is that the uniparental fraction s_G (30) at the ESS sex ratio (34) depends on Z:

$$s_G = \frac{s_H}{\widetilde{\sigma}_G Z - 1}. \tag{36}$$

Higher female viability Z reduces the uniparental fraction by increasing the proportion of obligately outcrossing reproductives:

$$\frac{ds_G}{dZ} = \frac{-\widetilde{\sigma}_G s_H}{(\widetilde{\sigma}_G Z - 1)^2} < 0.$$

An increase in the number of females tends to promote higher neutral genetic diversity under the ESF (4) through this effect:

$$\begin{aligned}\frac{\partial R}{\partial s_G}\frac{ds_G}{dZ} &= \frac{Z\widetilde{\sigma}_G s_H}{[P_G(1-P_G)(1-Z)^2+Z](\widetilde{\sigma}_G Z-1)^2} \\ &= \frac{2(1-P_G)s_H}{[P_G(1-P_G)(1-Z)^2+Z](\widetilde{\sigma}_G Z-1)} > 0.\end{aligned}$$

Under androdioecy, for which the uniparental fraction (s_A) is wholly determined by hermaphrodites, higher gonochore viability (Z) increases R only for Z between unity and the positive root of (29). Under gynodioecy, the first terms in (18),

$$\frac{\partial R}{\partial Z} + \frac{\partial R}{\partial P}\frac{dP}{dZ},$$

exhibit similar behavior. However, the effect of Z on R through the uniparental fraction,

$$\frac{\partial R}{\partial s}\frac{ds}{dZ} > 0$$

(for $s = s_G$), induces qualitatively new aspects into the relationship between R and Z. While the general finding (17) that relative effective number R cannot exceed $(1 - s_G/2)$ continues to hold, this upper bound increases with Z. For Z near the minimum value that permits the maintenance of females (35), both R and its upper bound $(1 - s_G/2)$ increase with Z. This increase of R with Z also occurs in the range in which female viability exceeds hermaphrodite viability ($Z > 1$), up to very large values of Z.

4 Discussion

Among the genome-wide implications of the system of reproduction is the level of neutral genetic diversity maintained in a population. In its original form, the Ewens Sampling Formula (ESF, Ewens 1972) provides a succinct expression for the level of neutral variation observed in a random sample of genes from a panmictic population evolving under the infinite-alleles model of mutation. Griffiths and Lessard (2005) showed that the ESF holds under a rescaling of time to structured populations, including those in which population size varies across the history of the sample.

A modified form of the ESF (4) also holds in a number of reproductive systems with regular inbreeding under a rescaling of time summarized by relative effective number R (3b), the large-population limit of the ratio of two forms of effective number (Appendix A). For a given per-generation mutation rate (u) and total number of reproductives (N), neutral genetic diversity increases with R.

We use relative effective number R as a standard of comparison across several reproductive systems. In addition to full gonochorism with random mating, we address regular inbreeding under full hermaphroditism, androdioecy, and gynodioecy. These reproductive systems imply distinct effective numbers and evolutionarily stable sex ratios, key determinants of R.

4.1 ESS sex ratios

An unbeatable (Hamilton 1967) or evolutionarily stable (Maynard Smith 1978) sex ratio corresponds to a population sex ratio at which rare modifiers of sex expression fail to increase at geometric rates. While mating type may represent a discrete trait, the proportion of the resources devoted to the production of each mating type can be regarded as a quantitative trait, evolving under the influence of modifiers (QTLs) throughout the genome.

Here, we have restricted consideration to population sex ratios that evolve under the influence of autosomal modifiers expressed in individuals that are uniformly related to zygotes of any mating type. In other cases, non-uniform relatedness to the mating types produced may affect the ESS sex ratio (Uyenoyama and Bengtsson 1981). Haplodiploidy provides a striking example: modifiers expressed in females that affect the sex of their siblings can

generate asymmetries in the unbeatable sex ratio that mirror asymmetries in relatedness to male and female siblings (Hamilton 1967).

As an ESS sex ratio is a property of the population, and not of any particular modifier locus, it is ideally independent of dominance or other locus-specific properties. Applying Fisher's (1958) heuristic (7) to a range of plant mating systems, Lloyd (1975) derived what would now be called ESS sex ratios by determining points of equality between the marginal values of the mating types. He further conjectured that the evolutionary dynamics among modifier alleles would cause convergence to those sex ratios. A number of workers (Ross and Weir 1975, 1976; Charlesworth and Charlesworth 1978; Wolf and Takebayashi 2004) have shown that the equilibria to which populations evolve depend on locus-specific properties, particularly dominance. While Lloyd (1977) explicitly acknowledged the failure of his secondary conjecture, the expressions he obtained by equating marginal values do in fact correspond to population sex ratios at which modifiers of sex expression fail to increase when rare, irrespective of dominance and other locus-specific properties (Uyenoyama and Takebayashi 2017).

Our analysis compares the level of neutral genetic diversity across systems of reproduction at their respective unbeatable strategies, population sex ratios at which all modifiers fail to increase at geometric rates. Conditions for the existence of an ESS are necessarily closely related to the conditions for the invasion or exclusion of modifiers with specified characteristics. For example, the minimum twofold advantage in the production of small gametes (sperm or pollen) that is required for the maintenance of males at the ESS (26) mirrors the condition for the invasion into a purely hermaphroditic population of a dominant or recessive modifier allele that causes female sterility (Equation (8) of Charlesworth and Charlesworth 1978). Similar correspondences occur under gynodioecy.

4.2 Empirical estimation of key evolutionary components

As noted earlier, the allele frequency spectrum in a sample (4) assumes independence among the sampled genes. For samples derived from multiple loci, lineages held by a given organism experience the most recent outcross event at precisely the same time. As a consequence, associations in homozygosity arise among unlinked loci throughout the genome. Cockerham and Weir (1968) introduced the term identity disequilibrium to describe these associations.

Given the number of generations since the most recent outcross, lineages ancestral to genes sampled from multiple loci within an individual become conditionally independent. The Bayesian analysis of Redelings *et al.* (2015) incorporated, for each organism sampled, a latent variable representing the number of generations since the most recent outcross. Further, information contained in the joint pattern of diversity across loci sampled from the same organism greatly facilitates the estimation of the uniparental fraction (s) relative to the single-locus case.

Each reproductive system studied here is characterized by a set of basic parameters: for example, rates of selfing by hermaphrodites (s^*), the intensity of inbreeding depres-

sion (τ), and rates of production of gametes by gonochores (females or males) relative to hermaphrodites (σ_G, σ_A). The generation of posterior distributions of the basic parameters of course permits generation of posterior distributions of functions of those parameters. Of particular interest here are the collective contribution to the gene pool of gonochores (P), relative effective number (R), and relative viability of gonochores (Z).

Figure 4, from Uyenoyama and Takebayashi (2017), depicts the posterior distribution of the proportion of the gene pool contributed by hermaphrodites ($1 - P$), inferred by Redelings *et al.* (2015) from patterns of variation observed at microsatellite loci in samples

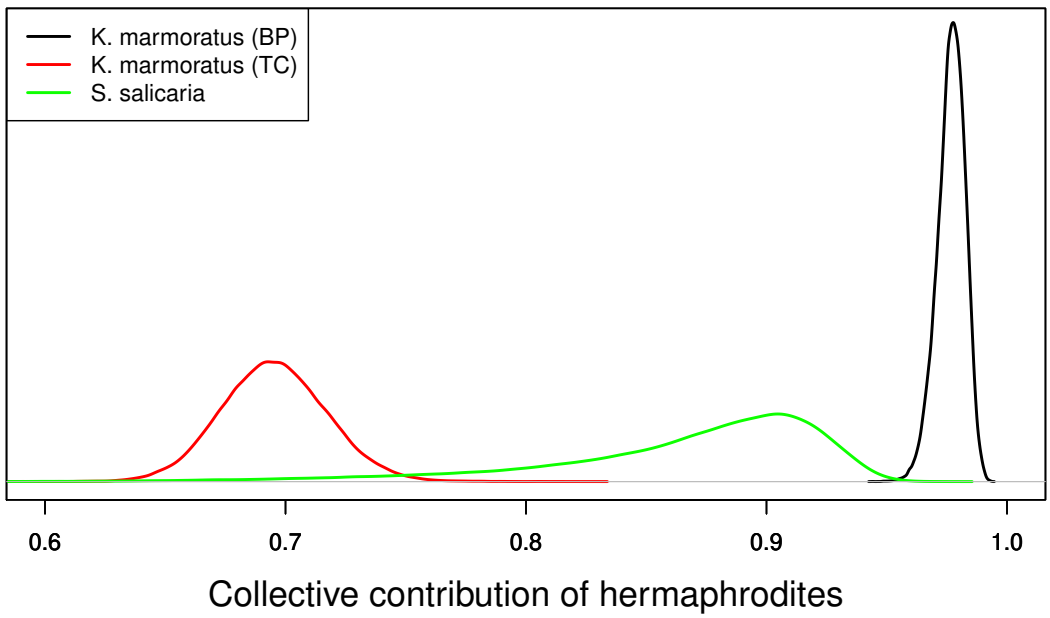

Figure 4: Posterior distributions of collective contribution of hermaphrodites to the gene pool ($1-P$), inferred from microsatellite data derived from the gynodioecious plant *Schiedea salicaria* and from 2 populations (BP and TC) of the androdioecious fish *Kryptolebias marmoratus* (Redelings *et al.* 2015).

derived from a population of the gynodioecious plant *Schiedea salicaria* and two populations of the androdioecious fish *Kryptolebias marmoratus*. The collective contribution of males in androdioecious populations corresponds to P_A (23), and the collective contribution of females in gynodioecious populations to P_G (32). Figure 4 indicates strong support for a greater collective contribution to the gene pool by hermaphrodites compared to gonochores. This empirical finding is consistent with our models: obligately outcrossing gonochores contribute only half of the genomes of their offspring, implying that P_G and P_A cannot exceed 1/2 (15). The posterior distributions for *K. marmoratus* suggest that the hermaphroditic contribution is higher in the BP population than in the TC population, concordant with the higher uniparental fraction s inferred for the BP population.

Central to the present analysis is relative effective number R (3b), the large-population limit of a ratio of two effective numbers:

$$R = \lim \frac{N_C}{N},$$

for N_C determined from the rate of coalescence and N corresponding to the total number of reproductives. While the estimation of either effective number would require additional information (*e.g.*, the per-generation rate of mutation u), their ratio can be estimated directly (Redelings *et al.* 2015). Figure 5 shows the posterior distribution of relative effective number (R) scaled to its maximum value ($1-s/2$), for s the uniparental fraction among reproductives under androdioecy (22) or gynodioecy (30). The abscissa axis label has been corrected from Redelings *et al.* (2015) and Uyenoyama and Takebayashi (2017).

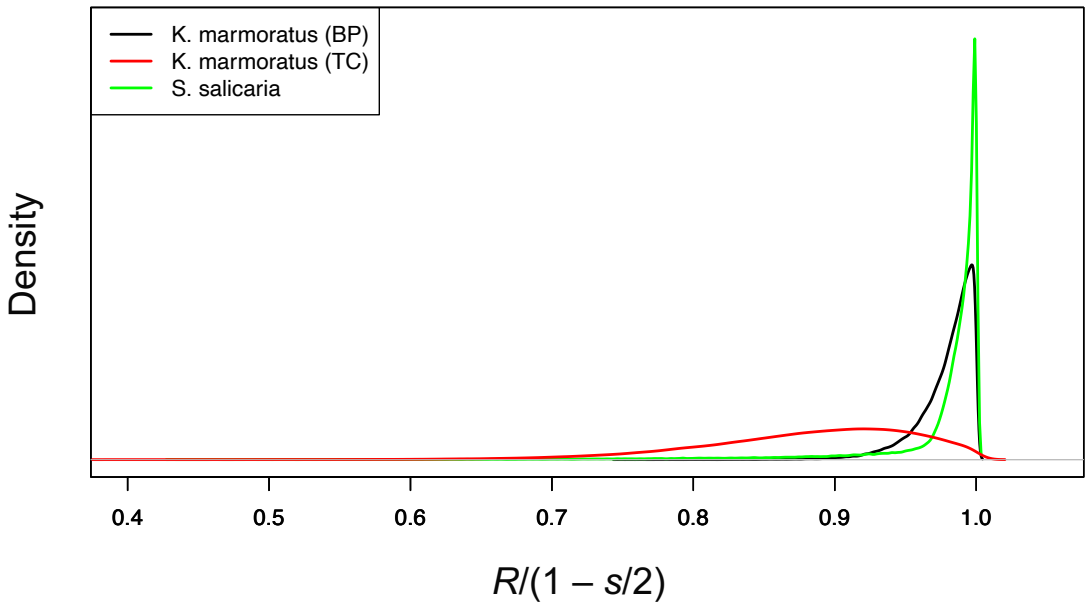

Figure 5: Posterior distributions of relative effective number R scaled to its maximum value ($1-s/2$), for s the uniparental fraction for the corresponding breeding system (modified from Uyenoyama and Takebayashi 2017). These distributions derive from the MCMC analysis of Redelings *et al.* (2015) based on observations of microsatellite variation in the gynodioecious plant *Schiedea salicaria* and from 2 populations (BP and TC) of the androdioecious fish *Kryptolebias marmoratus*, without assumptions on sex ratios.

Imposing the assumption that the sampled populations had evolved to the ESS sex ratio, Uyenoyama and Takebayashi (2017) obtained posterior distributions of the relative viability of gonochores (Z). Figure 6 indicates little support for viability differences between females and hermaphrodites in the gynodioecious *S. salicaria*, but strong support for a substantial (twofold) reduction in the viability of males relative to hermaphrodites in the androdioecious *K. marmoratus*. Of note is that relative male viability appears to be similar in the two

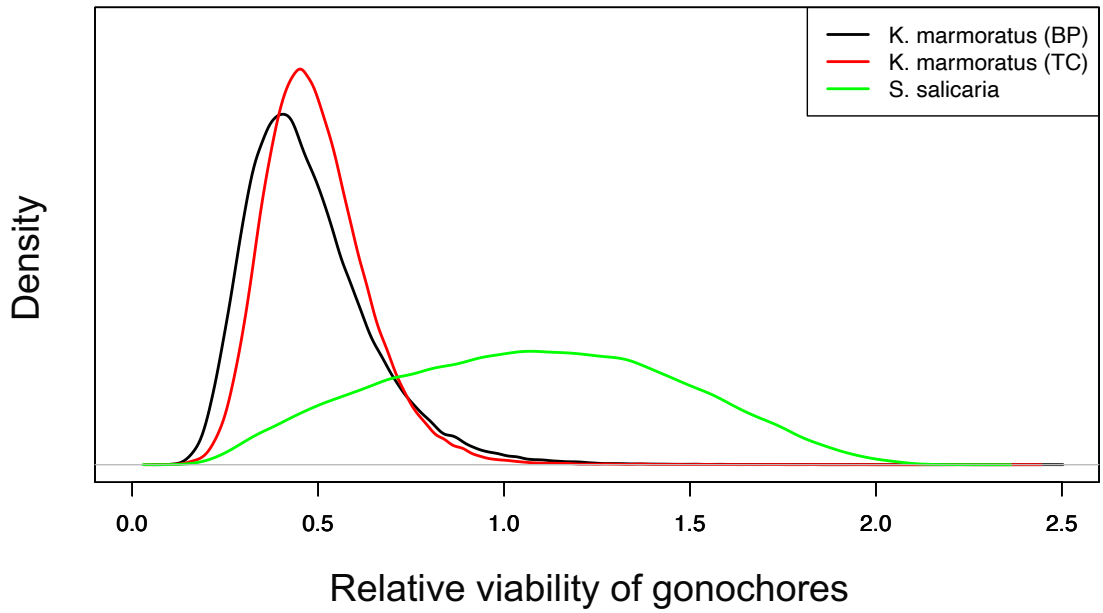

Figure 6: Posterior distributions of the viability of gonochores relative to hermaphrodites (Z) in the androdioecious fish *Kryptolebias marmoratus* and the gynodioecious plant *Schiedea salicaria* (from Uyenoyama and Takebayashi 2017).

Kryptolebias populations, in spite of marked differences in both male frequency and collective contribution of hermaphrodites (Figure 4). Turner *et al.* (2006) described *K. marmoratus* males as more conspicuous and reported higher frequencies of males in laboratory-reared populations than in natural populations. These findings are consistent with the view that predation in the wild may contribute to the reduction in male viability.

4.3 Neutral diversity across reproductive systems

For a given per-generation rate of mutation (u) and total number of reproductives (N), the level of neutral genetic diversity depends on relative effective number R (3b), which in turn reflects not only the population sex ratio but also the relative viability of mating types (Z). We have explored the nature of the dependence of R on Z across reproductive systems (18).

An important assumption of the models studied here is that any positive level of production of small gametes (sperm or pollen) is fully sufficient to fertilize the available large gametes. Under androdioecy, it is the production of small gametes by males relative to hermaphrodites (σ_A) that drives the evolution of the sex ratio and determines the level of neutral diversity in the population, not the relative investment in male versus female function in hermaphrodites. Similarly, for gynodioecy, a key determinant is the production of large gametes by females relative to hermaphrodites (σ_G). Among the implications of this assumption is that the relative viability of gonochores (Z) has qualitatively different effects

on the pattern of neutral genetic diversity under androdioecy versus gynodioecy.

4.3.1 Full gonochorism

In the classical case, without inbreeding and with reproduction restricted to mating between a pair of mating types, the effective numbers inferred from rates of parent-sharing and coalescence are identical ($N_P = N_C$), with relative effective number (3b) corresponding to

$$R = \lim \frac{N_P}{N_F + N_M}.$$

for N_F and N_F the numbers of reproductive females and males and N_P the harmonic mean of $2N_F$ and $2N_M$ (5). This expression reflects the effect on R of the structuring of the gene pool into mating types. In the absence of viability differences between the sexes ($Z = 1$), the ESS sex ratio (7) is unity ($N_F = N_M$), both at conception and at reproductive age, implying equal arithmetic and harmonic means and a relative effective number R of 1. Sex-specific differences in viability ($Z \neq 1$, Figure 6) cause the harmonic mean to decline relative to the arithmetic mean, diminishing relative effective number ($R < 1$).

4.3.2 Dioecy with selfing

Self-fertilization reduces the number of parental individuals contributing to the offspring generation. Each uniparental offspring derives from a single parent and each biparental offspring from two distinct parents, implying an average of

$$s + 2(1 - s) = 2 - s$$

parents per offspring, for s the probability that a random reproductive offspring is uniparental. In the dioecious systems of mating studied here, this effect of inbreeding combines with the partitioning of the gene pool into mating types to reduce further the relative effective number R and neutral genetic diversity compared to dioecy without inbreeding.

Our generalization to androdioecy and gynodioecy of the Fisherian heuristic (Section 2) indicates that the ESS sex ratio at conception corresponds to $P/(1 - P)$, for $(1 - P)$ the proportion of the gene pool derived from hermaphrodites (Figure 4). In reproductive systems that permit self-fertilization, the collective contribution to the gene pool of gonochores never exceeds that of hermaphrodites (15), reflecting only half the genome of biparental offspring derives from their gonochore parent. Accordingly, the ESS sex ratio at conception favors hermaphrodites for any positive level of production of uniparental offspring. That our results do not agree with those of Charnov *et al.* (1976) appears to reflect their assumption of equal collective contributions of mating types to the gene pool, with hermaphrodites regarded as fractional females and fractional males. Our determination of the unbeatable sex ratio reflects the evolutionary dynamics of genetic modifiers of sex expression (Uyenoyama and Takebayashi 2017).

Under androdioecy (Section 3.2), the uniparental fraction s_A (22) is independent of the frequency of males, and relative effective number R takes its maximum value of $(1 - s_A/2)$ only in the absence of sex-specific viability ($Z = 1$). As Z departs from 1 in either direction, R tends to decline. However, for Z quite low (less than the single positive root of (29)), R again approaches its maximum value of $(1 - s_A/2)$ as the frequency of male reproductives approaches zero.

Under gynodioecy (Section 3.3), R again takes its maximum value of $(1 - s_G/2)$ in the absence of sex-specific viability ($Z = 1$). In contrast with androdioecy, the presence of females reduces the uniparental fraction s_G (30). Departures of the relative viability of gonochores from unity also affect s_G and consequently the maximum value of R. Relative effective number R exhibits a non-monotone dependence on Z in multiple regions across its range under gynodioecy. By reducing s_G, the transition from full hermaphroditism to gynodioecy can increase the level of neutral genetic diversity, behavior that would not be expected in the transition to androdioecy.

4.4 Evolutionary responses to environmental change

Estimation of census or effective population numbers or their ratio has received much attention in the conservation biology literature (Luikart *et al.* 2010). Estimation of effective number generally poses challenges, with some approaches requiring sampling at multiple time points or observation of tightly linked loci (Waples 2006). Determining whether observation of levels of neutral genetic diversity can aid in identifying endangered species, detecting past bottlenecks in population size, or gauging resilience to major future shifts in habitat or population size (see Teixeira and Huber 2021; Garcia-Dorado and Caballero 2021) lies beyond the objectives of this study.

We use a ratio of effective numbers, R (3b), to compare a range of reproductive systems. Under a model-based approach, this index can be directly inferred from the pattern of neutral genetic diversity in a random sample (Figure 5). Relative effective number R depends on both the number of reproductives (N) and the sex ratio among them. Factors that influence the sex ratio among reproductives affect the level of neutral genetic diversity maintained in a population. An extrapolation of the non-monotone dependence of R on relative viability Z in gynodioecious systems (Section 4.3.2) may suggest that management policies that promote female-bias among reproductives in captive populations may reduce inbreeding but have complex implications for the level of neutral genetic diversity.

Evolutionarily stable sex ratios reflect the evolution of autosomal modifiers throughout the genome in response to the local sex ratio and conditions that affect the collective contribution to the gene pool of the mating types. In *Kryptolebias marmoratus*, an emerging model system for environmental sex determination (ESD, Kelley *et al.* 2016), functional hermaphrodites may transform into fertile males (secondary males). In addition, incubation at lower temperatures promotes the development of fertile primary males directly from self-fertilized eggs (Harrington 1967). Ellison *et al.* (2015) reported that interactions between

incubation temperature and methylation patterns affect the proportion of primary males, with the source population from which the laboratory lines were derived affecting the nature of the response to experimental treatments. This study appeared to suggest that ESD, mediated by methylation of genes controlling sex expression, may provide a means of regulating the rate of selfing.

An alternative view is that ESD may number among the mechanisms through which sex ratio evolves in response to the selfing rate. Our models assume that any positive level of male function is fully sufficient to fertilize available large gametes. Accordingly, increasing the proportion of obligately-outcrossing female reproductives does tend to reduce the rate of inbreeding (30), but increasing the proportion of male reproductives has no effect on the uniparental fraction s.

Acknowledgments

I thank two anonymous Reviewers for their comments and suggestions. Public Health Service grant GM 37841 provided partial funding for this research.

References

Barton, N. H., 2008. Identity and coalescence in structured populations: a commentary on 'Inbreeding coefficients and coalescence times' by Montgomery Slatkin. *Genet. Res.* **89**, 475–477.

Caballero, A., 1994. Developments in the prediction of effective population size. *Heredity* **73**, 657–679.

Caballero, A. and Hill, W. G., 1992. Effective size of nonrandom mating populations. *Genetics* **130**, 909–916.

Charlesworth, B. and Charlesworth, D., 1978. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997.

Charnov, E. L., Maynard Smith, J., and Bull, J. J., 1976. Why be a hermaphrodite. *Nature* **263**, 125–126.

Cockerham, C. C. and Weir, B. S., 1968. Sib mating with two linked loci. *Genetics* **60**, 629–640.

Crow, J. F., 1954. Breeding structure of populations II. Effective population number. In O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush, eds., *Statistics and mathematics in biology*, chapter 43, 543–556. The Iowa State College Press, Ames, IA.

Crow, J. F. and Denniston, C., 1988. Inbreeding and variance effective population numbers. *Evolution* **42**, 482–495.

Crow, J. F. and Kimura, M., 1970. An introduction to population genetics theory. Harper and Row, New York.

Ellison, A., Rodriguez López, C. M., Moran, P., Breen, J., Swain, M., Megias, M., Hegary, M., Wilkinson, M., Pauluk, R., and Consuegra, S., 2015. Epigenetic regulation of sex ratios may explain natural variation in self-fertilization rates. *Proc. R. Soc. Lond. B* **282**, 20151900.

Ewens, W. J., 1972. The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**, 87–112.

Ewens, W. J., 1982. On the concept of effective population size. *Theor. Pop. Biol.* **21**, 373–378.

Fisher, R. A., 1958. The Genetical Theory of Natural Selection. Dover, New York, second edition.

Fu, Y.-X., 1997. Coalescent theory for a partially selfing population. *Genetics* **146**, 1489–1499.

Ganapathy, G. and Uyenoyama, M. K., 2009. Site frequency spectra from genomic SNP surveys. *Theor. Pop. Biol.* **75**, 346–354.

Garcia-Dorado, A. and Caballero, A., 2021. Neutral genetic diversity as a useful tool for conservation biology. *Conservation Genetics* **22**, 541–545.

Griffiths, R. C. and Lessard, S., 2005. Ewens' sampling formula and related formulae: combinatorial proofs, extensions to variable population size and applications to ages of alleles. *Theor. Pop. Biol.* **68**, 167–177.

Hamilton, W. D., 1967. Extraordinary sex ratios. *Science* **156**, 477–488.

Harrington, Jr., R. W., 1967. Environmentally controlled induction of primary male gonochorists from eggs of the self-fertilizing hermaphroditic fish, *Rivulus marmoratus* Poey. *The Biological Bulletin* **132**, 174–199.

Karlin, S. and McGregor, J., 1972. Addendum to a paper of W. Ewens. *Theor. Pop. Biol.* **3**, 113–116.

Kelley, J. L., Yee, M.-C., Brown, A. P., Richardson, R. R., Tatarenkov, A., Lee, C. C., Harkins, T. T., Bustamante, C. D., and Earley, R. L., 2016. The genome of the self-fertilizing mangrove rivulus fish, *Kryptolebias marmoratus*: a model for studying phenotypic plasticity and adaptations to extreme environments. *Genome Biol. Evol.* **8**, 2145–2154.

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

Lloyd, D. G., 1975. The maintenance of gynodioecy and androdioecy in angiosperms. *Genetica* **45**, 325–339.

Lloyd, D. G., 1977. Genetic and phenotypic models of natural selection. *J. Theor. Biol.* **69**, 543–560.

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., and Allendorf, F. W., 2010. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics* **11**, 355–373.

Maynard Smith, J., 1978. The evolution of sex. Cambridge Univ. Press, Cambridge.

Möhle, M., 1998. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Prob.* **30**, 493–512.

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

Pollak, E., 1987. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117**, 353–360.

Redelings, B. D., Kumagai, S., Tatarenkov, A., Wang, L., Sakai, A. K., Weller, S. G., Culley, T. M., Avise, J. C., and Uyenoyama, M. K., 2015. A Bayesian approach to inferring rates of selfing and locus-specific mutation. *Genetics* **201**, 1171–1188.

Rivkin, L. R., Case, A. L., and Caruso, C. M., 2016. Why is gynodioecy a rare but widely distributed sexual system? Lessons from the Lamiaceae. *New Phytol.* **211**, 688–696.

Ross, M. D. and Weir, B. S., 1975. Maintenance of male sterility in plant populations. III. Mixed selfing and random mating. *Heredity* **35**, 21–29.

Ross, M. D. and Weir, B. S., 1976. Maintenance of males and females in hermaphrodite populations and the evolution of dioecy. *Evolution* **30**, 425–441.

Stephan, W., 2019. Selective sweeps. *Genetics* **211**, 5–13.

Steward, A. D. and Phillips, P. C., 2002. Selection and maintenance of androdioecy in *Caenorhabditis elegans*. *Genetics* **160**, 975–982.

Tavaré, S., 2004. Ancestral inference in population genetics. In S. Tavaré and O. Zeitouni, eds., *Lectures on Probability Theory and Statistics, Ecole d'Été de Probabilitiés de Saint-Flour XXXI – 2001*, 1–188. Springer-Verlag, New York.

Teixeira, J. C. and Huber, C. D., 2021. The inflated significance of neutral genetic diversity in conservation genetics. *Proc. Natl. Acad. Sci. (USA)* **118**, e2015096118.

Turner, B. J., Fisher, M. T., Taylor, D. S., Davis, W. P., and Jarrett, B. L., 2006. Evolution of 'maleness' and outcrossing in a population of the self-fertilizing killifish, *Kryptolebias marmoratus*. *Evolutionary Ecology Research* **8**, 1475–1486.

Uyenoyama, M. K. and Bengtsson, B. O., 1981. Towards a genetic theory for the evolution of the sex ratio II. Haplodiploid and diploid models with sibling and parental control of the brood sex ratio and brood size. *Theor. Pop. Biol.* **20**, 57–79.

Uyenoyama, M. K. and Takebayashi, N., 2017. Evolution of the sex ratio and effective number under gynodioecy and androdioecy. *Theor. Pop. Biol.* **118**, 27–45.

Uyenoyama, M. K., Takebayashi, N., and Kumagai, S., 2019. Inductive determination of allele frequency spectrum probabilities in structured populations. *Theor. Pop. Biol.* **129**, 148–159.

Waples, R. S., 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics* **7**, 167–184.

Wolf, D. E., Satkoski, J. A., White, K., and Rieseberg, L. H., 2001. Sex determination in the androdioecious plant *Datisca glomerata* and its dioecious sister species *D. cannabina*. *Genetics* **159**, 1243–1257.

Wolf, D. E. and Takebayashi, N., 2004. Pollen limitation and the evolution of androdioecy from dioecy. *Am. Nat.* **163**, 122–137.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Wright, S., 1939. Statistical genetics in relation to evolution. Actualités scientifiques et industrielles, 802: Exposés de Biométrie et de la statistique biologique XIII. Hermann & Cie, Paris.

Appendix A Labeled coalescent

In the course of proving the Ewens Sampling Formula (Ewens 1972), Karlin and McGregor (1972) described changes in the allele frequency spectrum (1) occurring across generations. Because this description includes the allelic states of genetic lineages, Uyenoyama *et al.* (2019) denoted this argument as the labeled coalescent. Here, we present a summary of the formulation presented by Uyenoyama *et al.* (2019), which addresses changes on the timescale of the most recent evolutionary event rather than generations.

A.1 Basic recursion

To determine the probability of observing a sample of labeled genes, we consider the relationship between the AFS of a sample of genes (descendant) to the AFS immediately prior to the most recent evolutionary event (ancestor). This most recent evolutionary event may for example correspond to coalescence, mutation, or migration. Let random variable D represent the state of the descendant, A the state of the ancestor, and E the most recent evolutionary

event. For a given evolutionary model with parameters $\boldsymbol{\Phi}$ (2), the likelihood corresponds to

$$\begin{aligned}\Pr[D = \boldsymbol{d}|\boldsymbol{\Phi}] &= \sum_{e\in\Omega_e} \Pr[D = \boldsymbol{d}, E = e|\boldsymbol{\Phi}] \\ &= \sum_{e\in\Omega_e} \Pr[D = \boldsymbol{d}|E = e, \boldsymbol{\Phi}] \Pr[E = e|\boldsymbol{\Phi}].\end{aligned}$$

in which $\boldsymbol{d}$ represents the state of the descendant AFS, e a particular realization of the most recent evolutionary event, and Ω_e the set of all such events. The law of total probability implies

$$\begin{aligned}\Pr[D = \boldsymbol{d}|\boldsymbol{\Phi}] &= \sum_{e\in\Omega_e} \sum_{\boldsymbol{a}\in\Omega_{\boldsymbol{a}}} \Pr[D = \boldsymbol{d}, A = \boldsymbol{a}|E = e, \boldsymbol{\Phi}] \Pr[E = e|\boldsymbol{\Phi}] \\ &= \sum_{e\in\Omega_e} \Pr[E = e|\boldsymbol{\Phi}] \sum_{\boldsymbol{a}\in\Omega_{\boldsymbol{a}}} \Pr[D = \boldsymbol{d}|A = \boldsymbol{a}, E = e, \boldsymbol{\Phi}] \Pr[A = \boldsymbol{a}|E = e, \boldsymbol{\Phi}],\end{aligned}$$

for $\boldsymbol{a}$ represents the state of the ancestor AFS and $\Omega_{\boldsymbol{a}}$ the set of all possible ancestor AFSs. Under the assumption that the state of the ancestor A is independent of E, the evolutionary event forward in time, we have

$$\Pr[A = \boldsymbol{a}|E = e, \boldsymbol{\Phi}] = \Pr[A = \boldsymbol{a}|\boldsymbol{\Phi}],$$

with

$$\Pr[D = \boldsymbol{d}|\boldsymbol{\Phi}] = \sum_{\boldsymbol{a}\in\Omega_{\boldsymbol{a}}} \Pr[A = \boldsymbol{a}|\boldsymbol{\Phi}] \sum_{e\in\Omega_e} \Pr[E = e|\boldsymbol{\Phi}] \Pr[D = \boldsymbol{d}|A = \boldsymbol{a}, E = e, \boldsymbol{\Phi}]. \tag{A.1}$$

This expression represents a recursion that relates the probability of the present state of the sample (D), to the probability of the state immediately prior to the most recent evolutionary event (A). In general, determination of the transition term

$$\Pr[D = \boldsymbol{d}|A = \boldsymbol{a}, E = e, \boldsymbol{\Phi}]$$

is straightforward: it proceeds from ancestor to descendant through a specified evolutionary event, with only a limited number of ancestor states $\boldsymbol{a}$ capable of generating descendant state $\boldsymbol{d}$ through event e.

A.2 Modified ESF

Among independent lineages, the most recent evolutionary event corresponds either to coalescence ($E = e_1$) or to mutation ($E = e_2$), both of which may depend on genomic location or system of reproduction.

For a sample comprising n genes, the probabilities that mutation or coalescence correspond to the most recent evolutionary event are

$$\begin{aligned}\Pr[E = e_1|\boldsymbol{\Phi}] &= \lim \frac{\binom{n}{2}/2N_C}{\binom{n}{2}/2N_C + nu} = \frac{n-1}{n-1+R\theta} \\ \Pr[E = e_2|\boldsymbol{\Phi}] &= \lim \frac{nu}{\binom{n}{2}/2N_C + nu} = \frac{R\theta}{n-1+R\theta},\end{aligned}$$

for θ and R defined in (3).

We first address the case in which coalescence ($E = e_1$) is the most recent event backward in time, which corresponds to the splitting or duplication of a lineage forward in time. The ancestor AFS ($A = \boldsymbol{a}$) comprises only $(n-1)$ genes and has the form

$$\boldsymbol{a} = \boldsymbol{d} + \boldsymbol{e_i} - \boldsymbol{e_{i+1}},$$

in which $\boldsymbol{e}_j$ denotes a unit vector, with unity in the jth position and zeros elsewhere, and i is the multiplicity in A of the allelic class involved in the splitting event. We define the probability of AFSs with negative elements or other unmeaningful characteristics as zero. A splitting event in an allelic class with multiplicity i increases the multiplicity of that allele to $(i+1)$ and decreases by 1 the number of alleles with multiplicity i. Using that the split occurs uniformly at random in any of the $(n-1)$ lineages of A, we have

$$\Pr(D = \boldsymbol{d}|A = \boldsymbol{a}, E = e_1, \boldsymbol{\Phi}) = i(d_i + 1)/(n-1),$$

reflecting that $(d_i + 1)$ allelic classes have multiplicity i in A.

Alternatively, the most recent event may correspond to mutation ($E = e_2$). For an ancestor with AFS identical to the descendant ($D = A = \boldsymbol{d}$), the mutation can only have occurred in a lineage representing a singleton allelic class ($i = 1$) in the ancestor AFS, with

$$\Pr(D = \boldsymbol{d}|A = \boldsymbol{d}, E = e_2, \boldsymbol{\Phi}) = d_1/n.$$

Given this ancestor A, a mutation in any other lineage would imply zero probability of generating the descendant D. Mutation in an allelic class with multiplicity greater than unity ($i > 1$) generates a new singleton allele and reduces the multiplicity of the allelic class in which the mutation arose. The AFS of ancestors with a positive probability of generating $D = \boldsymbol{d}$ through this route must have the form

$$\boldsymbol{a} = \boldsymbol{d} + \boldsymbol{e_i} - \boldsymbol{e_{i-1}} - \boldsymbol{e_1}.$$

In this case,

$$\Pr(D = \boldsymbol{d}|A = \boldsymbol{a}, E = e_2, \boldsymbol{\Phi}) = i(d_i + 1)/n,$$

for $i(d_i + 1)$ the total number of genes that represent an allele present in multiplicity i in the ancestor sample A. This expression also reflects that the mutation arises in any gene with equal probability.

At stationarity, the probability of a sample of size n with AFS $\boldsymbol{a}$ is independent of time period:

$$\Pr[A = \boldsymbol{a}] = \Pr[D = \boldsymbol{a}] = p_n(\boldsymbol{a})$$

The labeled coalescence recursion (A.1) then reduces to

$$\begin{aligned} p_n(\boldsymbol{a}) =& \frac{n-1}{n-1+R\theta} \sum_{i=1}^{n-1} p_{n-1}(\boldsymbol{a} + \boldsymbol{e}_i - \boldsymbol{e}_{i+1}) i(a_i + 1)/(n-1) \\ &+ \frac{R\theta}{n-1+R\theta} \left[p_n(\boldsymbol{a}) a_1/n + \sum_{i=2}^{n-1} p_n(\boldsymbol{a} + \boldsymbol{e_i} - \boldsymbol{e_{i-1}} - \boldsymbol{e_1}) i(a_i + 1)/n \right] \end{aligned} \tag{A.2}$$

(compare Tavaré 2004), in which unmeaningful expressions (*e.g.*, probability of spectra with negative elements) are defined as zero.

In the absence of population structure ($R = 1$), Karlin and McGregor (1972) proved the formula given by (Ewens 1972) by showing that it is a solution of (A.2). Some population structures, including those comprising multiple mating types and some forms of regular inbreeding, can be summarized by relative effective number R (3b). For such cases, the probability of a sample comprising n genes with AFS $\boldsymbol{a}$ corresponds to (4):

$$p_n(\boldsymbol{a}) = \frac{n!}{R\theta(R\theta+1)\dots(R\theta+n-1)} \prod_{i=1}^{n} \left(\frac{R\theta}{i}\right)^{a_i} \frac{1}{a_i!}.$$

This formula and the probability mass function of the number of alleles observed in a sample of size n correspond to Ewens's (1972) expressions under the substitution of $R\theta$ for θ. Other remarkable properties of the ESF that are preserved include that the probability that the nth-sampled gene represents a novel allele corresponds to

$$\frac{R\theta}{n-1+R\theta}$$

and that the distribution of allele multiplicities conditioned on a given number of alleles in the sample is independent of $R\theta$ (Ewens 1972).

Appendix B Parent-sharing under full hermaphroditism

The rate of parent-sharing corresponds to the probability that a pair of genes, each randomly sampled from distinct individuals in the offspring generation, derive from the same parent.

A pair of uniparental offspring share their parent with probability

$$\frac{\binom{1}{1}}{\binom{N}{1}} = \frac{1}{N},$$

for N the number of parental individuals. A uniparental offspring and a biparental offspring share a parent with probability

$$\frac{\binom{1}{1}\binom{N-1}{1}}{\binom{N}{2}} = \frac{2}{N},$$

corresponding to the probability that 1 of the parents of the biparental individual is identical to the parent of the uniparental individual. A pair of biparental offspring may have 2, 1, or 0 parents in common. The probability that the second individual shares both parents with the first individual is of order $1/N^2$,

$$\frac{\binom{2}{2}}{\binom{N}{2}} = \frac{2}{N(N-1)}.$$

It shares exactly one parent with probability

$$\frac{\binom{2}{1}\binom{N-2}{1}}{\binom{N}{2}} = \frac{4}{N} + o(1/N).$$

To order $1/N$, the probability that a random pair of offspring share exactly one parent corresponds to

$$s^2(1/N) + 2s(1-s)(2/N) + (1-s)^2(4/N) = (2-s)^2/N,$$

for s the probability that a random offspring of reproductive age is uniparental.

Given parent-sharing between the pair of offspring individuals, the probability that 2 genes, 1 sampled from each individual, derive from the shared parent is 1 for two uniparental individuals, 1/4 for two biparental individuals, and 1/2 in the remaining case. Consequently, the probability that 2 genes, each sampled from distinct offspring individuals, derive from the same parent is

$$1/N_P = s^2(1/N)(1) + 2s(1-s)(2/N)(1/2) + (1-s)^2(4/N)(1/4) = 1/N$$

(12). This index of effective number is independent of the uniparental fraction (s).

Figure captions

Figure 1. Offspring production under androdioecy. A proportion s of the offspring that survive to reproductive age are uniparental and $(1-s)$ biparental, with the biparental offspring fertilized from the pollen cloud. Each of the N_M males contributes pollen at rate σ_A relative to each of the N_H hermaphrodites.

Figure 2. Offspring production under gynodioecy. As under androdioecy (Figure 1), a proportion s of the offspring that survive to reproductive age are uniparental, with the complement biparental. Biparental offspring derive from seeds fertilized by the pollen cloud, to which hermaphrodites alone contribute. Relative to a hermaphrodite, each of the N_F females generates σ_G offspring of reproductive age through seeds, which are fertilized by pollen produced by hermaphrodites.

Figure 3. Ancestry of lineages borne by a single individual, for s the probability that a random reproductive individual is uniparental.

Figure 4. Posterior distributions of collective contribution of hermaphrodites to the gene pool $(1-P)$, inferred from microsatellite data derived from the gynodioecious plant *Schiedea salicaria* and from 2 populations (BP and TC) of the androdioecious fish *Kryptolebias marmoratus* (Redelings *et al.* 2015).

Figure 5. Posterior distributions of relative effective number R scaled to its maximum value $(1-s/2)$, for s the uniparental fraction for the corresponding breeding system (modified from Uyenoyama and Takebayashi 2017). These distributions derive from the MCMC analysis of Redelings *et al.* (2015) based on observations of microsatellite variation in the gynodioecious plant *Schiedea salicaria* and from 2 populations (BP and TC) of the androdioecious fish *Kryptolebias marmoratus*, without assumptions on sex ratios.

Figure 6. Posterior distributions of the viability of gonochores relative to hermaphrodites (Z) in the androdioecious fish *Kryptolebias marmoratus* and the gynodioecious plant *Schiedea salicaria* (from Uyenoyama and Takebayashi 2017).

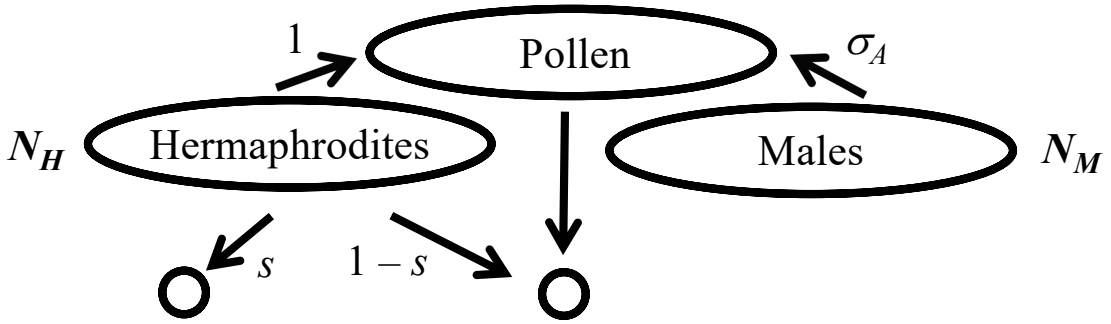

Figure 1: Offspring production under androdioecy. A proportion s of the offspring that survive to reproductive age are uniparental and $(1 - s)$ biparental, with the biparental offspring fertilized from the pollen cloud. Each of the N_M males contributes pollen at rate σ_A relative to each of the N_H hermaphrodites.

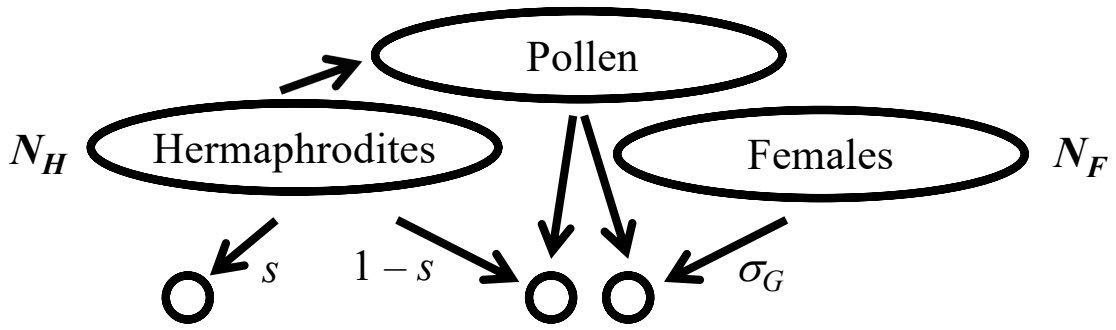

Figure 2: Offspring production under gynodioecy. As under androdioecy (Figure 1), a proportion s of the offspring that survive to reproductive age are uniparental, with the complement biparental. Biparental offspring derive from seeds fertilized by the pollen cloud, to which hermaphrodites alone contribute. Relative to a hermaphrodite, each of the N_F females generates σ_G offspring of reproductive age through seeds, which are fertilized by pollen produced by hermaphrodites.

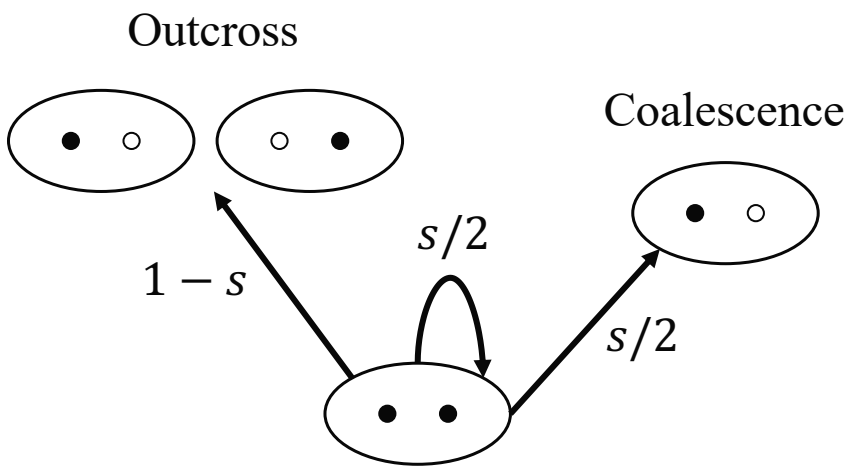

Figure 3: Ancestry of lineages borne by a single individual, for s the probability that a random reproductive individual is uniparental.

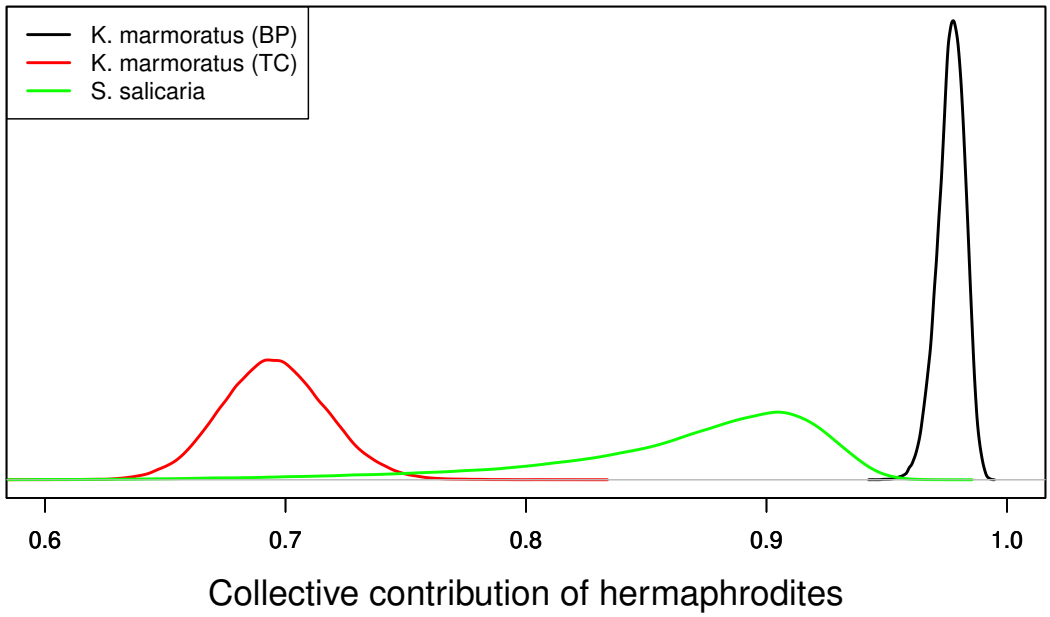

Figure 4: Posterior distributions of collective contribution of hermaphrodites to the gene pool $(1-P)$, inferred from microsatellite data derived from the gynodioecious plant *Schiedea salicaria* and from 2 populations (BP and TC) of the androdioecious fish *Kryptolebias marmoratus* (Redelings *et al.* 2015).

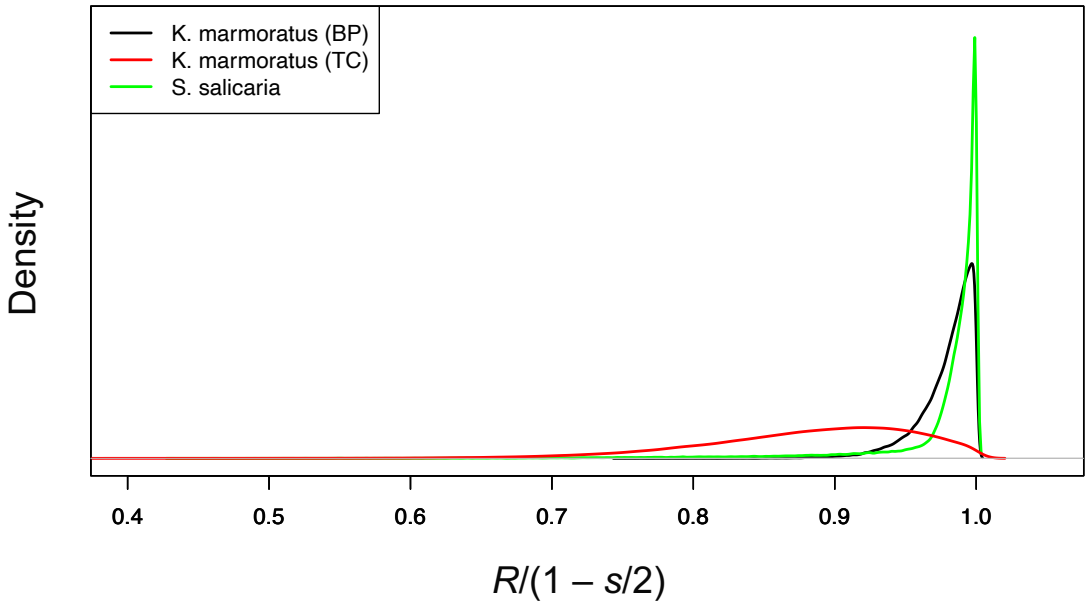

Figure 5: Posterior distributions of relative effective number R scaled to its maximum value $(1-s/2)$, for s the uniparental fraction for the corresponding breeding system (modified from Uyenoyama and Takebayashi 2017). These distributions derive from the MCMC analysis of Redelings *et al.* (2015) based on observations of microsatellite variation in the gynodioecious plant *Schiedea salicaria* and from 2 populations (BP and TC) of the androdioecious fish *Kryptolebias marmoratus*, without assumptions on sex ratios.

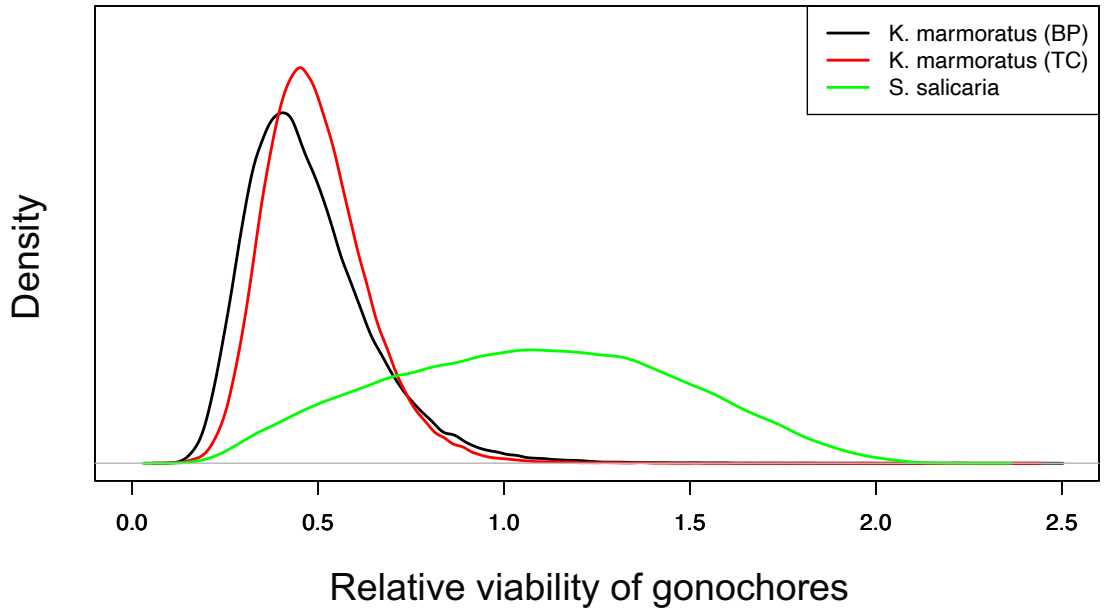

Figure 6: Posterior distributions of the viability of gonochores relative to hermaphrodites (Z) in the androdioecious fish *Kryptolebias marmoratus* and the gynodioecious plant *Schiedea salicaria* (from Uyenoyama and Takebayashi 2017).