Journal of Evolutionary Biology, 2024, **XX**, 1–16
https://doi.org/10.1093/jeb/voae104
Advance access publication 4 September 2024
Research Article

JOURNAL OF
Evolutionary Biology
eseb

F_{ST} between haploids and diploids in species with discrete ploidy phases

Kazuhiro Bessho[1], **Sarah P. Otto**[2]

[1]Medical Research Center, Saitama Medical University, 38 Morohongo Moroyama-machi, Iruma-gun, Saitama, Japan
[2]Department of Zoology, University of British Columbia, 6270 University Blvd., Vancouver, British Columbia, Canada

Handling editor: Luke Holman, Guest editor: Denis Roze

Corresponding author: Kazuhiro Bessho, Medical Research Center, Saitama Medical University, 38 Morohongo Moroyama-machi, Iruma-gun, Saitama 350-0495, Japan. Email: besshokazuhiro.reserch@gmail.com

Abstract

Many organisms alternate between distinct haploid and diploid phases, which generates population structure according to ploidy level. In this research, we consider a haploid–diploid population using statistical approaches developed for spatially subdivided populations, where haploids represent one "patch" and diploids another "patch." In species with alternating generations, sexual reproduction causes movement from diploids to haploids (by meiosis with recombination) and from haploids to diploids (by syngamy). Thus, an allele in one ploidy phase can be said to "migrate" to the other ploidy phase by sexual reproduction and to "remain" in the same ploidy phase by asexual reproduction. By analyzing a coalescent model of the probability of identity by descent and by state for a haploid–diploid system, we define F_{ST}-like measures of differentiation between haploids and diploids and show that these measures can be simplified as a function of the extent of sexuality in each ploidy phase. We conduct simulations with an infinite-alleles model and discuss a method for estimating the degree of effective sexuality from genetic data sets that use the observed F_{ST} measures of haploid–diploid species.

Keywords: haploid–diploid life cycle, Wright–Fisher model, fixation index, F-statistics, *FST*

Introduction

Many terrestrial plants, macroalgae, and fungi alternate generations between haploid and diploid phases (Mable & Otto, 1998), which each undergo mitotic growth and can experience selection. Modeling evolution in a population of a species with a haploid–diploid life cycle generates results, e.g., for the probability of fixation of an allele and the effective population size (Bessho & Otto, 2017), that differ from classic results in either haploid- or diploid-dominated species (e.g., Crow & Kimura, 1970).

Determining how haploid–diploidy affects the ecological and evolutionary dynamics of a target species not only provides fundamental biological insights but is also important from an applied perspective, for example, in the control of invasive macroalgal species (Krueger-Hadfield, 2020). The fitness of each phase, the extent of movement between ploidy phases and over space, and the sensitivity of each ploidy phase to management interventions together determine the effectiveness of such control efforts. The demographic and genetic structure of haploid–diploid species has been the subject of various studies (e.g., Destombe et al., 1989; Sosa et al., 1998; Engel et al., 2004; Thornber & Gaines, 2004; Stoeckel et al., 2021; Bessho, 2024; see review by Krueger-Hadfield et al., 2021). Advances in molecular genetic methods have led to increasing numbers of genetic studies of haploid–diploid species (e.g., marine macroalgae, Couceiro et al., 2015; Engel et al., 2004; Guillemin et al., 2008; Heiser et al., 2023; Krueger-Hadfield et al., 2011, 2013, 2016, 2017a, 2017b; Pardo et al., 2019; Sosa et al., 1998; van der Strate et al., 2002; Williams et al., 2024).

Sosa et al. (1998) was the first to infer reproductive systems from the genetic differentiation between haploids and diploids. They studied subpopulation differentiation among haploid and diploid gene frequencies in two red algal species in the genus *Gelidium* and discussed how asexual reproduction would lead to a reduction in gene flow between different ploidy phases. They also calculated the among-population fixation index (F_{ST}) from only diploid samples. Engel et al. (2004) extended their approach to examine genetic structure in the red alga, *Gracilaria gracilis*, across multiple habitats using both haploid and diploid samples. They measured spatial structure for different ploidy phases, comparing among-population F_{ST} values when sampling haploids to that when sampling diploids. Although allele frequencies were similar in haploids and diploids, haploids exhibited greater population structure (higher F_{ST} among sites) than diploids, which they attributed to a lag between migration and local breeding. Krueger-Hadfield et al. (2011; 2013) took a similar approach to examine spatial structure at different scales in the red alga *Chondrus crispus*. From the expected heterozygosities and allele frequencies, they infer the population studied is predominantly sexual (Krueger-Hadfield et al., 2011). They also estimated F_{ST} between populations at high and low shore levels, but not between haploids and diploids (Krueger-Hadfield et al., 2013). These previous studies thus used genetic measures in haploids and diploids to provide information about the reproductive system of a species but did not explicitly connect between ploidy F_{ST} to the degree of clonality.

Received January 14, 2024; revised July 29, 2024; accepted September 3, 2024

Inspired by these previous studies, Stoeckel et al. (2021) and Krueger-Hadfield et al. (2021) focused on the amount of genetic differentiation between ploidy phases using F_{ST} from samples of haploids and diploids from the same habitat. Stoeckel et al. (2021) numerically simulated a haploid–diploid Wright–Fisher model with different population sizes, ploidy ratios, frequencies of asexual reproduction, and mutation rates to examine the effects of these parameters on various genetic statistics, including F_{ST}. Krueger-Hadfield et al. (2021) tested these predictions by re-analyzing existing data from 14 species of moss and algae where genetic data from haploids and diploids had been collected, allowing F_{ST} between ploidy levels to be estimated. In their data analysis, "average F_{ST} among loci for all haploid and diploid subpopulations sampled at the same site" was used to assess the degree of genetic differentiation between gametophytes and sporophytes, exhibiting low genetic differentiation in many samples ($F_{ST} \sim 0.04$). Nevertheless, it remains unclear how to model F_{ST} in a haploid–diploid population such that it is appropriately connected to parameters of interest (e.g., population sizes and rates of asexual reproduction).

In this study, we address this gap by analyzing a haploid–diploid population as a class-structured population. Most populations are heterogeneous, structured by age, size, space, etc., which has important effects on the ecological and evolutionary dynamics of a species. When this structure can be described by discrete categories (e.g., haploids and diploids), they are best modeled as "class-structured" populations. Analytical models of this population structure can clarify the interplay between parameters (e.g., fertility, survival, and modes of reproduction in each phase) and help interpret empirical studies.

Here, we focus on the analogy between a haploid–diploid population and a spatially structured (subdivided) population to evaluate the genetic structure of a haploid–diploid population. Specifically, the evolution of a haploid–diploid species can be analyzed as a two-patch system, with sexual reproduction allowing migration to the other ploidy "patch" and offspring produced by asexual reproduction remaining in the parental patch (Bessho & Otto, 2017, 2022). This fact suggests that F_{ST} between haploid and diploid subpopulations may provide useful information to estimate the reproductive system in a haploid–diploid population. This approach has been explored in previous empirical studies (Engel et al., 2004; Krueger-Hadfield et al., 2011, 2013, 2021). However, an analytical prediction for the F_{ST} between ploidy phases is currently lacking. In this article, we set out to develop and analyze a coalescent model for a haploid–diploid system and define a fixation index (F_{ST}-like) measure of differentiation between haploids and diploids. We note that similar approaches have been developed recently to explore genetic differentiation between males and females in diploid systems (Cheng & Kirkpatrick, 2016; Kirkpatrick & Guerrero, 2014). In addition, we conducted simulations with an infinite-alleles model and proposed a method for estimating life history parameters that use the observed measures in a haploid–diploid species.

Models

Backward migration rates in a haploid–diploid system

To define F_{ST}-like indices, we first developed a coalescent model of a haploid–diploid system. To track the changes in allele frequencies of the system, we assumed that the life cycle can be described by a generalized Wright–Fisher model with free-living haploid and diploid individuals (Bessho & Otto, 2017, 2022). Hughes and Otto (1999) found that haploid–diploid life cycles were evolutionarily stable when haploids and diploids experience more competition within a ploidy level than between, consistent with the conditions for coexistence in the Lotka–Volterra model of competition. Thus, this result suggests that haploid–diploidy is expected when the ploidy phases differ in ecological niche (e.g., having different habitat requirements). Although few studies have empirically demonstrated this (Krueger-Hadfield, 2020), many ecological, biochemical, and morphological differences in gametophyte and sporophyte phenotypes have been reported (Thornber, 2006). These differences can, in turn, cause differences in density across different habitats. Here, we assume that the population sizes of haploids and diploids are independently regulated in a density-dependent manner (local regulation model in Bessho & Otto, 2022). The total numbers of haploid and diploid individuals are designated N_H and N_D, respectively.

Our haploid–diploid model also allows both haploid and diploid phases to reproduce asexually (e.g., vegetative reproduction, apomixis, and parthenogenesis), as well as sexually. Because the life history parameters can differ between haploids and diploids, we define the fertility w_p and the degree of asexuality a_p of haploids and diploids separately; here, the subscript p indicates the ploidy level ($p = H$ for haploid and D for diploid). For simplicity, we consider only the case of a monoicous species, where each gametophyte invests equal resources in male and female gametes. We also assume that male gametes are not limiting and that haploids and diploids can be morphologically distinguished (despite the fact that gametophytic selfing generates diploids that are fully homozygous and hard to distinguish genetically from haploids). The diagram of the haploid–diploid life cycle is represented in Figure 1.

At every time step t, both haploid and diploid individuals produce offspring and then die. In this reproductive process, all diploid individuals produce $(1 - a_D) w_D$ haploid spores and $a_D w_D$ diploid propagules via asexual reproduction, and all haploid individuals produce $[(1 - a_H) w_H]/2$ female gametes (the 1/2 accounting for equal investment in male gametes) and $a_H w_H$ haploid cells for asexual propagation. During syngamy, the female gametes are fertilized by male gametes at rate f and become diploid cells (zygotes), so that $f/2$ represents the reduction in fitness due to the costs of sex. The individuals of the next generation are randomly sampled from these reproductive cells (Wright–Fisher process).

Because $a_H w_H N_H$ haploid cells are produced by asexual haploid reproduction and $(1 - a_D) w_D N_D$ haploid cells are produced by diploids through meiosis, the probability that a haploid individual at the next time step $t + 1$ is the offspring of a diploid at the current time step t is

$$m_{H \leftarrow D} = \frac{(1 - a_D) w_D \rho_D}{a_H w_H \rho_H + (1 - a_D) w_D \rho_D} \tag{1a}$$

where $\rho_H = N_H/(N_H + N_D)$ is the frequency of haploids and $\rho_D = N_D/(N_H + N_D)$ is the frequency of diploids. Similarly, because $(1 - a_H)(f w_H N_H/2)$ diploid cells are produced by haploids through sexual reproduction and $a_D w_D N_D$ diploid cells are produced by diploids, the probability that a diploid individual at the next time step $t + 1$ is the offspring of a haploid at the current time step t is

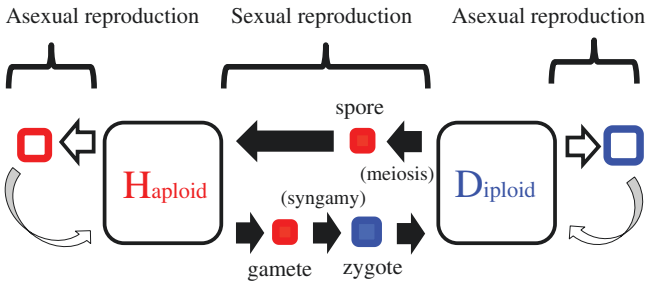

Figure 1. Diagram of the haploid–diploid life cycle we assume in this article. Free-living haploid and diploid populations are illustrated in black squares. The nuclear phase is described on the left for haploids and on the right for diploids. Sexual reproduction (solid arrows) allows the movement between haploids and diploids by meiosis with recombination and between diploids and haploids by syngamy. Haploids (diploids) may also asexually reproduce haploid (diploid) offspring (hollow arrows).

$$m_{D\leftarrow H} = \frac{(1-a_H)\,(fw_H\rho_H/2)}{(1-a_H)\,(fw_H\rho_H/2) + a_D w_D \rho_D} \tag{1b}$$

Thus, the parameters $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$ measure the probability that an allele moves between different ploidy phases. Because we are considering coalescent processes in a haploid–diploid population, these represent backward movement rates ($0 < m_{H\leftarrow D} < 1$ and $0 < m_{D\leftarrow H} < 1$).

F_{ST}-like measure in a haploid–diploid population

The fixation index, F_{ST}, is a commonly used index for evaluating the degree of genetic differentiation in a structured population (Nei & Chesser, 1983). Building upon his description of the inbreeding coefficient, Wright (1943) defined F_{ST} (= $\mathrm{Var}[p]/[p(1-p)]$, where p indicates the allele frequency), as a measure of genetic differentiation among spatial patches and showed that this measure describes the loss of heterozygosity due to the population structure. The expected value of F_{ST} has since been calculated explicitly for different models. In the classic island model with a homogeneous network of migration (at rate m) among an infinite number of diploid demes, each of size N_{deme}, F_{ST} comparing genetic diversity among populations to that within populations is approximately $1/(1+4mN_{deme})$, when N_{deme} is large and m is small, while $F_{ST} \approx 1/(1+8mN_{deme})$ when there are only two demes (Charlesworth, 1998; Eq. 4a).

Because a haploid–diploid population can be thought of as a two-island system where haploids represent one "patch" and diploids another, we can define an F_{ST}-like measure for the observed versus the expected variance between haploid and diploid organisms. The degree of sexuality, $1-a_p$, then represents migration between ploidy patches. Hence, F_{ST} between ploidy levels within a population can provide a valuable measure of asexuality within a haploid–diploid population that can potentially be used to estimate the degree of asexuality in wild populations (Engel et al., 2004; Krueger-Hadfield & Hoban, 2016; Krueger-Hadfield et al., 2011, 2013, 2021; Stoeckel et al., 2021).

F_{ST} is normally defined only for spatial structure among fully haploid or fully diploid populations, however. Therefore, we here derive an F_{ST}-like measure for populations that alternate generations between haploids and diploids within a given area, using the approach for estimating population structure described by Rousset (2004). This method relies on the probability, $Q_{ij}(t)$, that a gene chosen at random at time t from class i is identical by descent (IBD) to a gene from class j in the same generation, with no mutational differences between them. We describe the dynamics of changes in this probability for the ploidy classes, haploid (H) or diploid (D), with backward movement rates between different ploidy levels $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$ (see Appendix A). We initially assume that the IBD in diploids (Q_{DD}) is well approximated by assuming that the two sampled alleles come from different individuals, not from within a single diploid individual. This assumption is well justified if the population size of diploids is large enough that two alleles are unlikely to be sampled from the same individual. Later, we expand our model to consider both types of diploid samples, allowing the estimation of inbreeding (F_{IS}) as well (see *Discussion* section "Future perspectives," associated Supplementary Materials, and Appendix E). Using this probability, we follow Rousset (2004) and define an F_{ST}-like measure in a haploid–diploid population as,

$$\Phi_i = \lim_{\mu\to 0} \frac{\hat{Q}_{ii} - \hat{Q}_{HD}}{1-\hat{Q}_{HD}} \tag{2}$$

where μ is the mutation rate, $\hat{Q}_{ij}$ is the equilibrium value of Q_{ij}, and classes i and j indicate H or D. The excess IBD found within a subpopulation relative to that between subpopulations (Φ_H for haploids, Φ_D for diploids) is a measure of F_{ST} (Appendix A).

In the finite island model, F_{ST} has multiple definitions. We also evaluate F_{ST} for haploid–diploids using the approach of Whitlock and Barton (1997). For large population sizes and moderate amounts of sexual reproduction, F_{ST} defined using their approach is the same as that found by Rousset's (2004) method, although differences exist when sexual reproduction is very rare. We discuss the difference between these two approaches in Appendix C.

Individual-based simulation with an infinite-alleles model

To evaluate the dynamics of the probabilities of IBD and the consistency of the analytical results with the statistical expectations, we used an infinite-alleles model to numerically simulate the coalescent process of a haploid–diploid Wright–Fisher model with local population regulation (Bessho & Otto, 2022). To distinguish between Q_{ij} calculated in the numerical simulation and that calculated using the deterministic model (i.e., the coalescent model described in the previous sections), we denote the simulated values by q_{ij}. In this simulation, as the initial condition, the locus of all individuals is first set to an identical status, and then sexual and asexual reproduction of all individuals is simulated with non-overlapping generations, allowing all loci to mutate to a new state with probability μ. Because each mutation is unique in these simulations (mutation must result in a new allele), identity in state requires IBD, so that Q_{HH}, Q_{HD}, and Q_{DD} are calculated each generation from the probability of drawing the same allele twice in the appropriate haploid and/or diploid population. Then, the status of each locus is compared among all non-overlapping combinations of individuals, and the fraction of matches is calculated. This process is repeated until the probabilities q_{ij} converge to a steady state (100,000 time steps for burn-in). Normally, F_{ST} is calculated from multiple loci sampled across the genome, rather than a single locus. Therefore, we independently simulated 10 loci and averaged the values of q_{ij}

across them (ave $[q_{ij}]$) to calculate an F_{ST}-like measure for haploid–diploid populations:

$$\phi_i = \frac{\text{ave}\,[q_{ii}] - \text{ave}\,[q_{HD}]}{1 - \text{ave}\,[q_{HD}]} \tag{3}$$

Results

F_{ST} in the general case

By tracking IBD for individuals sampled from haploid and/or diploid populations, we obtain the following F_{ST} values (Appendix A):

$$\begin{aligned}\Phi_H = \{&(1-X)^3 + N_{tot}(1-X)[(2-X)X(2\rho_D) \\ &- m_{H\leftarrow D}(\rho_H + 2\rho_D)]\}/\{(1-X)^3 - N_{tot}(1-X) \\ &[(1-X)^2(\rho_H + 2\rho_D) - m_{D\leftarrow D}\rho_H - m_{H\leftarrow H}(2\rho_D)] \\ &+ N_{tot}^2 X^2 (2-X)(2\rho_D)\rho_H\},\end{aligned} \tag{4a}$$

$$\begin{aligned}\Phi_D = \{&(1-X)^3 + N_{tot}(1-X)[(2-X)X\rho_H \\ &- m_{D\leftarrow H}(\rho_H + 2\rho_D)]\}/\{(1-X)^3 - N_{tot}(1-X) \\ &[(1-X)^2(\rho_H + 2\rho_D) - m_{D\leftarrow D}\rho_H \\ &- m_{H\leftarrow H}(2\rho_D)] + N_{tot}^2 X^2 (2-X)(2\rho_D)\rho_H\}\end{aligned} \tag{4b}$$

where $N_{tot} = N_H + N_D$ and $X = m_{H\leftarrow D} + m_{D\leftarrow H}$, a measure of the total flow between the patches. Because diploids have twice the number of chromosomes as haploids, the frequency of diploids enters the above equations in a doubled form relative to haploids ($2\rho_D$ vs ρ_H). If the population size is large and none of the movement rates is very small, we can focus on the leading-order terms in N_{tot}:

$$\Phi_H = \frac{(1-X)[(2-X)X(2\rho_D) - m_{H\leftarrow D}(\rho_H + 2\rho_D)]}{N_{tot}X^2(2-X)\rho_H(2\rho_D)}, \tag{5a}$$

$$\Phi_D = \frac{(1-X)[(2-X)X\rho_H - m_{D\leftarrow H}(\rho_H + 2\rho_D)]}{N_{tot}X^2(2-X)\rho_H(2\rho_D)}, \tag{5b}$$

Using these general equations, we next focus on special cases of interest to obtain a stronger understanding of the genetic differences that develop between haploid and diploid populations.

F_{ST} for a population with symmetrical movement ($m_{H\leftarrow D} = m_{D\leftarrow H}$) between ploidy phases

We next consider the case of equal movement rates of alleles between haploid and diploid phases ($m = m_{H\leftarrow D} = m_{D\leftarrow H}$). If the total chromosome number of haploids is equal to that of diploids ($\rho_H = 2/3$), then F_{ST} is strictly equal for haploids and diploids (see Appendix A):

$$\Phi_H = \Phi_D = \frac{(1-2m)^2}{\frac{2}{3}N_{tot} - (1-2m)^2\left(\frac{2}{3}N_{tot} - 1\right)}. \tag{6}$$

When sexual reproduction is rare (m small), Equation 6 reduces to $1/(1 + 8mN_{deme})$, where N_{deme} is the size of each deme expressed in terms of a strictly diploid population (i.e., $N_{deme} = (1/2)N_H = N_D$, which is $N_{tot}/3$ when $\rho_H = 2/3$), matching the expectation for F_{ST} in a two-patch island model with equal migration rates (Charlesworth, 1998; Equation 4a).

As shown in Figure 2, Equation 4 is a cone-shape function of the backward movement rate m. F_{ST} takes a minimum value, $\Phi_H = \Phi_D = 0$, at $m = 1/2$ and a maximum value, $\Phi_H = \Phi_D = 1$, at $m = 0$ and $m = 1$. Intuitively, the genetic departure between the haploid and diploid phases by genetic drift is maximized when the species is rarely sexual ($a_H \approx 1$ and $a_D \approx 1$); in that case, a haploid–diploid population can be approximated as an independent two-patch system. Similarly, when the species is almost exclusively sexual ($a_H \approx 0$ and $a_D \approx 0$), most alleles in a haploid–diploid population undergo a strict alternation of generations; thus, the system is similar to two separate patches (a patch for those starting as haploids and a patch for those starting as diploids, with little movement between them), and F_{ST} is again maximized ($\Phi_H = \Phi_D \approx 1$).

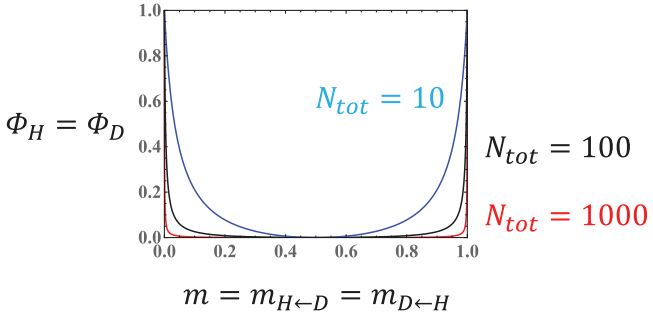

Figure 2. Changes in the F_{ST}-like measure $\Phi_H = \Phi_D$ in relation to m when the movement rates of alleles between haploid and diploid phases are symmetrical ($m = m_{H\leftarrow D} = m_{D\leftarrow H}$) and the total chromosome number in haploids is equal to that in diploids ($\rho_H = 2/3$) for a total population size ($N_{tot} = N_H + N_D$) of 1,000 (bottom red curve), 100 (middle black curve), and 10 (top blue curve).

However, when the ploidy ratio in a population is biased ($\rho_H \neq 2/3$), F_{ST} exhibits more complex behavior even when movement rates are equal (Figure 3a and b). As in the complete symmetry model (Figure 2), when the species is almost obligately sexual or obligately asexual ($a_H \approx 0$ and $a_D \approx 0$, or $a_H \approx 1$ and $a_D \approx 1$), F_{ST} is maximized ($\Phi_H = \Phi_D \approx 1$). However, with moderate values of the movement rate m, the distribution shape of F_{ST} is asymmetrical, and F_{ST} can even be negative when measured in haploids (Φ_H) or in diploids (Φ_D).

These trends can be explained as follows: (i) Because the probability of coalescence is $1/N_H$ for haploids and $1/(2N_D)$ for diploids, the coalescent event generally occurs in the ploidy phase with the smaller chromosome number. (ii) Because of Equation 2, large F_{ST} can be realized both when the probability of IBD for alleles from the same ploidy phase is large (Q_{ii} is high) and when the probability of IBD for alleles from different ploidy phases is small (Q_{HD} is low). (iii) F_{ST} can be negative when the probability of coalescence of alleles from patches with different ploidy is larger than that of alleles from the same ploidy ($Q_{ii} < Q_{HD}$).

For example, when we consider F_{ST} for the haploid class (Φ_H) in a population consisting primarily of haploid individuals ($\rho_H \gg 2/3$; red curves in Figure 3a), the greater number of chromosomes in haploids makes it more likely that coalescent events occur in diploids ($1/(2N_D)$ is high). In this case, a higher amount of sexual reproduction (higher m) generally increases Φ_H by increasing the probability of IBD of two alleles sampled from haploids (Q_{HH}) from the same diploid ancestor. Interestingly, when movement rates through sex are small, Φ_H can be negative, because the probability of IBD of two alleles sampled from haploids and diploids (Q_{HD}) can be higher than that of two alleles sampled from haploids (Q_{HH}), which mainly occurs once ancestors of both sampled lineages have moved (backwards in time) to the diploid phase, given

the smaller diploid population size. Finally, at very low movement rates, Φ_H becomes positive again because coalescence in the large haploid population (with probability $1/N_H$) is likely to occur before movement and so largely determines the probability of IBD.

In contrast, when we consider Φ_H in a diploid-dominant population ($\rho_H \ll 2/3$; blue curves in Figure 3a), F_{ST} can be negative when the backward movement rates are large (high m). In this case, the coalescent event frequently occurs in the smaller haploid population. If $m > 1/2$, however, two alleles sampled from different ploidy phases are more likely to have both descended from a parent in the smaller haploid population (with probability $m(1-m)$) than two alleles sampled from the haploid population (with probability $(1-m)^2$), so that the probability of IBD of two alleles sampled from haploids and diploids (Q_{HD}) can again be higher than that of two alleles sampled from the haploid population (Q_{HH}). When we consider F_{ST} for the diploid class (Φ_D), we get the opposite results (Figure 3b).

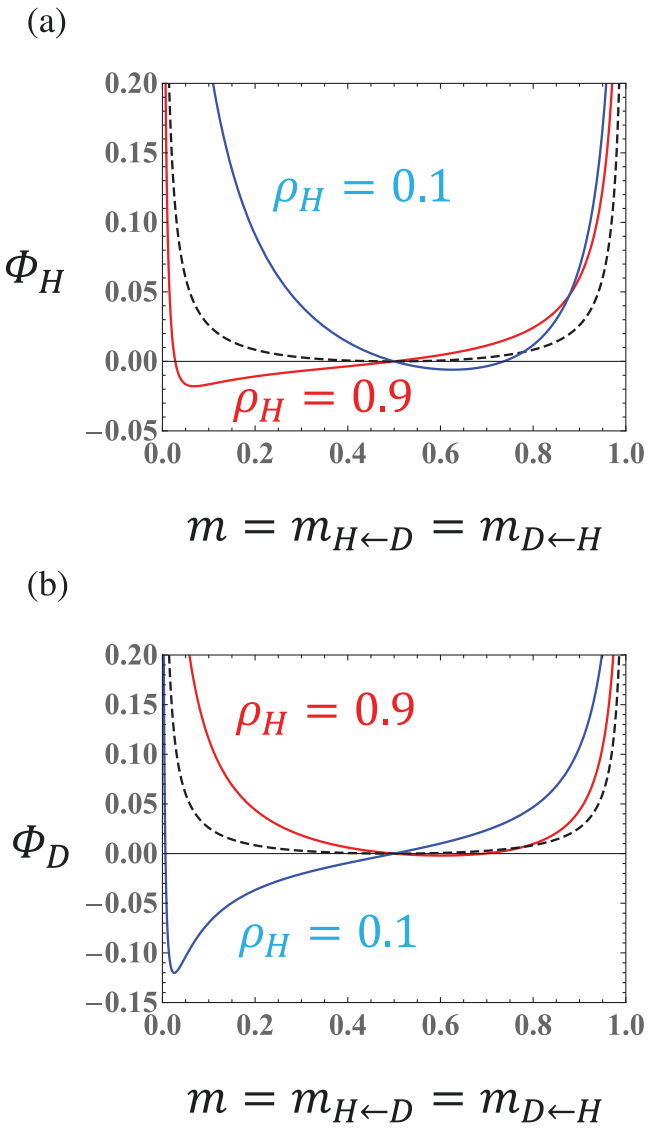

Figure 3. Changes in the F_{ST}-like measure for (a) haploids, Φ_H, and (b) diploids, Φ_D, in relation to m when the movement rates of alleles between haploid and diploid phases are symmetrical ($m = m_{H\leftarrow D} = m_{D\leftarrow H}$) but the ploidy ratio in a population is biased ($\rho_H \neq 2/3$) for a total population size of $N_{tot} = 100$. Red curves indicate a haploid-dominant population ($\rho_H = 0.9$), and blue curves indicate a diploid-dominant population ($\rho_H = 0.1$). The black dashed curves are the same as the curves in Figure 2 ($\rho_H = 2/3$) for $N_{tot} = 100$.

F_{ST} for a population with asymmetrical movement ($m_{H\leftarrow D} \neq m_{D\leftarrow H}$) between ploidy phases

We next relax the assumption of symmetrical movement. Figure 4 illustrates the F_{ST}-like measures when the backward movement rates between haploid and diploid patches are asymmetrical ($m_{H\leftarrow D} \neq m_{D\leftarrow H}$) and the number of chromosomes in haploids is equal to that in diploids ($\rho_H = 2/3$). First, we examine F_{ST} in the haploid class (Φ_H) when one movement rate is allowed to vary while the other movement rate remains fixed (Figure 4b and c). When movement from diploids to haploids is very rare ($m_{H\leftarrow D} \approx 0$), as $m_{D\leftarrow H}$ decreases, F_{ST} for haploids (Φ_H) increases (Figure 4b). Intuitively, this trend can be explained as follows: low $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$ indicate weak genetic mixing between the two patches; thus, the genetic differences between them, as measured by F_{ST}, becomes high. By contrast, when $m_{H\leftarrow D}$ is low and $m_{D\leftarrow H}$ is high, most haploid and diploid individuals descend recently from a haploid ancestor, and F_{ST} becomes low.

Moreover, when the movement rate from diploids to haploids is very high ($m_{H\leftarrow D} \approx 1$), F_{ST} for haploids (Φ_H) also increases as $m_{D\leftarrow H}$ increases (Figure 4b). This result can be explained by a similar logic as for $m_{H\leftarrow D} \approx 0$; when both $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$ are very high, the haploid–diploid population approximately consists of two independent genetic lines undergoing an alternation of generations and thus the genetic distance becomes high (large F_{ST}). If, however, $m_{H\leftarrow D}$ is high and $m_{D\leftarrow H}$ is very low, most haploid and diploid individuals descend recently from a diploid ancestor, and the genetic distance becomes low again (small F_{ST}).

When the movement from diploids to haploids ($m_{H\leftarrow D}$) is fixed (Figure 4c), F_{ST} of haploids (Φ_H) displays a peak when both movement rates are high (red curves and high $m_{D\leftarrow H}$ in Figure 4c) or low (blue curves and small $m_{D\leftarrow H}$). Intuitively, we can understand this as follows. When both $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$ are high, F_{ST} is high because of the nearly strict alternation of generations (red curves). When both $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$ are low, there is little gene flow between patches, and F_{ST} is again high, but it peaks and declines for very small $m_{D\leftarrow H}$ (blue curves). In this case, when $m_{H\leftarrow D}$ is small but $m_{D\leftarrow H}$ is even smaller (blue curve near zero in Figure 4c), the F_{ST} of haploids (Φ_H) dramatically decreases. In this case, a sample drawn from the haploid population is more likely to have descended from a diploid ancestor before coalescence (when the population size is large enough so that $m_{H\leftarrow D}N_{tot} \gg 1$), relative to the chance that a sample drawn from the diploid population descends from a haploid ancestor (because $m_{D\leftarrow H} \ll m_{H\leftarrow D}$). Thus, the IBD of two alleles sampled from different ploidy phases (Q_{HD}) is similar to that for two alleles sampled from a haploid population (Q_{HH}), and F_{ST} for haploids approaches zero. As in the case of symmetrical movement, F_{ST} in a haploid population (Φ_H) can be negative over a broad range of movement rates (Figure 4a). The results are similar for the F_{ST} of diploids (Φ_D) (Figure 4d–f). Note that Φ_H and Φ_D are never both expected to be negative.

When the number of chromosomes is different between haploids and diploids ($\rho_H \neq 2/3$) (Figure 5), the trends of F_{ST} for haploids (Φ_H) are similar to those when the number of chromosomes is similar between haploids and diploids

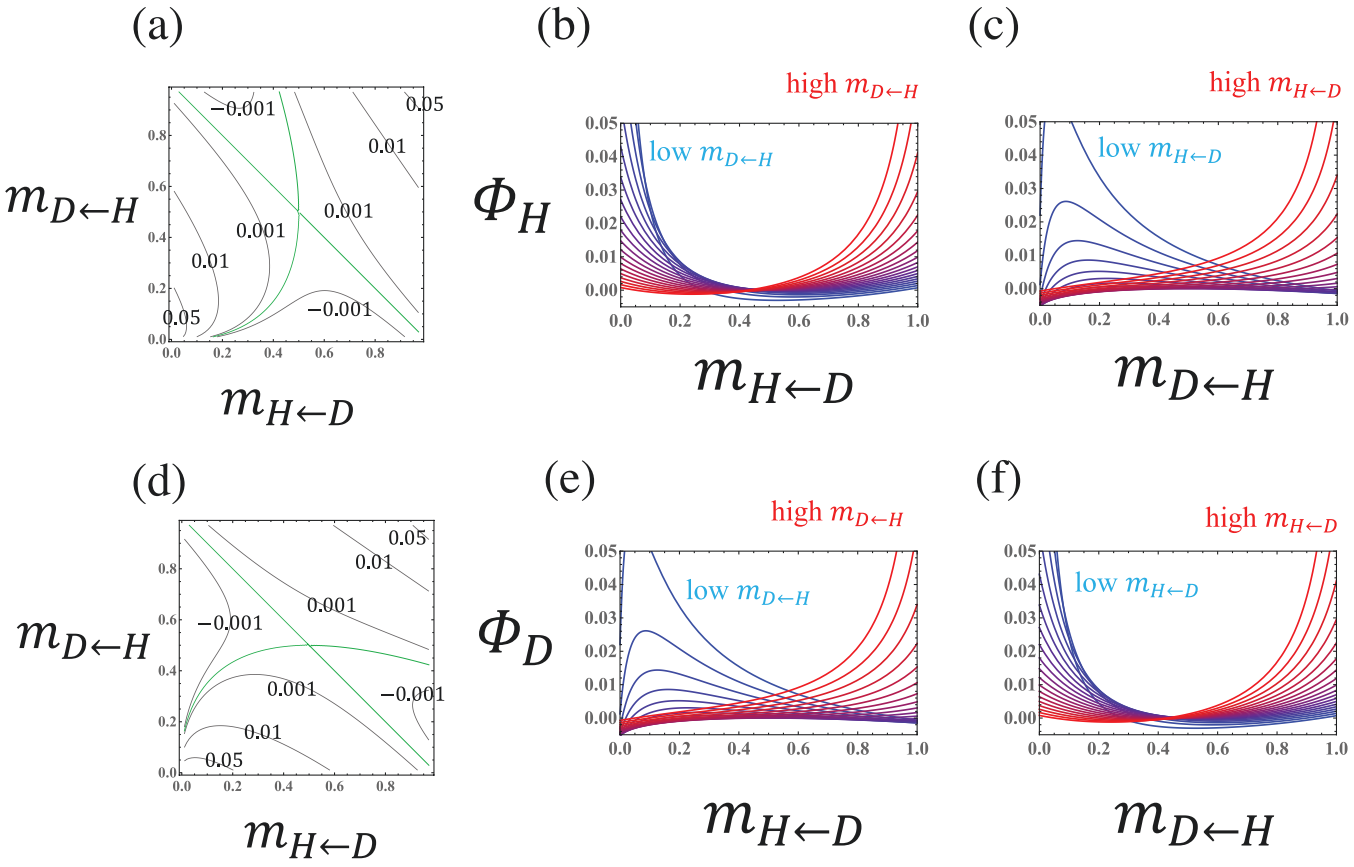

Figure 4. Changes in the F_{ST}-like measure in relation to m when the movement rates of alleles between haploid and diploid phases are asymmetrical ($m_{H\leftarrow D} \neq m_{D\leftarrow H}$) and the total chromosome number in haploids is equal to that in diploids ($\rho_H = 2/3$) for (a) haploids, Φ_H, and (d) diploids, Φ_D. The descending diagonal lines in green indicate where the F_{ST} measure, Φ_p, is zero. (b) Φ_H in relation to $m_{H\leftarrow D}$ when $m_{D\leftarrow H}$ is fixed and (c) in relation to $m_{D\leftarrow H}$ when $m_{H\leftarrow D}$ is fixed. (e) Φ_D in relation to $m_{H\leftarrow D}$ when $m_{D\leftarrow H}$ is fixed and (f) in relation to $m_{D\leftarrow H}$ when $m_{H\leftarrow D}$ is fixed. The total population size in all panels is $N_{tot} = 100$.

(Figure 4). When a population is dominated by haploids (high ρ_H) and the movement rate from haploids to diploids is low (low $m_{D\leftarrow H}$), however, F_{ST} in a haploid population (Φ_H) can be negative in a broad region (blue lines in Figure 5e). This result can be explained as follows. Because the coalescence event generally occurs in the diploid phase, with the smaller number of chromosomes, the probability of IBD for alleles from different ploidy phases (Q_{HD}) can be large when the movement rate from diploids to haploids ($m_{H\leftarrow D}$) is substantial, and F_{ST} for haploids (Φ_H) becomes negative when the probability of coalescence of alleles from different ploidy patches is larger than that of alleles from the same ploidy ($Q_{HH} < Q_{HD}$).

F_{ST} value for the entire haploid–diploid population

As defined above, F_{ST} measures the excess similarity when genes are drawn from the same ploidy state (Φ_H in haploids or Φ_D in diploids), relative to different ploidy states. We may also be interested in the average F_{ST} value for the entire metapopulation. For example, Krueger-Hadfield et al. (2021) used the "average F_{ST} among loci for all haploid and diploid subpopulations sampled at the same site" to assess the degree of genetic differentiation between gametophytes and sporophytes. Here, we calculate an overall F_{ST} by averaging Φ_H and Φ_D, weighting each by the probability of randomly drawing a chromosome from haploids or from diploids (and accounting for their being twice as many chromosomes in diploids):

$$\bar{\Phi} = \frac{\rho_H}{\rho_H + 2\rho_D}\Phi_H + \frac{2\rho_D}{\rho_H + 2\rho_D}\Phi_D. \quad (7)$$

When the number of chromosomes is equal between haploids and diploids ($\rho_H = 2/3$), $\bar{\Phi}$ becomes symmetric in the movement rates, $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$, and can be described as a function of the summed movement rates, $X = m_{H\leftarrow D} + m_{D\leftarrow H}$,

$$\bar{\Phi} = \frac{(1-X)^2}{\frac{2}{3}N_{tot} - (1-X)^2\left(\frac{2}{3}N_{tot} - 1\right)}. \quad (8)$$

This is identical to Equation 6 but with $X = m_{H\leftarrow D} + m_{D\leftarrow H}$ taking the place of $2m$ and allowing unequal movement rates between haploid and diploid populations. Figure 2 thus illustrates $\bar{\Phi}$, if we reinterpret the x-axis as the average movement rate ($X/2$). Figure 6 shows the complexity of $\bar{\Phi}$ when the number of chromosomes differs between haploids and diploids ($\rho_H \neq 2/3$).

F_{ST} in a large population of a species that rarely exhibits sexuality

Because haploid–diploid systems can easily exhibit asymmetry in both migratiohn rates and population sizes, distributions of F_{ST}-like measures for haploids and diploids are often complex. For tractability and to obtain simple equations for analyzing data sets, we here consider the special case of rare sexuality ($a_H \approx 1$ and $a_D \approx 1$) and a large population size (see

Appendix B for details), where we expect the most differentiation between haploids and diploids.

Assuming that both $m_{H\leftarrow D}N_{tot}$ and $m_{D\leftarrow H}N_{tot}$ are of order one, we obtain the following approximations:

$$\Phi_H = \{1 + (2\rho_D)\ (\theta_{HD} + \theta_{DH}) - [\rho_H\theta_{HD} - (2\rho_D)\,\theta_{DH}]\}\,/\,\{1 + 2\,(\theta_{HD} + \theta_{DH}) + 2\rho_H\,(2\rho_D)\,(\theta_{HD} + \theta_{DH})^2 - [\rho_H - (2\rho_D)]\,\theta_{DH}\} \quad (9a)$$

$$\Phi_D = \{1 + \rho_H\,(\theta_{HD} + \theta_{DH}) + [\rho_H\theta_{HD} - (2\rho_D)\,\theta_{DH}]\}/\{1 + 2\,(\theta_{HD} + \theta_{DH}) + 2\rho_H\,(2\rho_D)\,(\theta_{HD} + \theta_{DH})^2 - [\rho_H - (2\rho_D)]\,\theta_{DH}\} \quad (9b)$$

where $\theta_{HD} = m_{H\leftarrow D}N_{tot}$ and $\theta_{DH} = m_{D\leftarrow H}N_{tot}$ are population-scaled movement rates due to sexual reproduction.

Equations 9a and 9b approximate F_{ST} under the condition that most individuals reproduce asexually, leading to high F_{ST} values. By solving them for the two unknown parameters θ_{HD} and θ_{DH}, we obtain estimators for the amount of sex (see Supplementary Mathematica File),

$$\hat{\theta}_{HD} = \frac{(2\rho_D)\,(1-\phi_H)\,(\phi_H + 2\phi_D) - \rho_H\phi_H\,(1-\phi_D)}{2\rho_H\,(2\rho_D)\,(\phi_H + \phi_D)^2}, \quad (10a)$$

$$\hat{\theta}_{DH} = \frac{\rho_H\,(1-\phi_D)\,(2\phi_H + \phi_D) - (2\rho_D)\,\phi_D\,(1-\phi_H)}{2\rho_H\,(2\rho_D)\,(\phi_H + \phi_D)^2}, \quad (10b)$$

where ϕ_H and ϕ_D indicate observed F_{ST} values in genomic data sets (Equation 3). Using these equations, we can estimate the reproductive parameters (θ_{HD} and θ_{DH}) in a natural population with extensive asexuality. When sex is more frequent, we can also estimate the degree of sexuality using Equation 5, although the result is complicated (given in the Supplementary Mathematica File) and is less useful empirically because F_{ST} is then very low.

These estimators provide only the expected values, however. Because the coalescent process is probabilistic, the variance of the estimators must be evaluated. For this purpose, we conducted individual-based simulations with an infinite-alleles model.

Numerical simulation

In this section, we compare our analytical results with the results of individual-based simulations. For the simplest case, we set the chromosome numbers in haploid and diploid population to be equal ($\rho_H = 2/3$) and assume rare sexuality ($a_p \approx 1$). Supplementary Figure S1 illustrates the dynamics of the probabilities of IBD (q_{ij}), F_{st}-like measures (ϕ_p), and estimators of θ_{ij} ($\hat{\theta}_{ij}$) after burn-in for small ($N_{tot} = 900$) and large ($N_{tot} = 9,000$) populations. We note that we use the full population to calculate the probability of IBD in a simulation (full population sampling), representing the most powerful possible study. We fix the population-scaled movement rates by sexual reproduction at $\theta_{HD} = \theta_{DH} \approx 0.45$ and the

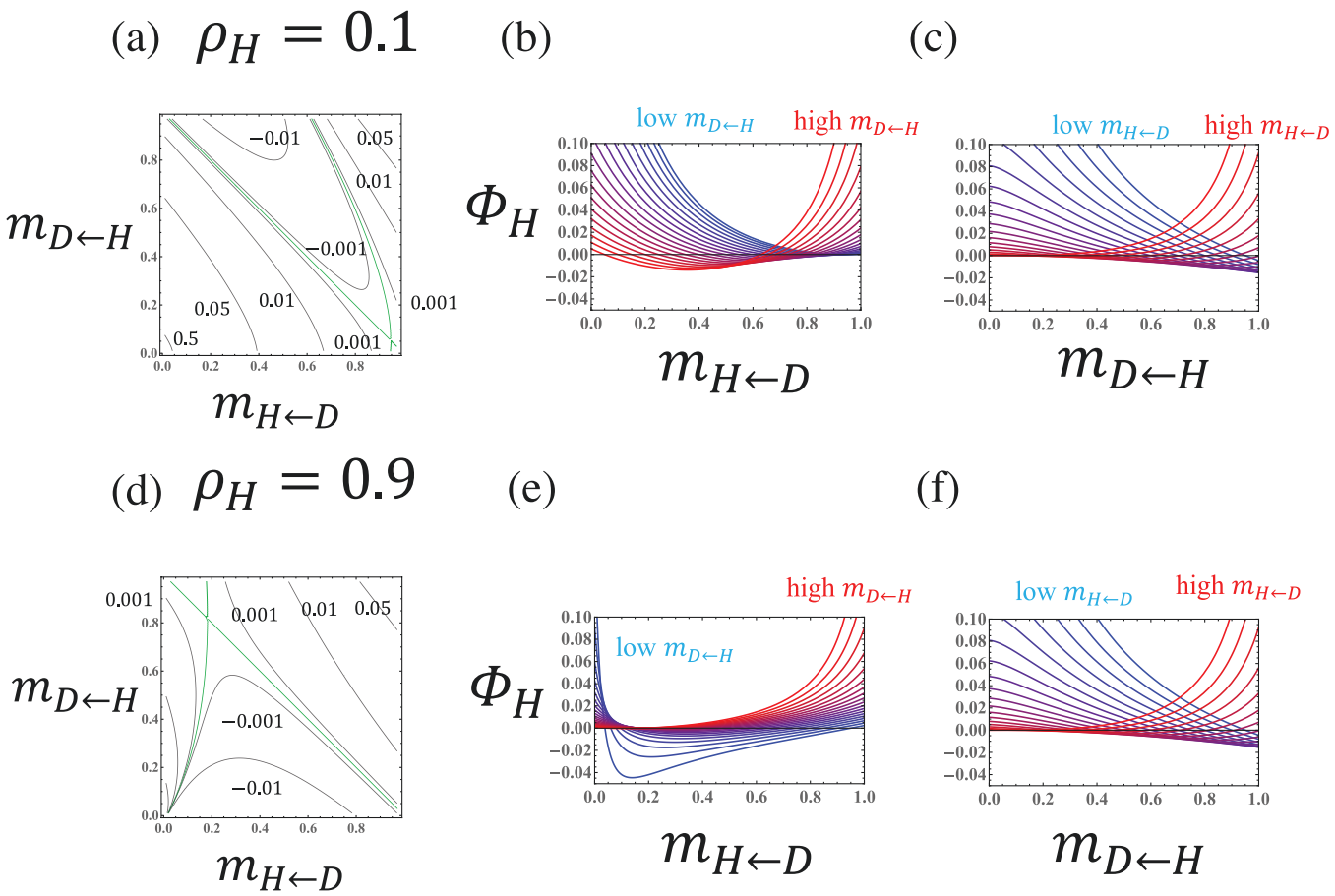

Figure 5. Changes in the F_{ST}-like measure for haploids (Φ_H) when the movement rates of alleles between the haploid and diploid phases are asymmetrical ($m_{H\leftarrow D} \neq m_{D\leftarrow H}$) and the ploidy ratio in a population is biased ($\rho_H \neq 2/3$). (a–c) Haploid-dominant population ($\rho_H = 0.9$) and (d–f) diploid-dominant population ($\rho_H = 0.1$). Other settings are the same as in Figure 4.

mutation rate at $\mu N_{tot} = 0.9$ for both large and small populations. Mutations are assumed to be unique, representing different haplotypes in one genomic region (e.g., a gene). Under these parameter settings, the relationship between haploids and diploids is symmetric (i.e., $\hat{Q}_{HH} = \hat{Q}_{DD}$, $\Phi_H = \Phi_D$, and $\theta_{HD} = \theta_{DH}$). Overall, the larger the population size, the more stable the observed dynamics tend to be (Supplementary Figure S1).

In practice, genomic sampling is likely to be performed at multiple sites across the genome; therefore, we present the distributions of various statistics for 10 sampled loci to evaluate the variability expected among samples (Figure 7). These distributions are unaffected by the population size (N_{tot}), if the population-scaled mutation and movement rates are held constant. Because the exact F_{ST} (Φ_p) is defined by assuming mutation rates are very low ($\mu \to 0$) but the simulations had appreciable mutation rates ($\mu N_{tot} = 0.9$), we find a better match to the simulations if we numerically calculate an F_{ST} -like index $\left(\left(\hat{Q}_{ii} - \hat{Q}_{HD}\right) / \left(1 - \hat{Q}_{HD}\right)\right)$ from Equation 2 but without taking the $\mu \to 0$ limit (black lines). Simulations with more biologically plausible mutation rates are expected to better match the analytical F_{ST} values given by Equation 4 (red lines) but would require more computational time to reach mutation-drift equilibrium. The distribution of the average probability of IBD (ave$[q_{ij}]$ in Figure 7a and b) and F_{ST} (ϕ_p in Figure 7c and d) across the 10 loci has a sample mean (black dots) that is close to the predictions of the deterministic model (solid lines). However, the estimated probability of IBD and F_{ST} for samples from diploids (ave$[q_{DD}]$ and ϕ_D) are slightly lower than that predicted by the deterministic model. We were able to confirm that this discrepancy in diploids is caused by ignoring sampling of two alleles from the same diploid individual when calculating IBD (i.e., the existence of F_{IS}; see details in the *Discussion* section), which has a substantial effect in populations with very rare sexual reproduction. Adjusted results, allowing for sampling within and among diploid individuals, are presented in Supplementary Figure S2.

We then explore the use of F_{ST} to estimate the population-scaled rates of sexual reproduction (θ_{HD} and θ_{DH}). The true population-scaled number of sexually produced haploids ($m_{H \leftarrow D} N_{tot}$) and diploids ($m_{D \leftarrow H} N_{tot}$) is 0.45 per generation, matching our predictions from Equation 10 (red lines in Figure 7e and f). The simulations, however, tended to overestimate the amount of sex, which again was due to the large mutation rates used in the simulations. We obtain an excellent estimate of the rate of sexuality when we use employ the F_{ST}-like index $\left(\left(\hat{Q}_{ii} - \hat{Q}_{HD}\right) / \left(1 - \hat{Q}_{HD}\right)\right)$ without taking the $\mu \to 0$ limit in Equation 10 (blue lines in Figure 7e and f). The estimator $\hat{\theta}_{DH}$ tends to be considerably larger than the true value, however, because we have ignored the chance of sampling alleles from the same diploid individual, which is substantial in the small populations sampled (corrected in Supplementary Figure S2).

Finally, we consider the average of the two effective movement rates $\left(\hat{\theta}_{HD} + \hat{\theta}_{DH}\right)/2$ from F_{ST}. Using the large population size estimator (Equation 10), we obtain:

$$\hat{\theta}_{ave} = \frac{\hat{\theta}_{HD} + \hat{\theta}_{DH}}{2} = \frac{(2\rho_D)(1-\phi_H) + \rho_H(1-\phi_D)}{4\rho_H(2\rho_D)(\phi_H + \phi_D)}. \quad (11)$$

The average is less noisy (Figure 7e and f), not only because it is the average of two estimators but also because when one movement rate is larger than expected, the other is likely to be smaller (Supplementary Figure S1e and f). Biologically, this average is also a meaningful estimator of how often a haploid–diploid organism reproduces sexually.

Discussion

Populations composed of individuals characterized by different numbers of chromosome sets, which Bessho and Otto (2022) called "ploidally structured" populations, exemplify a class-structured population and are analogous to metapopulations. Here we explore the estimation of reproductive systems using F_{ST}-like statistics based on this analogy.

The degree of genetic differentiation between haploid (gametophyte) and diploid (sporophyte) phases provides information about the mode of reproduction, as studied empirically in several macroalgal species (Engel et al., 2004; Krueger-Hadfield et al., 2011, 2013; Sosa et al., 1998). Although this relationship has been explored extensively in simulations (Stoeckel et al., 2021), theoretical models deriving the expected differentiation between ploidy phases in haploid–diploid systems had been lacking.

Thus, our study is an important step in building a theoretical foundation for studies linking genetic statistics in haploid–diploid populations to reproductive systems. We analyzed a coalescent model for a haploid–diploid population and defined an F_{ST}-like measure that allows the degree of sexuality to be estimated in haploid–diploid species. Our results thus provide a theoretical basis for interpreting genetic differences between haploids and diploids and relating these differences to the vital rates (survival and reproduction of haploids and diploids) and mode of reproduction of a species.

F_{ST}-like measure in a haploid–diploid population

The genetic structure of a subdivided population, especially genetic differentiation caused by genetic drift, is generally evaluated by using F_{ST}. The analogy between a haploid–diploid population and a subdivided population gave us the idea that it might be possible to estimate the reproductive mode (e.g., the degree of asexuality) from the degree of differentiation between haploid and diploid phases (Bessho & Otto, 2022). Our analytical results show, however, that F_{ST}-like measures must be interpreted with caution for a haploid–diploid population. Because movement and deme (local population) sizes can easily differ in a haploid–diploid population, the probability of IBD of two alleles can be different between haploid and diploid populations. Hence, we need to consider two types of F_{ST}: Φ_H comparing IBD expected among pairs of samples from haploids to pairs sampled from different ploidy phases and Φ_D comparing IBD for diploid samples compared to samples from different ploidy phases (Equations 5a and 5b). Furthermore, these coalescent-based F_{ST} indexes, derived using the method of Rousset (2004), differ from Wright's classical definition of F_{ST} $(= \mathrm{Var}[p] / [p(1-p)])$. While the latter can never be negative, we find that coalescent-based F_{ST} values are often negative in one (but not both) ploidy phase.

Intuitively, we expect a population that is primarily asexual to show more substantial genetic differentiation, resulting in high F_{ST} values, and this intuition is confirmed here. However, our analysis revealed that values of F_{ST} in a haploid–diploid population depend in a complex way on both the reproductive system (i.e., movement rates) and the population sizes of haploids and diploids (Equations 9a and 9b). A simulation study by Stoeckel et al. (2021) highlighted the effect of two parameters, the frequency of haploids and the degree of asexuality (clonality rate), on the distribution of genetic statistics,

including F_{ST} between haploids and diploids. They found a U-shaped pattern for F_{ST} as a function of the degree of asexuality (their Supplementary Figure S5), as confirmed here for sufficiently symmetric rates of clonality in haploids and diploids (our Figures 2 and 3). They also show that extreme biases in the ploidy ratio in a population increase the strength of genetic drift (lowering $\widehat{N}_e$, their Figure 3), as found theoretically by Bessho and Otto (2017). Further empirical studies are required to accumulate more data on the distribution patterns of these indicators in natural populations.

Interest in interactions among plants—their "behavioral ecology"—has been growing in recent years; in particular, a number of studies have examined the evolution of interactions between individuals, including cooperative behaviors (e.g., Yamawo & Mukai, 2017). Our study would reveal the relatedness between gametophytes and sporophytes, a quantity that strongly influences the evolution of social interactions (e.g., Trivers & Hare, 1976). Although different IBD comparisons may be appropriate for different questions, the approach developed here is readily extended to other measures (e.g., comparing individuals sampled from the same ploidy level to individuals sampled at random, rather than from different ploidy levels as done here; see Appendix C). Little research has yet been done on the evolution of interactions between gametophytes and sporophytes (but see Bessho & Sasaki, 2024); therefore, this is an important topic for future investigation.

Estimation of effective movement rates using F_{ST}

A direct application of F_{ST} is to the estimation of movement rates (degree of sexuality/asexuality) between haploid and diploid phases. We successfully derive a formula for estimating effective movement rates using F_{ST} when large populations are linked by rare sexual reproduction (Equations 6a and 6b). The degree of sexual reproduction can be quantified by applying these estimators to multiple neutral markers and then employing the mean of the IBD measures across loci. As yet, there are no examples of this approach, but future empirical studies are anticipated.

We note, however, that this new approach presents many challenges. These statistics ($\hat{\theta}_{DH}$, $\hat{\theta}_{HD}$) have the disadvantage that they are often noisy estimators, so the mean of a large number of loci should be calculated. One way to reduce noise is to focus on the average population-level movement rates (Equation 11, a measure of the degree of sexuality) rather than separately estimating rates from haploid to diploid phases and vice versa; a second way is to conduct repeated measures across multiple locations with similar biotic and abiotic attributes. Another challenge is that genetic associations will persist even for loci on different chromosomes when

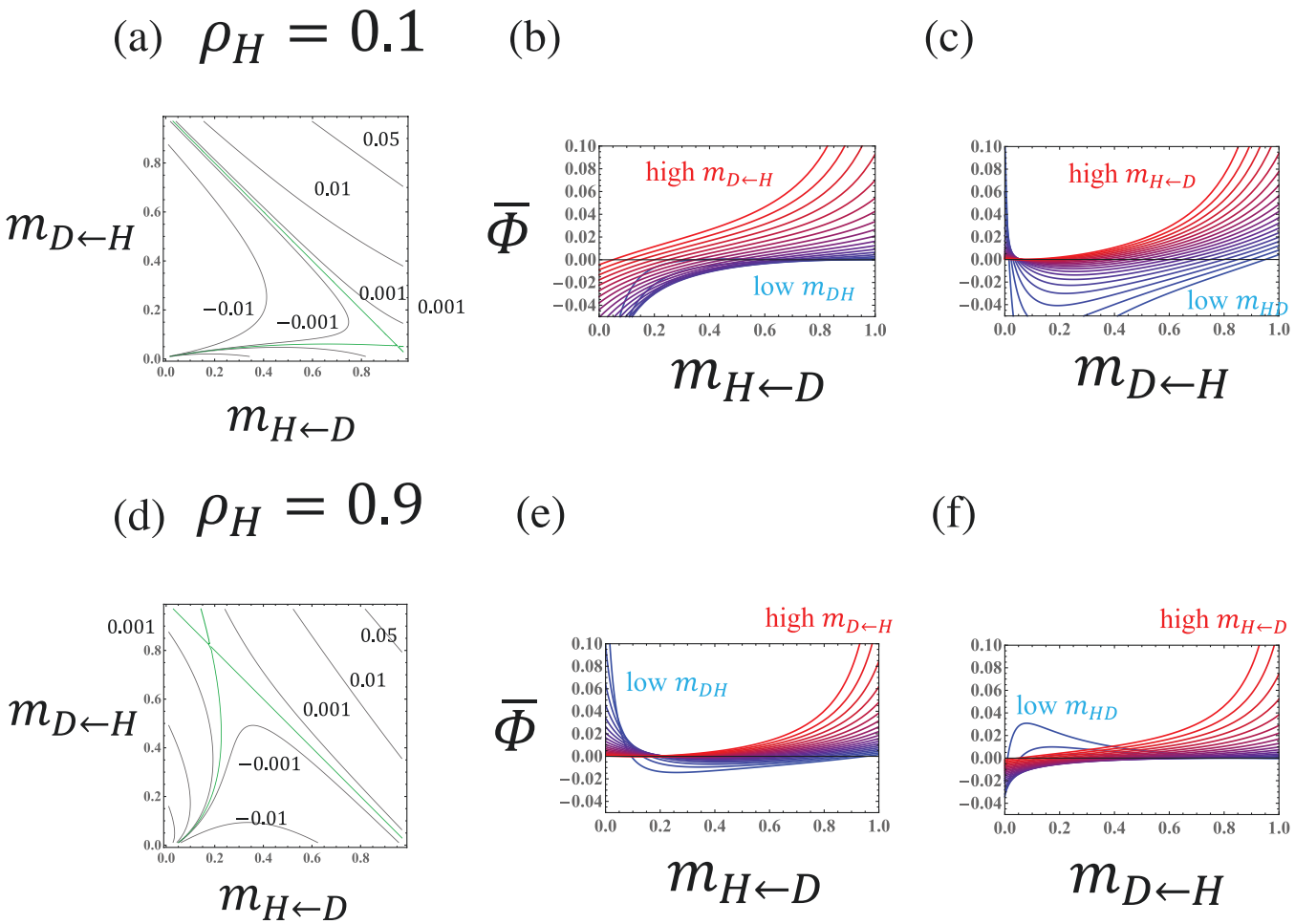

Figure 6. Changes in the average F_{ST}-like measure ($\bar{\Phi}$) when the ploidy ratio in a population is biased ($\rho_H \neq 2/3$). (a–c) Haploid-dominant population ($\rho_H = 0.9$) and (d–f) diploid-dominant population ($\rho_H = 0.1$). Other settings are the same as in Figure 4.

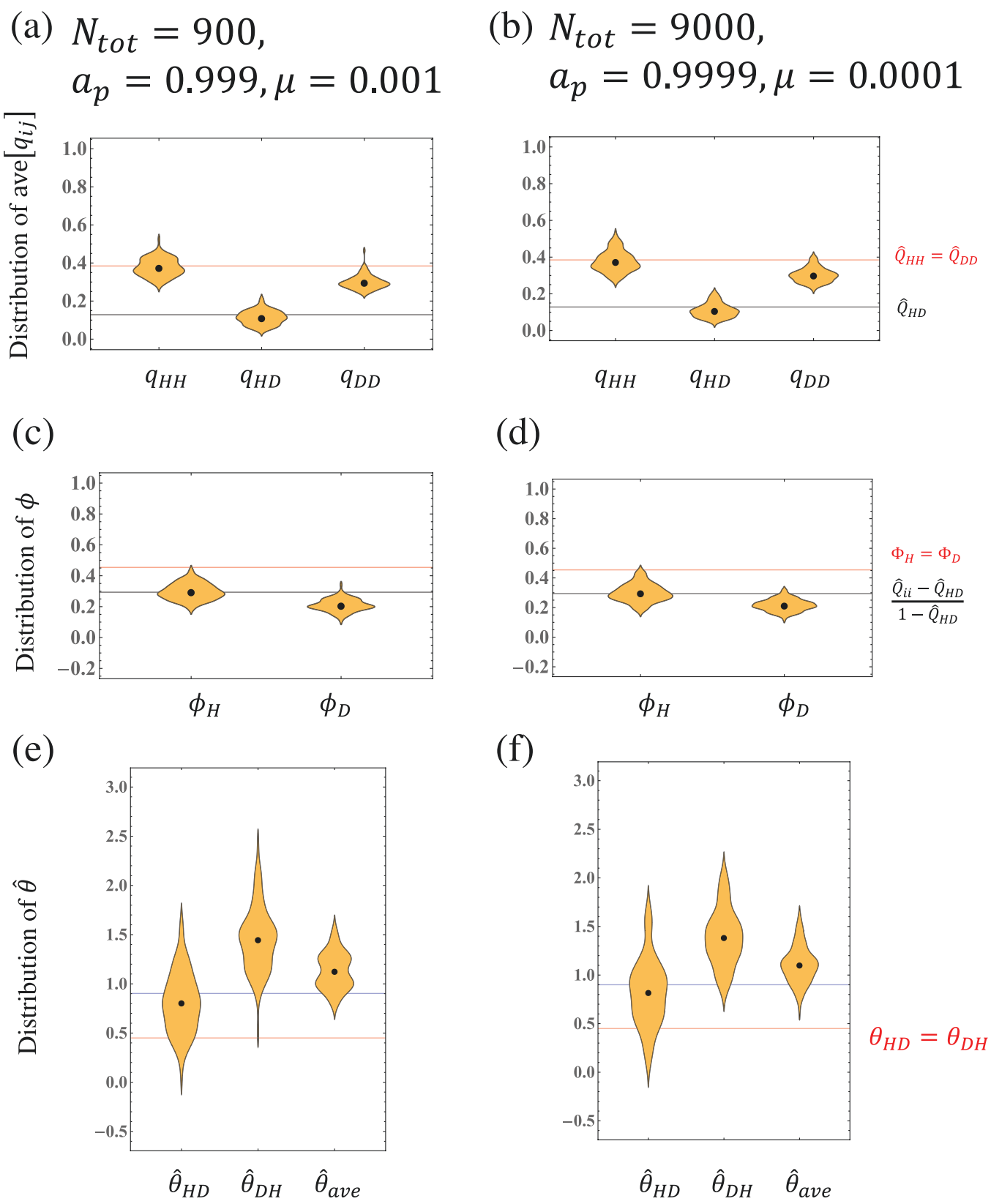

Figure 7. Distributions of 100 replicates of the (a, b) mean of IBD probabilities (avg $[q_{ij}]$), (c, d) F_{st}-like measures (ϕ_p), and (e, f) estimators $\hat{\theta}_{ij}$ and $\hat{\theta}_{ave}$ estimated from 10 genetic loci after burn-in. In (a) and (b), the top red line indicates the equilibrium of the deterministic model, $\hat{Q}_{HH} = \hat{Q}_{DD}$, and the bottom black line indicates $\hat{Q}_{HD}$. In (c) and (d), the top red line indicates the F_{st}-like measures defined by Equations 4a and 4b, $\Phi_H = \Phi_D$, and the bottom black line indicates $\left(\hat{Q}_{HH} - \hat{Q}_{HD}\right) / \left(1 - \hat{Q}_{HD}\right) = \left(\hat{Q}_{DD} - \hat{Q}_{HD}\right) / \left(1 - \hat{Q}_{HD}\right)$ with the mutation rate present in the simulations. In (e) and (f), bottom red lines indicate the true value of effective movement rate $\theta_{HD} = \theta_{DH}$, and top blue lines indicate the F_{ST}-like index $\left(\left(\hat{Q}_{ii} - \hat{Q}_{HD}\right) / \left(1 - \hat{Q}_{HD}\right)\right)$ without taking the $\mu \to 0$ limit in Equation 10. Black dots indicate the means of 100 replicates. In (a), (c), and (e), $N_{tot} = 900$, $a_H = a_D = 0.999$, and $\mu = 0.001$. In (b), (d), and (f), $N_{tot} = 9,000$, $a_H = a_D = 0.9999$, and $\mu = 0.0001$. Other parameters are $\rho_H = 2/3$, $w_H = w_D = 100$, and $f = 0.5$.

sexuality is rare, which is when our estimator is most usefully applied. When estimating the movement statistics ($\hat{\theta}_{DH}$, $\hat{\theta}_{HD}$), we assumed that the 10 sampled loci were independent (no linkage disequilibrium between loci), but simulations should be conducted with known map distances among loci to interpret empirical data estimates. Furthermore, we generally recommend that these estimators are used to predict the order of magnitude of the effective movement rate, not the precise value, given additional factors not considered here (separate sexes, selfing, spatial structure, etc.). Future work would benefit from close coordination with empirical researchers to tailor models and methods to the natural system and sampling design.

Future perspectives

This article is the first to derive F_{ST} for haploid–diploid populations and discuss its behavior and applications. Here, we discuss future directions. First, because our model conforms to the Wright–Fisher model, we do not consider overlapping generations (except to the extent that surviving parents can be counted as one of the asexual propagules competing for space in the next time step). Because in the Wright–Fisher model all individuals are replaced simultaneously, an oddity arises when the population is fully sexual (strict alternation of generations) in that the initial haploid and diploid populations are completely reproductively isolated (Appendix B). Because in reality generations may overlap, this strict alternation of genetic lines is unlikely to occur. To evaluate the genetic structure of haploid–diploid population with common sexuality, future work could develop a coalescent model based on the Moran model (Bessho & Otto, 2017).

Second, with asexual reproduction, the coalescent rates differ when both genes are sampled from the same individual versus different individuals because of their shared reproductive history, an effect that we have ignored. As shown in the Supplementary Mathematica File, however, it is possible to include sampling from the same (diploid) individual in the definition of F_{ST}. When the population size is large and sex frequent enough, this correction is negligible, so that Equations 5a and 5b are accurate. In small populations, however, the discrepancy between simulations and theory can be appreciable, especially when asexuality is very common, leading to better approximations when sampling within individuals is considered (Figure 7 and Supplementary Figure S2). Estimates of inbreeding, F_{IS}, can also be obtained by comparing samples from within versus between diploid individuals (we define it as Φ_{IS}, see Appendix E, Supplementary Mathematica File). In their simulations, Stoeckel et al. (2021) investigated the relationship between the degree of asexuality and F_{IS} (their Figure 6), finding that F_{IS} is typically near zero but becomes more variable and, on average, more negative as the frequency of asexuality approaches one. Φ_{IS} is consistent with this result but can be used to show that the expected value of F_{IS} is often positive, especially when haploidy is rare, with negative inbreeding levels of large magnitude only when the population reproduces largely clonally.

In classical population genetics, an island model with an infinitely large population and a homogeneous network of equal-sized demes has been used to investigate evolutionary dynamics. However, evolution in a metapopulation consisting of a finite population and a heterogeneous network has been less studied. As we show, understanding the population genetics of a metapopulation consisting of two demes of different sizes and with unequal gene flows between patches is complex but illuminating. The analogy with a spatially subdivided population is not only applicable to haploid–diploid species but also to other species with complex life cycles in which multiple phases have different genetic structures. For example, in polyploid species, different ploidy phases may be connected by rare sexuality. While crosses between diploid and tetraploid individuals are often thought to produce infertile triploids, increasing evidence suggests that those triploids serve as an important bridge for gene flow between diploid and tetraploid populations (Ramsey & Schemske, 1998). An example is the diploid–triploid–tetraploid system of *Carassius* fish. Mishina et al. (2021) revealed that triploids sometimes exhibit sexual reproduction with diploids and have expanded their distribution while acquiring genes. Their research suggests gene flow between diploids, triploids, and tetraploids (see Supplementary Figure S7 in their article). Normally, triploids reproduce triploid offspring asexually. However, the haploid sperm from sexual diploids sometimes fertilize an asexual triploid egg, producing tetraploid offspring. In such systems, because there is occasional movement of genes between different ploidies, we can treat it as a class-structured population. Our new definition of an F_{ST}-like parameter and the development of methods for estimating genetic structure using this F_{ST} will open new avenues of research for understanding the evolution of complex life cycles and reproductive strategies in such systems.

Supplementary material

Supplementary material is available at *Journal of Evolutionary Biology* online.

Data availability

Since this is a theoretical study, results are asserted by mathematical proof, not data. We note, however, that the Mathematica code used for the proof is available for download in the Supplementary file. The code written in C for the simulation is also available in the Supplementary file.

Author Contributions

Kazuhiro Bessho (Conceptualization [lead], Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization [equal], Writing—original draft [lead], Writing—review & editing [supporting]) and Sarah Otto (Conceptualization [supporting], Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization [equal], Writing—original draft [supporting], Writing—review & editing [lead])

Funding

We acknowledge financial support from a Grant-in-Aid from the Japan Society for the Promotion of Science to K.B. (16J05204; 19K16225; 22K06407) and NSERC Discovery Grant RGPIN-2022-03726 to S.P.O.

Acknowledgments

We are grateful for the helpful comments from Hisashi Ohtsuki and two anonymous reviewers. The first author (K.B.) used the free AI tools (DeepL, Google Translate, and Grammarly) to assist with language editing.

Conflicts of interest

None declared.

References

Bessho, K. (2024). Stable demographic ratios of haploid gametophyte to diploid sporophyte abundance in macroalgal populations. *PLoS One*, *19*(3), e0295409. https://doi.org/10.1371/journal.pone.0295409

Bessho, K., & Otto, S. P. (2017). Fixation probability in a haploid-diploid population. *Genetics*, *205*(1), 421–440. https://doi.org/10.1534/genetics.116.192856

Bessho, K., & Otto, S. P. (2022). Fixation and effective size in a haploid-diploid population with asexual reproduction. *Theoretical Population Biology*, *143*, 30–45. https://doi.org/10.1016/j.tpb.2021.11.002

Bessho, K., & Sasaki, A. (2024). Evolution of parental care in haploid–diploid plants. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *291*, 20232351. https://doi.org/10.1098/rspb.2023.2351

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*(5), 538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953

Cheng, C., & Kirkpatrick, M. (2016). Sex-specific selection and sex-biased gene expression in humans and flies. *PLoS Genetics*, *12*(9), e1006170. https://doi.org/10.1371/journal.pgen.1006170

Couceiro, L., Gac, M. L., Hunsperger, H. M., … Peters, A. F. (2015). Evolution and maintenance of haploid–diploid life cycles in natural populations: The case of the marine brown alga Ectocarpus. *Evolution*, *69*(7), 1808–1822. https://doi.org/10.1111/evo.12702

Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. The Blackburn Press.

Destombe, C., Valero, M., Vernet, P., & Couvet, D. (1989). What controls haploid–diploid ratio in the red alga, *Gracilaria verrucosa*? *Journal of Evolutionary Biology*, *2*(5), 317–338. https://doi.org/10.1046/j.1420-9101.1989.2050317.x

Engel, C. R., Destombe, C., & Valero, M. (2004). Mating system and gene flow in the red seaweed *Gracilaria gracilis*: Effect of haploid-diploid life history and intertidal rocky shore landscape on fine-scale genetic structure. *Heredity*, *92*(4), 289–298. https://doi.org/10.1038/sj.hdy.6800407

Guillemin, M. L., Faugeron, S., Destombe, C., … Valero, M. (2008). Genetic variation in wild and cultivated populations of the haploid-diploid red alga *Gracilaria chilensis*: How farming practices favor asexual reproduction and heterozygosity. *Evolution*, *62*(6), 1500–1519. https://doi.org/10.1111/j.1558-5646.2008.00373.x

Heiser, S., Amsler, C. D., Stoeckel, S., … Krueger-Hadfield, S. A. (2023). Tetrasporophytic bias coupled with heterozygote deficiency in Antarctic Plocamium sp. (Florideophyceae, Rhodophyta). *Journal of Phycology*, *59*(4), 681–697. https://doi.org/10.1111/jpy.13339

Hughes, J. S., & Otto, S. P. (1999). Ecology and the evolution of biphasic life cycles. *The American Naturalist*, *154*(3), 306–320. https://doi.org/10.1086/303241

Kirkpatrick, M., & Guerrero, R. F. (2014). Signatures of sex-antagonistic selection on recombining sex chromosomes. *Genetics*, *197*(2), 531–541. https://doi.org/10.1534/genetics.113.156026

Krueger-Hadfield, S. A. (2020). What's ploidy got to do with it? Understanding the evolutionary ecology of macroalgal invasions necessitates incorporating life cycle complexity. *Evolutionary Applications*, *13*(3), 486–499. https://doi.org/10.1111/eva.12843

Krueger-Hadfield, S. A., Collén, J., Daguin-Thiébaut, C., & Valero, M. (2011). Genetic population structure and mating system in *Chondrus crispus* (Rhodophyta). *Journal of Phycology*, *47*(3), 440–450. https://doi.org/10.1111/j.1529-8817.2011.00995.x

Krueger-Hadfield, S. A., Guillemin, M. L., Destombe, C., … Stoeckel, S. (2021). Exploring the genetic consequences of clonality in haplodiplontic taxa. *The Journal of Heredity*, *112*(1), 92–107. https://doi.org/10.1093/jhered/esaa063

Krueger-Hadfield, S. A., & Hoban, S. M. (2016). The importance of effective sampling for exploring the population dynamics of haploid–diploid seaweeds. *Journal of Phycology*, *52*(1), 1–9. https://doi.org/10.1111/jpy.12366

Krueger-Hadfield, S. A., Kollars, N. M., Byers, J. E., … Sotka, E. E. (2016). Invasion of novel habitats uncouples haplo-diplontic life cycles. *Molecular Ecology*, *25*(16), 3801–3816. https://doi.org/10.1111/mec.13718

Krueger-Hadfield, S. A., Kollars, N. M., Strand, A. E., … Sotka, E. E. (2017a). Genetic identification of source and likely vector of a widespread marine invader. *Ecology and Evolution*, *7*(12), 4432–4447. https://doi.org/10.1002/ece3.3001

Krueger-Hadfield, S. A., Magill, C. L., Bunker, F. S. P. D., … Maggs, C. A. (2017b). When invaders go unnoticed: The case of *Gracilaria vermiculophylla* in the British Isles. *Cryptogamie Algologie*, *38*(4), 379–400. https://doi.org/10.7872/crya/v38.iss4.2017.379

Krueger-Hadfield, S. A., Roze, D., Mauger, S., & Valero, M. (2013). Intergametophytic selfing and microgeographic genetic structure shape populations of the intertidal red seaweed *Chondrus crispus*. *Molecular Ecology*, *22*(12), 3242–3260. https://doi.org/10.1111/mec.12191

Mable, B. K., & Otto, S. P. (1998). The evolution of life cycles with haploid and diploid phases. *Bioessays*, *20*(6), 453–462. https://doi.org/10.1002/(sici)1521-1878(199806)20:6<453::aid-bies3>3.0.co;2-n

Mishina, T., Takeshima, H., Takada, M., … Watanabe, K. (2021). Interploidy gene flow involving the sexual-asexual cycle facilitates the diversification of gynogenetic triploid Carassius fish. *Scientific Reports*, *11*(1), 22485. https://doi.org/10.1038/s41598-021-01754-w

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, *47*(3), 253–259. https://doi.org/10.1111/j.1469-1809.1983.tb00993.x

Pardo, C., Guillemin, M., Peña, V., … Barreiro, R. (2019). Local coastal configuration rather than latitudinal gradient shape clonal diversity and genetic structure of *Phymatolithon calcareum* maerl beds in North European Atlantic. *Frontiers in Marine Science*, *6*, article 149. https://doi.org/10.3389/fmars.2019.00149

Ramsey, J., & Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, *29*(1), 467–501. https://doi.org/10.1146/annurev.ecolsys.29.1.467

Rousset, F. (2004). *Genetic structure and selection in subdivided populations*. Princeton University Press.

Sosa, P. A., Valero, M., Batista, F., & Gonzalez-Perez, M. A. (1998). Genetic structure of natural populations of Gelidium species: A re-evaluation of results. *Journal of Applied Phycology*, *10*, 279–284. https://doi.org/10.1023/A:1008092023549

Stoeckel, S., Arnaud-Haond, S., & Krueger-Hadfield, S. A. (2021). The combined effect of haplodiplonty and partial coloniality on genotypic and genetic diversity in a finite mutating population. *Journal of Heredity*, *112*(1), 78–91. https://doi.org/10.1093/jhered/esaa062

Thornber, C. S. (2006). Functional properties of the isomorphic biphasic algal life cycle. *Integrative and Comparative Biology*, *46*(5), 605–614. https://doi.org/10.1093/icb/icl018

Thornber, C. S., & Gaines, S. D. (2004). Populaiton demographics in species with biphasic life cycles. *Ecology*, *85*(6), 1661–1674. https://doi.org/10.1890/02-4101

Trivers, R. L., & Hare, H. (1976). Haploidploidy and the evolution of the social insect. *Science*, *191*(4224), 249–263. https://doi.org/10.1126/science.1108197

van der Strate, H. J., van de Zande, L., Stam, W. T., & Olsen, J. L. (2002). The contribution of haploids, diploids and clones to fine-scale population structure in the seaweed *Cladophoropsis membranacea* (Chlorophyta). *Molecular Ecology*, *11*(3), 329–345. https://doi.org/10.1046/j.1365-294x.2002.01448.x

Whitlock, M. C., & Barton, N. H. (1997). The effective size of a subdivided population. *Genetics*, *146*(1), 427–441. https://doi.org/10.1093/genetics/146.1.427

Williams, T. M., Krueger-Hadfield, S. A., Hill-Spanik, K. M., … Spalding, H. L. (2024). The reproductive system of the cryptogenic alga *Chondria tumulosa* (Florideophyceae) at Manawai, Papahānaumokuākea Marine National Monument. *Phycologia*, *63*(1), 36–44. https://doi.org/10.1080/00318884.2023.2284074

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/genetics/28.2.114

Yamawo, A., & Mukai, H. (2017). Seeds integrate biological information about consepecific and allospecific neighbours. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *284*, 20170800. https://doi.org/10.1098/rspb.2017.0800

Appendix A. F_{ST} in a haploid–diploid population

Following Rousset (2004), we describe recursion equations for the probabilities that a gene chosen at random at time t from ploidy class i is identical by descent (IBD) and alike in state to a gene from ploidy class j sampled in the same generation ($Q_{HD} = Q_{DH}$). To be IBD, the two sampled genes must have descended from the same gene without mutation in the intervening time interval. Assuming mutations occur at rate μ and that each mutation is unique, we have:

$$\begin{aligned} &Q_{HH}(t+1) = \\ &(1-\mu)^2\left\{m_{H\leftarrow H}^2\left[\frac{1}{N_H}+\left(1-\frac{1}{N_H}\right)Q_{HH}(t)\right]\right. \\ &+m_{H\leftarrow H}m_{H\leftarrow D}Q_{HD}(t)+m_{H\leftarrow D}m_{H\leftarrow H}Q_{DH}(t) \\ &\left.+m_{H\leftarrow D}^2\left[\frac{1}{2N_D}+\left(1-\frac{1}{2N_D}\right)Q_{DD}(t)\right]\right\}, \end{aligned} \tag{A.1a}$$

$$\begin{aligned} &Q_{HD}(t+1) = \\ &(1-\mu)^2\left\{m_{H\leftarrow H}m_{D\leftarrow H}\left[\frac{1}{N_H}+\left(1-\frac{1}{N_H}\right)Q_{HH}(t)\right]\right. \\ &+m_{H\leftarrow H}m_{D\leftarrow D}Q_{HD}(t)+m_{H\leftarrow D}m_{D\leftarrow H}Q_{DH}(t) \\ &\left.+m_{H\leftarrow D}m_{D\leftarrow D}\left[\frac{1}{2N_D}+\left(1-\frac{1}{2N_D}\right)Q_{DD}(t)\right]\right\}, \end{aligned} \tag{A.1b}$$

$$\begin{aligned} &Q_{DH}(t+1) = \\ &(1-\mu)^2\left\{m_{D\leftarrow H}m_{H\leftarrow H}\left[\frac{1}{N_H}+\left(1-\frac{1}{N_H}\right)Q_{HH}(t)\right]\right. \\ &+m_{D\leftarrow H}m_{H\leftarrow D}Q_{HD}(t)+m_{D\leftarrow D}m_{H\leftarrow H}Q_{DH}(t) \\ &\left.+m_{D\leftarrow D}m_{H\leftarrow D}\left[\frac{1}{2N_D}+\left(1-\frac{1}{2N_D}\right)Q_{DD}(t)\right]\right\}, \end{aligned} \tag{A.1c}$$

$$\begin{aligned} &Q_{DD}(t+1) = \\ &(1-\mu)^2\left\{m_{D\leftarrow D}^2\left[\frac{1}{2N_D}+\left(1-\frac{1}{2N_D}\right)Q_{DD}(t)\right]\right. \\ &+m_{D\leftarrow H}m_{D\leftarrow D}Q_{HD}(t)+m_{D\leftarrow D}m_{D\leftarrow H}Q_{DH}(t) \\ &\left.+m_{D\leftarrow H}^2\left[\frac{1}{N_H}+\left(1-\frac{1}{N_H}\right)Q_{HH}(t)\right]\right\}, \end{aligned} \tag{A.1d}$$

where we define the backward movement rates between different ploidy levels according to the chance that a gene that is currently haploid was produced by a diploid in the previous time step, $m_{H\leftarrow D}$, and similarly for a gene currently in a diploid having been produced by a haploid individual in the previous time step, $m_{D\leftarrow H}$ (Equations 1a and 1b). The chance of a gene remaining in the same ploidy class is $m_{H\leftarrow H} = 1 - m_{H\leftarrow D}$ and $m_{D\leftarrow D} = 1 - m_{D\leftarrow H}$. These recursion equations for Q_{ij} can be simplified by using matrix notation (Eq. (9.30) in Rousset, 2004):

$$\vec{Q}(t+1) = (1-\mu)^2\mathbf{A}\cdot\left(\vec{Q}(t)+\vec{c}\right) \tag{A.2}$$

where

$$\vec{Q} = \left(Q_{HH}, Q_{HD}, Q_{DH}, Q_{DD}\right)^T,$$

$$\vec{c} = \left(\left(1-Q_{HH}\right)/N_H, 0, 0, \left(1-Q_{DD}\right)/\left(2N_D\right)\right)^T, \text{ and}$$

$$\mathbf{A} = \begin{pmatrix} m_{H\leftarrow H}^2 & m_{H\leftarrow H}m_{H\leftarrow D} & m_{H\leftarrow D}m_{H\leftarrow H} & m_{H\leftarrow D}^2 \\ m_{H\leftarrow H}m_{D\leftarrow H} & m_{H\leftarrow H}m_{D\leftarrow D} & m_{H\leftarrow D}m_{D\leftarrow H} & m_{H\leftarrow D}m_{D\leftarrow D} \\ m_{D\leftarrow H}m_{H\leftarrow H} & m_{D\leftarrow H}m_{H\leftarrow D} & m_{D\leftarrow D}m_{H\leftarrow H} & m_{D\leftarrow D}m_{H\leftarrow D} \\ m_{D\leftarrow H}^2 & m_{D\leftarrow H}m_{D\leftarrow D} & m_{D\leftarrow D}m_{D\leftarrow H} & m_{D\leftarrow D}^2 \end{pmatrix}.$$

Furthermore, Equation A.2 can be represented as Eq. (4.6) in Rousset (2004),

$$(1-\mu)^2\vec{1} - \vec{Q}(t+1) = (1-\mu)^2\mathbf{G}\cdot\left(\vec{1}-\vec{Q}(t)\right) \tag{A.3}$$

where

$$\mathbf{G} = \begin{pmatrix} m_{H\leftarrow H}^2\left(1-\frac{1}{N_H}\right) & m_{H\leftarrow H}m_{H\leftarrow D} & m_{H\leftarrow D}m_{H\leftarrow H} & m_{H\leftarrow D}^2\left(1-\frac{1}{2N_D}\right) \\ m_{H\leftarrow H}m_{D\leftarrow H}\left(1-\frac{1}{N_H}\right) & m_{H\leftarrow H}m_{D\leftarrow D} & m_{H\leftarrow D}m_{D\leftarrow H} & m_{H\leftarrow D}m_{D\leftarrow D}\left(1-\frac{1}{2N_D}\right) \\ m_{D\leftarrow H}m_{H\leftarrow H}\left(1-\frac{1}{N_H}\right) & m_{D\leftarrow H}m_{H\leftarrow D} & m_{D\leftarrow D}m_{H\leftarrow H} & m_{D\leftarrow D}m_{H\leftarrow D}\left(1-\frac{1}{2N_D}\right) \\ m_{D\leftarrow H}^2\left(1-\frac{1}{N_H}\right) & m_{D\leftarrow H}m_{D\leftarrow D} & m_{D\leftarrow D}m_{D\leftarrow H} & m_{D\leftarrow D}^2\left(1-\frac{1}{2N_D}\right) \end{pmatrix}$$

To calculate F_{ST} in a haploid–diploid population, we substitute Equation A.2 into Equation 2 and calculate the limit as the mutation becomes increasingly rare ($\mu \to 0$) (see Supplementary Mathematica File).

From these calculations, we obtain measures of F_{ST} given by Equations 4a and 4b (see Supplementary Mathematica File). These equations satisfy a number of checks. (1) When a species exhibits complete asexuality or sexuality ($m_{H\leftarrow D} = m_{D\leftarrow H} = 0$ or $m_{H\leftarrow D} = m_{D\leftarrow H} = 1$), then $\Phi_H = \Phi_D = 1$. (2) When a haploid–diploid population is completely mixed by the movement of alleles ($m_{H\leftarrow D} = m_{D\leftarrow H} = 1/2$), we have $\Phi_H = \Phi_D = 0$. (3) More generally, when the sum of the movement rates is one ($m_{H\leftarrow D} + m_{D\leftarrow H} = 1$), then $\Phi_H = \Phi_D = 0$. (4) When one ploidy phase exhibits complete sexuality and the other ploidy phase complete asexuality ($m_{H\leftarrow D} = 0$ and $m_{D\leftarrow H} = 1$, or $m_{H\leftarrow D} = 1$ and $m_{D\leftarrow H} = 0$), then $\Phi_H = \Phi_D = 0$.

In the special case that both movement rates are the same and the number of chromosomes in haploids is equal to the number in diploids, F_{ST} in a haploid–diploid population reduces to Equation 6.

Appendix B. Approximation for F_{ST}

Although the equations, $\left(\hat{Q}_{ii} - \hat{Q}_{HD}\right)/\left(1-\hat{Q}_{HD}\right)$, can be numerically solved to give exact predictions for F_{ST}, approximations are needed to obtain simple formulae for F_{ST} that can be used to relate the frequency of asexuality within a haploid–diploid population to the degree of genetic differentiation (details are provided in the Supplementary Mathematica File). We next calculate leading-order estimates for F_{ST} (in terms of a small parameter ϵ), when sexuality is either very

rare or asexuality is very rare within a haploid–diploid population (see Supplementary Mathematica File).

(1) Sexuality rare in a large population: First, we approximate F_{ST} when sexual reproduction is very rare ($1-a_p=\mathcal{O}(\epsilon)$, so that $m_{H\leftarrow D}$ and $m_{D\leftarrow H}$ are $\mathcal{O}(\epsilon)$) and the population size is large ($N_{tot}=\mathcal{O}(1/\epsilon)$). We apply a Taylor series expansion to Equation 2 and derive Equations 9a and 9b.

(2) Asexuality rare in a large population: We next approximate F_{ST} when asexual reproduction is rare ($a_p=\mathcal{O}(\epsilon)$, so that $m_{H\leftarrow H}$ and $m_{D\leftarrow D}$ are $\mathcal{O}(\epsilon)$), and the population size is large ($N_{tot}=\mathcal{O}(1/\epsilon)$). In the Wright–Fisher model, because most alleles alternate between haploid and diploid classes (i.e., haploid–diploid Wright–Fisher model in Bessho & Otto, 2017), the difference in F_{ST} between haploid and diploid classes provides no information ($\Phi_H\approx\Phi_D$). A Taylor series expansion for F_{ST} gives:

$$\Phi_H=\Phi_D=\frac{\rho_H+(2\rho_D)}{\rho_H+(2\rho_D)+4\rho_H(2\rho_D)(\theta_{HH}+\theta_{DD})}, \quad \text{(B.1)}$$

where $\theta_{HH}=m_{H\leftarrow H}N_{tot}$ and $\theta_{DD}=m_{D\leftarrow D}N_{tot}$ indicate the effective rates of asexual reproduction. The observed value of F_{ST}, ϕ, can then be used to estimate these rates:

$$\frac{\hat{\theta}_{HH}+\hat{\theta}_{DD}}{2}=\frac{[\rho_H+(2\rho_D)](1-\phi)}{2\rho_H(2\rho_D)\phi}. \quad \text{(B.2)}$$

Appendix C. Different definitions of F_{ST}

In Appendix A, we define F_{ST} according to Rousset (2004). However, F_{ST} has various definitions. In this section, we define F_{ST} by applying the population genetics method of Whitlock and Barton (1997) and show its relationship to that obtained using Rousset's definition.

C.1. F_{ST} according to Whitlock & Barton (1997)

If the state of the gene is ignored (i.e., ignoring mutation, $\mu=0$), the probability that two genes are IBD and thus alike in state, Q_{ij}, becomes the probability that they are IBD, F_{ij}, which approaches one as time passes (Eq. (1) in Whitlock & Barton, 1997). Nevertheless, Whitlock and Barton (1997) observed that the asymptotic dynamics can be determined to find the rate at which F_{ij} approaches one and that these asymptotic dynamics can then be used to estimate F_{ST}. If we ignore changes in state by mutation, Q_{ij} becomes F_{ij}, and we have:

$$\vec{1}-\vec{F}(t+1)=\mathbf{G}\cdot\left(\vec{1}-\vec{F}(t)\right) \quad \text{(C.1)}$$

where $\vec{F}(t)$ represents $\left(F_{HH},\ F_{HD},\ F_{DH},\ F_{DD}\right)^T$ at time t (where $F_{HD}=F_{DH}$). Equation C.1 is the same as Eq. (1) of Whitlock and Barton (1997). This equation demonstrates that genetic diversity decreases, and the probability of IBD ultimately converges to one. Because Whitlock and Barton (1997) assume a fully diploid population, we can derive their parameters by replacing N_H in our model with $\tilde{N}_H=N_H/2$ (measuring the number of diploid individuals).

If we set $F_{ij}(0)=0$ as the initial condition of Equation C.1, the probability of IBD changes as follows:

$$F_{ij}(t)=1-\left[r_{ij}^{(1)}\left(\lambda^{(1)}\right)^t+r_{ij}^{(2)}\left(\lambda^{(2)}\right)^t+r_{ij}^{(3)}\left(\lambda^{(3)}\right)^t+r_{ij}^{(4)}\left(\lambda^{(4)}\right)^t\right] \quad \text{(C.2)}$$

where $\lambda^{(k)}$ and $r_{ij}^{(k)}$ are the kth eigenvalues and right eigenvectors of matrix $\mathbf{G}$, respectively (Whitlock & Barton, 1997, Eq. (A1)). Taking the limit as $t\to\infty$, we can neglect the smaller eigenvalues in Equation C.2 and approximate the dynamics according to the leading eigenvalue, whose right eigenvector is $\vec{r}^{(G)}=\left(r_{HH}^{(G)},\ r_{HD}^{(G)},\ r_{DH}^{(G)},\ r_{DD}^{(G)}\right)^T$. Whitlock and Barton (1997) compare a pair of chromosomes drawn from within a population ($F_{ii}(t)$) to a pair drawn at random across the full population ($F_{average}(t)$), taking the limit as time passes to estimate F_{ST}, a measure that we refer to as Ψ_i^{ave}.

Using Equation C.2, this approach yields:

$$\Psi_i^{ave}=\lim_{t\to\infty}\frac{F_{ii}-F_{ave}}{1-F_{ave}}=\frac{r_{ave}^{(G)}-r_{ii}^{(G)}}{r_{ave}^{(G)}} \quad \text{(C.3a)}$$

where

$$r_{ave}^{(G)}=\frac{N_H^2}{(N_H+2N_D)^2}r_{HH}^{(G)}+\frac{N_H(2N_D)}{(N_H+2N_D)^2}r_{HD}^{(G)}+\frac{(2N_D)N_H}{(N_H+2N_D)^2}r_{DH}^{(G)}+\frac{(2N_D)^2}{(N_H+2N_D)^2}r_{DD}^{(G)}$$

and

$$F_{ave}=\frac{N_H^2}{(N_H+2N_D)^2}F_{HH}+\frac{N_H(2N_D)}{(N_H+2N_D)^2}F_{HD}+\frac{(2N_D)N_H}{(N_H+2N_D)^2}F_{DH}+\frac{(2N_D)^2}{(N_H+2N_D)^2}F_{DD}$$

for a haploid–diploid population (see Eq. (5) in Whitlock & Barton, 1997 and their statement on p. 431).

Alternatively, we may define F_{ST} by comparison to the case where one chromosome is drawn from the haploid population and one from the diploid population, which is more analogous to the F_{ST} definition used by Rousset (2004; see Equation 2):

$$\Psi_i^{HD}=\lim_{t\to\infty}\frac{F_{ii}-F_{HD}}{1-F_{HD}}=\frac{r_{HD}^{(G)}-r_{ii}^{(G)}}{r_{HD}^{(G)}} \quad \text{(C.3b)}$$

Although the exact values of Ψ_i^{HD} and Ψ_i^{ave} differ from each other and from Φ_i (Equations 4a and 4b), we show in the Supplementary Mathematica File that Ψ_i^{HD} and Φ_i, but not Ψ_i^{ave}, approach the same value (Equations 5a and 5b) when the population size is very large ($N_{tot}=\mathcal{O}(1/\epsilon)$), although all three measures can differ if sexual reproduction is very rare or very common.

We next explore these measures in the symmetric case and then discuss the differences among the F_{ST} measures in the *Comparison of F_{ST} definitions* section.

C.2. Completely symmetric case

Here, we consider the special case when both movement rates and the number of chromosomes are equal in haploids and diploids ($m=m_{H\leftarrow D}=m_{D\leftarrow H}$ and $\rho_H=2/3$). In this case, we find F_{ST} in a haploid–diploid population (Supplementary Mathematica File) using the Whitlock and Barton approach to be:

$$\Psi_H^{ave}=\Psi_D^{ave}=\frac{\sqrt{9-4m(1-m)\left(9-m(1-m)(4N_{tot}-3)^2\right)}-2m(1-m)(4N_{tot}-3)}{3}, \quad \text{(C.4a)}$$

$$\Psi_H^{HD}=\Psi_D^{HD}=[3+2m(1-m)(4N_{tot}-9)-\sqrt{9-4m(1-m)\left(9-m(1-m)(4N_{tot}-3)^2\right)}]/[4m(1-m)(2N_{tot}-3)], \quad \text{(C.4b)}$$

These F_{ST} values reach a minimum of $\Psi_H=\Psi_D=0$ at m = 1/2 and a maximum of $\Psi_H=\Psi_D=1$ at m = 0 and 1 in extremely large populations.

Similar to the approximation of Φ_i in Appendix B, we obtain leading-order estimates for F_{ST} (in terms of a small parameter ϵ)

within a haploid–diploid population for the completely symmetric case. Here, we begin with the case of rare sexuality ($1-a_p=\mathcal{O}(\epsilon)$) in a large population ($N_{tot}=\mathcal{O}(1/\epsilon)$). In this case, we have

$$\Psi_H^{ave}=\Psi_D^{ave}=\frac{\sqrt{9+64m^2N_{tot}^2}-8mN_{tot}}{3}. \quad \text{(C.5a)}$$

Using $N_{deme}=N_{tot}/3$ for the deme size in diploid equivalents, we obtain $\Psi_H^{ave}=\Psi_D^{ave}=\sqrt{1+64m^2N_{deme}^2}-8mN_{deme}$. Comparing to samples drawn from different ploidy levels instead of the average, we obtain:

$$\Psi_H^{HD}=\Psi_D^{HD}=1-\frac{\sqrt{9+64m^2N_{tot}^2}-3}{8mN_{tot}}. \quad \text{(C.5b)}$$

Again using $N_{deme}=N_{tot}/3$, Equation C.5b may be written as $\Psi_H^{HD}=\Psi_D^{HD}=1-\left(\sqrt{1+64m^2N_{deme}^2}-1\right)/(8mN_{deme})$. As the level of sexual reproduction rises ($mN_{deme}>1$), $\Psi_H^{HD}=\Psi_D^{HD}$ approaches $1/(1+8mN_{deme})$, as obtained using Rousset's method (Equation 2), whereas $\Psi_H^{ave}=\Psi_D^{ave}$ approaches half this value, $1/(1+16mN_{deme})$. These results are consistent with F_{ST} reported by Charlesworth's (1998) for the two-patch island model, comparing within-deme diversity to diversity among chromosomes sampled in different patches (his Eq. 4a) or sampled across the entire population (his Eq. 4c).

By contrast, when sexual reproduction is very rare ($1-a_p=\mathcal{O}(\epsilon^2)$) and the population size is large ($N_{tot}=\mathcal{O}(1/\epsilon)$). Using $N_{deme}=N_{tot}/3$, we obtain $\Psi_H^{ave}=\Psi_D^{ave}=1-8mN_{deme}$, which now better matches that observed using the Rousset method, whereas $\Psi_H^{HD}=\Psi_D^{HD}=1-4mN_{deme}$ is larger than both other measures.

C.3. Comparison of F_{ST} definitions

Different results are obtained for F_{ST} when different definitions, Φ_i and Ψ_i, are used (e.g., Equations 9 and C.5 are different). Beyond differences in the comparison made (to a sample from different ploidy levels, HD, or to a randomly drawn sample from the entire population), these estimates of F_{ST} are also measured over different time frames. Rousset's measure reflects the more recent past and can be shown to describe the speed of coalescence for two alleles drawn from the same class i (T_{ii}) relative to different classes (Eq. (4.18) in Rousset, 2004): $\Phi_i=(T_{ii}-T_{HD})/T_{HD}$. By contrast, the method of Whitlock and Barton (1997) is asymptotic, and F_{ST} is defined in the more distant past using the leading eigenvalue of matrix $\mathbf{G}$, $\Psi_i^{ave}=\left(r_{ave}^{(G)}-r_{ii}^{(G)}\right)/r_{ave}^{(G)}$ or $\Psi_i^{HD}=\left(r_{HD}^{(G)}-r_{ii}^{(G)}\right)/r_{HD}^{(G)}$ (Equation C.3). We note that deriving F_{ST} of Rousset (2004) from a model requires all eigenvalue and eigenvectors of coalescence times, whereas only the leading eigenvalue and its right eigenvector are used by Whitlock and Barton (1997).

For moderate to large mixture rates, these approaches yield similar results ($\Psi_i^{HD}\approx\Phi_i$ given by Equations 5a and 5b) when using the same comparator (HD: sampling one haploid and one diploid chromosome), but not when using the average IBD across all pairs of samples as a comparator (Ψ_i^{ave}). By contrast, for very low rates of sex ($1-a_p=\mathcal{O}(\epsilon^2)$) in large populations ($N_{tot}=\mathcal{O}(1/\epsilon)$), comparing IBD to the average pair of samples better matches F_{ST} from Rousset's method ($\Psi_i^{ave}\approx\Phi_i$), presumably reflecting the longer time frame over which coalescent events occur (more distantly in the past with rare sex). When the population is nearly fully sexual, however, haploid and diploid phases alternate in a manner that isolates these two populations for long periods of time; in this case, we find that the asymptotic approach used by Whitlock and Barton (1997; focusing only on the leading eigenvalue) breaks down, yielding F_{ST} values outside of 0–1, whereas Rousset's method yields $F_{ST}\approx 1$, as expected for nearly isolated populations.

We note, interestingly, that different definitions give us the same effective population size, assuming a large population (Appendix D). Furthermore, this effective size is equivalent to the variance effective size in Bessho and Otto (2022).

Appendix D. Effective size of a haploid–diploid population

The effective size of a population is defined as equal to the degree of genetic drift experienced by a focal population compared to a population in the classical diploid Wright–Fisher model (Crow & Kimura, 1970). Because genetic drift affects diverse properties of a population, the effective population size can be defined by using different metrics, such as the rate at which heterozygosity is lost due to drift (inbreeding effective population size) or the rate at which allele frequencies fluctuate over time (variance effective population size). Here we show that the formulae for these two effective sizes are approximately equal to one another in a haploid–diploid Wright–Fisher model.

D.1. Asymptotic effective inbreeding size in a haploid–diploid population

We first consider the asymptotic effective inbreeding size in Rousset (2004). Because in the classical diploid Wright–Fisher model the probability of coalescence in the previous generation is $1/\left(2\tilde{N}_{tot}\right)$, Rousset (2004) defined the inbreeding effective size by the coalescence probability in a subdivided population as (Eq. (9.1) in Rousset, 2004),

$$\frac{1}{2N_e}=\lim_{\tau\to\infty}\frac{C_{ij}(\tau+1)}{1-\left(C_{ij}(1)+C_{ij}(2)+\cdots+C_{ij}(\tau)\right)}, \quad \text{(D.1)}$$

where $C_{ij}(\tau)$ indicates the probability of coalescence for two genes sampled in subpopulations i and j. In the limit of large time, this definition of effective size becomes independent of the combination of i and j, and becomes approximately $1-\lambda_G$, where λ_G is the leading eigenvalue of matrix $\mathbf{G}$ (Whitlock & Barton, 1997; p. 429; Rousset, 2004, Eq. (9.2)). Accordingly, the asymptotic inbreeding effective size becomes,

$$N_e=\frac{1}{2(1-\lambda_G)} \quad \text{(D.2)}$$

Because matrix $\mathbf{A}$ in Equation A.2 is a probability matrix, its leading eigenvalue is one, the associated right eigenvector is the unity vector, $\vec{1}$, and the associated left eigenvector is $\vec{v}^T=\left(v_H^2,\ v_Hv_D,\ v_Dv_H,\ v_D^2\right)$, where the elements of this left eigenvector are $v_H=m_{D\leftarrow H}/(m_{H\leftarrow D}+m_{D\leftarrow H})$ and $v_D=m_{H\leftarrow D}/(m_{H\leftarrow D}+m_{D\leftarrow H})$. We can thus approximate the leading eigenvalue of the matrix $\mathbf{G}$ by considering this matrix to be a perturbation of matrix $\mathbf{A}$, $\Delta\mathbf{A}=\mathbf{G}-\mathbf{A}$ (Rousset, 2004, Eq. (9.31)):

$$\lambda_G\approx 1+\frac{\vec{v}^T\,\Delta\mathbf{A}\,\vec{1}}{\vec{v}^T\vec{1}}, \quad \text{(D.3)}$$

where

$$\Delta\mathbf{A} = \begin{pmatrix} -m_{H\leftarrow H}^2/N_H & 0 & 0 & -m_{H\leftarrow D}^2/(2N_D) \\ -m_{H\leftarrow H}m_{D\leftarrow H}/N_H & 0 & 0 & -m_{H\leftarrow D}m_{D\leftarrow D}/(2N_D) \\ -m_{D\leftarrow H}m_{H\leftarrow H}/N_H & 0 & 0 & -m_{D\leftarrow D}m_{H\leftarrow D}/(2N_D) \\ -m_{D\leftarrow H}^2/N_H & 0 & 0 & -m_{D\leftarrow D}^2/(2N_D) \end{pmatrix}.$$

Assuming the population sizes are large (terms in $\Delta\mathbf{A}$ small) and substituting $\Delta\mathbf{A}$ and $\vec{v}^T$ into Equation D.3 and Equation D.2, we obtain

$$\frac{1}{2N_e} \approx \frac{v_H^2}{N_H} + \frac{v_D^2}{N_D} \tag{D.4}$$

Thus, the asymptotic inbreeding effective size from Equation D.4 is equal to the variance effective size in Bessho and Otto (2022) to this order of approximation (Supplementary Mathematica File).

D.2. Effective metapopulation size

We next consider the effective metapopulation size defined in Whitlock and Barton (1997, Eq. (9)). We note that we use $\tilde{N}_H = N_H/2$ as the local haploid population size in diploid equivalents because Whitlock and Barton (1997) assume a fully diploid population. We define a measure of the relative contribution of ploidy p to the next generation, $\left(\tilde{N}_H m_{H\leftarrow H} + N_D m_{D\leftarrow H}\right)/\tilde{N}_H$ and $\left(\tilde{N}_H m_{H\leftarrow D} + N_D m_{D\leftarrow D}\right)/N_D$.

In deriving the variance effective size (Bessho & Otto, 2022) and asymptotic effective inbreeding size (Rousset, 2004), we assumed the population size is large ($N_{tot} = \mathcal{O}(1/\epsilon)$). Making the same assumption here and plugging in Equation C.3, we find that the effective metapopulation population size given by Whitlock and Barton (1997, Eq. (9)) is also equivalent to Equation D.4 (Supplementary Mathematica File). Note that Whitlock and Barton's Eq. (9) for the effective population size is derived using F_{ST} measured relative to the average sample and so requires using Ψ_i^{ave} or adjusting their Eq. (9) for use with Ψ_i^{HD}.

Appendix E. Effective size of a haploid–diploid population

In the main manuscript (equations in Appendix A), we ignore the difference between the coalescent rates when both genes are sampled from the same individual versus different individuals in a diploid population. As shown in the Supplementary Mathematica File, however, we can identify and compute the difference between the probability that two alleles are IBD when sampled from the same diploid (Q_{DDw}) or different diploid individuals (Q_{DDb}). Using these probabilities, as in the main text, an F_{ST}-like measure can be defined for diploids as follows:

$$\Phi_{Db} = \lim_{\mu\to 0} \frac{\hat{Q}_{DDb} - \hat{Q}_{HD}}{1 - \hat{Q}_{HD}} \tag{E.1}$$

While this equation is not equal to Φ_D obtained without considering between and within diploid sampling, the results are nearly equivalent when the population size is large and sex frequent enough. In small populations, however, the discrepancy can be appreciable, especially when asexuality is very common, leading to better approximations when sampling within diploid individuals is considered (Figure 7 and Supplementary Figure S2).

Using these probabilities, we can also estimate the inbreeding level, F_{IS}, obtained by comparing samples from within versus between diploid individuals;

$$\begin{aligned} \Phi_{IS} = \lim_{\mu\to 0} \frac{\hat{Q}_{DDw} - \hat{Q}_{DDb}}{1 - \hat{Q}_{DDb}} \\ = \{ & m_{D\leftarrow D}X[m_{D\leftarrow H}(2\rho_D)\,N_{tot} - m_{D\leftarrow D}\rho_H N_{tot} \\ & + (1-X)^2(\rho_H N_{tot} - 1)]\}/\{m_{D\leftarrow D}X(1-X)^2 \\ & + m_{D\leftarrow D}X\left[m_{H\leftarrow D} + X(1-X)\right]\rho_H N_{tot} \\ & + m_{D\leftarrow H}(1-X)\left[2m_{D\leftarrow H} + X(1-X)\right](2\rho_D)\,N_{tot} \\ & + m_{D\leftarrow H}X^2(2-X)\rho_H(2\rho_D)N_{tot}^2\} \end{aligned} \tag{E.2}$$

Furthermore, when the population size is large and sex frequent enough, we have

$$\Phi_{IS} = \frac{m_{D\leftarrow D}\{-X(2-X)\rho_H + m_{D\leftarrow H}[\rho_H + (2\rho_D)]\}}{m_{D\leftarrow H}X(2-X)\,\rho_H\,(2\rho_D)\,N_{tot}} \tag{E.3}$$

For example, we consider the symmetrical case where asexuality occurs at the same rate in haploids and diploids (as in Stoeckel et al., 2021), plugging $m_{H\leftarrow D} = m_{D\leftarrow H} = 1 - c$ into Equation E.3, where c represents the amount of clonality. Φ_{IS} is thus very small when the population is nearly fully sexual ($(\rho_H + 2\rho_D)/(8\rho_H\rho_D N_{tot})$ when c is small) but tends to increase in magnitude as sex becomes rare ($c \approx 1$), approaching a value of $(2 - 5\rho_H)/[8(1-c)\,\rho_H\rho_D N_{tot}]$. Interestingly, our approximation suggests that Φ_{IS} can be negative only when $\rho_H > 2/5$. We note, however, different behavior occurs as c gets very near one and care must be taken with the orders of parameters when taking the Taylor series. Indeed, when the population size is large and clonality is very common (movement rates are $O(1/N_{tot}^2)$), Φ_{IS} becomes negative and approaches –1 (see details in Supplementary Mathematica File).