

Detecting frequency-dependent selection through the effects of genotype similarity on fitness components*

Yasuhiro Sato^{†1,2}, Yuma Takahashi³, Chongmeng Xu¹, and Kentaro K.
Shimizu^{‡1,4}

¹Department of Evolutionary Biology and Environmental Studies, University
of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

²Research Institute for Food and Agriculture, Ryukoku University, Yokotani
1-5, Seta Oe-cho, Otsu, Shiga 520-2194, Japan

³Graduate School of Science, Chiba University, Yayoi-cho 1-33, Inage-ku,
Chiba 263-8522, Japan

⁴Kihara Institute for Biological Research, Yokohama City University, Maioka
641-12, Totsuka-ward, Yokohama 244-0813, Japan

March 21, 2023

*This is the peer-reviewed version of the following article: Sato Y. et al., (2023) Detecting frequency-dependent selection through the effects of genotype similarity on fitness components. *Evolution*. <https://doi.org/10.1093/evolut/qpad028>; This post-print may be used for non-commercial purposes in accordance with Oxford University Press's Terms and Conditions for Use of Self-Archived Versions; This is not the published version. Please cite only the published version.

[†]Co-correspondence: yasuhiro.sato@uzh.ch (email); +41 44 635 4970 (phone)

[‡]Co-correspondence: kentaro.shimizu@uzh.ch (email); +41 44 635 6740 (phone)

Abstract

Frequency-dependent selection (FDS) is an evolutionary regime that can maintain or reduce polymorphisms. Despite the increasing availability of polymorphism data, few effective methods are available for estimating the gradient of FDS from the observed fitness components. We modeled the effects of genotype similarity on individual fitness to develop a selection gradient analysis of FDS. This modeling enabled us to estimate FDS by regressing fitness components on the genotype similarity among individuals. We detected known negative FDS on the visible polymorphism in a wild *Arabidopsis* and damselfly by applying this analysis to single-locus data. Further, we simulated genome-wide polymorphisms and fitness components to modify the single-locus analysis as a genome-wide association study (GWAS). The simulation showed that negative or positive FDS could be distinguished through the estimated effects of genotype similarity on simulated fitness. Moreover, we conducted the GWAS of the reproductive branch number in *Arabidopsis thaliana* and found that negative FDS was enriched among the top-associated polymorphisms of FDS. These results showed the potential applicability of the proposed method for FDS on both visible polymorphism and genome-wide polymorphisms. Overall, our study provides an effective method for selection gradient analysis to understand the maintenance or loss of polymorphism.

Running title: Detecting frequency-dependent selection

Keywords: Frequency-dependent selection, Genome-wide association study, Selection gradient analysis

1 Introduction

Wild organisms possess naturally occurring polymorphisms in their genomes. Polymorphisms, i.e., multiple alleles at a single locus, can be maintained by several forms of balancing selection such as through negative frequency-dependent selection (FDS), overdominance, and spatiotemporal variation in selection pressure (HEDRICK, 2007). Among these mechanisms, negative FDS can maintain multiple alleles at a locus by favoring rare alleles over common alleles; this regime has been reported to maintain polymorphisms in many traits such as coloration (GIGORD *et al.*, 2001; TAKAHASHI *et al.*, 2010; LE ROUZIC *et al.*, 2015; NOSIL *et al.*, 2018), self-incompatibility (LLAURENS *et al.*, 2008; JOLY and SCHOEN, 2011; SHIMIZU and TSUCHIMATSU, 2015), and resistance to natural enemies (ANTONOVICS and ELLSTRAND, 1984; BRUNET and MUNDT, 2000; SATO and KUDOH, 2017). Conversely, positive FDS unbalances polymorphisms by favoring common alleles over rare alleles (BORER *et al.*, 2010; GARRIDO *et al.*, 2016). These actions of FDS can modulate population-level processes such that the population-level mean fitness may increase or decrease as a consequence of negative or positive FDS, respectively (TAKAHASHI *et al.*, 2014; SVENSSON and CONNALLON, 2019).

Thus far, many researchers have attempted to quantify a gradient of natural selection by regressing fitness components on the traits of interests (e.g., LANDE and ARNOLD, 1983; HEISLER and DAMUTH, 1987; MITCHELL-OLDS and SHAW, 1987; CHONG *et al.*, 2018). This framework of selection gradient analysis (LANDE and ARNOLD, 1983) is also applicable to FDS. Several studies included population-level genotype frequencies as a covariate in the selection gradient analysis to reveal how observed fitness components depended on genotype frequencies (MCCAULEY and BROCK, 1998; CALSBEEK *et al.*, 2010; SATO and KUDOH, 2017). However, this approach is not technically efficient because we need to separately analyze the fitness-frequency relationship for each genotype. If the entire dataset is separated into subset data of each genotype, this reduces the total sample size and thus decrease statistical power to detect FDS. Therefore, a simpler regression model is required to quantify FDS in the framework of selection gradient analysis.

Besides its conventional usage, selection gradient analysis can also be applied for single nucleotide polymorphism (SNP) data for quantifying natural selection at a genome-wide scale (DURHAM *et al.*, 2014; FISHER *et al.*, 2016; EXPOSITO-ALONSO *et al.*, 2019). For example, EXPOSITO-ALONSO *et al.* (2019) quantified directional selection in the genome of *Arabidopsis thaliana* by conducting the selection gradient analysis for all SNPs. In technical contexts, the statistical genetic models of genome-wide association studies (GWASs) and quantitative trait locus (QTL) mapping are based on the linear regressions of a trait on genotype values

(BROMAN and SEN, 2009; GONDRO *et al.*, 2013). If the target trait is a direct component of fitness, the regression coefficient of each locus describes the relationship between individual's fitness and character — that is, selection gradient (LANDE and ARNOLD, 1983) — at a target locus. However, few methods, if any, can be used to detect FDS from GWAS or QTL mapping data.

Population genetics theory has long modeled genotype's fitness as a product of the frequency of a focal genotype and its encounter frequency with the other genotypes to develop a general model of FDS (WRIGHT, 1948; CLARKE and O'DONALD, 1964; SCHUTZ and USANIS, 1969; COCKERHAM *et al.*, 1972; ASMUSSEN and BASNAYAKE, 1990; TROTTER and SPENCER, 2007; SCHNEIDER, 2008). Such a population genetic model of FDS was indeed used in the empirical analyses of inter-genotype competition in plants (SCHUTZ and USA-NIS, 1969) and insects (ÁLVAREZ-CASTRO and ALVAREZ, 2005). Empirical evidence further suggests that FDS occurs not only through direct competitions but also through indirect interactions between genotypes (ANTONOVICS and ELLSTRAND, 1984; GIGORD *et al.*, 2001; TAKAHASHI *et al.*, 2010; SATO and KUDOH, 2017). For example, natural enemies mediate negative FDS on plant resistance when they attack undefended plants surrounding defended plants (ANTONOVICS and ELLSTRAND, 1984; BRUNET and MUNDT, 2000; SATO and KUDOH, 2017). In this case, FDS is expected to occur within a local plant patch or among neighboring plants (JANZEN, 1970; CONNELL, 1971; MOLOFSKY *et al.*, 2001; BROWNE and KARUBIAN, 2016, cf. Fig. 1a bottom). Further, mate choice in male insects mediates negative FDS on female color polymorphism when male insects learn about different colors of female individuals (VAN GOSSUM *et al.*, 2001; TAKAHASHI *et al.*, 2010). These individual insects are likely to move and interact within experimental cages or subpopulations under FDS (COSMIDIS *et al.*, 1999; FITZPATRICK *et al.*, 2007; TAKAHASHI *et al.*, 2014, cf. Fig. 1a top). If the population genetic model of FDS can be transformed as a general regression model, the selection gradient analysis of FDS will become feasible in diverse organisms under various spatial structures (Fig. 1a).

In this study, we propose an effective regression model that infers FDS based on observed fitness. The key idea behind our statistical modeling is the inclusion of genotype similarity as a covariate in selection gradient analyses (see methods "**Model development**"). We tested whether the proposed regression model (i) can accurately detect known FDS on a single-locus polymorphism (see the methods and results "**Single-locus examples**") and (ii) screen FDS-associated polymorphisms from genome-wide SNPs (see the methods and results "**GWAS extension**") to demonstrate the potential applicability of our methodology. For (i)

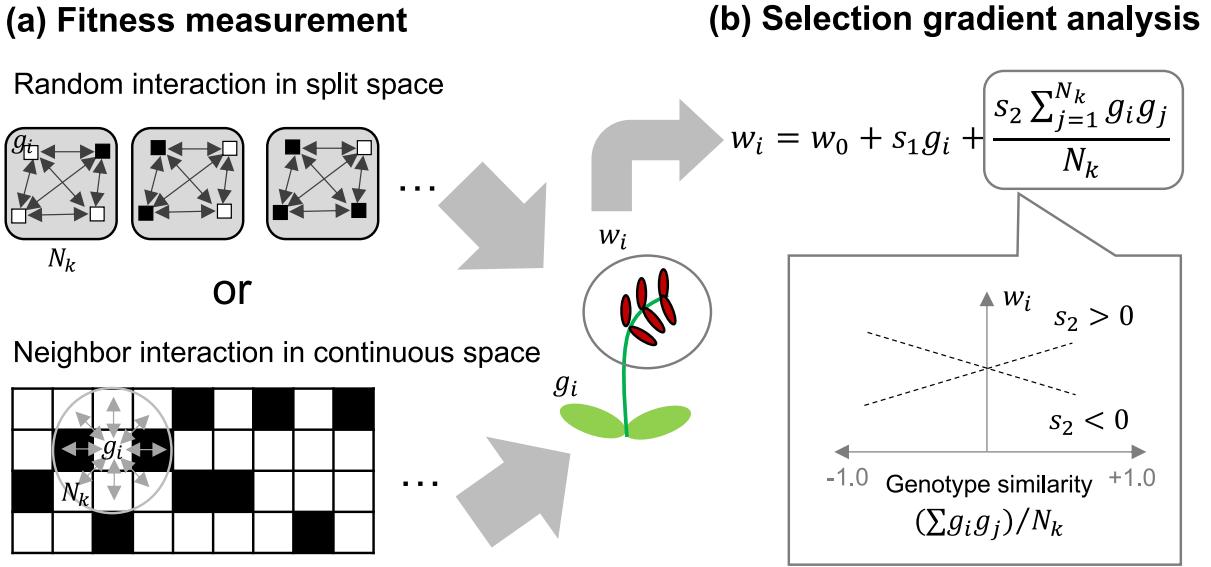


Figure 1: Workflow from the fitness measurement (a) to selection gradient analysis (b). (a) Individual fitness is observed as a consequence of interactions (two-way arrows) among individuals carrying different genotypes (black and white squares) within split subpopulations (top gray squares: cf. the case of *A. halleri* and *I. elegans* in this study) or a continuous space (bottom: cf. the case of *A. thaliana*). (b) The selection gradient is then analyzed based on regression of the observed fitness w_i on genotypes g_i . The second term in the equation is the same as that in Equation (1) and indicates how the second selection coefficient s_2 represents the effects of genotype similarity between the genotypes g_i and g_j . Summation \sum is made over N_k potentially interacting neighbors with the genotype g_j where $j = 1...N_k$.

the single-locus analyses, we reanalyzed data from previous studies on herbivore-mediated negative FDS on the trichome dimorphism of a wild herb *Arabidopsis halleri* (SATO and KUDOH, 2017) and male-mediated negative FDS on the female color polymorphism of a damselfly *Ischnura elegans* (TAKAHASHI *et al.*, 2014). For (ii) the genome-wide extension, we combined simulations and empirical *A. thaliana* data to showcase the GWAS of FDS. These empirical investigations on various systems will help illustrate the potential applicability of our regression method as a selection gradient analysis of FDS.

2 Methods

2.1 Model development

We analyzed the effects of genotype similarity on individual fitness to develop a simple regression model of FDS. We first modeled FDS based on genotype similarity among individuals (see "***FDS model based on genotype similarity***" below), and then, we redefined it as a regression model (see "***Regression model of FDS***" below). We finally analyzed fitness functions under the assumption of a panmictic population to illustrate how regression coefficients infer negative or positive FDS (see "***Fitness functions given by regression coefficients***" below).

2.1.1 FDS model based on genotype similarity

We focused on the effects of genotype similarity on individual fitness to model interactions between genotypes. Let us assume that diploid organisms interact within panmictic subpopulations and produce their offspring based on the realized fitness (Fig. 1a top). We assume that a subpopulation k belongs to the meta-population K as $k \in K$, where two individuals i and j belong to subpopulation k such that $i, j \in k$. We further assume that individuals i and j have two alleles at a locus, with the ancestral allele A being dominant over the derived allele a, where genotype values are encoded as $g_i \in \{AA, Aa, aa\} = \{+1, +1, -1\}$. This dominant encoding represents well-reported cases in which FDS often acts on a trait exhibiting dimorphism under Mendelian inheritance with complete dominance (e.g., TAKAHASHI *et al.*, 2010; SATO and KUDOH, 2017). We modeled fitness w for individual i by incorporating the genotype similarity between i and j as

$$w_i = w_0 + s_1 g_i + \frac{s_2}{N_k} \sum_{j=1}^{N_k} g_i g_j \quad (1)$$

where w_0 represents the base fitness; s_1 and s_2 represent the selection coefficients for individual genotype effects and genotype similarity, respectively; and N_k represents the total number of individuals within subpopulation k . The term $\sum_{j=1}^{N_k} g_i g_j$ represents the average similarity (or difference) of the genotype composition of the subpopulation (or neighbors) from the individual i . If two individuals share the *same* genotype values, $g_i g_j = (+1) \times (+1) = (-1) \times (-1) = +1$. If two individuals have *different* genotype values, $g_i g_j = (-1) \times (+1) = (+1) \times (-1) = -1$. Thus, the within-population genotype similarity $(\sum_{j=1}^{N_k} g_i g_j)/N_k$ ranges from -1.0 (perfect dissimilarity) to +1.0 (perfect similarity; Fig. 1b)

as a mean similarity across the number of interacting individuals within the subpopulation N_k . This line of modeling has a similar structure with the Ising model of statistical physics (Appendix S1; Table S1; and Fig. S1), where positive or negative s_2 favors the same or different genotypes between individuals i and j (SATO *et al.*, 2021b). Alternatively, we can also assume an additive expression for a trait responsible for FDS as $g_i \in \{\text{AA}, \text{Aa}, \text{aa}\} = \{+1, 0, -1\}$, where the products involving heterozygotes were considered $0 \times 0 = 0$ (neither similar nor dissimilar). This additive encoding assumes the intermediate strength of FDS for an intermediate morph (Appendix S2; Fig. S2). However, most empirical studies have reported FDS between two out of multiple morphs (e.g., GIGORD *et al.*, 2001; TAKAHASHI *et al.*, 2010; LE ROUZIC *et al.*, 2015; SATO and KUDOH, 2017; NOSIL *et al.*, 2018). As the intermediate FDS on an additive trait is still uncommon, we focused on dominant encoding.

2.1.2 Regression model of FDS

We modified Equation (1) as a regression model to fit the proposed model to empirical data. We redefined the individual fitness w_i as the response variable y_i ; the genotype g_i as the explanatory variables x_i ; the base fitness w_0 as the intercept β_0 ; and the selection coefficients s_1 and s_2 as the regression coefficients β_1 and β_2 , respectively. Further, we added a residual error e_i to Equation (1) to obtain the statistical model

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{N_k} \sum_{j=1}^{N_k} x_i x_j + e_i \quad (2)$$

where Equation (2) poses a regression analysis to estimate $\hat{\beta}_1$ and $\hat{\beta}_2$ from the given y_i and x_i . According to the inference from s_2 (Appendix S1), negative or positive β_2 represents a negative or positive FDS between two alleles as analyzed below (see "**Fitness functions given by regression coefficients**"). The negative fitness effects on one allele coincide with positive effects on another allele when y_i is a relative fitness, and therefore, it would not be necessary to consider the asymmetric FDS described in the next paragraph.

When y_i is measured as an absolute fitness, the FDS may act asymmetrically between the two alleles. For example, negative FDS on relative fitness is known to exist in the Hawk–Dove game, where both hawks and doves suffer from intensified competition with hawks. Therefore, a reduction of absolute fitness is observed along increasing frequencies of hawks (TAKAHASHI *et al.*, 2018, see also Fig. S3b). We modified Equation (2) to distinguish the asymmetric and symmetric FDS from absolute fitness. The asymmetric FDS can be modeled by a multiplicative interaction term between the second and third terms of Equation (2)

(SATO *et al.*, 2021b); it is expressed as

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{N_k} \sum_{j=1}^{N_k} x_i x_j + \frac{\beta_{12} x_i}{N_k} \sum_{j=1}^{N_k} x_i x_j + e_i \quad (3)$$

where β_{12} indicates the coefficient for the asymmetric effects of genotype similarity. If β_{12} was statistically significant, the slope coefficient of the genotype similarity differed among focal genotypes (SATO *et al.*, 2021b), which means that the strength or direction of FDS is asymmetric between genotypes. A practical usage of Equation (3) is the same as the test of an interaction term in a standard regression model: We should first test β_{12} using Equation (3) and then test β_2 using the linear model Equation (2) if β_{12} is not significant.

2.1.3 Fitness functions given by regression coefficients

We analyzed Equation (3) as a function of allele frequencies to clarify how the coefficients for genotype similarity effects β_2 and asymmetric effects β_{12} corresponded to FDS. If we suppose that all individuals uniformly interact in a sufficiently large population with random mating (i.e., $N_k \rightarrow \infty$), the probability of one genotype interacting with the other genotypes depends on genotype frequencies derived from an allele frequency within a population (Appendix S2). Let f be the frequency of the A allele within the panmictic infinite population. The ratio of genotype frequency on panmixia was given as follows: AA: Aa: aa = $f^2 : 2f(1-f) : (1-f)^2$. Assuming the complete dominance of the A allele over the a allele ($x_i \in \{AA, Aa, aa\} = \{+1, +1, -1\}$), we then calculated all the combinations among the three genotypes (Table S2) and consequent fitness y_i for AA, Aa, and aa genotypes as

$$y_{AA} = y_{Aa} = \beta_0 + (\beta_2 + \beta_{12})f^2 + 2(\beta_2 + \beta_{12})f(1-f) - (\beta_2 + \beta_{12})(1-f)^2 \quad (4a)$$

$$y_{aa} = \beta_0 - (\beta_2 - \beta_{12})f^2 - 2(\beta_2 - \beta_{12})f(1-f) + (\beta_2 - \beta_{12})(1-f)^2 \quad (4b)$$

where y_{AA} , y_{Aa} , and y_{aa} represent the fitness values for the AA, Aa, and aa genotypes, respectively (Appendix S2). The allele-level fitness was then defined by weighting the genotype fitness with the allele frequency as follows:

$$y_A = f y_{AA} + (1-f) y_{Aa} \quad (5a)$$

$$y_a = f y_{Aa} + (1-f) y_{aa} \quad (5b)$$

In Equations (4) and (5), the sign of the symmetric effects β_2 represents negative or positive FDS on relative fitness between two alleles, while the asymmetric effects β_{12} determine the slope of absolute fitness along the allele frequency f . Figure 2 shows how the fitness values Equations (5a) and (5b) vary in response to f . Symmetric negative FDS was exemplified by the negative β_2 without any asymmetric effects β_{12} (i.e., $\beta_{12} = 0$; Fig. 2a). Asymmetric negative FDS was described by the negative (Fig. 2b) or positive (Fig. 2c) asymmetric effect β_{12} , where the negative β_2 denoted negative FDS on the relative fitness between two alleles (Fig. 2b and c). In contrast, symmetric positive FDS was exemplified by the positive β_2 with no (Fig. 2d), negative (Fig. 2e), or positive (Fig. 2f) values of the asymmetric effect β_{12} . Thus, when we aim to detect FDS on relative fitness, we can focus on only the symmetric effect β_2 .

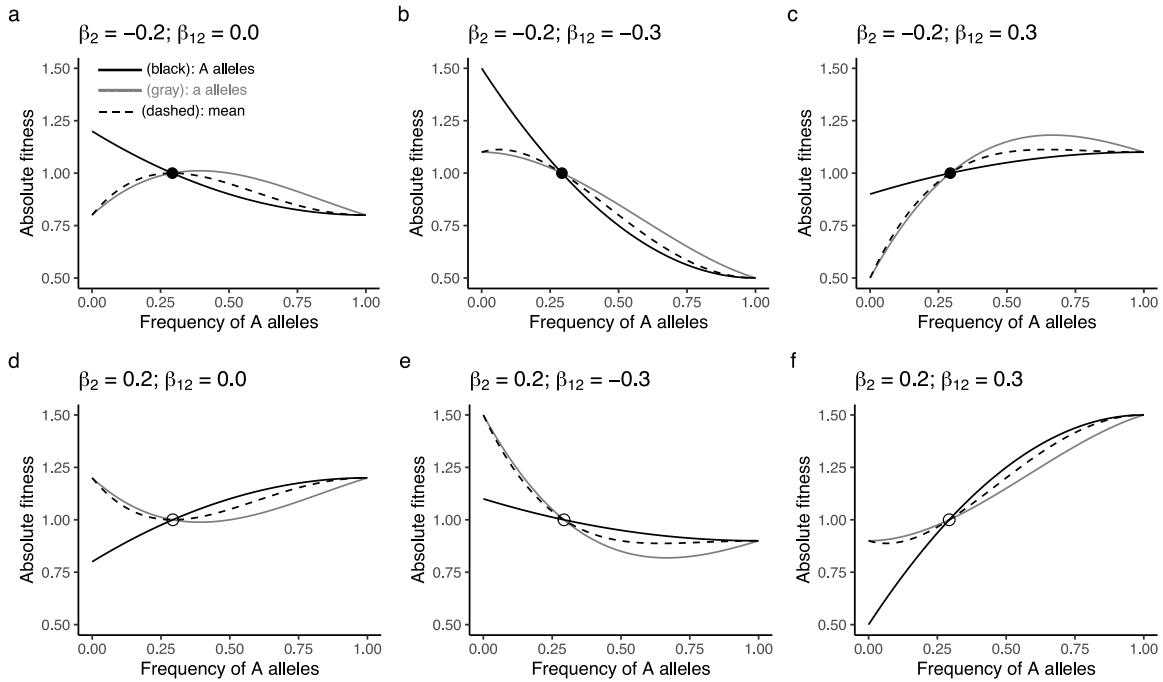


Figure 2: Numerical examples for the fitness values y_i in response to allele frequency when the A allele is completely dominant over the a allele. The black and gray lines indicate the allele-level fitness of A and a alleles; that is, Equations (5a) and (5b), respectively. Dashed curves indicate the population-level mean fitness Equation (6). (a) Symmetric negative FDS, (d) symmetric positive FDS, (b and c) asymmetric negative FDS, and (e and f) asymmetric positive FDS. Closed and open circles indicate stable or unstable states, respectively. No directional selection i.e., $\beta_0 = 1.0$ and $\beta_1 = 0$ are assumed for all panels to present the fitness functions of FDS.

Population-level mean fitness under FDS is another side of interests when FDS concerns absolute fitness (COCKERHAM *et al.*, 1972; SCHNEIDER, 2008; TAKAHASHI *et al.*, 2018). In a panmictic population, such a population-level mean fitness is given by the allele-level fitness weighted by its allele frequency (Appendix S2); that is, the weighted mean is given as follows:

$$\bar{y} = fy_A + (1 - f)y_a = f^2y_{AA} + 2f(1 - f)y_{Aa} + (1 - f)^2y_{aa} \quad (6)$$

This population-level mean fitness is maximized at an intermediate frequency under symmetric negative FDS (Fig. 2a) (SCHNEIDER, 2008). Even under an asymmetric negative FDS, the population-level mean fitness has a concave function against f (Fig. 2b and c), which indicates a larger population-level fitness than that expected by the frequency-independent selection (TAKAHASHI *et al.*, 2018). In contrast, the population-level mean fitness is minimized at an intermediate frequency under symmetric positive FDS (Fig. 2d). Under the asymmetric positive FDS, the population-level mean fitness has a convex function (Fig. 2e and f), and thus, it becomes less than expected by the frequency-independent selection (SCHNEIDER, 2008; TAKAHASHI *et al.*, 2018). When A and a alleles show additive expression as $x_i \in \{AA, Aa, aa\} = \{+1, 0, -1\}$, the fitness functions become so complicated that more than two equilibria may arise (Appendix S2; Fig. S2); however, they have rarely been reported empirically. Therefore, we assumed complete dominance for data analyses.

2.2 Single-locus examples

We determined whether the proposed regression models Equations (2) and (3) could detect negative FDS on single-locus dimorphism of hairy and glabrous phenotypes in a wild herb *Arabidopsis halleri* (see "**Flower production of hairy and glabrous plants**" below); and on single-locus polymorphism of female colors in a damselfly *Ischnura elegans* (see "**Egg production of andromorph and gynomorph damselflies**" below). These two systems from a plant and insect species share similar study designs. Both *A. halleri* and *I. elegans* cases involve absolute fitness values measured under the split subpopulation structure (Fig. 1a top). The fact that the trait expressions of both the trichome dimorphism and color polymorphism followed Mendelian inheritance with complete dominance (SHIMIZU, 2002; SÁNCHEZ-GUILLEN *et al.*, 2005; KAWAGOE *et al.*, 2011) allowed us to substitute phenotype frequencies for the genotype frequencies of the heterozygote and dominant homozygote. These similar study designs from two different systems provided strong evidence for the

rediscovery of negative FDS (see the results "Single-locus examples").

2.2.1 Flower production of hairy and glabrous plants

We applied single-locus analysis for the flower production data of *Arabidopsis halleri* subsp. *gemmaifera* to detect negative FDS mediated by herbivore attacks on hairy and glabrous plants (SATO and KUDOH, 2017). SATO and KUDOH (2017) recorded the trichome phenotype (hairy or glabrous), number of flowers, leaf damage score, and length of largest leaf for all individual plants after setting circular split plots (1 m in diameter) in a natural population at Japan (35°06'N, 134°56'E). Within these field plots, a flightless leaf beetle *Phaedon brassicae* naturally occurred and fed on *A. halleri* individuals. The total sample size comprised 3,070 individuals among 324 plots across four study years. Based on SATO and KUDOH (2017), we used a generalized linear mixed model (GLMM) with a Poisson error structure and a log-link function to analyze the number of flowers as a countable response variable. The fixed effects were the individual phenotype (hairy or glabrous), similarity between the two morphs, and total number of plants within each field plot. While the total number of plants was implicitly incorporated as a denominator N_k in Equations (2) and (3), this was also included as a fixed effect that represented plant density. Similar to that in SATO and KUDOH (2017), the random effects were the field plot IDs nested below the study years. The log-transformed length of the largest leaf (mm), which reflects the plant size, was included as an offset term. In *A. halleri*, hairy alleles are known to be dominant over glabrous alleles at the *GLABRA1* locus (SHIMIZU, 2002; KAWAGOE *et al.*, 2011). Given the complete dominance of hairy alleles, we assumed complete dominance at the *GLABRA1* locus with $x_i \in \{AA, Aa, aa\} = \{+1, +1, -1\}$ based on the trichome phenotype of an individual. The original data of SATO and KUDOH (2017) were downloaded from the Dryad repository (<https://doi.org/10.5061/dryad.53k2d>). The lme4 package (BATES *et al.*, 2015) in R version 4.0.3 (R CORE TEAM, 2019) was used for GLMM.

2.2.2 Egg production of andromorph and gynomorph damselflies

We applied the single-locus analysis for the egg production by the blue-tailed damselfly *Ischnura elegans* to detect negative FDS and the consequent increase in population-level mean fitness between an andromorph and a gynomorph (TAKAHASHI *et al.*, 2014). The original data were derived from TAKAHASHI *et al.* (2014) and consisted of 102 andromorphs and 79 *infuscans*-type gynomorphs. TAKAHASHI *et al.* (2014) assigned adult *I. elegans* with an andromorph frequency of 0.2, 0.5, or 0.8 into split-field cages under low- or high-density

conditions, and then counted the mature eggs. We used a GLMM with a Poisson error structure and a log-link function to analyze this countable response variable TAKAHASHI *et al.* (2014). The fixed effects were morph type and morph similarity within a cage. The cage ID and experimental ID were considered random effects, where the cage ID was nested below the experimental ID. The andromorph allele is known to be dominant over the *infuscans*-type gynomorph allele on an autosomal locus (SÁNCHEZ-GUILLÉN *et al.*, 2005). Therefore, on the basis of phenotype frequencies within the split cages, we assumed complete dominance with genotype values encoded as $x_i \in \{\text{AA}, \text{Aa}, \text{aa}\} = \{+1, +1, -1\}$ for homozygous andromorphs, heterozygous andromorphs, and homozygous gynomorphs, respectively. The interaction term between the morph type and similarity was considered in the line of GLMMs for testing the significance of asymmetric FDS between the two morphs. The coefficients of the main effect were estimated using GLMM without any interaction terms if it was not significant.

2.3 GWAS extension

We tested whether our method could aid in detecting FDS from genome-wide SNPs by extending the single-locus analysis to GWAS. We applied our regression methods for GWAS by implementing Equations (2) and (3) as linear mixed models (LMMs) that correct population genetic structure (see Appendix S3 "Mixed model extension"). To validate the performance of LMM implementation, we performed GWAS simulations under different selection regimes and subpopulation structures (see "**GWAS of simulated data**" below). We also applied the extended GWAS method to empirical data on reproductive branch number in *A. thaliana* accessions (see "**GWAS using field-grown *Arabidopsis thaliana***" below) to exemplify a GWAS-style experiment. This joint analysis using simulated and empirical data provides the first step to apply our method for a variety of GWAS data (see the results "**GWAS extension**").

2.3.1 GWAS of simulated data

We performed GWAS using simulated genomes and fitness values to test whether our regression method could distinguish FDS and other types of selection among genome-wide SNPs. In this simulation, we considered four scenarios of the selection process, which include negative FDS, positive FDS, overdominance, and spatiotemporally varying selection. Thirty independent rounds of three-step simulation ran as follows: (1) simulation of genomic structure under the four scenarios of selection, (2) virtual experiment to simulate

fitness from the simulated genomes under the split or continuous subpopulation structure, and (3) fitting regression methods to simulated fitness and genomes. First, we used SLiM version 3 (HALLER and MESSER, 2019) to simulate SNPs for 2000 individuals (see Fig. S4 for the simulated genomic structure). We simulated the four scenarios of selection together with stabilizing selection on individual fitness to assume individual fitness as a quantitative trait. The simulated SNPs were imported to R using the vcfR (KNAUS and GRÜNWALD, 2017) and the gaston package (PERDRY and DANDINE-ROULLAND, 2020). Second, we used the gaston (PERDRY and DANDINE-ROULLAND, 2020) and rNeighborGWAS (SATO *et al.*, 2021b) packages for fitness simulation. Out of the simulated individuals, 900 individuals are randomly assigned to split (Fig. 1a top) or continuous (Fig. 1a bottom) subpopulations. We simulated fitness values based simulated SNPs with the ratio of genetic variance on non-genetic variance kept at two third. Third, simulated fitness values were fitted using the LMM version of Equation (2) (Appendix S3), which estimated the genotype similarity effect $\hat{\beta}_2$ and its $-\log_{10}(p\text{-values})$ association score for all SNPs. For the fitness simulation and fitting, we focused on the symmetric FDS on relative fitness because the selection acted not on absolute but on relative fitness during the SLiM simulation above. We used the receiver operating characteristic (ROC) curve implemented in the pROC (ROBIN *et al.*, 2011) package to quantify the false vs. true positive rate of causal SNP detection by the $-\log_{10}(p\text{-values})$ association score. The model performance was evaluated by the area under the ROC curve (AUC). We performed the line of simulations using both LMs and LMMs to test the notion that linear mixed models (LMMs) usually outperform standard linear models (LMS) in GWAS (KANG *et al.*, 2008). Further, to examine the influence of heterozygosity on the model performance, we tested additive and dominant genotype encoding in the line of simulations. The details of these GWAS simulations are provided in Appendix S4.

2.3.2 GWAS using field-grown *Arabidopsis thaliana*

We conducted a pilot GWAS of the reproductive branch number in field-grown *A. thaliana* under a continuous setting to examine whether our method is applicable to the empirical GWAS dataset (Fig. 1a bottom). We first set three randomized blocks of 199 *A. thaliana* accessions in a field site at Zurich, Switzerland to observe the final reproduction. The experimental procedure followed that of SATO *et al.* (2021b), with slightly modifications for the present study (see Appendix S5 for details). From early summer (July 8, 2019), we recorded the length of the largest leaf (mm) at the beginning of the experiment, the presence of bolting after 2 weeks, and the reproductive branch number at the end of the

experiment (August 27, 2019). Further, we quantified a reproductive fitness component by the number of reproductive branches that had at least one trace of flowers because it was difficult to assess the exact number of flowers for diverse accessions under the harsh summer environment. Dead plants were recorded as a branch number of zero, *i.e.*, with no fecundity. The number of reproductive branches was $\log(x+1)$ -transformed to improve normality, and it was analyzed as a target trait of GWAS. With the cut-off threshold of minor allele frequency at >0.05 , we obtained 1,819,577 SNPs from genotype data registered in the AraGWAS catalog (TOGNINALLI *et al.*, 2018). The inbred lines of *A. thaliana* have either the AA or aa genotype, in which the qualitative interpretation of β_2 and β_{12} in this inbred case remains the same as in the case of complete dominance (Appendix S2; Fig. S3). We assumed that FDS arose from genetic interactions among neighboring plants in small *Arabidopsis*, and therefore, the genotype similarity was considered up to the nearest neighbors. The initial plant size, presence of bolting, experimental block ID, and edge of each plot (or not) were considered as nongenetic covariates. The rNeighborGWAS package version 1.2.4 (SATO *et al.*, 2021b) was used to implement Equations (2) and (3) as GWAS (Appendix S3). After association mapping, we focused on SNPs with a $-\log_{10}(p\text{-value})$ score > 4.0 . We searched candidate genes within ~ 10 kb around the target SNPs based on the Araport11 gene model with the annotation of The Arabidopsis Information Resource (TAIR; accessed on 31 December 2021). Accession names and phenotype data are presented in Table S3. The details of the field GWAS experiment are provided in Appendix S5.

3 Results

3.1 Single-locus examples

3.1.1 Negative FDS on trichome dimorphism in a wild *Arabidopsis*

We applied Poisson GLMM for the flower production data on hairy and glabrous plants of *A. halleri* within split field plots (see the methods "***Flower production of hairy and glabrous plants***" above) to test whether the single-locus analysis could detect the known negative FDS. The Poisson GLMM detected a negative and significant coefficient of the morph similarity $\hat{\beta}_2$ (Table 1a), which indicates a negative FDS where a focal plant produced more flowers as dissimilar morphs were grown within the same plot. This negative FDS was symmetric between hairy and glabrous plants, as indicated by the lack of significance of the interaction term between the trichomes and morph similarity (Table 1a). Further, we found

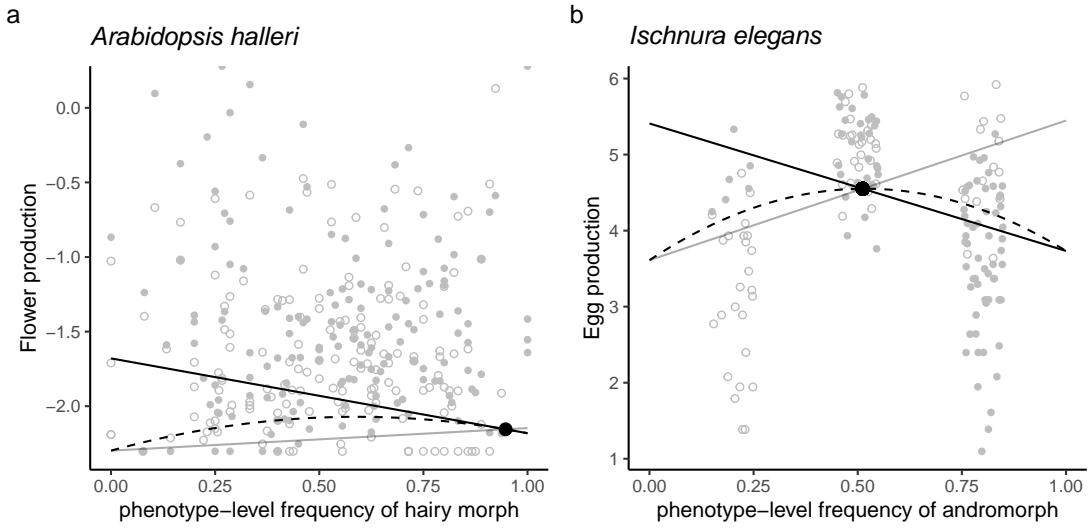


Figure 3: Negative frequency-dependent selection on (a) the flower production of hairy and glabrous plants in *A. halleri* and (b) the egg production of andromorph and gynomorph females in *I. elegans*. Estimated fitness function based on the fitted model (Table 1) are shown at a phenotype level (see case 3 in Appendix S2). Gray circles and the solid black line indicate observed and predicted fitness of dominant morphs (i.e., hairy and andromorph), respectively, while the white open circles and solid gray line indicate those of recessive morphs (glabrous and gynomorph), respectively. The dashed curve presents the population-level mean fitness based on the fitted model (Table 1), with a black dot indicating the stable equilibrium under negative FDS. A single circle in panel (a) corresponds to a field plot, where the flower production represents the plot-level average of $\log[(\text{no. of flowers} / \text{plant size (mm)}) + 0.1]$. A single circle in panel (b) corresponds to an individual morph, where the egg production represents $\log(\text{no. of mature eggs})$.

the same level of the individual morph coefficient $\hat{\beta}_1$ as the negative coefficient of morph similarity $\hat{\beta}_2$ (Table 1a); this indicates the simultaneous action of directional selection and negative FDS between the two morphs. The flower production was not dependent on the plant density as indicated by no significant effects from the total number of plants (Table 1a). The estimated fitness function based on the fitted model (Table 1) show that a stable equilibrium under negative FDS remains at a biased but not monomorphic frequency (Fig. 3a). This discrepancy between the observed frequency and expected equilibrium was likely because the present selection analysis could not incorporate another major component of fitness in *A. halleri*; i.e., clonal reproduction (SATO and KUDOH, 2017). These results provide qualitative evidence for negative FDS on trichome dimorphism through the fitness component of flower production.

Table 1: Poisson generalized linear mixed model (GLMM) applied to the number of flowers between hairy and glabrous *A. halleri* (a) or the number of mature eggs between the andromorph and gynomorph of *I. elegans* (b). Morph similarity was calculated from the genotype similarity defined by Equation (2). Estimated coefficients, their standard errors (SE), Z-values, and *p*-values are shown for multiple regressions. Bold letters indicate significance at *p* < 0.05 by Wald tests.

(a) *Arabidopsis halleri*

Fixed effects	Coefficient	SE	Z	<i>p</i>
Intercept $\hat{\beta}_0$	-2.076	0.17	-11.97	<2e-16
Individual morph $\hat{\beta}_1$	0.146	0.008	17.32	<2e-16
Morph similarity $\hat{\beta}_2$	-0.163	0.018	-9.22	<2e-16
Total no. of plants	-0.006	0.010	-0.63	0.53
Individual \times Similarity $\hat{\beta}_{12}$	-0.088	0.048	-1.81	0.07

(b) *Ischnura elegans*

Fixed effects	Coefficient	SE	Z	<i>p</i>
Intercept $\hat{\beta}_0$	4.55	0.191	23.8	<2e-16
Individual morph $\hat{\beta}_1$	-0.021	0.008	-2.56	0.01
Morph similarity $\hat{\beta}_2$	-0.878	0.022	-39.30	<2e-16
Density	0.097	0.237	0.41	0.68
Individual \times Similarity $\hat{\beta}_{12}$	-0.042	0.25	-0.17	0.87

3.1.2 Negative FDS on female color polymorphisms in a damselfly

To test whether the single-locus analysis could detect a known relationship between negative FDS and population-level mean fitness, we applied Poisson GLMM for the data on the number of mature eggs between the andromorph and gynomorph of *I. elegans* within split field cages (see the methods "**Egg production of andromorph and gynomorph damselflies**" above). Consistent with the previous evidence for negative FDS (VAN GOSSUM *et al.*, 2001; LE ROUZIC *et al.*, 2015), we found a significantly negative coefficient of morph similarity $\hat{\beta}_2$ (Table 1b). The interaction term between the morph type and similarity was not significant (Table 1b), indicating no significant asymmetry in the negative FDS between the two morphs. We also found a significant effect of the morph type (andromorph or gynomorph) on the egg production, but its effect was much less significant than that of morph similarity (Table 1b). As reported in a previous study (TAKAHASHI *et al.*, 2014), the density did not significantly affect the egg production (Table 1b). Estimated fitness function based on the fitted model (Table 1) shows that the negative FDS enabled the coexistence of the two morphs at an intermediate frequency (Fig. 3b). Consistent with the results of

TAKAHASHI *et al.* (2014), the population-level mean fitness increased at a stable equilibrium at the intermediate frequency under the negative FDS (Fig. 3b). These results support the action of symmetric negative FDS and the consequent increase in the population-level mean fitness. Taken together, *A. halleri* and *I. elegans* data suggest that our single-locus model can be applied to the selection gradient analyses of FDS (see also the discussion "**Selection gradient along genotype similarity**").

3.2 GWAS extension

3.2.1 Detection of simulated FDS across a genome

We simulated genotypes and fitness to test whether our method could distinguish negative FDS, positive FDS, overdominance, and spatiotemporally varying selection among genome-wide SNPs (see the methods "**GWAS of simulated data**" above; Appendix S4; and Figs. S5-S7). We further implemented the single-locus model Equation (2) as a linear mixed model (LMM) (Appendix S3) to correct the population genetic structure in GWAS. Assuming the complete dominance of A over a alleles, we first evaluated the performance of these LMMs in terms of causal polymorphism detection and effect size estimates (Fig. 4). The performance of LMMs to detect negative and positive FDS was high in both split and continuous settings (median performance > 0.8; Fig. 4b and e). The direction of the FDS matches the estimated sign of β_2 (Fig. 4c and f). In contrast, these LMMs showed little or weak performance to detect overdominance or spatiotemporally varying selection, respectively (median performance < 0.6; Fig. 4b and e), and their estimated coefficients were almost zero (Fig. 4c and f). Further, we tested another general assumption of GWAS, i.e., the additive effects of two allele on fitness, which resulted in the moderate to high performance of LMMs in detecting negative and positive FDS (Fig. S6; Appendix S4). These results indicate that our method retains the statistical power to detect negative and positive FDS in GWAS, with other types of balancing selection less likely to be confounded.

Then, we compared the performance of LMMs with that of standard LMs to inspect whether the correction of population genetic structure improved GWAS (Fig. 4, Fig. S5; Appendix S4). LMMs better detected negative FDS than LMs in terms of the model performance (Fig. 4b and e, Fig. S5b and e). Although LMs and LMMs exhibited similar performances for positive FDS, the LMs overestimated the true strength of positive FDS more than LMMs (Fig. 4c and f, Fig. S5c and f). Even when we assumed additive fitness effects from two alleles, the LMMs better detected negative FDS and less overestimated

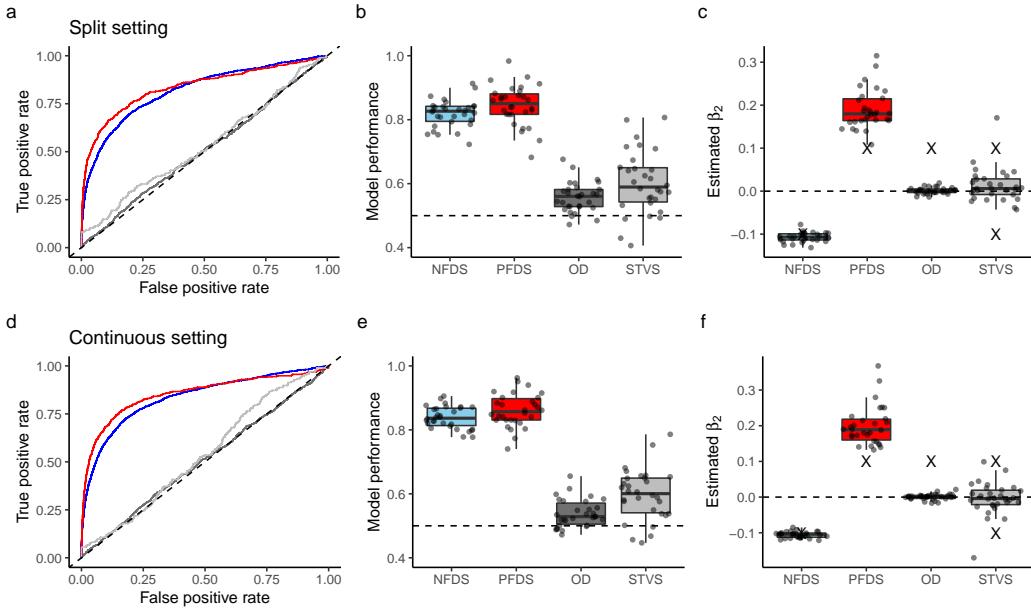


Figure 4: Performance of linear mixed models for detecting four types of simulated selection among genome-wide polymorphisms: NFDS, negative frequency-dependent selection (blue); PFDS, positive frequency-dependent selection (red); OD, overdominance (dark gray); and STVS, spatiotemporally varying selection (light gray). The top and bottom panels show the results of the split and continuous settings of subpopulation structure, respectively (Fig. 1a). (a and d) The receiver operating characteristic (ROC) curve showing the relationships between the true and false positive rates. (b and e) Model performance evaluated by the area under the ROC curve (AUC). Dashed lines at 0.5 indicate no statistical power to detect causal single nucleotide polymorphisms (SNPs). (c and f) Estimated β_2 of causal SNPs, where the negative and positive values indicate negative and positive FDS, respectively. Cross marks indicate the true simulated magnitude of β_2 . Boxplots show the median by a center line, upper and lower quartiles by box limits, and $1.5 \times$ interquartile range by whiskers.

positive FDS than that with LMs (Fig. S6 and Fig. S7; Appendix S4). These results suggest that LMMs are better suited to the proposed method because they can more efficiently capture negative FDS or prevent overestimation of the strength of positive FDS.

3.2.2 Patterns of FDS in the *Arabidopsis thaliana* genome

We applied an LMM for the GWAS of the reproductive branch number in *A. thaliana* (Fig. 5) under the continuous setting (Fig. 1a bottom) to demonstrate feasibility using an empirical GWAS dataset (see the methods "**GWAS using field-grown *Arabidopsis thaliana***" above; Appendix S5; and Figs. S8-S10). To diagnose our data, we examined

the individual genotype effects β_1 using a standard GWAS. The association mapping of the individual genotype effects β_1 detected no significant SNPs above the Bonferroni threshold; however, a weak peak was observed on the top of chromosome 4 (Fig. 5a; Table S4a; see also Fig. S8 for quantile-quantile plots). The top of chromosome 4 is known to encompass a flowering QTL near the *FRIGIDA* gene (ATWELL *et al.*, 2010). Our results of individual genotype effects agree with previous evidence and support the usefulness of our GWAS data.

To examine the relevance of symmetric FDS, we next tested the significance of the genotype similarity effects β_2 for all SNPs. To elucidate the genome-wide patterns of β_2 and candidate genes, we focused on SNPs exhibiting p -values of $< 10^{-4}$. Of the 254 SNPs selected, 195 and 59 showed negative and positive β_2 , respectively (Fig. 5d), which indicates the prevalence of negative FDS in the *A. thaliana* genome. This excess of negative $\hat{\beta}_2$ was consistently observed even when we used LMs (Fig. S9d). To obtain a functional inference, we focused on candidate genes near top-scoring SNPs. Within 10 kbp near the top-scoring SNP on chromosome 3 (Fig. 5b), we observed the *GAE6* gene, which encodes a UDP-D-glucuronate 4-epimerase involved in pectin biosynthesis, cell wall integrity, and immunity to pathogens (BETHKE *et al.*, 2016). Genes involved in plant immunity and resistance, such as *GAE6*, *ARGONAUTE 4* (*AGO4*), and *ACTIVATED DISEASE RESISTANCE 1* (*ADR1*), were observed within 10 kb near the SNPs showing negative β_2 at p -value $< 10^{-4}$ (Table S4b). In contrast, *AVRRPT2-INDUCED GENE 1* (*AIG1*) was only a resistance-related gene observed near the SNPs showing positive $\hat{\beta}_2$ at p -values of $< 10^{-4}$ (Table S4b); these results suggest the potential relevance of disease resistance genes to the widespread pattern of negative FDS.

We then tested the asymmetric effects β_{12} for all SNPs to examine the relevance of asymmetric FDS. We detected a QTL on chromosome 5 near the Bonferroni threshold (Fig. 5c). These SNPs exhibited positive β_2 and β_{12} (i.e., $\hat{\beta}_2 = 0.78$ and $\hat{\beta}_{12} = 1.13$ on chromosome 5 at position 22382673; Table S4c), indicating positive effects of a reference allele on absolute fitness together with positive FDS on relative fitness. Genes potentially related to growth were located near this top-scoring SNP, including *SUMO2* and *SUMO3*. In the line of GWAS analyses, the individual genotype effect β_1 , symmetric FDS β_2 , and asymmetric β_{12} did not share any SNPs that exhibited $p < 10^{-4}$. The sort of top-scoring SNPs and candidate genes infer different functions among individual genotype effects, symmetric negative FDS, and asymmetric positive FDS.

We examined a potential bias of $-\log_{10}(p\text{-values})$ association scores based on quantile-quantile plots (Fig. S8 and Fig. S10) to inspect whether LMMs yields better results than

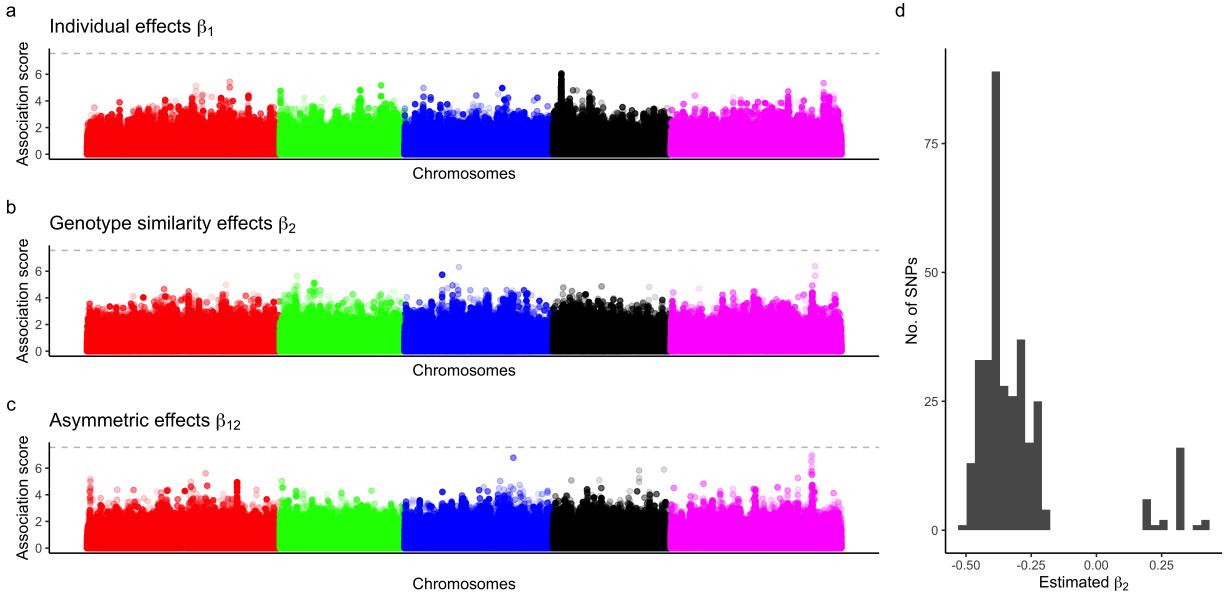


Figure 5: Genome-wide association studies of reproductive branch number in field-grown *Arabidopsis thaliana* under the continuous population setting (Fig. 1a bottom). The results of linear mixed models are shown. (a, b, and c) Manhattan plots showing the association score of $-\log_{10}(p\text{-values})$ for individual genotype effects, genotype similarity effects, and asymmetric effects, respectively. Horizontal dashed lines indicate $p\text{-value} = 0.05$, after Bonferroni correction. (d) The histogram of estimated β_2 among SNPs exhibiting $p\text{-values} < 10^{-4}$. The negative and positive β_2 infer loci responsible for negative and positive FDS, respectively.

LMs. Quantile-quantile plots of LMMs exhibited a slight inflation of the observed association score against the expected score for the genotype similarity effects β_2 and asymmetric effects β_{12} (Fig. S8b-c), whereas those of LMs showed slightly inflated scores (Fig. S10b-c). Combined with the simulation above, these empirical results suggest that LMMs are more suitable for GWAS than LMs. The line of GWAS simulation and application suggests that our method can also be used to screen polymorphisms associated with FDS (see also the discussion "Applicability for GWAS").

4 Discussion

4.1 Selection gradient along genotype similarity

Selection gradient analysis is a powerful approach for empirical studies to quantify selection in action (LANDE and ARNOLD, 1983; MITCHELL-OLDS and SHAW, 1987; CHONG *et al.*,

2018). The incorporation of population-level covariates offers further opportunity to reveal how selection depends on population contexts (HEISLER and DAMUTH, 1987). To analyze FDS, empirical studies have often incorporated morph frequency such as the population-level covariate and regressed fitness components on the morph frequency (GIGORD *et al.*, 2001; MCCUALEY and BROCK, 1998; CALSBEEK *et al.*, 2010; SATO and KUDOH, 2017). We propose a simple regression model where the two regression coefficients β_2 and β_{12} represent FDS on relative and absolute fitness, respectively, by incorporating genotype similarity as a covariate. As the covariate of genotype similarity $(\sum_{j=1}^{N_k} x_i x_j)/N_k$ denotes how similar (or dissimilar) the neighbor compositions are with the focal individual, conclusions are expected to be the same as we regress fitness components on the frequency of other morphs. In fact, the single-locus analysis detected known negative FDS in *A. halleri* and *I. elegans*, and it was attributed to the negative sign of β_2 . These reanalyses exemplify the simple and wide usability of our regression model for plants and animals in natural or semi-natural fields.

The population-level mean fitness is another side of interests in a general model of FDS when FDS acts on absolute fitness (COCKERHAM *et al.*, 1972; ASMUSSEN and BASNAYAKE, 1990; SCHNEIDER, 2008). Based on the regression model of FDS, our method provides an additional inference of the relationships between the population-level mean fitness and allele frequency (Fig. 2). For example, symmetric negative FDS lead the population-level mean fitness to increase at an intermediate allele frequency (Fig. 2a). Empirical studies on *I. elegans* reported such an increased population-level mean fitness at an intermediate frequency (TAKAHASHI *et al.*, 2014) as well as negative FDS on female color polymorphisms (LE ROUZIC *et al.*, 2015). This previous evidence is supported by our reanalysis showing symmetric negative FDS (i.e., $\hat{\beta}_2 < 0$; Table 1b) and a consequent increase in population-level mean fitness at an intermediate frequency in *I. elegans* (Fig. 3b). Although the reanalysis of *A. halleri* data shows the biased frequency on its equilibrium, negative FDS still maintained the dimorphism and very slightly increased population-level mean fitness at the equilibrium frequency (Fig. 3a). In addition to the efficient detection of FDS, the present method provides an empirical approach to understand how FDS increases population-level mean fitness.

4.2 Applicability for GWAS

Selection gradient analysis is used for genome-wide studies to quantify the ongoing selection (e.g., EXPOSITO-ALONSO *et al.*, 2019; GROEN *et al.*, 2020). We extended our regression method to LMMs that have often been used in GWAS by incorporating the population

genetic structure as random effects (e.g., KANG *et al.*, 2008; GONDRO *et al.*, 2013, see also Appendix S3). Our simulations suggest that LMMs improve the statistical power to detect causal polymorphisms or prevent us from exaggerating effect-size estimates of FDS. We should also focus on the genetic structure of loci underlying positive or negative FDS to correctly understand the power of LMMs. Because positive FDS reduces polymorphisms, their selected loci likely showed low heterozygosity and strong population differentiation (Fig. S4). Although LMMs can deal with the population genetic structure to some extent, their effect-size estimates are still larger than the true signals (Fig. 4c and f). This seems consistent with the general notion that a systematic bias, such as the incomplete correction of population genetic structure, leads GWAS to inflate association estimates (GONDRO *et al.*, 2013). The overestimation of positive FDS may thus be because of a strong correlation in the genetic structure between neutral and selected loci under positive FDS. In contrast, negative FDS results in high heterozygosity by maintaining polymorphisms within a population (Fig. S4) where LMMs can perform better than LMs by separating the neutral genetic structure and loci subject to negative FDS. While the current application is limited to homozygous *A. thaliana*, our simulations clarified what should be considered to analyze heterozygous individuals.

The application to *A. thaliana* accessions found a genome-wide excess of negative FDS and a few loci underlying asymmetric positive FDS. This result seems plausible because polymorphic loci are more likely to persist under negative FDS than positive FDS. In the present study, we found candidate genes that might be involved in conferring plant resistance to pathogens. This finding supports the hypothesis that negative FDS is likely to act on the polymorphism of pathogen resistance (ANTONOVICS and ELLSTRAND, 1984; BRUNET and MUNDT, 2000). In contrast, such resistance genes were not found near the loci responsible for asymmetric positive FDS. We instead found growth-related candidate genes near the loci associated with asymmetric positive FDS. The asymmetric and positive frequency-dependent effects from tall to short plants is known to occur in plant competition (WEINER, 1990), where growth-related loci are more likely to be observed than defense-related loci. However, our field experiment provides a one-year snapshot of FDS in a common garden, and it does not indicate the long-term importance of FDS compared to other selective pressures responsible for genomic variation among *A. thaliana* accessions. These accessions were originally collected from different natural populations (HORTON *et al.*, 2012; ALONSO-BLANCO *et al.*, 2016), and thus, the extent to which the common garden experiment simulates naturally co-occurring alleles is also elusive. Further, the branch number remains a qualitative measure

that presumably has a strong association with fitness in *A. thaliana* (CHONG *et al.*, 2018). Further studies using an accurate fitness data in varying space and time will be required to verify the patterns of FDS in the *A. thaliana* genome. Limited datasets are currently available to reveal general patterns of FDS because our methods require exact spatial positions or subpopulation structure for individuals. If open data are registered with the spatial positions as well as individual genotypes and fitness, these datasets suit the analysis of FDS using our method.

4.3 Potential limitation

In addition to the model of one locus with two alleles (COCKERHAM *et al.*, 1972; ASMUSSEN and BASNAYAKE, 1990; SCHNEIDER, 2008), FDS have been theoretically analyzed in relation to one locus with multi-alleles (SCHNEIDER, 2006; TROTTER and SPENCER, 2007) and multi-loci with multiple alleles (SCHNEIDER, 2010). Even a one-locus two-allele model can make fitness functions have multiple equilibria when it involves the additive action of FDS on a quantitative trait (SCHNEIDER, 2008, Fig. S2; Appendix S2); however, more complex outcomes are possible in the realistic genetic architecture. Since our method is customized to detect simple instances of FDS, other more complicated actions of FDS are unlikely to be captured by the proposed regression model. For example, FDS can act on more than two co-dominant alleles at a single locus, such as the S-allele system in plant self-incompatibility (HATAKEYAMA *et al.*, 1998; SHIMIZU and TSUCHIMATSU, 2015). The multi-allele extension of our regression model would be required to widen its applicability for the well-reported instances of FDS.

4.4 Conclusion

This study offers an effective method to model the patterns of fitness-genotype associations with respect to FDS. Our phenotype-driven approach can distinguish between positive and negative FDS based on the direct observations of fitness components. Although considerable effort is required to measure individual fitness, a selection gradient analysis has an advantage in quantifying ongoing FDS. Such a snapshot of the ongoing selection may not always match the past signatures of selection; however, a joint study using the genomic signature and selection gradient analysis will reveal consistent patterns of FDS throughout the genome. In molecular population genetics, phenotype-free methods are available for detecting the signatures of past balancing selection (SIEWERT and VOIGHT, 2017). Now that genome

information is accumulating in wild organisms (LEWIN *et al.*, 2018), organismal biologists may utilize genome data for a deeper understanding of the maintenance of polymorphism from the past to present. Our pilot test using *A. thaliana* accessions is expected to provide the first step to reveal genome-wide patterns of FDS in the field.

Data Availability Statement

The source codes and original data generated by this study are available in the GitHub repository (<https://github.com/yassato/RegressionFDS>) and its published version is deposited on Dryad (<https://doi.org/10.5061/dryad.zs7h44jdv>). A vignette using toy data is available at <https://yassato.github.io/RegressionFDS>, and its source files are included in the GitHub and the Dryad repository.

Author Contributions

Y.S. developed the model, analyzed the data, and wrote a draft. Y.T. provided the damselfly data, discussed the results, and contributed to the conceptualization regarding FDS. C.X. and Y.S. conducted the field experiment using *A. thaliana*. Y.S. and K.K.S. designed the study, and revised the manuscript with inputs from the co-authors.

Funding

This study was supported by the Japan Society for the Promotion of Science (JSPS) Kakenhi (Grant No. 20K15880) and Japan Science and Technology Agency (JST) Precursory Research for the Embryonic Science and Technology (PRESTO) (JPMJPR17Q4) grants (Y.S.) and Swiss National Science Foundation (31003A_182318, 31003A_212551) and JSPS Kakenhi (22H05179) and Japan Science and Technology Agency (JST) Core Research for Evolutional Science and Technology (CREST) (JPMJCR16O3) grants (K.K.S.). The fieldwork at Zurich was supported by the University of Zurich via the University Research Priority Program for "Global Change and Biodiversity."

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors thank M. Yamazaki, A. Morishima, and all members of the Shimizu group for their assistance with the setup of the *A. thaliana* experiment; H. Kudoh for discussions during the *A. halleri* study; J. Bascompte, S.E. Wuest, T. Bataillon, and anonymous reviewers for comments on the manuscript; and Editage (<http://www.editage.com>) for English language editing. Y.S. thanks A.J. Nagano for hosting him as a guest researcher at Ryukoku University.

Supplementary Materials

Appendix S1. Ising model behind genetic interactions among individuals

The idea to include genotype similarity as a covariate was inspired by a model of ferromagnetism, which is widely known as the Ising model (CIPRA, 1987). The Ising model addresses how physical interactions between two magnets that attract or repel each other can shape spatial patterns of total energy. This physical model provides an analogy to genetic interactions between two alleles of neighboring individuals. If two genotypes benefit from their similarity, these positive interactions force the two individuals to have similar genotypes. If two genotypes benefit from their dissimilarity, these negative interactions force the two individuals to have different genotypes. Using this analogy of the Ising model, our regression model provides an approach for estimating the coefficient of physical interactions from individual energy (SATO *et al.*, 2021b,a). In this appendix, we investigate how our regression method can mimic biological evolution in its analogy to the Ising model.

Here, note that our regression model possesses a comparable structure with the Ising model. Individual fitness is defined by Equation (1) in the main text as $w_i = w_0 + s_1 g_i + \frac{s_2}{N_k} \sum_{j=1}^{N_k} g_i g_j$ (see the methods "**FDS model based on genotype similarity**"). The evolutionary process in Equation (1) is to select genotypes g_i based on the fitness w_i under selection pressures s_1 and s_2 . Similarly, the magnetic interaction is defined in the Ising model as $H = -J \sum_{\langle i,j \rangle} g_i g_j + \eta \sum g_i$. The total energy H can be regarded as the population sum of fitness $\sum w_i$, magnetic interaction coefficient J as the interaction strength $s_2 N_k$, the allelic status of individual g_i as N or S dipole, and external magnetic force η as the strength of directional selection $s_1 N_k$. Ising model's simulation process involves updating the magnetic status g_i on the basis of the two coefficients η and J ; thus, it is comparable with the process of inheritance after selection.

To simulate selection and inheritance, we updated the genotype g_i and fitness w_i based on the Mendelian inheritance in Equation (1). The individual fitness w_i was updated following the Metropolis algorithm (METROPOLIS *et al.*, 1953), which has often been used in a series of stochastic sampling methods, such as the Markov chain Monte Carlo method (BISHOP, 2006). Here, we consider a change in the genotype g_i from generation t to $t+1$. Let g'_i be a proposed genotype for $t+1$, and let $w(g'_i | g_{i,t})$ be a conditional fitness with a given genotype $g_{i,t}$ at t . Then, g'_i can be accepted/rejected based on its likelihood ratio on the current fitness

$w(g_{i,t})$ as

$$g_{i,t+1} = \begin{cases} g'_i & \exp(w(g'_i|g_{i,t}) - w(g_{i,t})) > p \\ g_{i,t} & \text{otherwise.} \end{cases}$$

where scalar p is sampled from a uniform distribution as $p \sim \text{Unif}(0, 1)$. This update process simulates the selection to some extent of stochasticity. Transition from $g_{i,t}$ to $g_{i,t+1}$ was weighted by genotype segregation among the offspring to further incorporate Mendelian inheritance into the process of inheritance. When two genotypes $g_{i,t}$ and $g_{j,t}$ crossed each other at generation t , we could expect nine possible combinations among parental genotypes, as shown in Table S1.

Table S1: Cross tables among AA, Aa, and aa genotypes.

g_i / g_j	AA		A		aa	
AA	AA	AA	AA	Aa	Aa	Aa
Aa	AA	AA	AA	Aa	Aa	Aa
aa	Aa	Aa	Aa	aa	aa	aa
	Aa	Aa	Aa	aa	aa	aa

The probability of sampling AA, Aa, and aa at $t+1$ generation is denoted as $P_{AA,t+1}$, $P_{Aa,t+1}$, and $P_{aa,t+1}$, respectively. Let f_{AA} , f_{Aa} , and f_{aa} be the frequencies of genotypes AA, Aa, and aa within a population with f_{AA} , f_{Aa} , and f_{aa} . In summary, the transition from generation t to $t+1$ is expressed as

$$\begin{pmatrix} P_{AA,t+1} \\ P_{Aa,t+1} \\ P_{aa,t+1} \end{pmatrix} = \begin{pmatrix} (f_{AA} + 0.5f_{Aa}) & (0.5f_{AA} + 0.25f_{Aa}) & 0 \\ (f_{aa} + 0.5f_{Aa}) & (0.5f_{AA} + 0.5f_{Aa} + 0.5f_{aa}) & (f_{AA} + 0.5f_{Aa}) \\ 0 & (0.25f_{Aa} + 0.5f_{aa}) & (0.5f_{Aa} + f_{aa}) \end{pmatrix} \begin{pmatrix} P_{AA,t} \\ P_{Aa,t} \\ P_{aa,t} \end{pmatrix},$$

where the elements of the transition matrix were calculated from the three genotype frequencies and the segregation ratio of Mendelian inheritance. The zero elements pose a constraint where one homozygote cannot turn into another homozygote in a single generation. Given that the sum of the three genotype frequencies was 1 as $f_{AA} + f_{Aa} + f_{aa} = 1$, the probability of remaining as a heterozygote was 0.5. Therefore, the outcome from the modified Metropolis algorithm was expected to be qualitatively the same as a random proposal of three genotypes, but quantitatively different in the excess of heterozygotes within a population.

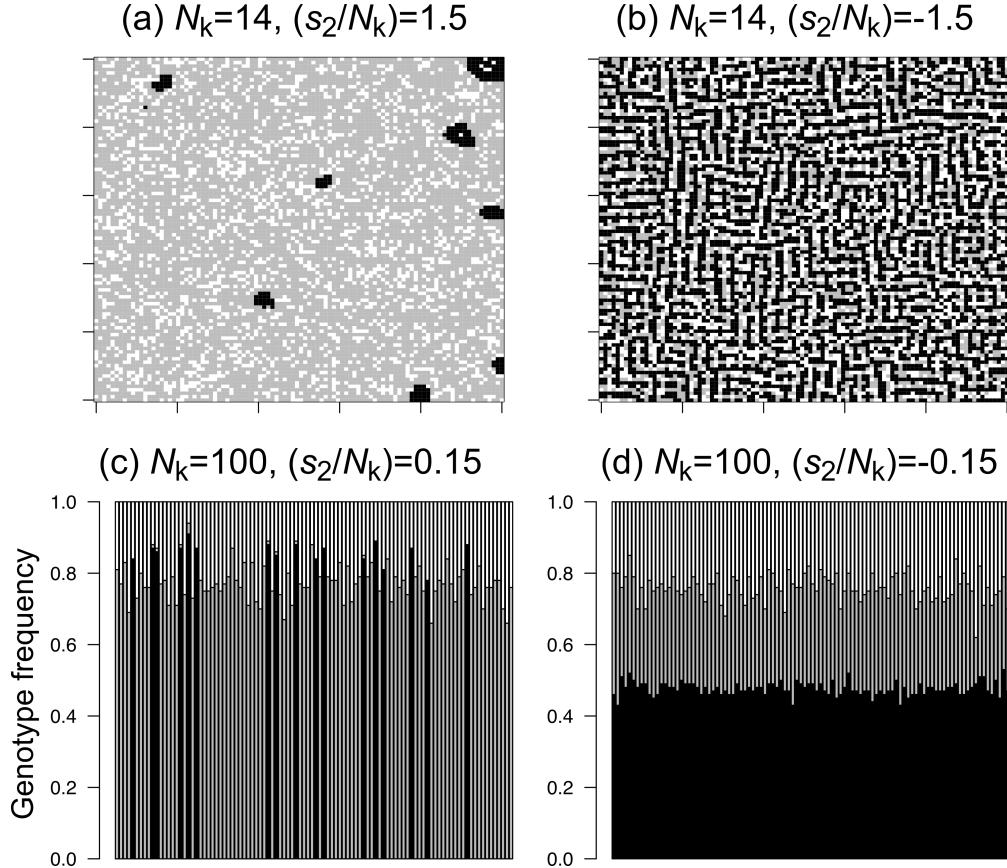


Figure S1: Numerical simulations maximizing the fitness $w_1 = w_0 + s_1 g_i + (s_2/N_k) \sum_{j=i}^{N_k} g_i g_j$ (Appendix S1) for 100 generations. White, gray, and black indicate the AA, Aa, and aa genotypes, respectively. Top panels (a, b) represent plant genotype distributions across a 100×100 continuous lattice space when interactions are restricted to the second nearest neighbors ($N_k = 14$). Each grid corresponds to an individual. Bottom panels (c, d) represent the genotype frequencies among 100 split populations composed of 100 individuals each. Each vertical bar corresponds to a population. Left panels (a, c) simulate positive frequency-dependent selection (FDS), while right panels (b, d) simulate negative FDS. No directional selection were assumed as $w_0 = 0$ and $s_1 = 10^{-4}$ for all simulations, whereas the subpopulation size N_k and interaction strength per individual s_2/N_k were changed.

Figure S1 shows the results of numerical simulations under continuous and split sub-populations with complete dominance as $g_i \in \{\text{AA}, \text{Aa}, \text{aa}\} = \{+1, +1, -1\}$. The three genotypes were well mixed and maintained in a continuous space when $s_2 < 0$ (Fig. S1b), whereas several clusters of the aa genotype were observed when $s_2 > 0$ (Fig. S1a). Three genotypes were also maintained at an intermediate frequency in a split space when $s_2 < 0$ (Fig. S1d), whereas the allele frequency was heavily biased when $s_2 > 0$ (Fig. S1c). The numerical simulations show that the sign of s_2 likely corresponded to the direction of the frequency-dependent selection (FDS) in a continuous and split space.

In summary, our Equation (1) based on genotype similarity represents negative and positive FDS with its selection coefficient s_2 . By estimating s_2 from the individual fitness w_i , the regression model Equation (2) provided an empirical framework to detect FDS in the main text (see the methods "**Regression model of FDS**").

Appendix S2. Fitness function under symmetric and asymmetric FDSs

In this appendix, we show how our regression model can represent FDS under three cases of genotype encoding: complete dominance with random mating; additive effects with random mating; and asexual or inbred lines without mating. To analyze the regression model Equation (3) as a fitness function of allele frequency, we considered a single diallelic locus in an ideal population where diploid individuals were randomly mating and uniformly interacting in a sufficiently large population (i.e., $N \rightarrow \infty$). We also assumed no maternal or paternal effects on fitness such that the genotypes Aa and aA could not be distinguished. To concentrate on the mean trends of the model, we neglected the residuals as $e_i = 0$. Replacing N_k into N , we redefined Equation (3) as follows:

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{N} \sum_{j=1}^N x_i x_j + \frac{\beta_{12}}{N} \sum_{j=1}^N x_i x_j \quad (\text{S1})$$

where the trait value y_i corresponds to the fitness value w_i for individual i , the intercept β_0 corresponds to the base fitness w_0 , the individual genotype coefficient β_1 corresponds to the directional selection coefficient s_1 , and the coefficients β_2 and β_{12} correspond to the selection coefficients related to FDS. The second term can also be transformed for simplicity as $\beta_2(\sum_{j=1}^N x_i x_j)/N = \beta_2 x_i (\sum_{j=1}^N x_j)/N$, and the third term, $\beta_{12} x_i (\sum_{j=1}^N x_i x_j)/N = \beta_{12} x_i^2 (\sum_{j=1}^N x_j)/N$. Provided $x_i^2 = (-1)^2 = 1^2 = 1$, the third term represents a selection gradient due to the relative abundance of one allele in a neighborhood, irrespective of the individual genotype of a focal individual i .

We then transformed Equation (S1) into a function of the frequency of A alleles f , where the frequency of a alleles was defined conversely as $1 - f$. When all the individuals were randomly interacting in a panmictic population, the interaction strength between i and j depended on the frequencies of AA, Aa, or aa genotypes derived from allele frequency. Thus, the fitness values of the three genotypes, $y_{AA}(f)$, $y_{Aa}(f)$, and $y_{aa}(f)$, are functions of allele frequency f . For convenience, we suppressed the dependence on f , unless necessary. The ratio of AA, Aa, and aa genotypes within the panmictic population was given by AA:Aa:aa = $f^2 : 2f(1-f) : (1-f)^2$. Assuming the aforementioned ideal population, we designated all combinations of interactions among AA, Aa, and aa genotypes and weighted the interactions based on their genotype frequencies. Table S2 lists the interaction strength weighted by the genotype frequency for symmetric and asymmetric effects.

Table S2: Cross tables showing the strength of interactions between the focal genotype x_i and counterpart x_j in a randomly interacting and mating population.

(a) Symmetric effects

$x_j \setminus x_i$	x_{AA}	x_{Aa}	x_{aa}
x_{AA}	$\beta_2 f^2 x_{AA} x_{AA}$	$\beta_2 f^2 x_{Aa} x_{AA}$	$\beta_2 f^2 x_{aa} x_{AA}$
x_{Aa}	$2\beta_2 f(1-f) x_{AA} x_{Aa}$	$2\beta_2 f(1-f) x_{Aa} x_{Aa}$	$2\beta_2 f(1-f) x_{aa} x_{Aa}$
x_{aa}	$\beta_2 (1-f)^2 x_{AA} x_{aa}$	$\beta_2 (1-f)^2 x_{Aa} x_{aa}$	$\beta_2 (1-f)^2 x_{aa} x_{aa}$

(b) Asymmetric effects

$x_j \setminus x_i$	x_{AA}	x_{Aa}	x_{aa}
x_{AA}	$\beta_{12} f^2 x_{AA}^2 x_{AA}$	$\beta_{12} f^2 x_{Aa}^2 x_{AA}$	$\beta_{12} f^2 x_{aa}^2 x_{AA}$
x_{Aa}	$2\beta_{12} f(1-f) x_{AA}^2 x_{Aa}$	$2\beta_{12} f(1-f) x_{Aa}^2 x_{Aa}$	$2\beta_{12} f(1-f) x_{aa}^2 x_{Aa}$
x_{aa}	$\beta_{12} (1-f)^2 x_{AA}^2 x_{aa}$	$\beta_{12} (1-f)^2 x_{Aa}^2 x_{aa}$	$\beta_{12} (1-f)^2 x_{aa}^2 x_{aa}$

Based on these cross tables (Table S2), we redefined Equation (S1) for the three genotypes as:

$$y_{AA} = \beta_0 + \beta_1 x_{AA} + \beta_2 f^2 x_{AA} x_{AA} + 2\beta_2 f(1-f) x_{AA} x_{Aa} + \beta_2 (1-f)^2 x_{AA} x_{aa} + \beta_{12} f^2 x_{AA}^2 x_{AA} + 2\beta_{12} f(1-f) x_{AA}^2 x_{Aa} + \beta_{12} (1-f)^2 x_{AA}^2 x_{aa} \quad (\text{S2a})$$

$$y_{Aa} = \beta_0 + \beta_1 x_{Aa} + \beta_2 f^2 x_{Aa} x_{AA} + 2\beta_2 f(1-f) x_{Aa} x_{Aa} + \beta_2 (1-f)^2 x_{Aa} x_{aa} + \beta_{12} f^2 x_{Aa}^2 x_{AA} + 2\beta_{12} f(1-f) x_{Aa}^2 x_{Aa} + \beta_{12} (1-f)^2 x_{Aa}^2 x_{aa} \quad (\text{S2b})$$

$$y_{aa} = \beta_0 + \beta_1 x_{aa} + \beta_2 f^2 x_{aa} x_{AA} + 2\beta_2 f(1-f) x_{aa} x_{Aa} + \beta_2 (1-f)^2 x_{aa} x_{aa} + \beta_{12} f^2 x_{aa}^2 x_{AA} + 2\beta_{12} f(1-f) x_{aa}^2 x_{Aa} + \beta_{12} (1-f)^2 x_{aa}^2 x_{aa} \quad (\text{S2c})$$

We further weighted the genotype-level fitness values by allele frequency. The allele-level fitness for A or a alleles is then given by:

$$y_A = f y_{AA} + (1-f) y_{Aa} \quad (\text{S3a})$$

$$y_a = f y_{Aa} + (1-f) y_{aa} \quad (\text{S3b})$$

The population-level mean fitness was finally defined by the weighted mean of the allele-level

fitness as follows:

$$\begin{aligned}\bar{y} &= fy_A + (1 - f)y_a \\ &= f^2 y_{AA} + 2f(1 - f)y_{Aa} + (1 - f)^2 y_{aa}\end{aligned}\tag{S4}$$

Consequently, we could analyze fitness functions by inputting specific values in the three genotype values x_{AA} , x_{Aa} , and x_{aa} throughout Equations (S2) and (S3). In the following, we present three specific cases of complete dominance, additive effects, and asexual populations.

Case 1. Complete dominance with random mating: Novel mutations are expected to be recessive during adaptive evolution. Empirical studies have reported FDS on dimorphic traits that often exhibit complete dominance of one over another allele (e.g., TAKAHASHI *et al.*, 2010; SATO and KUDOH, 2017; GOLDBERG *et al.*, 2020). First, we considered the case in which the A alleles were completely dominant over a alleles, as encoded by $x_i \in \{AA, Aa, aa\} = \{+1, +1, -1\}$. Here, we neglected the directional selection as $\beta_1 = 0$ in Equation (S2) to focus on the fitness functions under FDS alone. Replacing the genotypes (x_{AA} , x_{Aa} , and x_{aa}) according to their genotype values in Equation (S2) gave the fitness value to the three genotypes as

$$\begin{aligned}y_{AA} = y_{Aa} &= \beta_0 + \beta_2[(+1) \times (+1)]f^2 + \beta_2[(+1) \times (+1)]2f(1 - f) + \beta_2[(+1) \times (-1)](1 - f)^2 \\ &\quad + \beta_{12}[(+1)^2 \times (+1)]f^2 + \beta_{12}[(+1)^2 \times (+1)]2f(1 - f) + \beta_{12}[(+1)^2 \times (-1)](1 - f)^2 \\ &= \beta_0 + \beta_2(2f - 1 + 2f - 2f^2) + \beta_{12}(2f - 1 + 2f - 2f^2) \\ &= \beta_0 + (\beta_{12} + \beta_2)(4f - 2f^2 - 1) \quad \text{where } x_{AA} = x_{Aa} = +1\end{aligned}\tag{S5a}$$

$$\begin{aligned}y_{aa} &= \beta_0 + \beta_2[(-1) \times (+1)]f^2 + \beta_2[(-1) \times (+1)]2f(1 - f) + \beta_2[(-1) \times (-1)](1 - f)^2 \\ &\quad + \beta_{12}[(-1)^2 \times (+1)]f^2 + \beta_{12}[(-1)^2 \times (+1)]2f(1 - f) + \beta_{12}[(-1)^2 \times (-1)](1 - f)^2 \\ &= \beta_0 - \beta_2(2f - 1 + 2f - 2f^2) + \beta_{12}(2f - 1 + 2f - 2f^2) \\ &= \beta_0 + (\beta_{12} - \beta_2)(4f - 2f^2 - 1) \quad \text{where } x_{aa} = +1\end{aligned}\tag{S5b}$$

The allele-level fitness following Equations (S3a) and (S3b) was given by

$$y_A = fy_{AA} + (1 - f)y_{Aa} = fy_{AA} + (1 - f)y_{AA} = y_{AA} \quad (\text{S6a})$$

$$y_a = fy_{Aa} + (1 - f)y_{aa} = y_{aa} + 2f\beta_2(4f - 2f^2 - 1) \quad (\text{S6b})$$

The relative fitness between A and a alleles was calculated as: $y_A - y_a = y_{AA} - y_{aa} - 2f\beta_2(4f - 2f^2 - 1) = 2\beta_2(1 - f)(4f - 2f^2 - 1)$. Solving $y_A - y_a = 0$ with respect to f within the range of $(0, 1)$ resulted in $f^* = -0.5\sqrt{2} + 1$, which showed a single stable or unstable state within $f = (0, 1)$ in the case of complete dominance under FDS. The population-level mean fitness is finally given by

$$\begin{aligned} \bar{y} &= fy_A + (1 - f)y_a \\ &= \beta_0 + (4f\beta_2 - 2f^2\beta_2 - \beta_2 + \beta_{12})(4f - 2f^2 - 1) \end{aligned} \quad (\text{S7})$$

Figure 2 in the main text (see the methods "***Fitness function given by regression coefficients***") shows the allele-level fitness [Equations (S6a) and (S6b)] and population-level mean fitness [Equation (S7)]. The mean fitness was maximized at a stable equilibrium at the intermediate allele frequency under symmetric negative FDS, whereas it was minimized at an unstable equilibrium under symmetric positive FDS. When asymmetric FDS and complete dominance are involved, equilibria do not always match the maxima or minima of the mean fitness because of the nonlinearity of the allele-level fitness in response to f . Still, the mean fitness at the stable or unstable point was higher or lower than expected compared with the weighted mean of the two monomorphic populations. A similar notion was suggested by a general model of FDS in population genetics (COCKERHAM *et al.*, 1972; SCHNEIDER, 2008).

Case 2. Additive effects with random mating: Although few empirical studies have reported FDS on quantitative traits, this case is of theoretical interest in the general model of FDS (e.g., SCHNEIDER, 2008). Considering the fitness value as a quantitative trait, we then analyzed the additive effects of A and a alleles on y_i as encoded by $x_i \in \{\text{AA}, \text{Aa}, \text{aa}\} = \{+1, 0, -1\}$. Replacing the genotypes (x_{AA} , x_{Aa} , and x_{aa}) according to their genotype

values in Equations (S2) gave the fitness value to the three genotypes as:

$$\begin{aligned}
y_{AA} &= \beta_0 + \beta_2[(+1) \times (+1)]f^2 + \beta_2[(+1) \times (-1)](1-f)^2 \\
&\quad + \beta_{12}[(+1)^2 \times (+1)]f^2 + \beta_{12}[(+1)^2 \times (-1)](1-f)^2 \\
&= \beta_0 + \beta_2f^2 - \beta_2(1-2f+f^2) + \beta_{12}f^2 - \beta_{12}(1-2f+f^2) \\
&= \beta_0 + \beta_2(2f-1) + \beta_{12}(2f-1) \\
&= \beta_0 + (\beta_{12} + \beta_2)(2f-1) \quad \text{where } x_{AA} = +1
\end{aligned} \tag{S8a}$$

$$y_{Aa} = \beta_0 \quad \text{where } x_{Aa} = 0 \tag{S8b}$$

$$\begin{aligned}
y_{aa} &= \beta_0 + \beta_2[(-1) \times (+1)]f^2 + \beta_2[(-1) \times (+1)](1-f)^2 \\
&\quad + \beta_{12}[(-1)^2 \times (+1)]f^2 + \beta_{12}[(-1)^2 \times (-1)](1-f)^2 \\
&= \beta_0 - \beta_2f^2 + \beta_2(1-2f+f^2) + \beta_{12}f^2 - \beta_{12}(1-2f+f^2) \\
&= \beta_0 - \beta_2(2f-1) + \beta_{12}(2f-1) \\
&= \beta_0 + (\beta_{12} - \beta_2)(2f-1) \quad \text{where } x_{aa} = -1
\end{aligned} \tag{S8c}$$

The fitness function for the two homozygotes AA and aa [i.e., Equations (S2a) and (S2c)] turned out to be linear in response to f . The allele-level fitness following Equations (S3a) and (S3b) is then given by:

$$\begin{aligned}
y_A &= fy_{AA} + (1-f)y_{Aa} \\
&= \beta_0 + f(\beta_{12} + \beta_2)(2f-1)
\end{aligned} \tag{S9a}$$

$$\begin{aligned}
y_a &= fy_{Aa} + (1-f)y_{aa} \\
&= \beta_0 + (1-f)(\beta_{12} - \beta_2)(2f-1)
\end{aligned} \tag{S9b}$$

The relative fitness was calculated as $y_A - y_a = (2f-1)[f(\beta_{12} + \beta_2) - (1-f)(\beta_{12} - \beta_2)] = (2f-1)(2\beta_{12}f - \beta_{12} + \beta_2)$. Solving $y_A - y_a = 0$ with respect to f within a range of (0,1) yields $f^* = 0.5$ and $f^* = 0.5(\beta_{12} - \beta_2)/\beta_{12}$, showing that the additive action of FDS made multiple equilibria possible at the intermediate allele frequency within $f = (0, 1)$. The mean

fitness following Equation (S4) is finally given by:

$$\begin{aligned}
\bar{y} &= fy_A + (1 - f)y_a \\
&= \beta_0 + f^2(\beta_{12} + \beta_2)(2f - 1) + (1 - f)^2(\beta_{12} - \beta_2)(2f - 1) \\
&= \beta_0 + (2f - 1)[2f^2(\beta_{12} + \beta_2) - 2f(\beta_{12} - \beta_2) + \beta_{12} - \beta_2]
\end{aligned} \tag{S10}$$

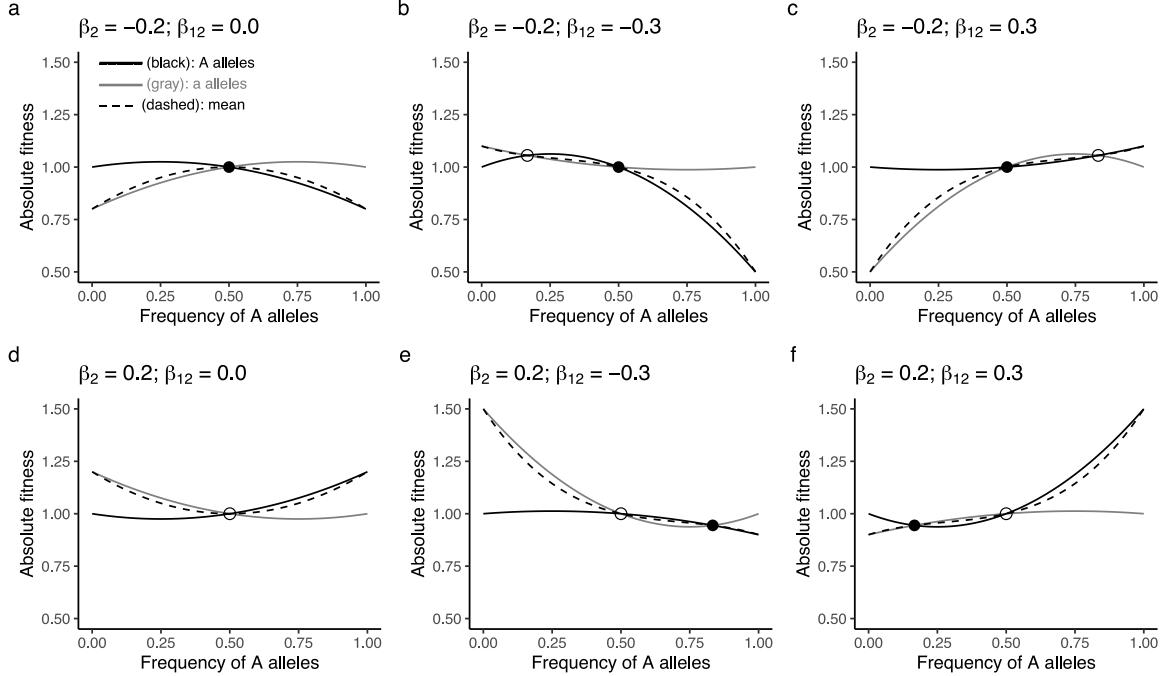


Figure S2: Numerical examples for the fitness values y_i in response to allele frequency when A and a alleles have additive effects on the fitness; that is, Equations (S9a) and (S9b) in Appendix S2. Black and gray lines indicate the allele-level fitness of the A or a allele, respectively. The dashed curves indicate the mean fitness per population; that is, Equation (S10) in Appendix S2. (a) Symmetric negative frequency-dependent selection (FDS); (d) symmetric positive FDS; (b and c) asymmetric negative FDS; and (e and f) asymmetric positive FDS. Closed and open circles indicate a stable or unstable state, respectively. No directional selection i.e., $\beta_0 = 1.0$ and $\beta_1 = 0$ are assumed for all panels to present fitness functions of FDS.

Figure S2 shows numerical examples of the allele-level fitness [Equations (S9a) and (S9b)] and the mean fitness [Equation (S10)] in response to f . Similar to the case of complete dominance, the mean fitness was maximized or minimized at a stable or unstable equilibrium under the symmetric FDS (Fig. S2a and d). In contrast, a stable and unstable equilibrium

occurred simultaneously under asymmetric FDS (Fig. S2b-c and e-f), where the maxima or minima of mean fitness did not always match the equilibria. This potential of multiple equilibria was also suggested by a one-locus two-allele model of FDS when it involved asymmetric FDS (SCHNEIDER, 2008).

Case 3. Asexual or inbred lines without mating: In common gardens or laboratory experiments, researchers arbitrarily distribute inbred accessions in space and retrieve individuals before mating (e.g., SATO *et al.*, 2021b). This was also the case for the field GWAS experiment in the main text. Furthermore, ecological studies often focus on FDS at the phenotype level with asexual reproduction assumed (e.g., TAKAHASHI *et al.*, 2018). In these cases, heterozygotes were negligible, and the two homozygotes were encoded as $x_i \in \{\text{AA}, \text{aa}\} = \{+1, -1\}$ (SATO *et al.*, 2021b). Let f_{AA} and f_{aa} be the frequency of AA and aa genotypes within a population, where $f_{\text{AA}} + f_{\text{aa}} = 1$. The fitness function for the AA or aa genotype is given by:

$$\begin{aligned} y_{\text{AA}} &= \beta_0 + \beta_2 f_{\text{AA}}[(+1) \times (+1)] + \beta_2(1 - f_{\text{AA}})[(+1) \times (-1)] \\ &\quad + \beta_{12} f_{\text{AA}}[(-1)^2 \times (+1)] + \beta_{12}(1 - f_{\text{AA}})[(-1)^2 \times (-1)] \\ &= \beta_0 + \beta_2(2f_{\text{AA}} - 1) + \beta_{12}(2f_{\text{AA}} - 1) \\ &= \beta_0 + (\beta_{12} + \beta_2)(2f_{\text{AA}} - 1) \quad \text{where } x_{\text{AA}} = +1 \end{aligned} \tag{S11a}$$

$$\begin{aligned} y_{\text{aa}} &= \beta_0 + \beta_2 f_{\text{AA}}[(-1) \times (+1)] + \beta_2(1 - f_{\text{AA}})[(-1) \times (-1)] \\ &\quad + \beta_{12} f_{\text{AA}}[(-1)^2 \times (+1)] + \beta_{12}(1 - f_{\text{AA}})[(-1)^2 \times (-1)] \\ &= \beta_0 - \beta_2(2f_{\text{AA}} - 1) + \beta_{12}(2f_{\text{AA}} - 1) \\ &= \beta_0 + (\beta_{12} - \beta_2)(2f_{\text{AA}} - 1) \quad \text{where } x_{\text{aa}} = -1 \end{aligned} \tag{S11b}$$

where the population-level mean fitness was given by its weighted mean as follows.

$$\begin{aligned} \bar{y} &= f_{\text{AA}} y_{\text{AA}} + (1 - f_{\text{AA}}) y_{\text{aa}} \\ &= f_{\text{AA}}(y_{\text{AA}} - y_{\text{aa}}) + y_{\text{aa}} \\ &= 2\beta_2 f_{\text{AA}}(2f_{\text{AA}} - 1) + (\beta_{12} - \beta_2)(2f_{\text{AA}} - 1) + \beta_0 \end{aligned} \tag{S12}$$

Figure S3 shows the fitness function of the AA or aa genotype [Equations (S11a) and (S11b)] and population mean [Equation S12]. The fitness function of the two homozygotes

was the same as that of the additive case described above. The mean fitness became simpler as the allele frequency corresponded to the genotype frequency. As this inbred case represented two genotypes with asexual reproduction, its conclusion was basically the same as that derived from game theoretical models (TAKAHASHI *et al.*, 2018).

In Figure 3, we present this inbred case with $\beta_1 \neq 0$, where the genotype fitness Equations (S11a) and (S11b) is rewritten as

$$y_{AA} = \beta_0 + \beta_1 + (\beta_{12} + \beta_2)(2f_{AA} - 1) \quad (\text{S13a})$$

$$y_{aa} = \beta_0 - \beta_1 + (\beta_{12} - \beta_2)(2f_{AA} - 1) \quad (\text{S13b})$$

where the relative fitness is given by $y_{AA} - y_{aa} = 2\beta_1 + 2\beta_2(2f_{AA} - 1)$. Solving $y_{AA} - y_{aa} = 0$ with respect to $f_{AA} = (0, 1)$ yields $f_{AA}^* = 1 - 0.5\beta_1/\beta_2$. Therefore, the directional selection coefficient β_1 may modify the equilibrium. When $\beta_1 \neq 0$, the mean fitness Equation (S12) can also be rewritten as:

$$\begin{aligned} \bar{y} &= f_{AA}y_{AA} + (1 - f_{AA})y_{aa} \\ &= 2\beta_2 f_{AA}(2f_{AA} - 1) + (\beta_{12} - \beta_2)(2f_{AA} - 1) + 2f\beta_1 - \beta_1 + \beta_0 \end{aligned} \quad (\text{S14})$$

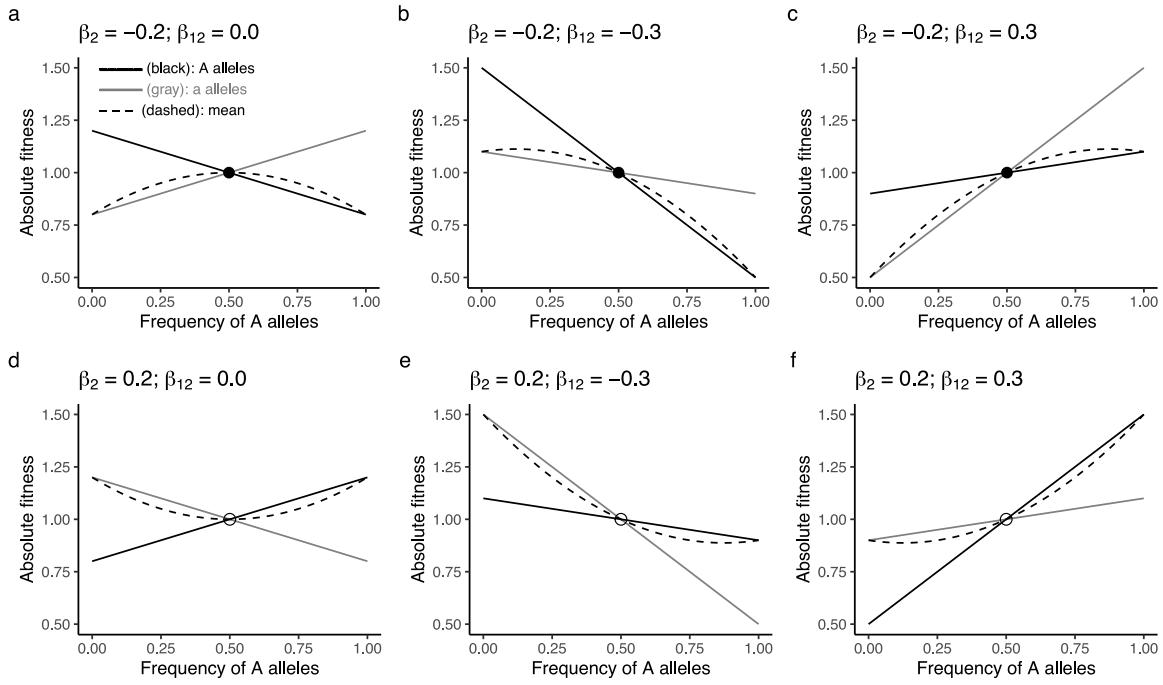


Figure S3: Numerical examples of fitness values y_i in response to allele frequency when only the AA and aa genotypes exist without mating [Equations (S11a) and (S11b) in Appendix S2]. Black and gray lines indicate the fitness functions for the AA and aa genotypes, respectively. The dashed curves indicate the mean fitness per population; that is, Equation (S12) in Appendix S2. (a) Symmetric negative frequency-dependent selection (FDS); (d) symmetric positive FDS; (b and c) asymmetric negative FDS; and (e and f) asymmetric positive FDS. Closed and open circles indicate a stable or unstable state, respectively. No directional selection i.e., $\beta_0 = 1.0$ and $\beta_1 = 0$ are assumed for all panels to present fitness functions of FDS.

Appendix S3. Mixed model extension

To implement GWAS, we modified Equations (2) and (3) as a linear mixed model (LMM) that considered genetic relatedness as a random effect (KANG *et al.*, 2008). In terms of Henderson's mixed model (HENDERSON *et al.*, 1959), such GWAS models have the same structure as phylogenetic comparative methods that analyze interspecific phenotypic variation among phylogenetic trees (KANG *et al.*, 2008; HADFIELD and NAKAGAWA, 2010). To modify Equation (2) as a LMM, we designated the genetic-related matrix as \mathbf{A} and introduced random effects u_i to Equation (2) as follows:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 \sum_{j=1}^{N_k} x_i x_j + u_i + e_i \quad (\text{S15})$$

where a vector including u_i for n individuals followed a normal distribution as $u_i \in \mathbf{u}$ and $\mathbf{u} \sim \text{Norm}(\mathbf{0}, \sigma_1^2 \mathbf{A}_1 + \sigma_2^2 \mathbf{A}_2)$. The residual e_i is expressed as $e_i \in \mathbf{e}$ and $\mathbf{e} \sim \text{Norm}(\mathbf{0}, \sigma_e^2 \mathbf{I})$. The $n \times n$ variance-covariance matrices denote the individual genetic relatedness or the entire genotype similarity among n individuals as $\mathbf{A}_1 = \frac{1}{2(q-1)} \mathbf{X}_1^\top \mathbf{X}_1 + \frac{1}{2}$ and $\mathbf{A}_2 = \frac{1}{(q-1)} \mathbf{X}_2^\top \mathbf{X}_2$, where q denotes the number of loci. The elements of n individuals $\times q$ loci matrix \mathbf{X}_1 consist of explanatory variables of the individual genotype values. As we defined $x_i = (-1, 1)$, the genetic-related matrix \mathbf{A}_1 was scaled to represent the proportion of loci shared among $n \times n$ individuals. The elements of n individuals $\times q$ loci matrix \mathbf{X}_2 consist of the genotype similarity as:

$$\mathbf{X}_2 = \begin{pmatrix} (\sum_{j=1}^{N_k} x_{1,1} x_j)/N_k & (\sum_{j=1}^{N_k} x_{1,2} x_j)/N_k & \dots & (\sum_{j=1}^{N_k} x_{1,n} x_j)/N_k \\ (\sum_{j=1}^{N_k} x_{1,2} x_j)/N_k & (\sum_{j=1}^{N_k} x_{2,2} x_j)/N_k & \dots & (\sum_{j=1}^{N_k} x_{2,n} x_j)/N_k \\ \dots & \dots & \dots & \dots \\ (\sum_{j=1}^{N_k} x_{q,1} x_j)/N_k & (\sum_{j=1}^{N_k} x_{q,2} x_j)/N_k & \dots & (\sum_{j=1}^{N_k} x_{q,n} x_j)/N_k \end{pmatrix}$$

where the $n \times n$ matrix \mathbf{A}_2 indicates a sample structure related to genotype similarity. The variance component parameters σ_1^2 and σ_2^2 determine the relative contributions of \mathbf{A}_1 and \mathbf{A}_2 to the vector of random effects \mathbf{u} .

To incorporate asymmetric FDS into GWAS, we considered a sample structure based on the asymmetric FDS in LMM. Here, we extended Equation (3) into LMM as

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{N_k} \sum_{j=1}^{N_k} x_i x_j + \frac{\beta_{12} x_i}{N_k} \sum_{j=1}^{N_k} x_i x_j + u_i + e_i \quad (\text{S16})$$

where the random effect u_i is redefined as $u_i \in \mathbf{u}$ and $\mathbf{u} \sim \text{Norm}(\mathbf{0}, \sigma_1^2 \mathbf{A}_1 + \sigma_2^2 \mathbf{A}_2 + \sigma_{12}^2 \mathbf{A}_{12})$. The additional $n \times n$ variance-covariance matrix \mathbf{A}_{12} denotes a sample structure because of the asymmetric effects among n individuals as $\mathbf{A}_{12} = \frac{1}{(q-1)} \mathbf{X}_{12}^\top \mathbf{X}_{12}$. The elements of n individuals $\times q$ loci matrix \mathbf{X}_{12} consist of explanatory variables for the asymmetric effects as follows:

$$\mathbf{X}_{12} = \begin{pmatrix} (x_{1,1} \sum_{j=1}^{N_k} x_{1,1} x_j)/N_k & (x_{1,2} \sum_{j=1}^{N_k} x_{1,2} x_j)/N_k & \dots & (x_{1,n} \sum_{j=1}^{N_k} x_{1,n} x_j)/N_k \\ (x_{1,2} \sum_{j=1}^{N_k} x_{1,2} x_j)/N_k & (x_{2,2} \sum_{j=1}^{N_k} x_{2,2} x_j)/N_k & \dots & (x_{2,n} \sum_{j=1}^{N_k} x_{2,n} x_j)/N_k \\ \dots & \dots & \dots & \dots \\ (x_{q,1} \sum_{j=1}^{N_k} x_{q,1} x_j)/N_k & (x_{q,2} \sum_{j=1}^{N_k} x_{q,2} x_j)/N_k & \dots & (x_{q,n} \sum_{j=1}^{N_k} x_{q,n} x_j)/N_k \end{pmatrix}$$

The additional parameter of the variance component σ_{12}^2 compares the relative importance of the asymmetric effects with those of the individual genotype effects σ_1^2 and genotype similarity effects σ_2^2 . Additionally, the individual-level formula Equation (S16) can also be converted into a common matrix form (HENDERSON *et al.*, 1959) as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (\text{S17})$$

where \mathbf{y} is $n \times 1$ fitness vector with $y_i \in \mathbf{y}$; \mathbf{X} is a matrix of fixed effects, including a unit vector, individual genotype x_i , genotype similarity covariate $(\sum_{j=1}^{N_k} x_i x_j)/N_k$, and other covariates for n individuals; $\boldsymbol{\beta}$ is a vector that included coefficients of the fixed effects; \mathbf{Z} is a design matrix allocating individuals to a genotype; \mathbf{u} is the random effect as $\text{Var}(\mathbf{u}) = \sigma_1^2 \mathbf{A}_1 + \sigma_2^2 \mathbf{A}_2 + \sigma_{12}^2 \mathbf{A}_{12}$; and \mathbf{e} is the residual as $\text{Var}(\mathbf{e}) = \sigma_e^2 \mathbf{I}$.

To efficiently solve LMMs, we first estimated σ_1^2 and σ_2^2 without any fixed effects (i.e., null model) using the average-information restricted maximum likelihood (AI-REML) method implemented in the *gaston* package (PERDRY and DANDINE-ROULLAND, 2020). Then, to compare the significance of β_1 and β_2 to the null model, we solved Equation (S17) using fast approximation by eigenvalue decomposition on a random effect matrix $\hat{\sigma}_1^2 \mathbf{A}_1 + \hat{\sigma}_2^2 \mathbf{A}_2 + \hat{\sigma}_{12}^2 \mathbf{A}_{12}$. The likelihood ratio test was used to compare the models with and without β_2 . The standard GWAS is a subset of Equation (S15) when $\beta_2 = 0$ and $\sigma_2^2 = 0$ (SATO *et al.*, 2021b); thus, we set β_2 and σ_2^2 to zero when testing β_1 . From results of this standard GWAS, we can also estimate the genomic heritability as $h^2 = \hat{\sigma}_1^2 / (\hat{\sigma}_1^2 + \hat{\sigma}_e^2)$. Because the individual genotype variable x_i and the genotype similarity variable $(\sum_{j=1}^{N_k} x_i x_j)/N_k$ were strongly correlated when the MAF was very small (SATO *et al.*, 2021b), likelihood ratio tests should be performed one-by-one for each parameter in Equations (S15) and (S16). Such stepwise likelihood ratio

tests were implemented using the rNeighborGWAS package version 1.2.4 (SATO *et al.*, 2021b). The `nei_lmm()` function was used for both the GWAS simulation and data analysis. The "asym = TRUE" option was used when testing the asymmetric effects β_{12} .

Appendix S4. Details of the GWAS simulation

We performed simulations as summarized in the main text to examine the statistical power of the linear mixed model Equation (S15) in GWAS (see the methods "**GWAS of simulated data**"). The entire procedure consisted of three steps: We simulated genomic structure under four different scenarios of selection (see "**Simulated genomes**" below), conducted virtual experiments to simulate fitness from the simulated genomes (see "**Virtual experiments**" below), and applied our method for the GWAS of simulated fitness and genomes (see "**GWAS using simulated genomes and fitness**" below).

Simulated genomes

To generate a realistic genome structure, we performed population genetic simulations using SLiM version 3 (HALLER and MESSER, 2019). By running 30 independent iterations for 2000 generations, we simulated 50 kb nucleotide \times three chromosomes \times 10 subpopulations \times 200 individuals. Base parameters were set as follows: mutation rate $\mu = 10^{-6}$, selection coefficient $s_1 = s_2 = 0.1$ for non-neutral mutations, and recombination rate $r = 10^{-5}$. Ten subpopulations were distributed in a circle with a low migration rate $m = 10^{-4}$ between neighboring populations. In this simulation, we decomposed the total fitness as $w_i = w_{i,1} + w_{i,2}$. The first fitness component $w_{i,1}$ involves individual genotype effects β_1 , and $w_{i,2}$ is the second fitness component subject to the power analysis of genotype similarity effects β_2 . For individual genotype effects β_1 , we define stabilizing selection as $w_{i,1} = 1.5 - [(z_{i,1} - z_1^*)^2 / s_1 N_k]$, where z_1^* is the optimum number of QTLs responsible for individual genotype effects per genome per population and z_i is the number of QTLs for individual i . We set z_1^* to 5 and assumed additive effects by the QTLs. Following the standard way to simulate polygenic selection (HALLER and MESSER, 2019), we did not substitute QTLs for $w_{1,i}$ even after they were fixed. For the genotype similarity effects β_2 , we assumed four specific scenarios of selection on the second fitness component $w_{i,2}$ as follows:

Scenario 1. Negative frequency-dependent selection: For the test of β_2 , we first simulated negative FDS on the second fitness component $w_{i,2}$. We simulated negative FDS for the second fitness component as $w_{i,2} = 1.5 - s_2 g_i f_k$, where f_k indicates the frequency of the mutation within a subpopulation k . Novel mutations were assumed to be recessive to ancestral alleles with genotype g_i redefined as $g_i \in \{\text{AA}, \text{Aa}, \text{aa}\} = \{1, 1, 0\}$. Polymorphisms were likely balanced under negative FDS; thus, the mutation rate was set at half of the base parameter to maintain the number of causal SNPs in the same order as that in the other scenario.

Scenario 2. Positive frequency-dependent selection: We also simulated the opposite regime to negative FDS i.e., positive FDS on the second fitness component $w_{i,2}$. It is known that locally acting positive FDS within a subpopulation can lead to global coexistence of two alleles among subpopulations (MOLOFSKY *et al.*, 2001). Therefore, we separated the entire population into four panmictic subpopulations to simulate polymorphic loci. Similar to the simulation of negative FDS, we simulated positive FDS as $w_{i,2} = 1.5 + s_2 g_i f_k$. Novel mutations were assumed to be recessive to ancestral alleles with genotype g_i redefined as $g_i \in \{\text{AA, Aa, aa}\} = \{1, 1, 0\}$.

Scenario 3. Overdominance: To test whether β_2 confounded frequency-independent types of balancing selection, we simulated genomes under overdominance selection. The second fitness component is defined as $w_{i,2} = 1.5 + s_2 h g_i$, where h is the dominance coefficient expressed on the basis of genotypes as $\{h_{\text{AA}}, h_{\text{Aa}}, h_{\text{aa}}\} = \{1.0, 2.0, 1.0\}$.

Scenario 4. Spatiotemporally varying selection: To test another type of frequency-independent balancing selection, we simulated genomes under spatiotemporally varying selection. The second fitness component is defined as $w_{i,2} = 1.5 + s g_i$, where s varies in space and time. We assumed $s_2 = 0.1$ for two subpopulations, and $s_2 = -0.1$ for the other two subpopulations. For six of the four subpopulations, we changed the selection coefficient in time to $s_2 = 0.1$ for odd generations and $s_2 = -0.1$ for even generations. We reset $m = 0.001$ to allow a higher migration rate that could ensure spatially varying selection. In this scenario, novel mutations were assumed to be recessive to ancestral alleles.

As a result, the simulated genomes had 2,000 to 4,500 SNPs with MAFs > 0.01 across 50 kbp nucleotide sequences (Fig. S4). They exhibited low heterozygosity ($H_t < 0.1$) and moderate to strong differentiation among 10 populations ($G_{st} < 0.5$; Fig. S4c), where approximately 200 SNPs were involved in polygenic stabilizing selection (Fig. S4a). Regarding the causal SNPs, SNPs responsible for positive FDS showed strong population genetic structures ($G_{st} > 0.8$) with low heterozygosity ($H_t < 0.05$; Fig. S4b) because positive FDS reduces polymorphisms within a population. In contrast, since negative FDS maintains polymorphisms within a population, SNPs responsible for negative FDS had weak population genetic structures ($G_{st} < 0.2$) with high heterozygosity ($H_t > 0.35$; Fig. S4b).

Virtual experiments

To simulate individual fitness, we sampled the simulated genomes and generated fitness values from the genotype data. The simulated genomes were exported in variant call format (.vcf) and loaded into R with the gaston (PERDRY and DANDINE-ROULLAND, 2020) and the

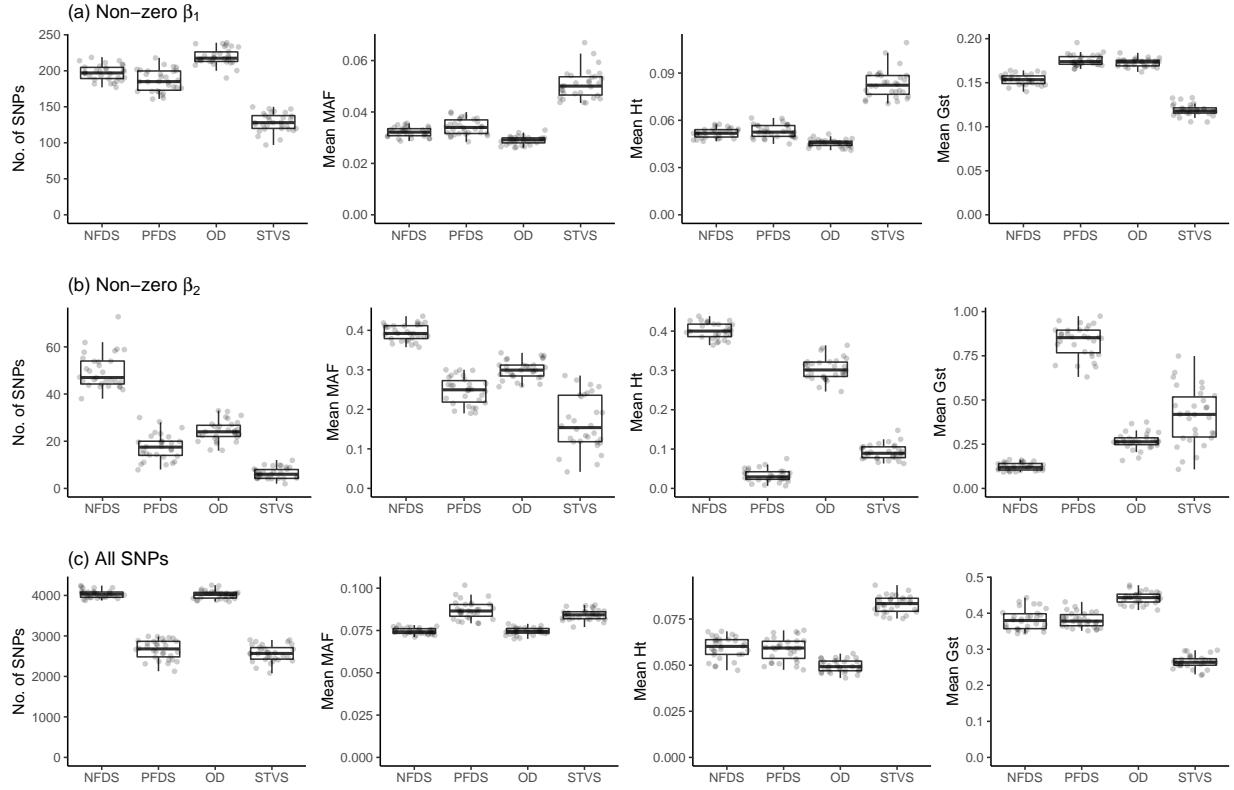


Figure S4: Structure of simulated genomes regarding the loci responsible for stabilizing selection (a), the other forms of selection (b), and genome-wide single nucleotide polymorphisms [SNPs](c). Number of SNPs, mean minor allele frequency (MAF), mean heterozygosity (H_t), and mean fixation indices (G_{st}) are shown among 30 iterations for four scenarios of selection: NFDS, negative frequency-dependent selection; PFDS, positive frequency-dependent selection; OD, overdominance; STVS, spatiotemporally varying selection.

vcfR package (KNAUS and GRÜNWALD, 2017). SNPs were filtered with a cut-off threshold of a minor allele frequency (MAF) of 0.01. We tested two experimental settings: split and continuous populations (Fig. 1a). To describe the two distinct cases, 900 individuals were randomly sampled from each simulation without replacement and assigned to a 30×30 lattice space for the continuous setting, or 10 individuals each to 90 split cages for the split setting.

Fitness values were then simulated from the simulated genomes and their spatial arrangement. To calculate the individual genotype component $\beta_1 x_i$ in Equation (2), we assigned 0.1 (which corresponded to the selection coefficient s_1 in the population genetic simulation) to β_1 of causal SNPs or zero to those of the other SNPs. The second fitness component at causal

SNPs was generated with $\beta_2 = 0.1$ (corresponding to the strength of balancing selection s_2 in the population genetic simulation) for negative or positive FDS as $(\pm \beta_2 \sum_{j=1}^{N_k} x_i x_j)/2N_k$, for overdominance as $x_i \in \{\text{AA, Aa, aa}\} = \{1 + \beta_2, 1 + \beta_2 h, 1\}$, and for spatiotemporally varying selection as a random assignment of $\pm \beta_2$ to $\beta_2 x_i$. Fitness variance was not adjusted for the first and second fitness components since the number of causal SNPs and their effect sizes were controlled during the population genetic simulation above. Gaussian residual errors were finally added to the simulated fitness such that approximately one-third of the total phenotypic variation was attributed to the environmental variance as $\text{Var}(\mathbf{e}) = (0.75)^2 \times \text{Var}(\mathbf{w})$. This range of the number of causal SNPs and the proportion of phenotypic variation explained by the model were based on parameter settings where the model performance was well differentiated (SATO *et al.*, 2021b).

GWAS using simulated genomes and fitness

We performed association mapping with respect to β_1 and β_2 to detect FDS from the simulated fitness and genomes. The rNeighborGWAS package (SATO *et al.*, 2021b) was used to implement the regression model in Equation (2) as an LMM (Appendix S2). The false vs. true positive rates were analyzed using the receiver operating characteristic (ROC) curve. Like the generative model, we assumed the dominant encoding for the three genotypes, $x_i \in \{\text{AA, Aa, aa}\} = \{+1, +1, -1\}$. We calculated the area under the ROC curve (AUC) for the $-\log_{10}(p\text{-values})$ of β_1 or β_2 to quantify the efficiency of causal polymorphism detection. The AUC ranged from 0.5 (no power to detect causal SNPs) to 1.0 (perfect matching between the top p -value score and causal SNPs). We compared the true and estimated values of β_2 to quantify the accuracy of the effect size estimates. We also compared AUCs and $\hat{\beta}_2$ between LMM and LM to test whether LMMs could outperform standard LMs.

Figure 4 presented the results when we assumed the complete dominance of A alleles over a allele. Figure 4 showed the effectiveness of linear mixed models (LMMs) for estimating the positive and negative FDS. Further, we find that LMMs outperform standard linear models (LMs) in terms of the efficient detection of causal SNPs (cf. Fig. 4b, e and Fig. S5b, e) and the correct estimation of β_2 (cf. Fig. 4c, f and Fig. S5c, f). Yet, researchers often assume the additive effects of two alleles on quantitative traits such as fitness, in GWAS (e.g, GONDRO *et al.*, 2013; PERDRY and DANDINE-ROULLAND, 2020). Therefore, we examined the case of additive genotype encoding in this appendix. With the same simulation setting above, we fitted Equation (2) to the simulated fitness with the additive genotype encoding as $g_i \in \{\text{AA, Aa, aa}\} = \{+1, 0, -1\}$. Under this assumption of the additive effects, Figure S6 or

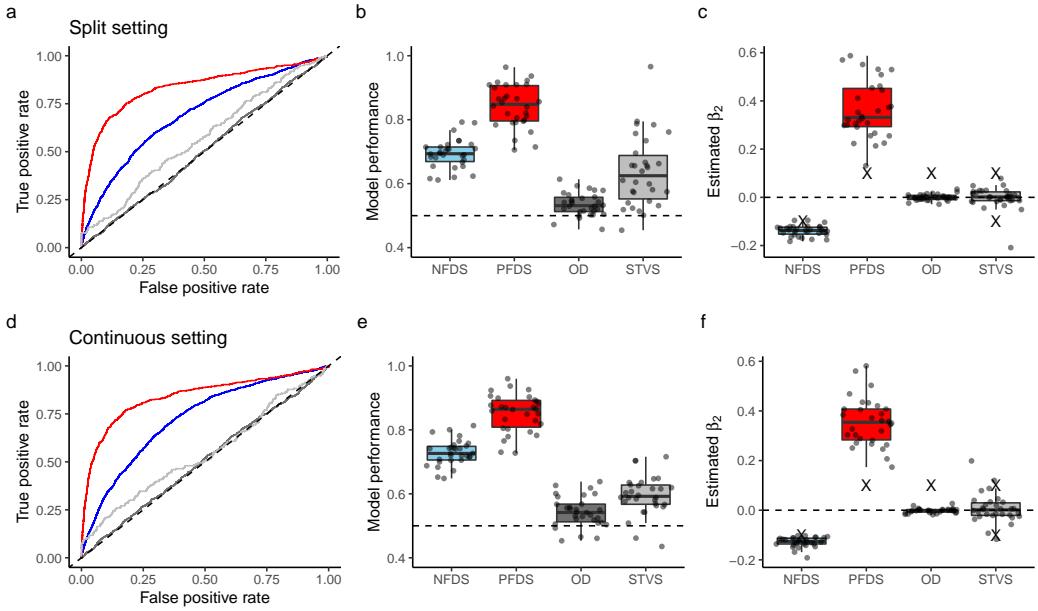


Figure S5: Performance of standard linear models to estimate the four types of simulated selections: NFDS, negative frequency-dependent selection; PFDS, positive frequency-dependent selection; OD, overdominance; and STVS, spatiotemporally varying selection. The complete dominance of A alleles over a alleles was assumed in this simulation. The top and bottom panels show the results of the split and continuous settings, respectively (Fig. 1a). (a and d) The receiver operating characteristic (ROC) curve shows the relationship between the true and false positive rate. Line colors indicate different simulation scenarios (blue, NFDS; red, PFDS; black, OD; gray, STVS). (b and e) Model performance evaluated by the area under the ROC curve (AUC). Dashed lines at 0.5 indicate no power to detect causal single nucleotide polymorphisms (SNPs). (c and f) Estimated β_2 of causal SNPs, where negative and positive values indicate negative and positive FDS, respectively. Cross marks indicate the true simulated magnitude of β_2 .

S7 present the results of LMMs or LMs, respectively. The LMMs retain high performance to detect both negative and positive FDS (median performance > 0.8 : Fig. S6) and show better performance than LMs to detect negative FDS (Fig. S6b, e; S7b, e). Although LMs and LMMs show similar performances in detecting positive FDS (Fig. S6b, e; S7b, e), the positive coefficient of $\hat{\beta}_2$ is more overestimated in LMs than in LMMs (Fig. S6c, f; S7c, f). These additional simulations suggest that LMMs are still preferred over LMs to suppress the overestimation and false positives.

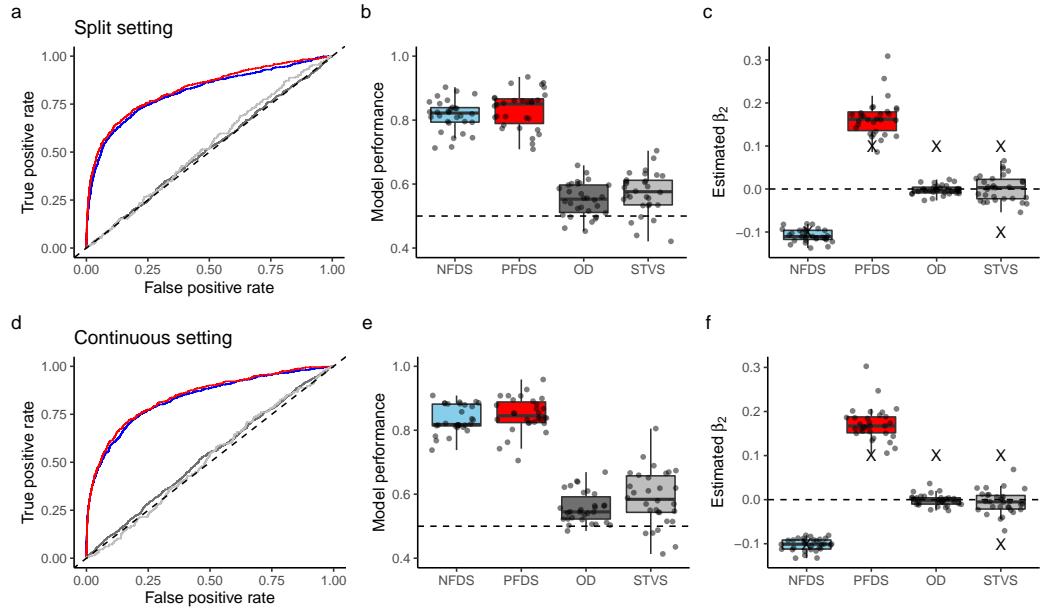


Figure S6: Performance of linear mixed models to estimate the four types of simulated selections: NFDS, negative frequency-dependent selection; PFDS, positive frequency-dependent selection; OD, overdominance; and STVS, spatiotemporally varying selection. The additive effects of two alleles on fitness were assumed in this simulation. The top and bottom panels show the results of the split and continuous settings, respectively (Fig. 1a). (a and d) Receiver operating characteristic (ROC) curves showing the relationship between the true and false positive rates. Line colors indicate different simulation scenarios (blue, NFDS; red, PFDS; black, OD; gray, STVS). (b and e) Model performance evaluated by the area under the ROC curve (AUC). Dashed lines at $AUC = 0.5$ indicate no power to detect causal single nucleotide polymorphisms (SNPs). (c and f) Estimated β_2 of causal SNPs, where negative and positive values indicate negative and positive FDS, respectively. Cross marks indicate the true simulated magnitude of β_2 .

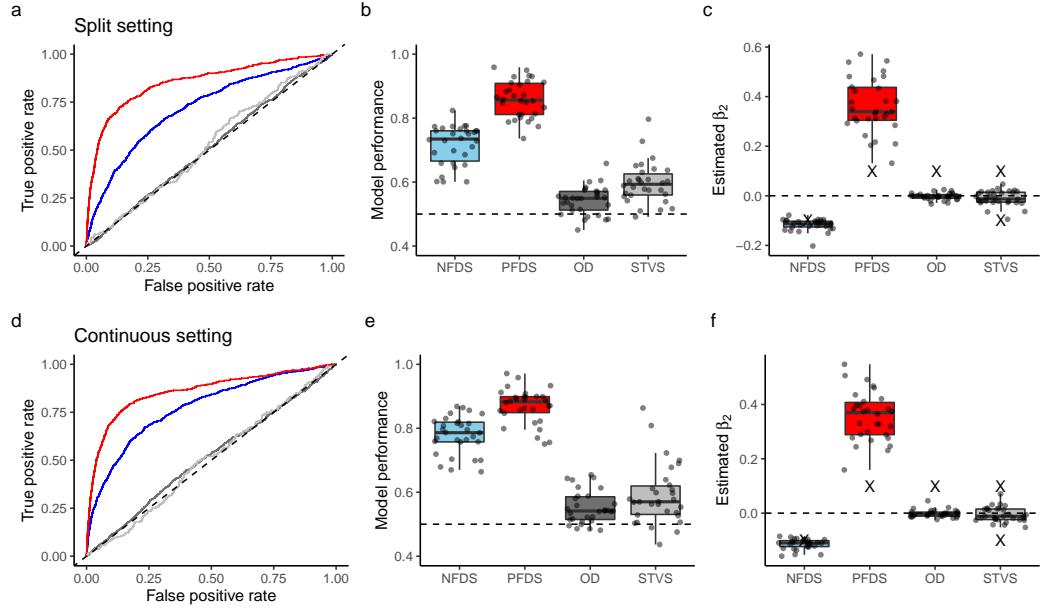


Figure S7: Performance of standard linear models to estimate the four types of simulated selections: NFDS, negative frequency-dependent selection; PFDS, positive frequency-dependent selection; OD, overdominance; and STVS, spatiotemporally varying selection. The additive effects of two alleles on fitness were assumed in this simulation. The top and bottom panels show the results of the split and continuous settings, respectively (Fig. 1a). (a and d) Receiver operating characteristic (ROC) curve, which indicates the relationship between the true and false positive rate. Line colors indicate different simulation scenarios (blue, NFDS; red, PFDS; black, OD; gray, STVS). (b and e) Model performance evaluated by the area under the ROC curve (AUC). Dashed lines at $AUC = 0.5$ indicate no power to detect causal single nucleotide polymorphisms (SNPs). (c and f) Estimated β_2 of causal SNPs, where negative and positive values indicate negative and positive FDS, respectively. Cross marks indicate the true simulated magnitude of β_2 .

Appendix S5. Details of GWAS experiment using *A. thaliana*

We measured the number of reproductive branches of *Arabidopsis thaliana* accessions to perform GWAS using empirical data, as summarized in the main text (see the methods "**GWAS using field-grown *Arabidopsis thaliana***"). According to SATO *et al.* (2019), we established a summer cohort composed of natural accessions with various life cycles to investigate the survival and reproduction under stressful environments. We used 199 worldwide accessions (Table S3) out of 2029 inbred lines genotyped by the RegMap (HORTON *et al.*, 2012) and 1001 Genomes project (ALONSO-BLANCO *et al.*, 2016). For the 199 accessions, 1,819,577 SNPs were selected at a cut-off threshold of $\text{MAF} > 0.05$. The seeds of the 199 accessions were sown on Jiffy-seven (33 mm in diameter) and stratified under constant dark conditions at a temperature of 4C° for 1 week. Germinated seedlings were then grown under a short-day condition (8 hours light: 16 h dark, 20C°) for 6 weeks to keep all the accession in the vegetative stage. Individual plants with Jiffy-pots were then potted into a new plastic pot (6 cm in diameter) filled with mixed soils of agricultural compost (Profi Substrat Classic CL ED73, Einheitserde Co.) and perlite with a 3:1 L ratio of perlite. Potted plants were transferred to the common garden at the Irchel Campus of the University of Zurich (Zurich, Switzerland: 47°23'N, 08°33'E) on 8 July 2019. In the field setting, a set of 199 accessions and an additional Col-0 accession were randomly assigned to each block without replacement. The 200 plants were set in plastic trays (10 × 40 cells in a continuous space) in a checkered pattern, which made each individual have four neighbors (i.e., $N_k = 4$). Three replicates of each randomized block were set > 1.5 m apart from each other. Of 561 plants, 181 bolted 2 weeks after they were transferred to the field. The genomic heritability of the reproductive branch number was as high as $h^2 = 0.68$, which indicates a genetic control of this trait.

We depicted quantile-quantile plots (QQ plots) between the expected and observed association score of $-\log_{10}(p\text{-value})$ to inspect whether the population genetic structure was corrected well. An ideal pattern is that only the top-scoring SNPs have far higher association score than expected. For example, S8c exhibited an almost flat association score as randomly expected, but only several highest-scoring SNPs had higher observed score than expected. If the observed association scores are higher than that randomly expected all across the SNPs, there is likely a confounding variable such as a population genetic structure that is responsible for such an inflation of p -values (e.g., Fig. S10a). In this context, QQ plots of LMMs are as flat as observed association scores exhibit almost randomly expected association scores for β_2 and β_{12} at $-\log_{10}(p\text{-value}) < 5$ (Fig. S8b, c). In contrast, LMs show

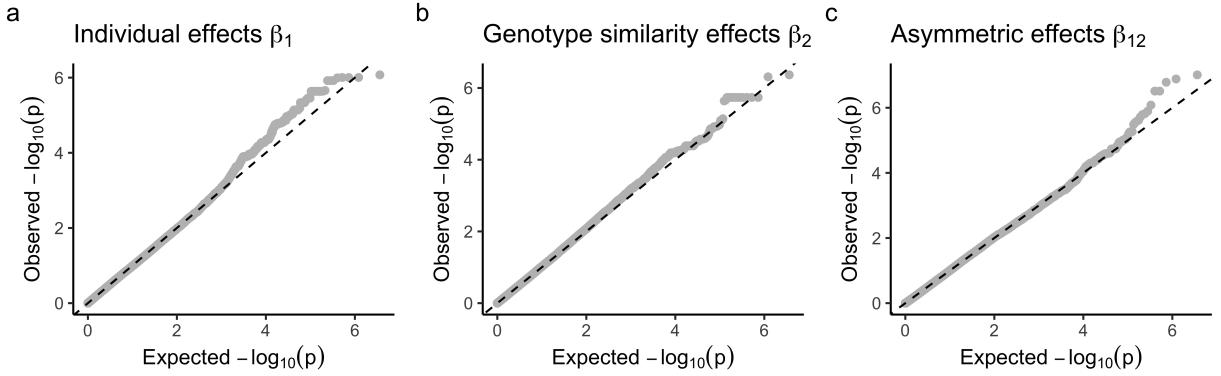


Figure S8: Quantile-quantile plots showing the observed and expected association score of the $-\log_{10}(p\text{-values})$ for the genome-wide association studies of the reproductive branch number in field-grown *A. thaliana*. The results obtained using the linear mixed models are shown. The left (a), middle (b), and right (c) panels display individual genotype, genotype similarity, and asymmetric effects, respectively. Dashed lines indicate the identity between the observed and expected association scores.

larger scores of $-\log_{10}(p\text{-value})$ than LMMs (Fig. S9a-c; Fig. 5a-c) and their QQ plot exhibits a large or slight inflation of the association score (Fig. S10). The association score of the individual genotype effects β_1 are severely inflated than randomly expected (Fig. S10a). These diagnostic analyses support the notion that LMMs are better than LMs to analyze empirical GWAS data.

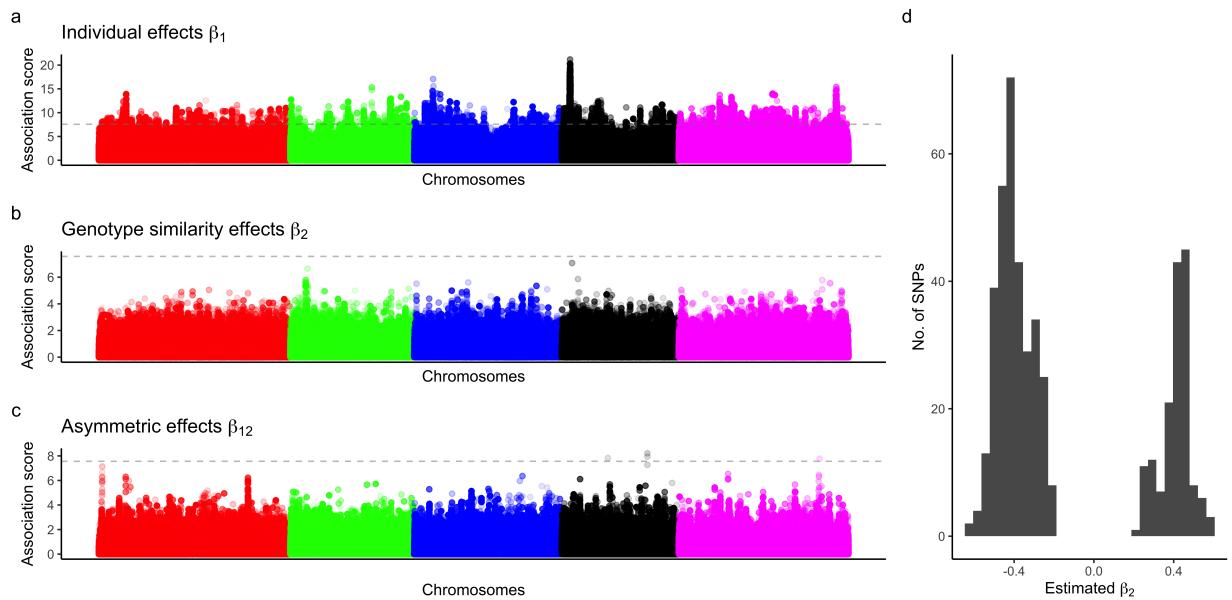


Figure S9: Genome-wide association studies of the reproductive branch number in field-grown *A. thaliana*. The results of standard linear models are presented. (a, b, and c) Manhattan plots showing the $-\log_{10}(p\text{-values})$ association scores for individual genotype effects, genotype similarity effects, and asymmetric effects, respectively. Horizontal dashed lines indicate p -value of < 0.05 , after Bonferroni correction. (d) Histogram of estimated β_2 among single nucleotide polymorphisms (SNPs) exhibiting p -values of $< 10^{-4}$. Negative and positive β_2 infer loci responsible for negative and positive frequency-dependent selection, respectively.

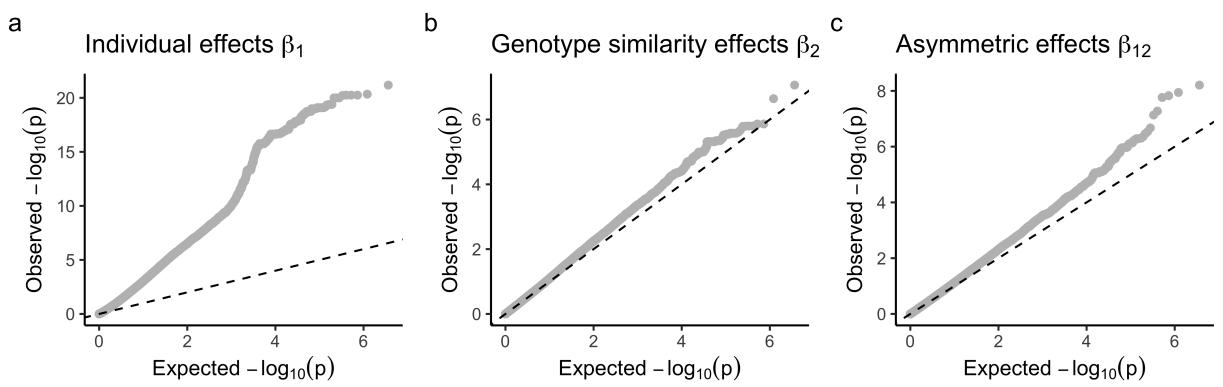


Figure S10: Quantile-quantile plots showing the observed and expected association score of the $-\log_{10}(p\text{-values})$ for the genome-wide association studies of branch number in field-grown *A. thaliana*. The results of standard linear models are presented. The left (a), middle (b), and right (c) panels display individual genotype, genotype similarity, and asymmetric effects, respectively. Dashed lines indicate the identity between the observed and expected association scores.

Supplementary Tables S3-S4 (see SuppTablesS3–S4.xlsx)

Table S3: List of plant accessions used for genome-wide association studies (GWAS) and their phenotypes.

Table S4: List of candidate genes related to individual genotype effects (a), genotype similarity effects (b), and asymmetric effects (c) on the reproductive branch number in *Arabidopsis thaliana*.

References

- ALONSO-BLANCO, C., J. ANDRADE, C. BECKER, F. BEMM, J. BERGELSON *et al.*, 2016
1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**: 481–491.
- ÁLVAREZ-CASTRO, J. M. and G. ÁLVAREZ, 2005 Models of general frequency-dependent selection and mating-interaction effects and the analysis of selection patterns in *Drosophila* inversion polymorphisms. *Genetics* **170**: 1167–1179.
- ANTONOVICS, J. and N. C. ELLSTRAND, 1984 Experimental studies of the evolutionary significance of sexual reproduction. I. a test of the frequency-dependent selection hypothesis. *Evolution* **38**: 103–115.
- ASMUSSEN, M. A. and E. BASNAYAKE, 1990 Frequency-dependent selection: the high potential for permanent genetic variation in the diallelic, pairwise interaction model. *Genetics* **125**: 215–230.
- ATWELL, S., Y. S. HUANG, B. J. VILHJALMSSON, G. WILLEMS, M. HORTON *et al.*, 2010
Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631.
- BATES, D., M. MÄCHLER, B. BOLKER and S. WALKER, 2015 Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**: 1–48.

- BETHKE, G., A. THAO, G. XIONG, B. LI, N. E. SOLTIS *et al.*, 2016 Pectin biosynthesis is critical for cell wall integrity and immunity in *Arabidopsis thaliana*. *The Plant Cell* **28**: 537–556.
- BISHOP, C. M., 2006 *Pattern Recognition and Machine Learning*, chapter 11: Sampling Methods. Springer-Verlag, pp. 523–558.
- BORER, M., T. VAN NOORT, M. RAHIER and R. E. NAISBIT, 2010 Positive frequency-dependent selection on warning color in alpine leaf beetles. *Evolution* **64**: 3629–3633.
- BROMAN, K. W. and Š. SEN, 2009 Single-QTL analysis. In *A Guide to QTL Mapping with R/qtl*. Springer, 75–133.
- BROWNE, L. and J. KARUBIAN, 2016 Frequency-dependent selection for rare genotypes promotes genetic diversity of a tropical palm. *Ecology Letters* **19**: 1439–1447.
- BRUNET, J. and C. C. MUNDT, 2000 Disease, frequency-dependent selection, and genetic polymorphisms: experiments with stripe rust and wheat. *Evolution* **54**: 406–415.
- CALSBEEK, R., L. BONVINI and R. M. COX, 2010 Geographic variation, frequency-dependent selection, and the maintenance of a female-limited polymorphism. *Evolution: International Journal of Organic Evolution* **64**: 116–125.
- CHONG, V. K., H. F. FUNG and J. R. STINCHCOMBE, 2018 A note on measuring natural selection on principal component scores. *Evolution Letters* **2**: 272–280.
- CIPRA, B. A., 1987 An introduction to the Ising model. *The American Mathematical Monthly* **94**: 937–959.
- CLARKE, B. and P. O'DONALD, 1964 Frequency-dependent selection. *Heredity* **19**: 201–206.
- COCKERHAM, C. C., P. BURROWS, S. YOUNG and T. PROUT, 1972 Frequency-dependent selection in randomly mating populations. *The American Naturalist* **106**: 493–515.
- CONNELL, J., 1971 On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In P. den Boer and G. G.R., editors, *Dynamics of numbers in populations*. Center for Agricultural Publishing and Documentation, Wageningen, The Netherlands, 298–312.

COSMIDIS, N., M. LOUKAS and E. ZOUROS, 1999 Rarer need not be better if commoner is worse: frequency-dependent selection for developmental time at the alcohol dehydrogenase locus of the olive fruit fly, *Bactrocera oleae*. *Evolution* **53**: 518–526.

DURHAM, M. F., M. M. MAGWIRE, E. A. STONE and J. LEIPS, 2014 Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. *Nature Communications* **5**: 4338.

EXPOSITO-ALONSO, M., 500 GENOMES FIELD EXPERIMENT TEAM, H. A. BURBANO, O. BOSSDORF, R. NIELSEN *et al.*, 2019 Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature* **573**: 126–129.

FISHER, H. S., E. JACOBS-PALMER, J.-M. LASSANCE and H. E. HOEKSTRA, 2016 The genetic basis and fitness consequences of sperm midpiece size in deer mice. *Nature Communications* **7**: 13652.

FITZPATRICK, M. J., E. FEDER, L. ROWE and M. B. SOKOLOWSKI, 2007 Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature* **447**: 210–212.

GARRIDO, E., L. LLAMAS-GUZMÁN and J. FORNONI, 2016 The effect of frequency-dependent selection on resistance and tolerance to herbivory. *Journal of Evolutionary Biology* **29**: 483–489.

GIGORD, L. D., M. R. MACNAIR and A. SMITHSON, 2001 Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina* (L.) Soo. *Proceedings of the National Academy of Sciences USA* **98**: 6253–6255.

GOLDBERG, J. K., C. M. LIVELY, S. R. STERNLIEB, G. PINTEL, J. D. HARE *et al.*, 2020 Herbivore-mediated negative frequency-dependent selection underlies a trichome dimorphism in nature. *Evolution Letters* **4**: 83–90.

GONDRO, C., J. VAN DER WERF and B. HAYES, 2013 *Genome-wide association studies and genomic prediction*. Humana Press.

GROEN, S. C., I. CALIC, Z. JOLY-LOPEZ, A. E. PLATTS, J. Y. CHOI *et al.*, 2020 The strength and pattern of natural selection on gene expression in rice. *Nature* **578**: 572–576.

- HADFIELD, J. and S. NAKAGAWA, 2010 General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology* **23**: 494–508.
- HALLER, B. C. and P. W. MESSER, 2019 SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution* **36**: 632–637.
- HATAKEYAMA, K., M. WATANABE, T. TAKASAKI, K. OJIMA and K. HINATA, 1998 Dominance relationships between s-alleles in self-incompatible *Brassica campestris* l. *Heredity* **80**: 241–247.
- HEDRICK, P. W., 2007 Balancing selection. *Current Biology* **17**: R230–R231.
- HEISLER, I. L. and J. DAMUTH, 1987 A method for analyzing selection in hierarchically structured populations. *The American Naturalist* **130**: 582–602.
- HENDERSON, C. R., O. KEMPTHORNE, S. R. SEARLE and C. VON KROSGK, 1959 The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**: 192–218.
- HORTON, M. W., A. M. HANCOCK, Y. S. HUANG, C. TOOMAJIAN, S. ATWELL *et al.*, 2012 Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics* **44**: 212–216.
- JANZEN, D. H., 1970 Herbivores and the number of tree species in tropical forests. *The American Naturalist* **104**: 501–528.
- JOLY, S. and D. J. SCHOEN, 2011 Migration rates, frequency-dependent selection and the self-incompatibility locus in *Leavenworthia* (Brassicaceae). *Evolution* **65**: 2357–2369.
- KANG, H. M., N. A. ZAITLEN, C. M. WADE, A. KIRBY, D. HECKERMAN *et al.*, 2008 Efficient control of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- KAWAGOE, T., K. K. SHIMIZU, T. KAKUTANI and H. KUDOH, 2011 Coexistence of trichome variation in a natural plant population: a combined study using ecological and candidate gene approaches. *PLoS ONE* **6**: e22184.
- KNAUS, B. J. and N. J. GRÜNWALD, 2017 vcfR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17**: 44–53.

- LANDE, R. and S. J. ARNOLD, 1983 The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.
- LE ROUZIC, A., T. F. HANSEN, T. P. GOSDEN and E. I. SVENSSON, 2015 Evolutionary time-series analysis reveals the signature of frequency-dependent selection on a female mating polymorphism. *The American Naturalist* **185**: E182–E196.
- LEWIN, H. A., G. E. ROBINSON, W. J. KRESS, W. J. BAKER, J. CODDINGTON *et al.*, 2018 Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences USA* **115**: 4325–4333.
- LLAURENS, V., S. BILLIARD, J.-B. LEDUCQ, V. CASTRIC, E. K. KLEIN *et al.*, 2008 Does frequency-dependent selection with complex dominance interactions accurately predict allelic frequencies at the self-incompatibility locus in *Arabidopsis halleri*? *Evolution* **62**: 2545–2557.
- MCCAULEY, D. E. and M. T. BROCK, 1998 Frequency-dependent fitness in *Silene vulgaris*, a gynodioecious plant. *Evolution* **52**: 30–36.
- METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER and E. TELLER, 1953 Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**: 1087–1092.
- MITCHELL-OLDS, T. and R. G. SHAW, 1987 Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* **41**: 1149–1161.
- MOLOFSKY, J., J. D. BEVER and J. ANTONOVICS, 2001 Coexistence under positive frequency dependence. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**: 273–277.
- NOSIL, P., R. VILLOUTREIX, C. F. DE CARVALHO, T. E. FARKAS, V. SORIA-CARRASCO *et al.*, 2018 Natural selection and the predictability of evolution in *Timema* stick insects. *Science* **359**: 765–770.
- PERDRY, H. and C. DANDINE-ROULLAND, 2020 *gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models*. R package version 1.5.6.
- R CORE TEAM, 2019 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- ROBIN, X., N. TURCK, A. HAINARD, N. TIBERTI, F. LISACEK *et al.*, 2011 pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**: 77.
- SÁNCHEZ-GUILLÉN, R., H. VAN GOSSUM and A. CORDERO RIVERA, 2005 Hybridization and the inheritance of female colour polymorphism in two ischnurid damselflies (Odonata: Coenagrionidae). *Biological Journal of the Linnean Society* **85**: 471–481.
- SATO, Y. and H. KUDOH, 2017 Herbivore-mediated interaction promotes the maintenance of trichome dimorphism through negative frequency-dependent selection. *The American Naturalist* **190**: E67–E77.
- SATO, Y., R. SHIMIZU-INATSUGI, M. YAMAZAKI, K. K. SHIMIZU and A. J. NAGANO, 2019 Plant trichomes and a single gene *GLABRA1* contribute to insect community composition on field-grown *Arabidopsis thaliana*. *BMC Plant Biology* **19**: 163.
- SATO, Y., K. TAKEDA and A. J. NAGANO, 2021a Neighbor QTL: an interval mapping method for quantitative trait loci underlying plant neighborhood effects. *G3; Genes|Genomes|Genetics* **11**: jkab017.
- SATO, Y., E. YAMAMOTO, K. K. SHIMIZU and A. J. NAGANO, 2021b Neighbor GWAS: incorporating neighbor genotypic identity into genome-wide association studies of field herbivory. *Heredity* **126**: 597–614.
- SCHNEIDER, K. A., 2006 A multilocus-multiallele analysis of frequency-dependent selection induced by intraspecific competition. *Journal of Mathematical Biology* **52**: 483–523.
- SCHNEIDER, K. A., 2008 Maximization principles for frequency-dependent selection I: the one-locus two-allele case. *Theoretical Population Biology* **74**: 251–262.
- SCHNEIDER, K. A., 2010 Maximization principles for frequency-dependent selection II: the one-locus multiallele case. *Journal of Mathematical Biology* **61**: 95–132.
- SCHUTZ, W. and S. USANIS, 1969 Inter-genotypic competition in plant populations II. maintenance of allelic polymorphisms with frequency-dependent selection and mixed selfing and random mating. *Genetics* **61**: 875.
- SHIMIZU, K. K., 2002 Ecology meets molecular genetics in *Arabidopsis*. *Population Ecology* **44**: 0221–0233.

- SHIMIZU, K. K. and T. TSUCHIMATSU, 2015 Evolution of selfing: recurrent patterns in molecular adaptation. *Annual Review of Ecology, Evolution, and Systematics* **46**: 593–622.
- SIEWERT, K. M. and B. F. VOIGHT, 2017 Detecting long-term balancing selection using allele frequency correlation. *Molecular Biology and Evolution* **34**: 2996–3005.
- SVENSSON, E. I. and T. CONNALLON, 2019 How frequency-dependent selection affects population fitness, maladaptation and evolutionary rescue. *Evolutionary Applications* **12**: 1243–1258.
- TAKAHASHI, Y., K. KAGAWA, E. I. SVENSSON and M. KAWATA, 2014 Evolution of increased phenotypic diversity enhances population performance by reducing sexual harassment in damselflies. *Nature Communications* **5**: 4468.
- TAKAHASHI, Y., R. TANAKA, D. YAMAMOTO, S. NORIYUKI and M. KAWATA, 2018 Balanced genetic diversity improves population fitness. *Proceedings of the Royal Society B: Biological Sciences* **285**: 20172045.
- TAKAHASHI, Y., J. YOSHIMURA, S. MORITA and M. WATANABE, 2010 Negative frequency-dependent selection in female color polymorphism of a damselfly. *Evolution* **64**: 3620–3628.
- TOGNINALLI, M., U. SEREN, D. MENG, J. FITZ, M. NORDBORG *et al.*, 2018 The AraG-WAS catalog: a curated and standardized *Arabidopsis thaliana* GWAS catalog. *Nucleic Acids Research* **46**: D1150–D1156.
- TROTTER, M. V. and H. G. SPENCER, 2007 Frequency-dependent selection and the maintenance of genetic variation: exploring the parameter space of the multiallelic pairwise interaction model. *Genetics* **176**: 1729–1740.
- VAN GOSSUM, H., R. STOKS and L. DE BRUYN, 2001 Frequency-dependent male mate harassment and intra-specific variation in its avoidance by females of the damselfly *Ischnura elegans*. *Behavioral Ecology and Sociobiology* **51**: 69–75.
- WEINER, J., 1990 Asymmetric competition in plant populations. *Trends in Ecology & Evolution* **5**: 360–364.
- WRIGHT, S., 1948 On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution* **2**: 279–294.