

Reviewed Preprint

v1 • June 30, 2026

Not revised

✉ For correspondence:

yassato@ees.hokudai.ac.jphamazaki@ut-biomet.org

Competing interests: No competing interests declared

Funding: See [page 23](#)

Reviewing editor: Jonathan Flint, University of California, Los Angeles, United States

© 2026, Sato & Hamazaki. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Dissecting oligogenic and polygenic indirect genetic effects through the lens of neighbor genotypic identity

Yasuhiro Sato¹✉, Kosuke Hamazaki²✉

¹Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan • ²Institute for Advanced Academic Research (IAAR), Chiba University, Chiba, Japan

eLife Assessment

This **useful** study introduces a statistical model and accompanying software for jointly analysing how an organism's own genotype, and those of its neighbors, shape its traits (assessing both direct and indirect genetic effects), based on simulations and three datasets from plants. The implementation and its behavior on simulated data are **solid**, but the evidence that the approach is more powerful, more interpretable, or more novel than established alternatives is **incomplete**, because the authors do not benchmark against existing methods, nor validate the candidate genes they identify, nor test realistic scenarios in which neighbor effects are weaker than direct effects. The work will be of interest to quantitative geneticists and plant breeders studying competition among neighboring genotypes.

<https://doi.org/10.7554/eLife.111650.1.sa3>

Abstract

Individual phenotypes often depend on the genotypes of other individuals within a group. These phenomena are termed ‘indirect genetic effects’ (IGEs) and have been distinguished from direct genetic effects (DGEs) using quantitative genetic models. Recent studies have utilized high-resolution polymorphism data to enable genomic prediction (GP) and genome-wide association study (GWAS) of IGEs, but unified methods remain limited. Here we integrate polygenic and oligogenic IGEs using a multi-kernel mixed model incorporating two random effects with a single covariance parameter. Underlying this implementation, the Ising model of ferromagnetics enabled us to simplify locus-wise and background IGEs for GWAS and GP, respectively. Our simulations demonstrated that, while the previous and present models exhibited similar performance, the present model can infer a trade-off between DGEs and IGEs. By applying this method to three species of woody plants, we found evidence for intergenotypic competition in aspen and apple trees, but limited evidence in climbing grapevines. Based on GWAS, we also detected significant variants associated with the competitive IGEs on the apple trunk growth. Our study offers a flexible implementation for GWAS/GP of IGEs, thereby providing an effective tool to dissect the genetic architecture of group performance.

Introduction

Individual phenotypes often depend not only on an individual's own genotype but also on the genotypes of other individuals within a group. This phenomenon, termed ‘indirect genetic effects’ (IGEs), is critical for evolutionary trajectories, as it may constrain the response of direct genetic effects (DGEs) to selection [1–3]. If DGEs and IGEs coevolve, this offers a valuable perspective for plant and animal breeding to optimize group performance by mitigating intraspecific competition [4–6]. However, functional traits responsible for IGEs remain elusive [7]. For example, a set of crop traits designed to reduce competition, termed an ideotype [8], includes stem height, leaf angle, root

length and other dimensions of plant morphology [9,10]. To address the difficulty in identifying intermediate traits responsible for IGEs, many researchers have utilized genotype-based IGE models, which allow for trait-blind analyses of IGEs [1,5,7,11].

The application of genotype-based IGE models has been enhanced by the increasing availability of high-resolution and population-wide data of single nucleotide polymorphisms (SNPs) [3,12]. For instance, recent studies have conducted genome-wide association study (GWAS) and genomic prediction (GP) to dissect the genetic architecture of IGEs in humans [13,14], mice [7,15], *Arabidopsis thaliana* [11], and crops [16,17]. In laboratory mice, IGEs from cage mates have been analyzed for over 100 phenotypes and transcriptomes [7], where GWAS identified 24 quantitative trait loci (QTLs) underlying IGEs [15]. Besides mobile animals, IGEs have been analyzed for the model plant *A. thaliana* and crops to estimate mixing abilities among neighboring genotypes [16,18,19]. These pioneering studies provide the first step to implement GWAS and GP of IGEs. Yet, flexible implementation has not been available for joint modeling of polygenic and oligogenic IGEs.

In plants, IGEs occur among neighboring individuals as they cannot escape from their neighbors. These neighboring IGEs include multiple types of ecological interactions such as competition and facilitation [20–23]. To detect IGEs at a locus level, our previous study integrated the Ising model of ferromagnetics into a linear mixed model, named ‘Neighbor GWAS’ [24]. Analogous to physical interactions between two magnets, Neighbor GWAS formulated frequency-dependent selection at each SNP locus [25]. To reduce false positive signals, we employed a multi-kernel mixed model to correct for the population structure shaped by additional factors such as IGEs. This multi-kernel mixed model enables an efficient implementation called Reliable Association INference By Optimizing Weights (RAINBOW) [26], thereby allowing us to perform GP and GWAS by partitioning variance components and association scores between DGEs and IGEs.

In this study, we aimed to develop a flexible model of polygenic and oligogenic IGEs on a complex phenotype. We first presented a genotype-based IGE model by integrating Neighbor GWAS with RAINBOW. Subsequently, we conducted simulations to validate the performance of the proposed model against the original model. The proposed model was finally applied to three datasets of woody plant species. Based on the simulations and applications, we illustrated the feasibility of GWAS/GP of IGEs on a complex phenotype.

Results

Model development

To enable flexible modeling of DGEs and IGEs, we developed a multi-kernel mixed model encompassing oligogenic and polygenic effects on a complex phenotype. These oligogenic and polygenic effects were modeled as fixed and random effects, where the former and latter realized GWAS and GP, respectively. To model neighbor genotypic interactions, let us assume that a focal individual i are interacting with neighboring individuals j in a local space k , which are designated as k_i and k_j (Fig. 1). Following the original model of Neighbor GWAS [24], we then formulate IGEs of neighboring genotypes at q -th SNP locus on focal k_i -th individual phenotype y_{k_i} as follows.

$$y_{k_i} = \beta_0 + \beta_{q,1}x_{q,k_i} + \beta_{q,12} \sum_{j=1}^{J_k} x_{q,k_j} / J_k + \beta_{q,2} \sum_{k_j=1}^J (x_{q,k_i}x_{q,k_j}) / J_k + u_{k_i,1} + u_{k_i,2} + e_{k_i}$$

, or

$$\mathbf{y} = \mathbf{X}_0\beta_0 + \mathbf{X}_1\beta_1 + \mathbf{X}_{12}\beta_{12} + \mathbf{X}_2\beta_2 + \mathbf{Z}\mathbf{u}_1 + \mathbf{Z}\mathbf{u}_2 + \mathbf{e}$$

, where

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2\mathbf{K}_1 & \rho_{12}\sigma_1\sigma_2\mathbf{K}_{12} \\ \rho_{12}\sigma_1\sigma_2\mathbf{K}_{21} & \sigma_2^2\mathbf{K}_2 \end{pmatrix} \right] \tag{Eq.1}$$

As for the fixed effects, x_{q,k_i} and x_{q,k_j} represent the genotype of focal i -th and neighbor j -th individuals in the local space k (where J_k is the total number of neighboring individuals). At the focal locus q , the two homozygotes are encoded as $aa = -1$ and $AA = +1$ with heterozygote as $Aa = 0$, i.e., $x_{q,k_i} \in \{AA, Aa, aa\} = \{+1, 0, -1\}$. Accordingly, the fixed-effect coefficient $\beta_{q,2}$ indicates the

influence of mean allelic (di)similarity [i.e., $(-1) \times (+1) = (+1) \times (-1) = -1$ or $(+1) \times (+1) = (-1) \times (-1) = +1$] between focal and all neighboring individuals on y_{k_i} . This formulation is analogous to the Ising model of ferromagnetics [27] when inter-genotypic interactions occur among neighboring individuals in a lattice space (e.g., sessile plants). In this analogy, $\beta_{q,2}$ represents physical interactions among the south/north dipole of neighboring magnets while $\beta_{q,1}$ corresponds to an external energy (Appendix S1 [24]; Fig. S1 [24]). In the context of evolutionary biology, the coefficient $\beta_{q,2}$ represents positive or negative frequency-dependent selection when inter-genotypic interactions occur within a group [25] (Appendix S2 [24] and S3 [24]; Fig. S2 [24]), which gives a specific case of population genetic models of frequency-dependent selection [28]. In addition to $\beta_{q,1}$ and $\beta_{q,2}$, $\beta_{q,12}$ is defined as an interaction term between $\beta_{q,1}$ and $\beta_{q,2}$ because $x_{q,k_i} \sum_{k_j=1}^{J_k} (x_{q,k_i} x_{q,k_j}) / J_k = x_{q,k_i}^2 \sum_{k_j=1}^{J_k} (x_{q,k_j}) / J_k = \sum_{k_j=1}^{J_k} (x_{q,k_j}) / J_k$ due to $\{AA, Aa, aa\} = \{+1, 0, -1\}$ where $x_{q,k_i}^2 = 1$ or $x_{q,k_i}^2 = 0$ when x_{q,k_i} is heterozygous (Appendix S2 [24]). Theoretically, the sign of $\beta_{q,2}$ determines whether the weighted mean of phenotypic values \bar{y} can be concave or convex against neighboring allele frequency (Fig. S2a and d [24]), whereas $\beta_{q,12}$ determines the slope of \bar{y} along the allele frequency (Fig. S2b-c and e-f [24]). Although bistable evolutionary equilibria become possible in the presence of heterozygote (Fig. S2b and e [24]; cf. Fig. S2 of Sato et al. [25]), this property may not be relevant to cultivars for which we can manipulate genotype frequencies and their spatial arrangement. Therefore, the three coefficients $\beta_{q,1}$, $\beta_{q,2}$, $\beta_{q,12}$ can be used as key parameters to distinguish between SNP-wise DEGs and IGEs in GWAS, which underpin the mean behavior of phenotypic values and consequent group performance under allelic mixture at a given locus.

Besides the fixed effects, Eq. 1 [24] includes random effects that can represent polygenic effects (or best linear unbiased predictor: BLUP) for DGEs and IGEs. In the multi-variate normal (MVN) distribution, $\sigma_1^2 (> 0)$ and $\sigma_2^2 (> 0)$ denote genetic variance of DGEs and IGEs, respectively. The design matrix \mathbf{Z} assigns individuals and genotypes, which should be an identify matrix \mathbf{I} because inter-genotypic interactions should occur among individuals. The $N \times N$ kinship matrix \mathbf{K}_1 represents polygenic DGEs (where N means the total number of individuals), whereas one of the other variance-covariance matrices \mathbf{K}_2 represents polygenic IGEs. These two $N \times N$ variance-covariance matrices are defined by the cross-products $\mathbf{K}_1 = \frac{1}{2(Q-1)} \mathbf{X}_1 \mathbf{X}_1^\top + \frac{1}{2}$ and $\mathbf{K}_2 = \frac{1}{(Q-1)} \mathbf{X}_2 \mathbf{X}_2^\top$ (where Q means the number of SNP loci). The n -th and q -th element of the SNP matrix \mathbf{X}_1 corresponds to the self-genotypic value as $(\mathbf{X}_1)_{q,n} = x_{q,k_i}$. As we define $x_{q,k_i} \in (-1, 0, 1)$, \mathbf{K}_1 is scaled so that its element represents the proportion of SNP loci shared among individuals. The n -th and q -th element of N individuals $\times Q$ SNPs matrix \mathbf{X}_2 is given as $(\mathbf{X}_2)_{n,q} = (\sum_{j=1}^{J_n} x_{q,n_i} x_{q,n_j}) / J_n$. Accordingly, the k -th and l -th element of $N \times N$ matrix \mathbf{K}_2 is given as $(\mathbf{K}_2)_{k,l} = \sum_{q=1}^Q [(x_{q,k_i} x_{q,l_i}) \sum_{j=1}^{J_k} (x_{q,k_j}) \sum_{j=1}^{J_l} (x_{q,l_j})] / (Q-1)$. This indicates the polygenic effects (i.e., individual genetic similarity $x_{q,k_i} x_{q,l_i}$) weighted by global allelic similarity between two local spaces k and l (i.e., a similarity metric of the mean allelic similarity in two local spaces $\sum_{j=1}^{J_k} (x_{q,k_j}) \sum_{j=1}^{J_l} (x_{q,l_j})$) (Fig. 1b [24]), which is expected to correct for a spatial bias in allelic distribution in the GWAS of the neighbor genotypic effects $\beta_{q,2}$.

When not only a single locus but also many loci involve both DGE and IGE, these non-independent QTLs result in interactions between polygenic DGEs and IGEs (i.e., $\mathbf{X}_1 \boldsymbol{\beta}_1$ and $\mathbf{X}_2 \boldsymbol{\beta}_2$), thereby shaping covariance between $\mathbf{X}_1 \boldsymbol{\beta}_1$ and $\mathbf{X}_2 \boldsymbol{\beta}_2$ as $\text{Cov}(\mathbf{X}_1 \boldsymbol{\beta}_1, \mathbf{X}_2 \boldsymbol{\beta}_2)$. Given that mean SNP effects are expected to be zero as $E(\boldsymbol{\beta}_1) = E(\boldsymbol{\beta}_2) = 0$, the DGE-IGE covariance is given as $\text{Cov}(\mathbf{X}_1 \boldsymbol{\beta}_1, \mathbf{X}_2 \boldsymbol{\beta}_2) = \mathbf{X}_1 E(\boldsymbol{\beta}_1 \boldsymbol{\beta}_2^\top) \mathbf{X}_2^\top = \mathbf{X}_1 \text{Cov}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \mathbf{X}_2^\top$. When $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are not independent, $\text{Cov}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \neq 0$ and this covariance can be redefined as the MVN parameter $\rho_{12} \sigma_1 \sigma_2 = \text{Cov}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Meanwhile, $\mathbf{X}_1 \mathbf{X}_1^\top$ can be scaled and redefined as $\mathbf{K}_{12} = \frac{1}{(Q-1)} \mathbf{X}_1 \mathbf{X}_2^\top$. The k -th and l -th element of \mathbf{K}_{12} is $(\mathbf{K}_{12})_{k,l} = \sum_{q=1}^Q \sum_{j=1}^{J_l} x_{q,k_i} x_{q,l_i} x_{q,l_j} = \sum_{q=1}^Q [(x_{q,k_i} x_{q,l_i}) \sum_{j=1}^{J_l} x_{q,l_j}] / (Q-1)$, indicating the polygenic effects weighted by the local-space allele frequency in focal individual's neighborhood. Similarly, $\text{Cov}(\mathbf{X}_2 \boldsymbol{\beta}_2, \mathbf{X}_1 \boldsymbol{\beta}_1) = \mathbf{X}_2 \text{Cov}(\boldsymbol{\beta}_2, \boldsymbol{\beta}_1) \mathbf{X}_1^\top$ where $\mathbf{K}_{21} = \frac{1}{(Q-1)} \mathbf{X}_2 \mathbf{X}_1^\top$. The k -th and l -th element of \mathbf{K}_{21} is $(\mathbf{K}_{21})_{k,l} = \sum_{q=1}^Q [(x_{q,l_i} x_{q,k_i}) \sum_{j=1}^{J_k} x_{q,k_j}] / (Q-1)$. This indicates $\mathbf{K}_{12} \neq \mathbf{K}_{21}$, but $(\mathbf{K}_{12})^\top = \frac{1}{(Q-1)} (\mathbf{X}_1 \mathbf{X}_2^\top)^\top = \frac{1}{(Q-1)} \mathbf{X}_2 \mathbf{X}_1^\top = \mathbf{K}_{21}$. Furthermore, the local-space allele frequencies $\sum_{j=1}^{J_l} x_{q,l_j}$ and $\sum_{j=1}^{J_l} x_{q,l_j}$ refer to the frequency of one allele at the same SNP locus. Thus, we can assume the same effect from the focal allele at a given locus as $\text{Cov}(\boldsymbol{\beta})$

$\mathbf{1}, \boldsymbol{\beta}_2) = \text{Cov}(\boldsymbol{\beta}_2, \boldsymbol{\beta}_1) = \rho_{12}\sigma_1\sigma_2\mathbf{I}$. Combined with the two asymmetric variance-covariance matrices (i.e., $\mathbf{K}_{12} = \mathbf{K}_{21}^T$), $\rho_{12}\sigma_1\sigma_2\mathbf{K}_{12}$ and $\rho_{12}\sigma_1\sigma_2\mathbf{K}_{21}$ represent asymmetric polygenic effects between focal and neighboring genotypes on y_{k_i} . In GWAS, this random-effect covariance can be used to correct for population structure due to the locus-wise asymmetric neighbor effects $\beta_{q,12}$.

In the case of Eq. 1, \mathbf{u}_1 and \mathbf{u}_2 are not independent due to their covariance. Therefore, the total phenotypic variance $\text{Var}(\mathbf{y})$ can be decomposed as:

$$\begin{aligned} \text{Var}(\mathbf{y}) &= \text{Var}(\mathbf{u}_1 + \mathbf{u}_2) + \text{Var}(\mathbf{e}) \\ &= \text{Var}(\mathbf{u}_1) + \text{Var}(\mathbf{u}_2) + \text{Cov}(\mathbf{u}_1, \mathbf{u}_2) + \text{Cov}(\mathbf{u}_2, \mathbf{u}_1) + \text{Var}(\mathbf{e}) \\ &= \sigma_1^2\mathbf{K}_1 + \sigma_2^2\mathbf{K}_2 + \rho_{12}\sigma_1\sigma_2\mathbf{K}_{12} + \rho_{12}\sigma_2\sigma_1\mathbf{K}_{21} + \sigma_e^2\mathbf{I} \\ &= \sigma_1^2\mathbf{K}_1 + \sigma_2^2\mathbf{K}_2 + \rho_{12}\sigma_1\sigma_2(\mathbf{K}_{12} + \mathbf{K}_{21}) + \sigma_e^2\mathbf{I} \quad (\text{Eq. 2}) \end{aligned}$$

This Eq. 2 indicates that phenotypic variations can be partitioned into those explained by DGE, IGE, and their covariance in addition to residuals, where the three parameters take the range of $\sigma_1^2 > 0$, $\sigma_2^2 > 0$, and $\rho_{12} = (-\infty, \infty)$. The sign of ρ_{12} therefore determines whether the DGE-IGE covariance increases or decreases the phenotypic variation $\text{Var}(\mathbf{y})$, which would in turn affect response to selection.

We implemented the multi-kernel mixed model (Eq. 1) based on the RAINBOW algorithm [26]. This method works by optimizing weights $w_1 + w_2 = 1$ for \mathbf{K}_1 and \mathbf{K}_2 . In the present model, the variance component parameters are redefined as $\sigma_1^2 = \sigma_u^2 w_1$ and $\sigma_2^2 = \sigma_u^2 w_2$, where σ_u^2 represents the total genetic variance due to the two random effects. Accordingly, the proportion of phenotypic variation explained (PVE) by DGE, IGE, and their covariance can be partitioned respectively into $\text{PVE}_{\text{DGE}} = w_1\text{PVE}_u$, $\text{PVE}_{\text{IGE}} = w_2\text{PVE}_u$, and $\text{PVE}_{\text{cov}} = \rho_{12}\sqrt{w_1 w_2}\text{PVE}_u$, where $\text{PVE}_u = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$. Estimating variance component parameters yields a weighted kernel as $\mathbf{K}' = w_1\mathbf{K}_1 + w_2\mathbf{K}_2 + \rho_{12}\sqrt{w_1 w_2}(\mathbf{K}_{12} + \mathbf{K}_{21})$. This weighted kernel allows us to GWAS by comparing models with or without each SNP-wise DGEs or IGEs for $\beta_{q,1}$, $\beta_{q,2}$, or $\beta_{q,12}$.

Benchmark simulation

Using simulated genomes and phenotypes, we tested the performance of the proposed model in variation partitioning, GP, and GWAS. To first address how correctly variance component parameters can be estimated, we performed independent iterations 30 times with varying covariance between DGE and IGE (Fig. 2; Fig. S3). These simulations showed that the medians of the covariance parameter $\hat{\rho}_{12}$ close to those of true models (Fig. 2a and g; Fig. S3a, e and i), validating the capability of the proposed model in distinguishing the sign of ρ_{12} . In a quantitative term, the scenarios of positive DGE-IGE covariance exhibited smaller $\hat{\rho}_{12}$ variations as the absolute value of true ρ_{12} became large (MAE = 0.099, 0.074, and 0.022 for $\rho_{12} = 0, 0.3, 0.6$, respectively; Fig. 2a; Fig. S3a and e). In contrast, the scenarios of negative DGE-IGE covariance showed larger $\hat{\rho}_{12}$ variations as the absolute value of true ρ_{12} became large (MAE = 0.210, 0.176 and 0.099 for $\rho_{12} = -0.6, -0.3, 0$, respectively; Fig. 2g; Fig. S3e and i). Similar trends were observed for the estimated IGE weight \hat{w}_2 , such that the medians of \hat{w}_2 were close to the true value regardless of the incorporation of ρ_{12} but showed different extents of variations (MAE = 0.12, 0.076, 0.072, 0.050, 0.026 for $\rho_{12} = -0.6, -0.3, 0, 0.3, 0.6$, respectively; model = 'cov' in Fig. 2b and h; Fig. S3b, f and j). These results indicate that while the case of negative covariance was more difficult to estimate than that of positive ones, the proposed model is able to estimate polygenic DGE and IGE with the sign of their covariance.

We further tested whether the proposed model can be used as a tool of GP. To evaluate the predictability for simulated phenotypes, we left one on the tenth of simulated data as a test dataset while the rest as a training dataset under varying DGE-IGE covariances. In the training models, total BLUPs were highly correlated between models with and without ρ_{12} ($R^2 > 0.92$ for \hat{y} between the two models; Fig. S4). Nonetheless, when the covariance was positive, the inclusion of DGE-IGE covariance ρ_{12} slightly but better predicted the test dataset in terms of R^2 and MAE (Fig. 2c-d; Fig. S3c-d). When the covariance was negative, models with the DGE-IGE covariance ρ_{12} did not perform better than those without ρ_{12} or even slightly worse in terms of R^2 and MAE (Fig.

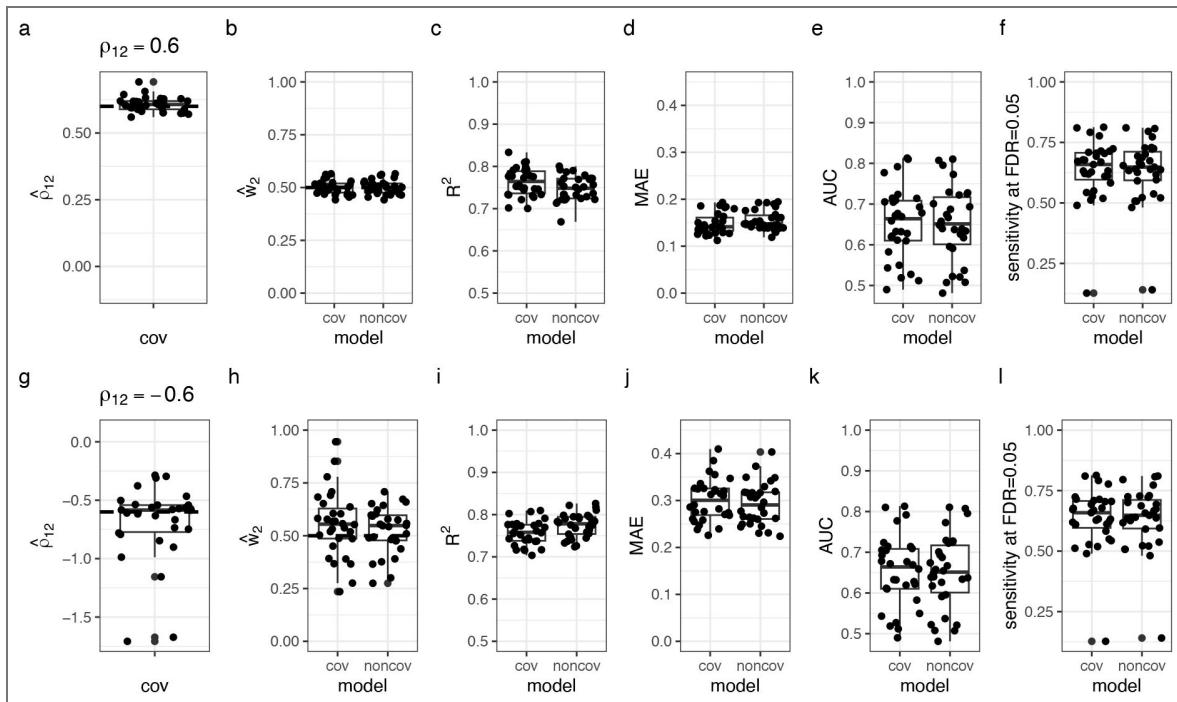


Figure 2. Benchmark simulations under the scenarios of positive or negative covariance between DGE and IGE.

The upper panels (a-f) show the results of positive covariance ($\rho = 0.6$) while the lower panels (g-l) show those of negative covariance ($\rho = -0.6$). (a and g) Estimated covariance $\hat{\rho}_{12}$ between polygenic DGE and IGE. Horizontal dashed lines indicate the true ρ_{12} value. (b and h) Estimated weight of IGE \hat{w}_2 relative to that of DGE $\hat{w}_1 = 1 - \hat{w}_2$. Models with or without the covariance $\hat{\rho}_{12}$ (cov or noncov) were fitted to simulated phenotypes. Horizontal dashed lines indicate the true w_2 value. (c-d and i-j) Genomic prediction based on the models with or without the covariance $\hat{\rho}_{12}$ (cov or noncov). The model performance was evaluated using the coefficient of determination (R^2 : c and i) and mean absolute error (MAE: d and j) between the predicted phenotypic values \hat{y} and the observed phenotypic values of the test dataset y . (e-f and k-l) GWAS performance for the models with or without the covariance $\hat{\rho}_{12}$ (cov or noncov) to detect causative IGE SNPs. The model performance was evaluated using the area under the ROC curve (AUC: e and k) and sensitivity at the stringent threshold at false discovery rate (FDR) = 0.05 (f and l). Box plots display the median and quartiles among 30 independent iterations, with whiskers extending to $1.5 \times$ inter-quartiles.

2i-j; Fig. S3k-l). For the case of $\rho_{12} = 0$ (null model), we observed almost no differences between the models with and without ρ_{12} (Fig. S3g and h). These results revealed the conditions under which the proposed model could perform better in GP.

We then tested whether the inclusion of the covariance ρ_{12} could improve the GWAS performance to detect causative IGE SNPs (i.e., non-zero $\beta_{q,2}$). Under the distinct scenarios of positive or negative covariance $\rho_{12} = \pm 0.6$, we evaluated the model performance using the receiver operating characteristic curve (ROC). In terms of the area under the ROC curve (AUC), the cases of negative and positive covariance both showed the moderate power to detect the causative SNPs (AUC > 0.6 and AUC < 0.8; Fig. 2e and k). This moderate power was also supported by the sensitivity at the stringent specificity (at FDR < 0.05; Fig. 2f and l). These two metrics confirmed the feasibility of GWAS of IGE but its performance did not differ considerably regardless of the DGE-IGE covariance (Fig. 2e-f and k-l). Thus, the both models with and without the DGE-IGE covariance work enough for GWAS.

Application to real data

We applied the proposed model to the three datasets — trembling aspen (*Populus tremuloides*), apple (*Malus × domestica*), and grape (*Vitis vinifera*) — to demonstrate its usability. All three datasets involve woody species and provide individual-level data on phenotypes, genotypes, and spatial arrangement. Based on the spatial layout of the study species, we analyzed the influence of nearest neighbors on each trait.

We first analyzed life-history and defensive traits of aspen genotypes grown in a field garden at Wisconsin of USA, called WisAsp project [29–31]. Among 18 trait records collected over two years, basal area increment at the later growth stage (BAI2024-2015) alone showed a significant negative covariance between DGE and IGE ($\rho_{12} = -0.599$ and -0.629 , $\chi^2 = 3.875$ and 4.727 , $p = 0.049$ and 0.030 with likelihood ratio tests for 2014 and 2015, respectively; Table S1), suggesting IGEs on growth-related traits out of diverse traits. For the late basal area increment, we conducted GP based on 10-fold cross-validation. This GP showed the similar performance between models with and without the covariance in terms of R^2 and MAE (Fig. S5), providing consistent evidence with the simulations (Fig. 2). GWAS of the 18 traits detected two significant IGE SNPs for the first bud burst and size growth (FDR < 0.05; Table S1), but none for the other traits including defensive chemicals (FDR > 0.05). Besides the plant traits, we analyzed the two quantitative indices of arthropod community compositions, but found neither significant DGE-IGE covariance nor IGE SNPs (Table S1). These results led us to focus on intergenotypic competition in the other tree species.

Second, we analyzed apple REFPOP data collected across multiple years and sites in Europe (Fig. 4; Fig. S6). This dataset included 20,280 individuals of 534 accessions across six European sites for 30 traits [32,33]. We quantified PVEs by DGEs, IGEs, and their covariance for these traits (Fig. 4a, c and e; Fig. S6) and focused on three traits with less missing records, i.e., flowering intensity, trunk diameter, and trunk increment (see also Table S2). The flowering intensity was the most largely regulated by DGEs among these three traits (Fig. 4a), whereas the trunk diameter and increment showed relatively large influence of IGEs (Fig. 4c and e). These results suggest that growth traits were more likely to be subjected to IGEs rather than phenological traits.

To further examine the intensity of competition among apple trees, we analyzed relationships between PVE by the DGE-IGE covariance (i.e., PVE_{cov}) and average tree size (Fig. 4b, d and f). The flowering intensity showed a positive but non-significant relationship between PVE_{cov} and mean trunk diameter (LMM, slope coef. = 0.0007, $t = 0.295$, $p = 0.613$ by Wald test; Fig. 4b), where we did not detect any significant $\hat{\rho}_{12}$ (Table S2). In contrast, the trunk diameter showed a negative but non-significant relationship between PVE_{cov} and mean trunk diameter (slope coef. = -0.002 , $t = -1.078$, $p = 0.150$; Fig. 4d), where significant $\hat{\rho}_{12}$ were detected at four sites (Table S2). Remarkably, PVE_{cov} of the trunk increment significantly decreased as the mean trunk diameter became large (slope coef. = -0.006 , $t = -2.61$, $p = 0.013$; Fig. 4f), indicating that intergenotypic competition turned intense as trees grew bigger. Regarding the trunk increment, GWAS also detected two significant IGE SNPs on the seventh chromosome at Belgium site in 2019

(Fig. 4g [↗](#)), which were located on the seventh chromosome at 8743248 and 8744866 base pair positions. Both SNPs had negative $\beta_{q,2}$ estimates, indicating the negative influence of allelic mixtures on the trunk increment ($\hat{\beta}_{q,2} = 0.813$: see also Fig. S1 [↗](#) and S2 [↗](#) for the interpretation of $\pm\beta_{q,2}$). In total, the apple data provide supportive evidence for competitive genetic interactions among neighboring trees.

Third, we additionally analyzed the grape INNOVINE project data [\[34\]](#) to examine a climbing woody plant as another test case. This dataset included 3,642 individuals and 90,007 SNPs for 19 traits [\[34\]](#). In this dataset, most traits did not show large influence of neighboring genotypes compared to DGEs ($\hat{w}_2 < 0.1$ for 17 out of the 19 traits: Fig. 5 [↗](#)). Although the pruning weight and citric acid was influenced by IGEs, these two traits showed not negative but positive DGE-IGE covariance ($\hat{w}_2 = 0.200$ and 0.718 , $\hat{\rho}_{12} = 0.662$ and 0.767 for the pruning weight and citric acid, respectively: Table S3 [↗](#)). GWAS of the the citric acid detected over a thousand of significant IGE SNPs but quantile-quantile plot showed the inflation of p -values (Fig. S7 [↗](#)), indicating the risk of false positive signals. Only the shikimic acid did we find a significantly negative covariance ($\hat{\rho}_{12} = -1.211$, $\chi^2 = 5.207$, $p = 0.023$), but this trait showed the small contribution of IGEs compared to DGEs ($\hat{w}_2 = 0.026$: Table S3 [↗](#)). These results from climbing grapevines provide limited evidence for competitive interactions among neighboring individuals, highlighting the case in which intergenotypic competitions are difficult to detect.

Discussion

We demonstrated the feasibility of GWAS/GP of IGE based on the simulations and applications. Especially, we focused on the DGE-IGE covariance ρ_{12} as a key parameter underlying the genetic trade-off between individual and group performance. Below, we first discuss methodological advantages and potential limitations in the proposed model. Then we discuss biological findings from the real data on woody plant species.

Model performance and limitation

By formulating SNP-wise IGEs based on the Ising model, we integrated oligogenic and polygenic IGEs into a multi-kernel mixed model. The present model can be exchanged into previous IGE models using the design matrix \mathbf{Z} instead of \mathbf{K}_2 and $\beta_{q,2}$. For example, Costa e Silva et al. [\[5\]](#) defined interacting neighbors by the design matrix, or incidence matrix \mathbf{Z} . This approach can be implemented using the RAINBOWR package as it allows the flexible use of \mathbf{Z} (Appendix S4 [↗](#)). Furthermore, it seems plausible for plants and sessile organisms to assume weaker IGEs from distant neighbors [\[5,35\]](#). With 0-1 weights, the design matrix may also accommodate the spatial distance decay of neighbor genotypic effects [\[35\]](#). Examples of such distance decay include the inverse, exponential, or any other decay functions of Euclidean distance. Therefore, the proposed model can implement preceding IGE models once the neighbor incidence and its distance decay are made explicit.

While we applied our models to neighboring interactions in land plants, this approach can also be applied to animal herds or subpopulations. In this case, spatial genetic interactions based on the Ising model correspond to a population genetic model of frequency-dependent selection [\[25\]](#) (Appendix S3 [↗](#); Fig. S2 [↗](#)). By applying the proposed model to cage-reared damselflies, previous studies indeed detected negative frequency-dependent selection on female color polymorphism [\[25\]](#). Practically, this type of grouping can be achieved by setting neighbor spatial scales to infinity and subpopulations as a grouping factor in Neighbor GWAS [see 25 and its vignette]. Given that IGEs have often been analyzed to mitigate intraspecific competition in livestock [\[4,6\]](#), our method may offer a valuable opportunity to dissect the genetic architecture of group performance in both animals and plants.

Our simulations demonstrated the performance of the proposed model in GWAS and GP. In GWAS, models with or without the covariance ρ_{12} showed a similar performance to detect causative IGE SNPs. This is likely because total BLUPs are very similar regardless of ρ_{12} ($R^2 > 0.92$: Fig. S2 [↗](#)), which may not make a large difference in the weighted kernel \mathbf{K} . In GP, the consideration of covariance improved the predictability especially when DGE and IGE are positively correlated, but

not when they are negatively correlated. This might be due to modeling issues and technical limitation in the multi-kernel mixed model. When ρ_{12} is positive, the weighted kernel \mathbf{K}' is likely to be positive semi-definite because all the variance component parameters σ_1^2 , σ_2^2 , and ρ_{12} were all positive. When ρ_{12} is negative, however, \mathbf{K}' may not be positive semi-definite due to $\rho_{12} < 0$. To anticipate this issue, we approximated \mathbf{K}' to the nearest positive-definite matrix in this study, which might have resulted in the unstable estimation of negative ρ_{12} . Further theoretical studies are needed for the quantitative improvement of the present model and its estimation.

Genetic architecture of tree competition

To demonstrate the model applicability, we analyzed the three woody species: aspen, apple, and grape. The aspen and apple exhibited negative genetic covariance between DGEs and IGEs on their growth. Notably, these negative effects of neighboring genotypes were amplified as trees matured. These results are consistent with the previous studies on *Eucalyptus* trees, which detected significant negative DGE-IGE covariance at the late growth stage [5]. In contrast, the climbing woody plants, grapevines, did not exhibit large negative effects of neighboring genotypes on their growth traits. Like other climbing plants, grapes grow vertically and may therefore experience weak competition with horizontal neighbors. Furthermore, grapes possess considerable phenotypic plasticity in response to light quality, such as ultraviolet and far-red wavelengths, to alter their stem length, leaf morphology, and metabolite content [36,37]. Together with the exceptional but plausible case of grapes, the parallel evidence from aspen and apples exemplifies inter-genotypic competition within a plant species.

The present data encompass phenological, chemical, and growth traits, though we did not observe a significantly negative DGE-IGE covariance for these traits. Regarding phenological traits, flowering time is often regulated by self-genotypes as shown by high heritability within a field site [38,39]. We detected substantial phenotypic variance explained by DGEs, but limited variance explained by IGEs, for flowering time across the three woody species (e.g., “BS, bud set” in Fig. 3 [3]; “Harvesting date” in Fig. S6 [3]; and “days.upto.veraison” in Fig. 5 [3]). This suggests that phenological traits are unlikely to be substantially explained by IGEs. In contrast to phenological traits, some chemical traits exhibited large phenotypic variance explained by IGEs, accompanied by a positive DGE-IGE covariance, suggesting synergistic effects of DGEs and IGEs on chemical properties. Specifically, metabolite accumulation in ripening fruits may be facilitated by volatile chemicals, such as ethylene [40,41]. Similarly, for leaf secondary metabolites, herbivore damage often activates jasmonate signaling and the production of green leaf volatiles, which can induce defense responses in neighboring trees [42,43]. These IGEs on specific metabolites provide testable hypotheses for future studies.

In addition to variation partitioning, GWAS detected two site-specific but significant IGE SNPs with negative $\hat{\beta}_{q,2}$ for the apple trunk increment. The negative sign of these SNPs suggests that neighbors with dissimilar alleles reduce the trunk increment of focal individuals, providing evidence for intergenotypic competition at allelic levels. These two significant SNPs exhibited a complete linkage disequilibrium ($R^2 = 1$) and were closely located on the top-to-center positions of seventh chromosome. For this chromosomal region, previous publications documented QTLs underlying DGEs on fruit size [44] and basal trunk increment [45]. More specifically, our significant SNPs range between 8743248 and 8744866 bp positions on the seventh chromosome, which include a locus encoding a protein kinase superfamily protein (MD07G1084900). This locus is homologous to AT5G56790 in *A. thaliana*, but its biological function remains unknown. The nearest loci of MD07G1084900 include the homologs of MYB33 (MD07G1085000; homolog of AT5G06100) and SUCROSE PHOSPHATE SYNTHASE 1F (MD07G1084800; homolog of AT5G20280). Although these two genes AT5G06100 and AT5G20280 are involved in the vegetative phase change [46] and leaf CO₂ assimilation [47] in *A. thaliana*, respectively, their functions have yet to be validated in apples. To this end, our study uncovered a QTL and candidate genes underlying intraspecific competition.

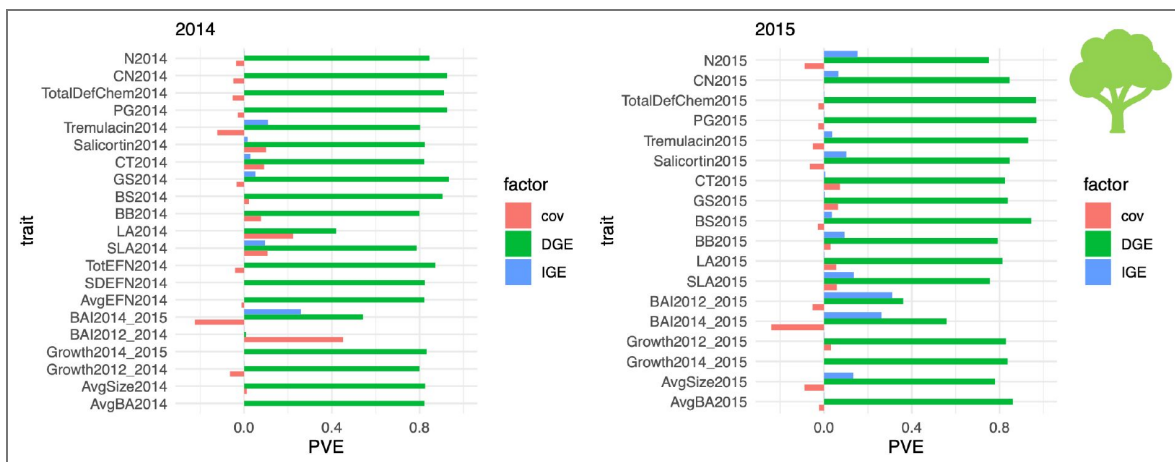


Figure 3. Proportion of phenotypic variation explained (PVE) by direct genetic effects (DGEs), indirect genetic effects (IGEs), and their covariance (cov) in aspen.

Blue (upper), green (middle), and red (lower) bars respectively indicate PVE by IGEs, DGEs, and covariance for each trait (see PVE_{IGE} , PVE_{DGE} , and PVE_{cov} calculations in the main text subsection “Model development”). The left and right panels present results from data collected in 2014 and 2015, respectively. Trait abbreviations: AvgBA, average basal area; AvgSize, average tree volume; GrowthYEAR1_YEAR2, growth of AvgSize from YEAR1 to YEAR2; BAI, basal area increment; AvgEFN, average extra-floral nectary; SDEFN, standard deviation of extra-floral nectary; TotEFN, total extra-floral nectary; SLA, specific leaf area; LA, leaf area; BB, first bud break; BS, terminal bud set; GS, growing season (i.e., BB to BS); CT, condensed tannin; PG, phenolic glycoside; TotalDefChem, total defense chemicals; CN, carbon-nitrogen ratio; and N, nitrogen.

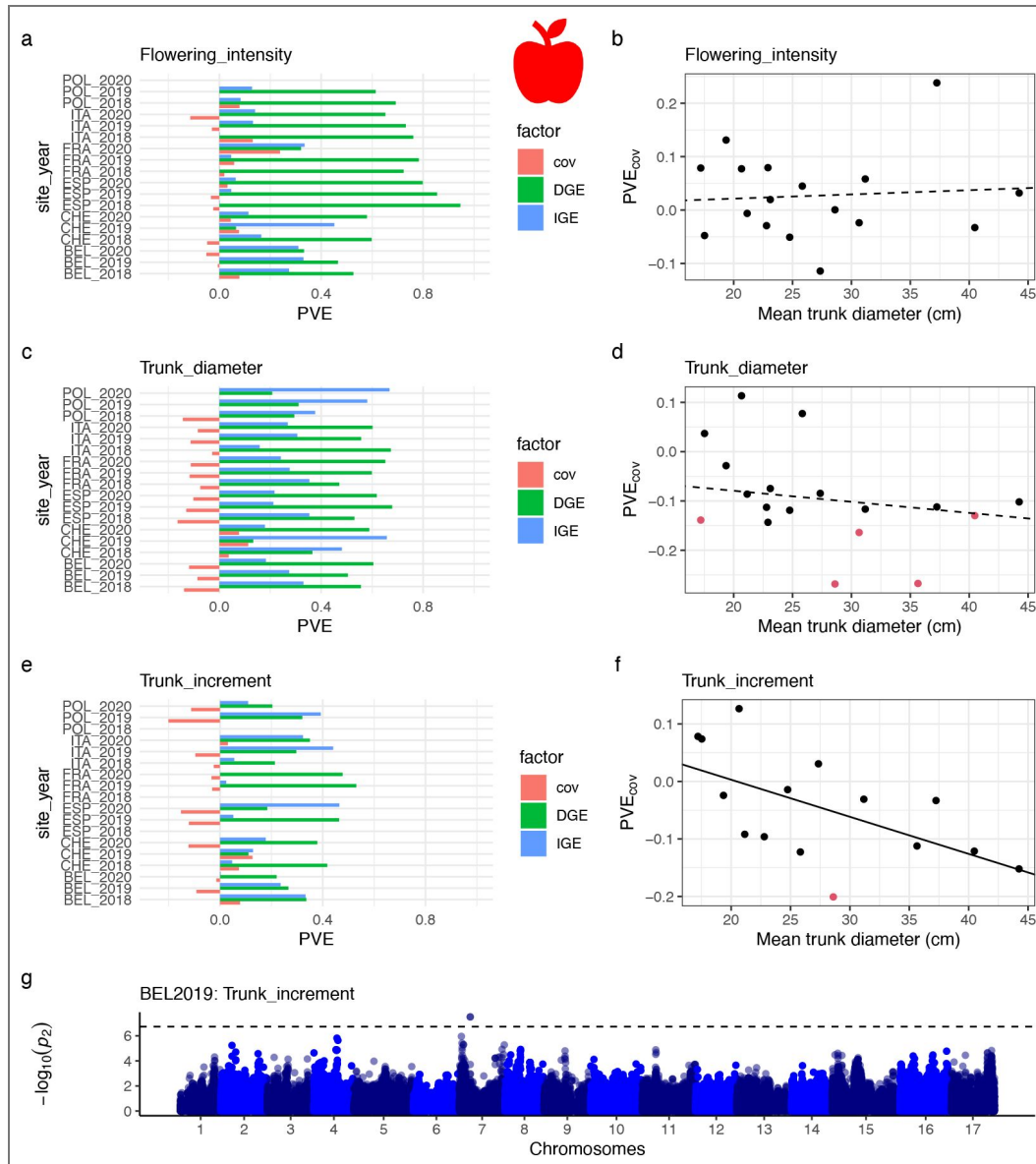


Figure 4. Indirect genetic effects of conspecific neighbors on apple traits across the six sites (Belgium, BEL; France, FRA; Italy, ITA; Poland, POL; Spain, ESP; and Switzerland, CHE) and three years (2018, 2019, and 2020).

(a-f) Variation partitioning of the flowering intensity (a-b), trunk diameter (c-d), and trunk increment (e-f). The left panels (a, c and e) show the proportion of phenotypic variation explained (PVE) by direct genetic effects (DGEs), indirect genetic effects (IGEs), and their covariance (cov) following the aspen example (see Fig. 3). The right panels (b, d, and f) display PVE by the covariance against the mean trunk diameter per trial. A single data point corresponds to each field trial. Red circles highlights trials with significant $\hat{\rho}_{12}$. Solid and dashed lines indicate significant and non-significant trends detected by linear mixed models that include the sites and years as random effects. (g) GWAS manhattan plot of IGEs on the trunk increment at the Belgium site in 2019 (BEL2019). The y-axis $[-\log_{10}(p_2)]$ indicates the association score of $\beta_{q,2}$, which is plotted against 17 chromosomes of apple. The horizontal dashed line indicates the genome-wide threshold at FDR = 0.05.

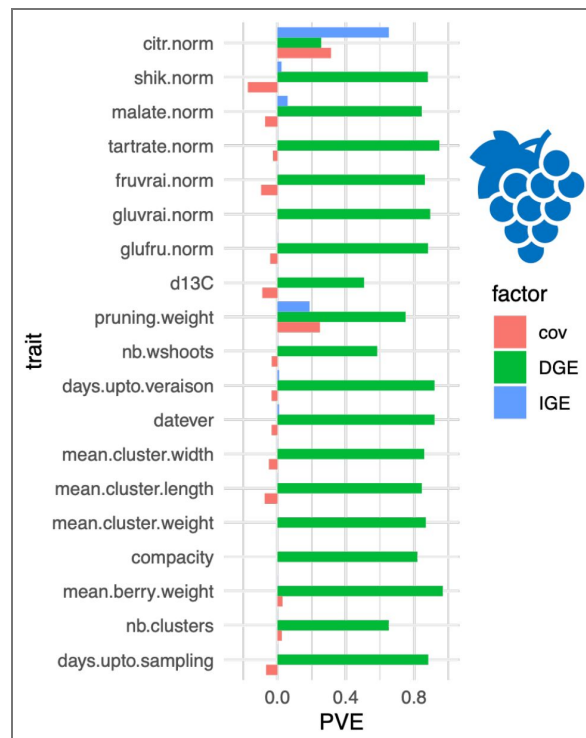


Figure 5. Proportion of phenotypic variation explained (PVE) by direct genetic effects (DGEs), indirect genetic effects (IGEs), and their covariance (cov) in grape.

Blue (upper), green (middle), and red (lower) bars respectively indicate PVE by IGEs, DGEs, and covariance for each trait (see also Fig. 3). Trait abbreviations: “nb.clusters”, the number of clusters; “datever”, date of veraison; “nb.wshoots”, the number of woody shoots, “d13C”, $\delta^{13}C$; “glufru.norm”, the sum of glucose and fructose; “gluvrai.norm”, glucose; “fruvrai.norm”, fructose; “tartrate.norm”, tartaric acid; “malate.norm”, malic acid; “shik.norm”, shikimic acid; and “citr.norm”, citric acid.

Neighbor GWAS is widely applicable to a broad range of phenotypes and species as far as spatial information is available alongside phenotype and genotype data. While the present data include individual trees as the unit of genotype [29,33,34], our method allows for defining a genotype unit as multiple individuals of a single genotype, such as a single plot of field crops [48]. It is important to note, however, that spatial metadata may not always be complete. Even with available spatial information, it is necessary to interpret whether intergenotypic interactions operate at a given spatial scale. Although our study found little evidence of grapevine competition, this lack of evidence may have resulted from sufficient spacing to preclude competition. Further research is needed to assess the generality and specificity of intraspecific competition mediated by IGEs.

Conclusion

The present study provides a flexible model to accommodate polygenic and locus-wise effects of neighboring genotypes on a complex phenotype. This single model enables both GP and GWAS of IGEs. The theory of IGEs has frequently involved the evolution of social interactions [2,49], while empirical applications have predominantly been conducted in domesticated animals and plants [6,50,51]. Consequently, genomics of IGEs promises to contribute to both fundamental and applied issues, including multi-level selection and agricultural applications. Recent research has highlighted the growing need for an interdisciplinary approach to IGEs [3,12,22]. With the increasing availability of genome-wide polymorphism data, our methodology may enhance interdisciplinary practice by facilitating GWAS/GP of IGEs across diverse organisms.

Methods

GWAS/GP implementation

We used the RAINBOWR (>v0.1.41) [26] and rNeighborGWAS (>v1.2.4) [24] packages to implement the multi-kernel mixed model (Eq. 1 [1]). Multi-kernel mixed models Eq. 1 [1] were solved using restricted maximum likelihood (REML) method without the fixed effects $\beta_{q,1}$, $\beta_{q,2}$, and $\beta_{q,12}$. The statistical significance was determined using likelihood ratio tests by dropping ρ_{12} out of the full model. The mixed models with or without ρ_{12} were solved using the EM3.cpp or EM3.cov functions of the RAINBOWR package [26], respectively. When the kernels did not become semi-positive definite, they were approximated by the nearest positive definite matrix using the nearPD function in R (which was enabled with the “forceApproxK=TRUE” option in the EM3.cov function). For GWAS, each SNP locus was subjected to a likelihood ratio test using an eigenvalue decomposition and diagonalization method for the weighted kernel \mathbf{K}' , which was implemented in the lmm.diago function of the gaston package [52]. This GWAS procedure followed the nei_imm function of the rNeighborGWAS package [24]. All source codes and input data are deposited on the GitHub and Zenodo repositories (see “Data accessibility” section), which includes the vignette and subroutine functions for GWAS/GP implementations. An easy R package to implement Eq. 1 [1] is also available at GitHub and Zenodo repositories (see “Data accessibility” section).

Simulation

We conducted simulations to examine the performance of the proposed model. This involved (i) simulating genomes and phenotypes, (ii) assessing trait predictability and causal variant detection, and (iii) comparing these abilities between the previous and newer model. We utilized previously simulated genotype data [25]. Specifically, by simulating spatial and temporally varying selection over 2,000 generations among ten subpopulations, Sato et al. [25] generated 30 independent datasets, each comprising 2,000 diploid individuals with 50 kb nucleotide sequences across three chromosomes (see also Appendix S4 of Sato et al. [25]). This simulation was performed using SLiM version 3 [53] and its outputs were deposited as variant call format (.vcf) on Dryad (<https://doi.org/10.5061/dryad.zs7h44jdv>). Depending on the iterations, each genotype dataset contained approximately ten causal DGE SNPs out of ca. 2,500 SNPs (see Figure S4b and c of [25]). Out of wider parameter range and combinations than described in the present study, our previous study [24] identified conditions under which the model performance differentiated. Our base parameters therefore followed those of our previous study [24].

Variance component and genomic prediction

To test the model capability in the inference and prediction of polygenic IGEs, we simulated phenotypes under two scenarios, i.e., negative and positive covariance between polygenic DGEs and IGEs. Specifically, we assumed the true parameters of $\rho_{12} = 0, +0.3, \text{ or } \pm 0.6$ with $\text{PVE}_u = 0.6$; $w_1 = 0.5$; $w_2 = 0.5$; and $\sigma_1^2 = \sigma_2^2 = 1.0$. With these true parameters, we randomly assigned 1,600 individuals to a 40×40 lattice space. Then we simulated a phenotype subjected to DEGs and IGEs from the nearest neighbors along diagonals and off diagonals (i.e., $\sqrt{2}$ Euclidean distance from a focal individual). We repeated this line of simulations for the thirty genotype datasets as described above. After fitting the proposed model to the simulated genotypes and phenotypes, we calculated the median and mean absolute errors (MAEs) of the variance component parameters among the 30-times simulations to test whether these parameters could be estimated well. To further test the applicability to genomic prediction, we divided the simulated data into a training and test dataset. One on the tenth individuals were randomly left out of the simulated data as a test dataset, whereas the rest was fitted by the proposed model as a training dataset. To determine the trait predictability using the test dataset, we calculated the coefficient of determination (R^2) and MAEs between the estimated and simulated phenotypic values.

Genome-wide association study

We also tested the model performance to detect causative IGE SNPs in GWAS. Our procedure followed Sato et al. [25] while simulating the negative or positive covariance between DGEs and IGEs as $\rho_{12} = \pm 0.6$. The other parameters were the same as described above for GP. We also assumed total 20 causative IGE SNPs, of which ten had a positive coefficient as $\beta_{q,2} = 0.1$ and the rest ten had a negative coefficient as $\beta_{q,2} = -0.1$. To make DGE and IGE correlated, we further assumed that the causative IGE SNPs overlapped with DGE SNPs (i.e., $\beta_{q,1} = \beta_{q,2} = \pm 0.1$). In addition to the non-zero $\beta_{q,1}$ and $\beta_{q,2}$, a half of the causative IGE SNPs was assumed to have non-zero $\beta_{q,12}$ as $\beta_{q,12} = \pm 0.1$. With or without ρ_{12} , the proposed model Eq. 1 [25] was fitted to simulated phenotype y to estimate the weighted kinship matrix \mathbf{K}' . Based on the RAINBOW algorithm [26], we applied the eigenvalue decomposition to \mathbf{K}' and thereby performed GWAS for $\beta_{q,2}$ using a likelihood ratio test. After model fitting, GWAS performance was evaluated based on the false positive and negative rate to detect causative IGE SNPs. The balance between false and true positive rates was calculated using the receiver-operating characteristic curve (ROC) and its area under the curve (AUC). Based on the ROC curves, we also determined the power to capture true positives at the stringent false discovery rate (FDR) = 0.05. We utilized the pROC package implemented in R [54] for the ROC and AUC analysis.

Application

Aspen

For the WisAsp data, we curated phenotype and genotype data from Barker et al. [29] and [30], respectively. These data were deposited at Dryad (<https://doi.org/10.5061/dryad.fr045hv> and <https://doi.org/10.5061/dryad.st463>). The annotation of genotype-phenotype link was obtained from Riehl et al. [31]. We set the cut-off value of minor allele frequency (MAF) at 0.05. In total, we analyzed data from 813 individuals in 2014, comprising 169,150 SNPs, and data from 826 individuals in 2015, comprising 169,111 SNPs. In this SNP matrix, 73% of the loci were homozygous while the rest 27% were heterozygous.

Using the curated aspen data, we performed variance component analysis and GWAS of IGEs. The influential area of neighbors was assumed for the nearest neighbors along the diagonal and off-diagonal direction, i.e., 1 Euclidean distance from a focal individual. We included experimental blocks as non-genetic covariates in all the mixed models. These analyses were separately performed for the year 2014 and 2015. For each trait per year, we tested the statistical significance of the covariance parameter ρ_{12} over the model with $\rho_{12} = 0$ based on the likelihood ratio test. According to the estimated variance components, we then conducted GWAS of $\beta_{q,2}$ based on the weighted kinship matrix **K**. Unless otherwise stated, we determined the significance threshold by FDR for GWAS. We used the CalcThreshold function implemented in the RAINBOWR package [26] to calculate FDR for GWAS.

Apple

We curated the apple REFPOP data [32,33] available via the Data INRAE repository (<https://doi.org/10.15454/IOPGYF>, <https://doi.org/10.15454/1ERHGX>, and <https://doi.org/10.15454/VARJYJ>). The phenotype data deposited by Jung et al. [33] included 30 traits for 20,280 individuals of 534 accessions across six European sites (Belgium, France, Italy, Poland, Spain, and Switzerland). Ten of these traits were measured at one location, while the remaining 20 traits were available from at least two locations (cf. Fig. 1 of Jung et al. [32]). Genotype information of the 534 accessions is available as imputed SNP data between 20K and 480K array [32], yielding 281,140 SNPs at MAF > 0.05. In the SNP matrix, 66% of the loci were homozygous while the rest 34% were heterozygous. The imputed genotype data are available upon reasonable request to Michaela Jung and Helene Muranty as described in Jung et al. [32]. To screen candidate genes in GWAS, we referred to the whole genome assembly and annotation of *Malus x domestica* GDDH13 v1.1 [55] through the Apple Genome and Epigenome project website (<https://iris.angers.inra.fr/gddh13/>; Last accessed on 18-Feb-2026).

Using the curated apple data, we conducted variance component analysis and GWAS per year and site. In this dataset, individuals were assigned along rows and positions. Depending on the site, planting distance within and between tree rows ranged 0.9-1.3 m and 3.2-3.6 m, respectively [32]. Therefore, we considered the nearest neighboring interactions along the position within the tree rows. For each year and site, we then performed likelihood ratio test for ρ_{12} and GWAS of $\beta_{q,2}$ as described for the aspen data above. To test yearly trends in PVE by DGE-IGE covariance (PVE_{cov}), we used linear mixed models in which the study years and sites were considered fixed and random effects, respectively.

Grape

We downloaded the set of grape phenotype and genotype dataset from the Data INRAE repository (<https://doi.org/10.15454/8DHKGL>), which was all archived by Flutre et al. [34]. These data comprise 3,642 individuals of 270 accessions with 90,007 SNPs at MAF > 0.05. Of these SNP loci, 79% were homozygous while the rest 21% were heterozygous. Nineteen traits were available, some exhibited numerous missing data points (Table S3) (see also Supplementary Fig. 1 of Flutre et al. [34]).

Using the retrieved grape data, we conducted variance component analysis and GWAS for each trait. In this dataset, individuals were assigned along ranks and locations in the field, where the ranks were separated 2.5 m each other while the locations were 1 m apart between individuals

within a rank [34]. Therefore, we assumed the nearest neighboring interactions along locations with ranks considered a grouping factor. We also included the experimental blocks as non-genetic covariate in all the mixed models. Then we performed likelihood ratio test for ρ_{12} and GWAS of $\beta_{q,2}$ as described for the aspen data above.

Supplementary Tables

Table S1. Variance component estimates and significant GWAS SNPs of IGEs in aspen. For 2014 or 2015, shown are the effective number of individuals with trait records (n), the total genetic variance $\hat{\sigma}_u^2$, environmental variance $\hat{\sigma}_e^2$, DGE and IGE weights (\hat{w}_1 and \hat{w}_2), DGE-IGE covariance ($\hat{\rho}_{12}$), likelihood ratio test and its p-values over a model without the covariance ρ_{12} (χ^2 and p), and the number of significant IGE SNPs (#sig. SNPs). See also the caption of Figure 3 for the trait abbreviations.

Table S2. Variance component estimates and significant GWAS SNPs of IGEs in apple. Shown are the effective number of individuals with trait records (n), the total genetic variance $\hat{\sigma}_u^2$, environmental variance $\hat{\sigma}_e^2$, DGE and IGE weights (\hat{w}_1 and \hat{w}_2), DGE-IGE covariance ($\hat{\rho}_{12}$), likelihood ratio test and its p-values over a model without the covariance ρ_{12} (χ^2 and p), and the number of significant IGE SNPs (#sig. SNPs). The site codes are shown as follows: Belgium, BEL; France, FRA; Italy, ITA; Poland, POL; Spain, ESP; and Switzerland, CHE. NA means not available. †Not available due to false convergence. ‡Option changed as “optimizer = nlmn” in the EM3.cov function.

Table S3. Variance component estimates and significant GWAS SNPs of IGEs in grape. Shown are the effective number of individuals with trait records (n), the total genetic variance $\hat{\sigma}_u^2$, environmental variance $\hat{\sigma}_e^2$, DGE and IGE weights (\hat{w}_1 and \hat{w}_2), DGE-IGE covariance ($\hat{\rho}_{12}$), likelihood ratio test and its p-values over a model without the covariance ρ_{12} (χ^2 and p), and the number of significant IGE SNPs (#sig. SNPs). †Not available due to false convergence.

Supplementary Figures

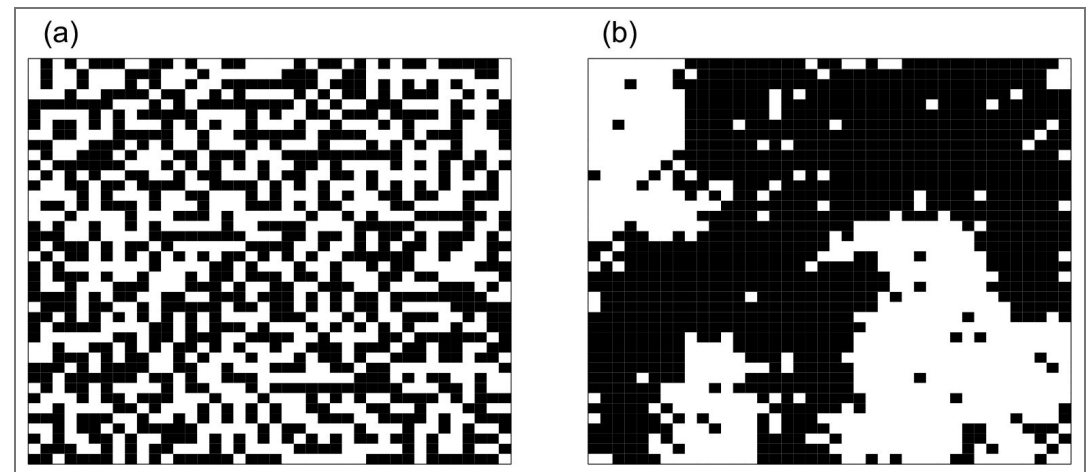


Figure S1. Ising model-based simulation of spatial patterns shaped by the nearest neighbor interactions.

The lattice space with 40 x 40 cells shows the outcome of 100-times update of the genotype value x_{k_i} after a random configuration of two genotypes (black and white cells) following the fixed effects of Eq. S1.1, i.e., $y_{k_i} = \beta_0 + \beta_{q,1}x_{q,k_i} + \beta_{q,2} \sum_{k_j=1}^{J_k} (x_{q,k_i}x_{q,k_j}) / J$ (a) The case of $\beta_{q,2} = -0.2$, $\beta_{q,1} = 0$, $\beta_0 = 0$. (b) The case of positive $\beta_{q,2} = 0.2$, $\beta_{q,1} = 0$, $\beta_0 = 0$.

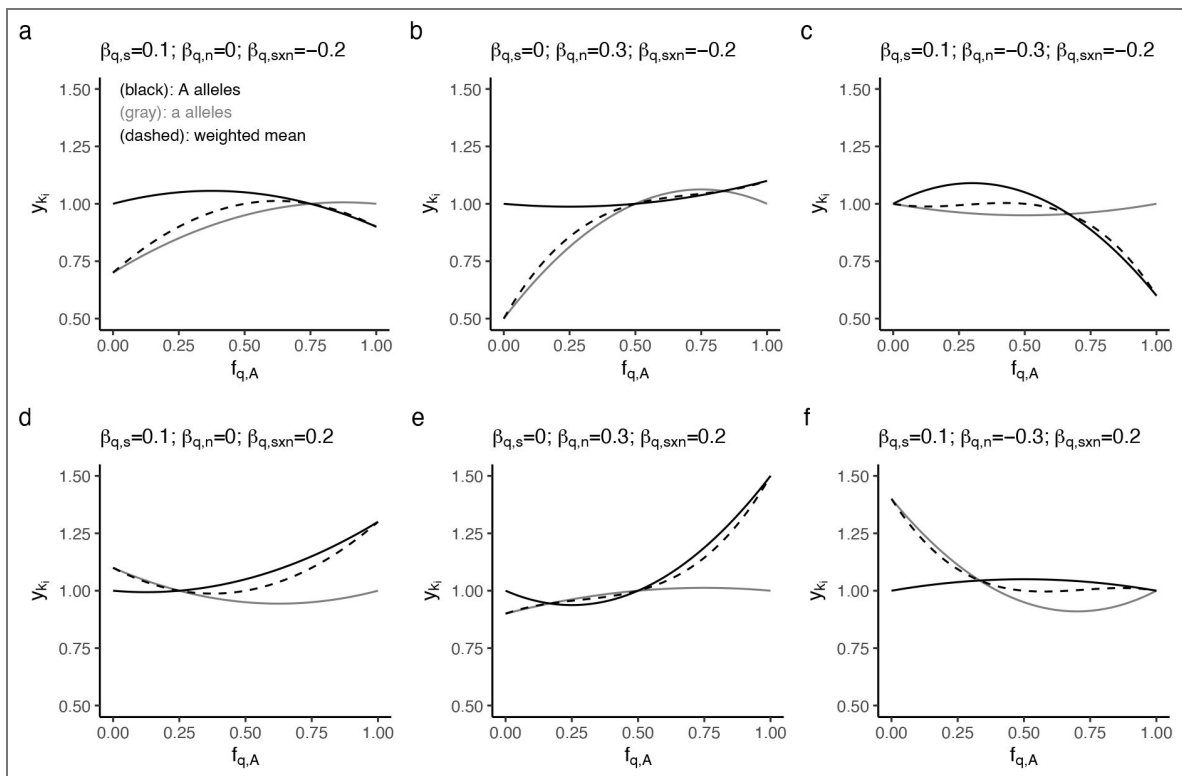


Figure S2. Numerical examples for the theoretical expectation between the phenotypic value y_{k_i} and the neighboring frequency of A alleles $f_{q,A}$ at q -th SNP locus within a local space k .

Black and grey lines indicate the phenotypic values of A and a alleles, respectively. Dashed curves represent a group mean defined by the weighted mean of the phenotypic values between A and a alleles. Panels (a-d) exemplify respective cases of positive and/or negative $\beta_{q,n}(= \beta_{q,12})$ and $\beta_{q,sxn}(= \beta_{q,2})$ with zero or non-zero $\beta_{q,s}(= \beta_{q,1})$ and $\beta_0 = 1.0$.

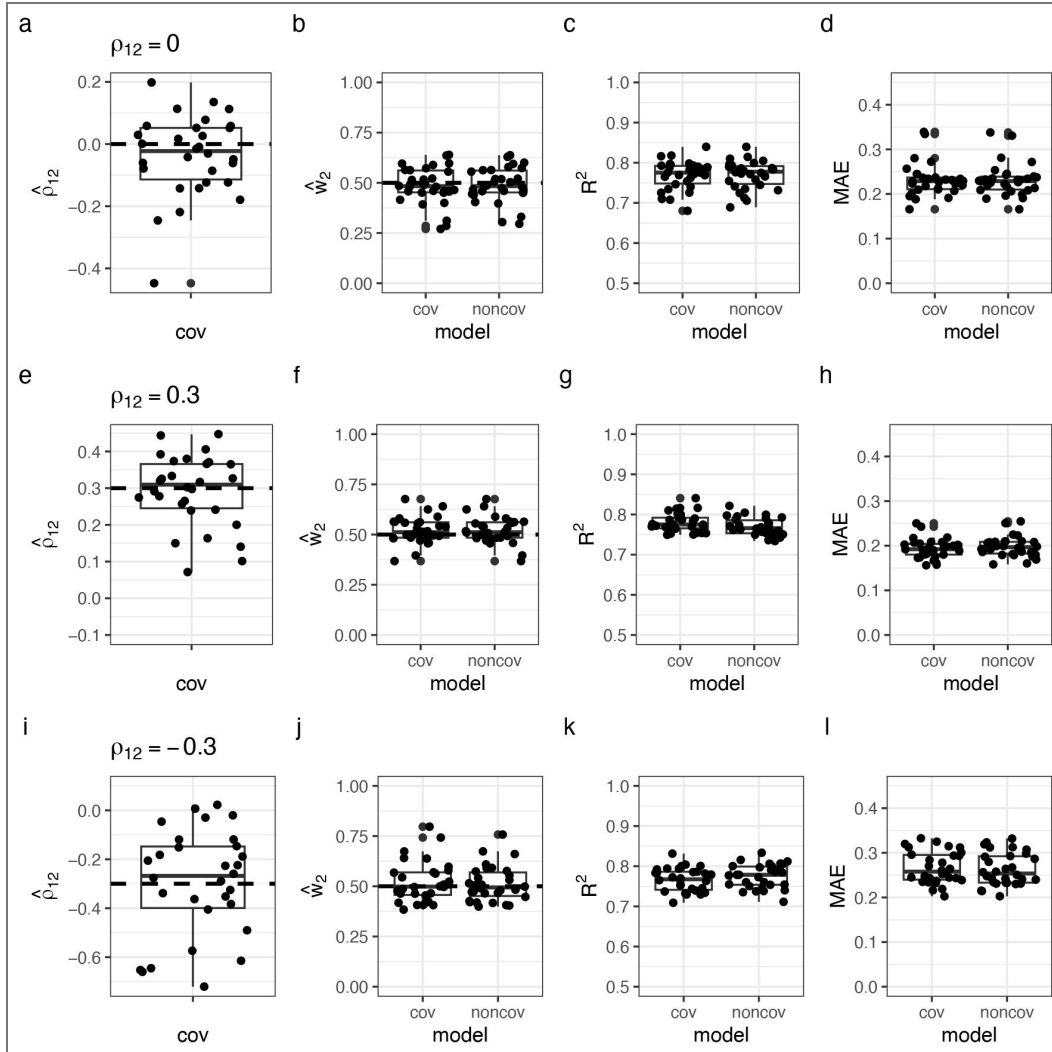


Figure S3. Variance components and genomic prediction of simulated data under the scenarios of moderate positive ($\rho = 0.3$: a-d), null ($\rho = 0$: e-h), or moderate negative ($\rho = -0.3$: i-l) covariance between DGE and IGE.

(a, e and i) Estimated covariance $\hat{\rho}_{12}$ between polygenic DGE and IGE. Horizontal dashed lines indicate the true ρ_{12} value. (b, f, and j) Estimated weight of IGE \hat{w}_2 relative to that of DGE $\hat{w}_1 = 1 - \hat{w}_2$. Models with or without the covariance $\hat{\rho}_{12}$ (cov or noncov) were fitted to simulated phenotypes. Horizontal dashed lines indicate the true w_2 value. (c-d, g-h and k-l) Genomic prediction based on the models with or without the covariance $\hat{\rho}_{12}$ (cov or noncov). The model performance was evaluated using the coefficient of determination (R^2 : c, g and k) and mean absolute error (MAE: d, h and l) between the predicted phenotypic values $\hat{\mathbf{y}}$ and the observed trait values of the test dataset \mathbf{y} . Box plots display the median and quartiles among 30 independent iterations, with whiskers extending to $1.5 \times$ inter-quartiles.

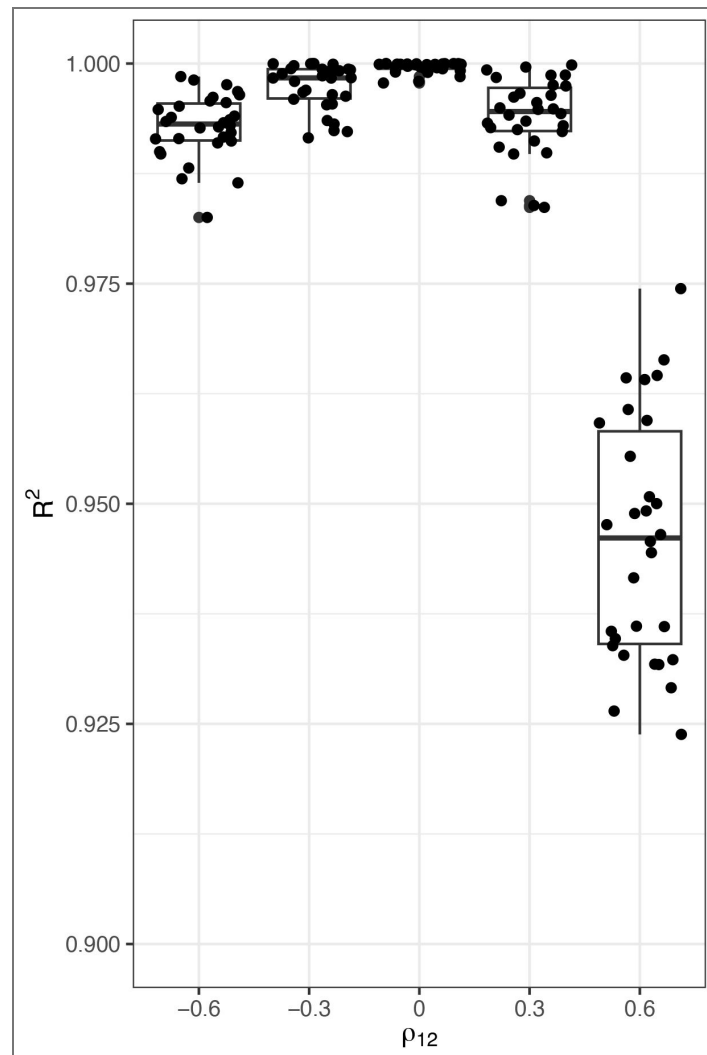


Figure S4. Coefficient of determination (R^2) for the estimated phenotypic values \hat{y} between the models with or without the covariance parameter ρ_{12} .

Box plots are shown for the simulation scenarios with varying covariance between DGE and IGE ($\rho = -0.6, -0.3, 0, 0.3, 0.6$). The high correlations ($R^2 > 0.9$) indicate the very similar \hat{y} between the two models.

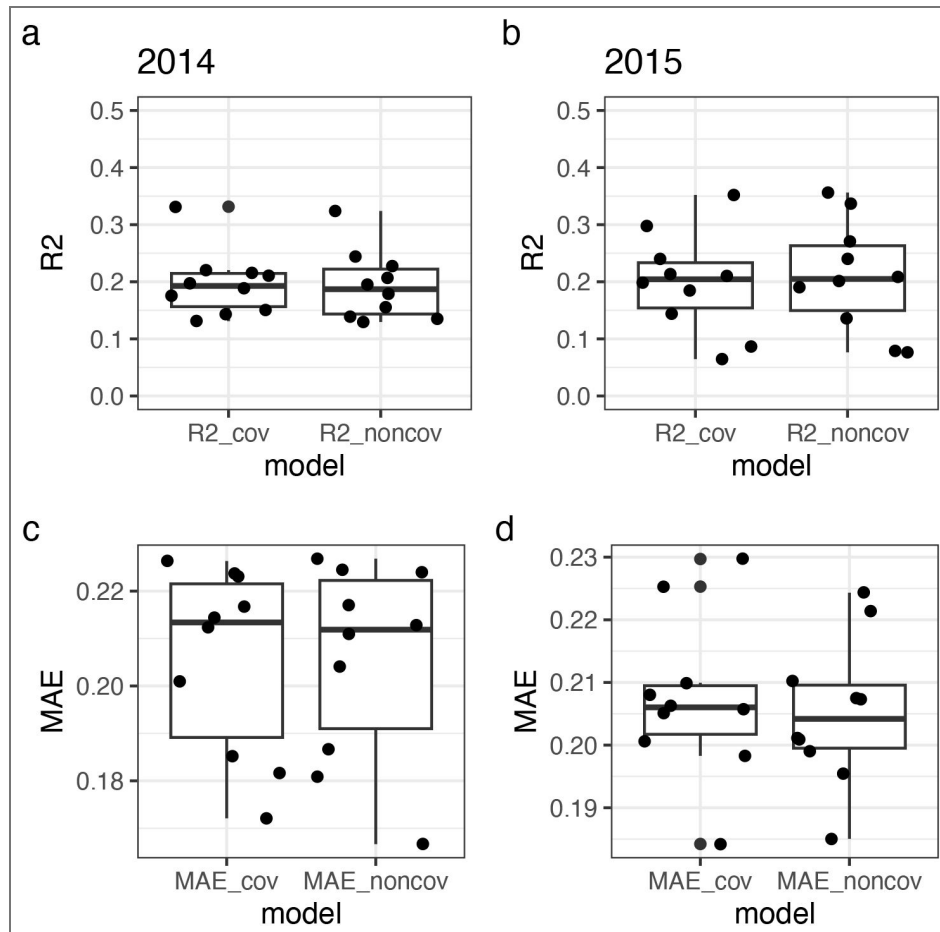


Figure S5. Genomic prediction of the basal area increment from 2014 to 2015 (BAI2014_2015) in aspen.

The coefficient of determination (R^2 : a and b) and mean absolute error (MAE: c and d) of predicted phenotypic values \hat{y} among 10-fold cross-validations are shown for the data collected in 2014 (left) and 2015 (right).

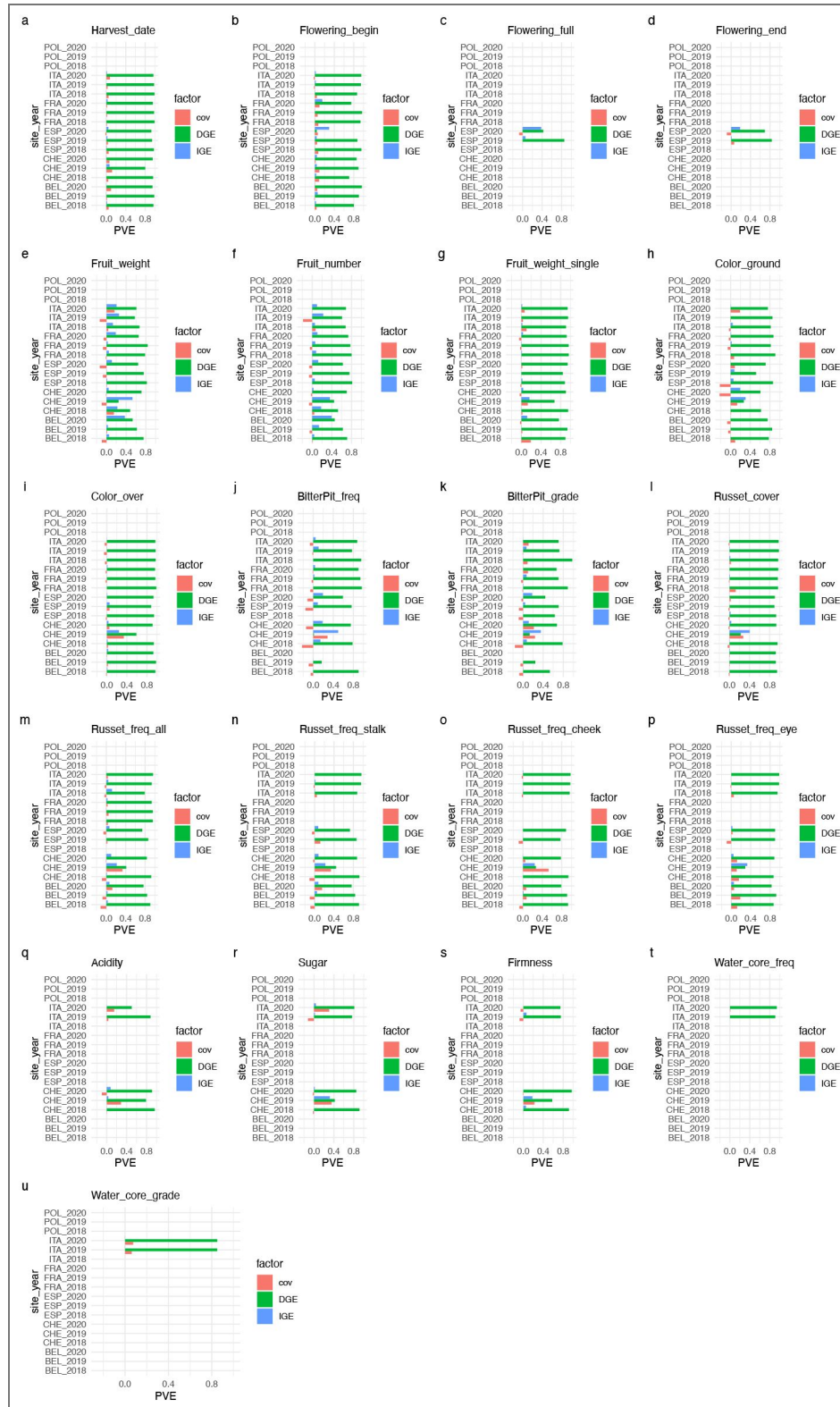


Figure S6. Proportion of phenotypic variation explained (PVE) by indirect genetic effects (IGEs), direct genetic effects (DGEs), and their covariance (cov) in 20 apple traits other than the flowering intensity, trunk diameter, and trunk increment given in the main text.

Blue (upper), green (middle), and red (lower) bars respectively indicate PVE by IGEs, DGEs, and their covariance for each trait. Blank bars indicate the absence of data.

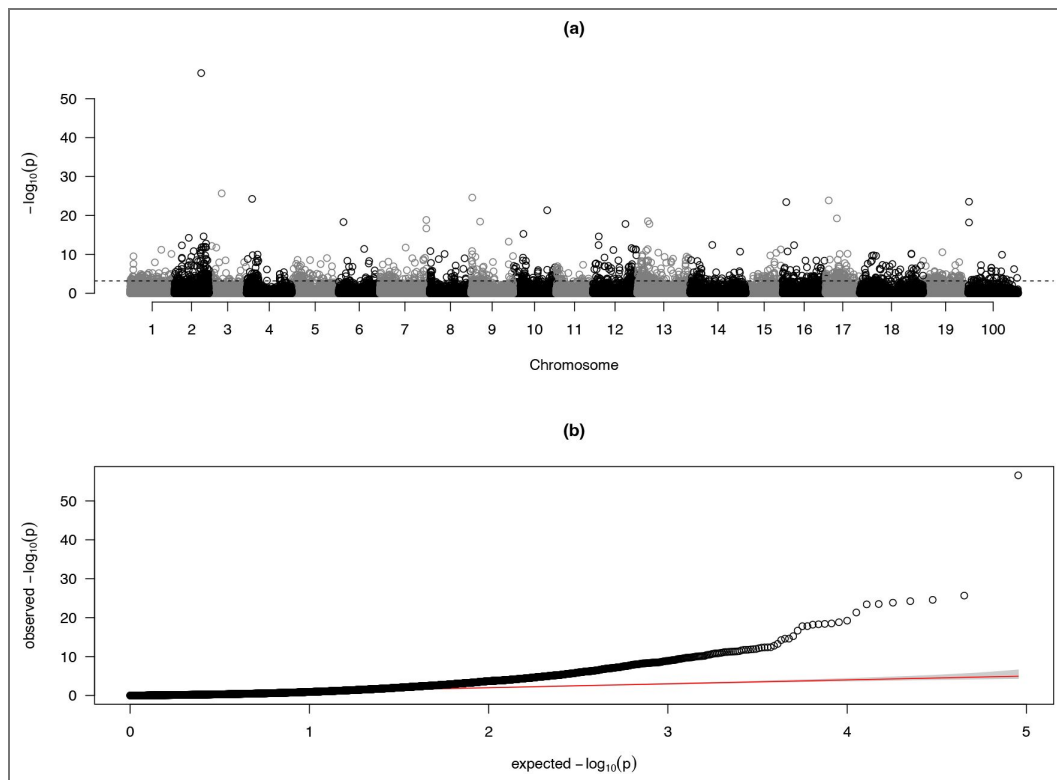


Figure S7. GWAS Manhattan (a) and quantile-quantile (b) plots for the citric acid ("citr.norm" in Fig. 5) in grape.

(a) The association score of $-\log_{10}(p)$ plotted against the 19 chromosomes of grape (where chromosome 100 means unanchored chromosomes). Horizontal dashed line indicates the genome-wide significance threshold at FDR = 0.05. (b) The observed association scores $-\log_{10}(p)$ plotted against randomly expected values. Trend line and shade indicate $y = x$ and its 95% confidence interval.

Data availability

All source codes and input data are available in the GitHub (<https://github.com/yassato/ngwas2>) and Zenodo (<https://doi.org/10.5281/zenodo.19362850>) repositories. These repositories include a vignette that demonstrates how to implement relevant models using RAINBOWR, which can be viewed at <https://yassato.github.io/ngwas2/>. To make the present modeling easy, we provide an R package named 'rNeighborLMM' at GitHub (<https://github.com/yassato/rNeighborLMM>) and Zenodo (<https://doi.org/10.5281/zenodo.19362843>).

Acknowledgements

We thank M. Jung and H. Muranty for providing complete genotype data of the apple REFPOP project; H. Barker and J. Riehl for addressing our questions about the aspen field arrangement and data annotation; and H. Iwata, S.E. Wuest, and M. Roth for discussions during the study. We acknowledge that the computing resource was provided by Human Genome Center at the University of Tokyo, Japan (<http://sc.hgc.jp/shirokane.html>). We used Paperpal by Editage (<https://paperpal.com/>) to correct English, but did not use any generative AI to draft the manuscript.

Additional information

Funding

This study was supported by Japan Science and Technology Agency (Grant no. JPMJFR233L to YS) and by Japan Society for the Promotion of Science (JP25H00928 to YS and KH). KH was supported by RIKEN Special Postdoctoral Researchers (SPDR) program during the study.

Author contributions

YS: Conceptualization, Methodology, Software, Data curation, Investigation, Formal analysis, Validation, Visualization, Funding acquisition, Writing - Original Draft; **KH:** Conceptualization, Methodology, Software, Funding acquisition, Writing - Review & Editing.

Funding

Funder	Grant reference number	Author
MEXT Japan Science and Technology Agency (JST)	https://doi.org/10.52926/JPMJFR233L	Yasuhiro Sato
MEXT Japan Society for the Promotion of Science (JSPS)	JP25H00928	Kosuke Hamazaki

Author ORCID iDs

Yasuhiro Sato:  <https://orcid.org/0000-0002-6466-723X>

Additional files

[Supplementary Tables S1-S3](#)

Appendix S1

Original model of Neighbor GWAS

In the main text, we modified the original Neighbor GWAS model to implement a flexible mixed model including polygenic and oligogenic IGEs. Here we present the original model [24] to highlight how to model locus-wise IGEs. Following the main text (Fig. 1), we modeled IGEs of

neighboring genotypes at q -th SNP locus on focal k_i -th individual phenotype y_{k_i} at the local space k as follows:

$$y_{k_i} = \beta_0 + \beta_{q,1}x_{q,k_i} + \beta_{q,2} \sum_{k_j=1}^{J_k} (x_{q,k_i}x_{q,k_j}) / J + u_{k_i} + e_{k_i} \tag{Eq.S1.1}$$

where $x_{q,i}$ and $x_{q,j}$ represent the genotype of focal i -th and neighbor j -th individuals in a local space k (where J_k is the total number of neighboring individuals). For simplicity, let us assume inbred lines with two homozygotes are encoded as $aa = -1$ and $AA = +1$. Then, the fixed effect $\beta_{q,2}$ indicates the influence of mean allelic (di)similarity [i.e., $(-1) \times (+1) = (+1) \times (-1) = -1$ or $(+1) \times (+1) = (-1) \times (-1) = +1$] between focal and all neighboring individuals on y_{k_i} . This is analogous to the Ising model of ferromagnetics [27], in which $\beta_{q,2}$ represents physical interactions among the south/north dipole of neighboring magnets while $\beta_{q,1}$ is an external energy. By focusing on the fixed effects of Eq. S1.1, Figure S1 exemplifies spatial patterns shaped by neighbor genetic interactions over time, in which negative and positive $\beta_{q,2}$ favors spatial mixing and separation of two genotypes in a neighborhood (Fig. S1a and b, respectively). In real data, we aim to estimate $\beta_{q,1}$ and $\beta_{q,2}$ by regressing observed y_{k_i} on x_{q,k_i} .

In addition to the fixed effects $\beta_{q,1}$ and $\beta_{q,2}$, u_{k_i} denotes random effects that represent polygenic self and neighbor genotypic effects, respectively. Using vectors and matrices, Eq. S1.1 can also be expressed as a mixed model as follows:

$$\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_1\mathbf{u} + \mathbf{e} \tag{Eq.S1.2}$$

In the original model [24], \mathbf{u} consists of additive polygenic effects from self and neighbor genotypes as $u_{k_i} \in \mathbf{u} \sim N(\mathbf{0}, \sigma_1^2\mathbf{K}_1 + \sigma_2^2\mathbf{K}_2)$. The residual e_i is defined as $e_i \in \mathbf{e}$ and $\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2\mathbf{I})$. \mathbf{Z} denotes a design matrix and can be an identity matrix \mathbf{I} when all genotypes are different. The $N \times N$ variance-covariance matrices \mathbf{K}_1 and \mathbf{K}_2 are the same as defined in the main text (see “Model development”).

When the two random effects \mathbf{u}_1 and \mathbf{u}_2 are additive and independent, the total phenotypic variance can be decomposed as $\text{Var}(\mathbf{y}) = \text{Var}(\mathbf{u}_1) + \text{Var}(\mathbf{u}_2) + \text{Var}(\mathbf{e}) = \sigma_1^2\mathbf{K}_1 + \sigma_2^2\mathbf{K}_2 + \sigma_e^2\mathbf{I}$. Then, the total genetic and environmental variance can be partitioned as $V_g = (\sigma_1^2 + \sigma_2^2)$ and $V_e = \sigma_e^2$. The proportion of phenotypic variation explained by additive DEGs and IGEs (PVE₁₊₂) is then given as $\text{PVE}_{1+2} = V_g / (V_g + V_e)$. Furthermore, the net contribution of IGEs to the phenotypic variation should be calculated as $\text{PVE}_2 = \text{PVE}_{1+2} - h_{\text{SNP}}^2$, where h_{SNP}^2 denotes the conventional SNP heritability. Although this conservative metric was needed due to a strong element-wise correlation between \mathbf{K}_1 and \mathbf{K}_2 [24], this issue would be addressed by allowing the covariance parameter $\rho_{12} = (-\infty, \infty)$ to decrease or increase the total genetic variance in the present study (see Eq. 2).

Appendix S2

Interaction term between SNP-wise self and neighbor effects

The original model Eq. S1.1 implicitly assumes symmetric effects between focal and neighboring individuals at the locus q by $\beta_{q,2} \sum_{j_k=1}^{J_k} (x_{q,k_i}x_{q,k_j}) / J_k$. This assumption can be relaxed by introducing an interaction term between DGE and IGE as follows:

$$y_{k_i} = \beta_0 + \beta_{q,1}x_{q,k_i} + \beta_{q,12}x_{q,k_i} \sum_{j_k=1}^{J_k} (x_{q,i_k}x_{q,j_k}) / J_k + \beta_{q,2} \sum_{j_k=1}^{J_k} (x_{q,k_i}x_{q,k_j}) / J_k + u_{k_i} + e_{k_i} \tag{Eq.S2.1}$$

When genotypes values are encoded as {AA, Aa, aa} = {+1, 0, -1}, squared-genotype values x_{q,k_i}^2 turn +1 or 0. This leads to $\beta_{q,12}x_{q,k_i} \sum_{k_j=1}^{J_k} (x_{q,k_i}x_{q,k_j}) / J_k = \beta_{q,12}x_{q,k_i}^2 \sum_{k_j=1}^{J_k} (x_{q,k_j}) / J_k = \beta_{q,12} \sum_{k_j=1}^{J_k} (x_{q,k_j}) / J_k$, thereby making Eq. S2.1 simpler as follows:

$$y_{k_i} = \beta_0 + \beta_{q,1}x_{q,k_i} + \beta_{q,12} \sum_{j=1}^{J_k} x_{q,k_j} / J_k + \beta_{q,2} \sum_{k_j=1}^J (x_{q,k_i}x_{q,k_j}) / J_k + u_{k_i} + e_{k_i} \quad (\text{Eq.S2.2})$$

This Eq. S2.2 means that in the context of the Ising model, the statistical interaction term $\beta_{q,12}$ can be exchanged as a main effect [56]. For clarity, we redefine $\beta_{q,1} = \beta_{q,s}$; $\beta_{q,12} = \beta_{q,n}$; and $\beta_{q,2} = \beta_{q,s \times n}$, which yields:

$$y_{k_i} = \beta_0 + \beta_{q,s}x_{q,k_i} + \beta_{q,n} \sum_{k_j=1}^J x_{q,k_j} / J_k + \beta_{q,s \times n} \sum_{k_j=1}^J (x_{q,k_i}x_{q,k_j}) / J_k + u_{k_i} + e_{k_i} \quad (\text{Eq.S2.3})$$

Similarly, Eq. S2.3 can be extended as

$$\mathbf{y} = \mathbf{X}_0\beta_0 + \mathbf{X}_s\beta_s + \mathbf{X}_n\beta_n + \mathbf{X}_{s \times n}\beta_{s \times n} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (\text{Eq.S2.4})$$

where the three random effects \mathbf{u}_s , \mathbf{u}_n , and $\mathbf{u}_{s \times n}$ can be respectively defined as $\mathbf{u} \sim N(\mathbf{0}, \sigma_s^2\mathbf{K}_s + \sigma_n^2\mathbf{K}_n + \sigma_{s \times n}^2\mathbf{K}_{s \times n})$. To correct for the population structure due to asymmetric neighbor effects, Eq S2.4 introduced the third random effect the original model. Similar to $\mathbf{K}_s = \mathbf{K}_1$ and $\mathbf{K}_{s \times n} = \mathbf{K}_2$, we can define \mathbf{K}_n as the $N \times N$ variance-covariance matrix that represents a polygenic self-by-neighbor genotypic effects on y_{k_i} . This self-by-neighbor matrix \mathbf{K}_n is also given by the cross-product as $\mathbf{K}_n = \frac{1}{(Q-1)}\mathbf{X}_n\mathbf{X}_n^T$. The element of the N individuals $\times Q$ SNPs matrix \mathbf{X}_n consists of covariates for the self-by-neighbor genotypic effects as $(\mathbf{X}_n)_{n,q} = (\sum_{j=1}^{J_n} x_{q,n_j}) / J_n$. After estimating the three variance component parameters σ_s^2 , σ_n^2 , and $\sigma_{s \times n}^2$, we can obtain a weighted kernel as $\mathbf{K}' = \hat{\sigma}_s^2\mathbf{K}_s + \hat{\sigma}_n^2\mathbf{K}_n + \hat{\sigma}_{s \times n}^2\mathbf{K}_{s \times n}$. Similar to GWAS based on Eq. 1 (see Methods), this weighted kernel matrix represents the total population structure due to polygenic self, neighbor, and self-by-neighbor effects. GWAS can be performed using an eigenvalue decomposition of the weighted kernel \mathbf{K}' and diagonalization method implemented in the lmm.diago function of the gaston package [52].

Appendix S3

Theoretical expectation of mean phenotypic values

In terms of evolutionary biology, the Neighbor GWAS model offers a straightforward regression method for population genetic models of frequency-dependent selection [25,56], also known as pairwise interaction models [28]. Specifically, the sign of $\beta_{q,2}$ represents positive or negative frequency-dependent selection when y_{k_i} is a fitness-related phenotype [25]. In other words, a sort of Eq. S1.1 and S2.1 poses a selection gradient analysis of frequency-dependent selection to estimate $\beta_{q,2}$ by regressing phenotypes on genotypes and their spatial similarity [25].

By focusing on allele frequency within a local space k , we can clarify how the three locus-wise coefficients $\beta_{q,s}$, $\beta_{q,n}$, and $\beta_{q,s \times n}$ affect the focal phenotypic value y_{k_i} [25]. Given that all the three datasets analyzed in this study included non-negligible fraction of heterozygous loci (see Methods), we assumed additive effects and encoded genotype values as {AA, Aa, aa} = {+1, 0, -1}. This holds the case 2 in Appendix S2 of Sato et al. [25]. According to Sato et al. [25], let us assume random mating and interactions among AA, Aa, and aa genotypes in the local space k where $f_{q,A}$ denotes the frequency of A alleles at q -th SNP locus. The phenotypic values for AA, Aa, and aa genotypes are then given as follows:

$$\begin{aligned}
 y_{AA} &= \beta_0 + \beta_{q,s} + (\beta_{q,n} + \beta_{q,s \times n})(2f_{q,A} - 1) \\
 y_{Aa} &= \beta_0 \\
 y_{aa} &= \beta_0 - \beta_{q,s} + (\beta_{q,n} - \beta_{q,s \times n})(2f_{q,A} - 1)
 \end{aligned}
 \tag{Eq.S3}$$

. Then, the marginal phenotypic values at allele levels are given by their weighted mean as $y_A = \beta_0 + f_{q,A}R\beta_{q,n} + \beta_{q,s \times n}SR2f_{q,A} - 1S + \beta_{q,s}f_{q,A}$ and $y_a = \beta_0 + R1 - f_{q,A}SR\beta_{q,n} - \beta_{q,s \times n}SR2f_{q,A} - 1S - \beta_{q,s}R1 - f_{q,A}S$ (see also Appendix S2 of Sato et al. [25]). Furthermore, the group mean of y_{k_i} is given by the weighted mean between the two genotypes as $\bar{y} = y_A f_{q,A} + y_a (1 - f_{q,A})$.

Figure S2 presents numerical examples of y_A, y_a and \bar{y} in response to $f_{q,A}$. When $\beta_{q,n} < 0$, the group mean \bar{y} decreases as increasing $f_{q,A}$ (Fig. S2c and f). Contrarily, \bar{y} increases as increasing $f_{q,A}$ when $\beta_{q,n} > 0$ (Fig. S2b and e). These two cases indicate that the sign of $\beta_{q,n}$ determines a positive or negative response of the group mean to the selection for A alleles. Regarding the other coefficients, $\beta_{q,s}$ modulates the intercept while $\beta_{q,s \times n}$ determines relative difference (i.e., frequency-dependent selection) between y_A and y_a and resulting shape of \bar{y} (i.e., convex or concave: Fig. S2a and d). These examples showcase the way in which the group mean \bar{y} decreases or increases in response to frequency-dependent selection on y_{k_i} .

The proposed model Eq. 1 extends $\beta_{q,s}, \beta_{q,n}$, and $\beta_{q,s \times n}$ to accommodate their polygenic effects. As shown by Figure S2, the sign of $\beta_{q,n}$ ($= \beta_{q,12}$) determines whether the group mean decreases or increases when the AA genotypes are selected to become rare or abundant in a group k . According to the interpretation of $\beta_{q,n}$ ($= \beta_{q,12}$), the covariance ρ_{12} determines negative or positive response of self-polygenic effects x_{k_i}, x_{l_i} to the neighboring allele frequency $f_{q,A} = \sum_{j=1}^{J_k} x_{q,k_j} / J_k$ (see also the main text Eq. 1). Taken together, the covariance ρ_{12} can be interpreted as a key parameter that represents negatively or positively correlated responses between polygenic DGE and IGE. This interpretation aligns with DGE-IGE covariance in the ordinal IGE model (see Appendix S4 below).

Appendix S4

Designating indirect genetic effects by design matrices

Other than our models, neighboring individuals can be specified using a design matrix or incidence matrix \mathbf{Z} . Ordinal IGE models adopted this way to formulate polygenic DGEs and IGEs [5,11,57]. The ordinal IGE model can also be implemented using the RAINBOWR package as it allows users to define \mathbf{Z} . Here, let the two design matrices be \mathbf{Z}_1 and \mathbf{Z}_2 for polygenic DGEs and IGEs, respectively. The ordinal model is then given as follows.

$$\mathbf{y} = \mathbf{X}_0\beta_0 + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}$$

where

$$\begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \otimes \mathbf{K}_1 \right]
 \tag{Eq.S4}$$

This ordinal model defines polygenic DGE and IGE by random effects \mathbf{u}_1 and \mathbf{u}_2 without fixed effects for each SNP. In the random effects, the kinship matrix \mathbf{K}_1 is shared among the variance and covariance parameters σ_1^2, σ_2^2 , and ρ_{12} as denoted by the Kronecker product \otimes . In this context, the sign of the covariance ρ_{12} determines antagonistic (negative) or synergistic (positive) responses of focal and neighboring individuals to selection [57]. To achieve GWAS, previous studies tested SNP-wise fixed effects from focal and neighboring individuals with Eq. S4 set as a null model [e.g., 11,15]. Compared to the ordinal model (Eq. S4), the proposed model (Eq. 1) is more extensible to incorporate SNP-wise fixed effects $\beta_{q,1}, \beta_{q,2}$, and $\beta_{q,12}$ while distinguishing variance-covariance matrices among $\mathbf{K}_1, \mathbf{K}_2, \mathbf{K}_{12}$ and \mathbf{K}_{21} .

References

1. **Bijma P**, Muir WM, Ellen ED, Wolf JB, Van Arendonk JAM (2007) Multilevel Selection 2: Estimating the Genetic Parameters Determining Inheritance and Response to Selection. *Genetics* **175**:289-299 <https://doi.org/10.1534/genetics.106.062729> | PubMed
2. **Bijma P** (2020) The Price equation as a bridge between animal breeding and evolutionary biology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**:20190360 <https://doi.org/10.1098/rstb.2019.0360> | PubMed
3. **Baud A**, McPeck S, Chen N, Hughes KA (2022) Indirect Genetic Effects: A Cross-disciplinary Perspective on Empirical Studies. *Journal of Heredity* **113**:1-15 <https://doi.org/10.1093/jhered/esab059> | PubMed
4. **Bijma P**, Muir WM, Van Arendonk JAM (2007) Multilevel Selection 1: Quantitative Genetics of Inheritance and Response to Selection. *Genetics* **175**:277-288 <https://doi.org/10.1534/genetics.106.062711> | PubMed
5. **Costa e Silva J**, Potts BM, Gilmour AR, Kerr RJ (2017) Genetic-based interactions among tree neighbors: Identification of the most influential neighbors, and estimation of correlations among direct and indirect genetic effects for leaf disease and growth in *Eucalyptus globulus*. *Heredity* **119**:125-135 <https://doi.org/10.1038/hdy.2017.25> | PubMed
6. **Bijma P** (2020) Social genetic effects in livestock: Current status and future avenues of research. *Journal of Animal Breeding and Genetics* **137**:261-262 <https://doi.org/10.1111/jbg.12477> | PubMed
7. **Baud A**, Mulligan MK, Casale FP, Ingels JF, Bohl CJ, Callebert J, et al. (2017) Genetic Variation in the Social Environment Contributes to Health and Disease. *PLOS Genetics* **13**:e1006498 <https://doi.org/10.1371/journal.pgen.1006498> | PubMed
8. **Donald CM** (1968) The breeding of crop ideotypes. *Euphytica* **17**:385-403 <https://doi.org/10.1007/BF00056241>
9. **Weiner J** (2019) Looking in the Wrong Direction for Higher-Yielding Crop Genotypes. *Trends in Plant Science* **24**:927-933 <https://doi.org/10.1016/j.tplants.2019.07.001> | PubMed
10. **Wuest SE**, Peter R, Niklaus PA (2021) Ecological and evolutionary approaches to improving crop variety mixtures. *Nature Ecology & Evolution* **5**:1068-1077 <https://doi.org/10.1038/s41559-021-01497-x> | PubMed
11. **Montazeaud G**, Helleu Q, Wuest SE, Keller L (2023) Indirect genetic effects are shaped by demographic history and ecology in *Arabidopsis thaliana*. *Nature Ecology & Evolution* **7**:1878-1891 <https://doi.org/10.1038/s41559-023-02189-4> | PubMed
12. **Sato Y**, Akram I (2026) Extending genomic prediction from individuals to a population and community. In: Rusch A (Ed). *Advances in Ecological Research* Elsevier. <https://doi.org/10.1016/bs.aecr.2026.02.001>
13. **Warrington NM**, Hwang L-D, Nivard MG, Evans DM (2021) Estimating direct and indirect genetic effects on offspring phenotypes using genome-wide summary results data. *Nature Communications* **12**:5420 <https://doi.org/10.1038/s41467-021-25723-z> | PubMed
14. **Howe LJ**, Nivard MG, Morris TT, Hansen AF, Rasheed H, Cho Y, et al. (2022) Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics* **54**:581-592 <https://doi.org/10.1038/s41588-022-01062-7> | PubMed
15. **Baud A**, Casale FP, Barkley-Levenson AM, Farhadi N, Montillot C, Yalcin B, et al. (2021) Dissecting indirect genetic effects from peers in laboratory mice. *Genome Biology* **22**:216 <https://doi.org/10.1186/s13059-021-02415-x> | PubMed
16. **Montazeaud G**, Flutre T, Ballini E, Morel J, David J, Girodolle J, et al. (2022) From cultivar mixtures to allelic mixtures: Opposite effects of allelic richness between genotypes and genotype richness in wheat. *New Phytologist* **233**:2573-2584 <https://doi.org/10.1111/nph.17915> | PubMed

17. Mathieu L, Montazeaud G, Delmond M, Brousse A, Fréville H, Méteignier L-V, et al. (2024) Co-GWAS unveils the genetic architecture of inter-individual epistasis affecting biomass and disease severity in wheat binary mixtures. *Research Square* <https://doi.org/10.21203/rs.3.rs-5227322/v1>
18. Wuest SE, Pires ND, Luo S, Vasseur F, Messier J, Grossniklaus U, et al. (2022) Increasing plant group productivity through latent genetic variation for cooperation. *PLOS Biology* **20**:e3001842 <https://doi.org/10.1371/journal.pbio.3001842> | [PubMed](#)
19. Sato Y, Shimizu-Inatsugi R, Takeda K, Schmid B, Nagano AJ, Shimizu KK (2024) Reducing herbivory in mixed planting by genomic prediction of neighbor effects in the field. *Nature Communications* **15**:8467 <https://doi.org/10.1038/s41467-024-52374-7> | [PubMed](#)
20. Subrahmaniam HJ, Libourel C, Journet E-P, Morel J-B, Muñoz S, Niebel A, et al. (2018) The genetics underlying natural variation of plant-plant interactions, a beloved but forgotten member of the family of biotic interactions. *The Plant Journal* **93**:747-770 <https://doi.org/10.1111/tpj.13799> | [PubMed](#)
21. Becker C, Berthomé R, Delavault P, Flutre T, Fréville H, Gibot-Leclerc S, et al. (2023) The ecologically relevant genetics of plant-plant interactions. *Trends in Plant Science* **28**:31-42 <https://doi.org/10.1016/j.tplants.2022.08.014> | [PubMed](#)
22. Santamarina C, Mathieu L, Bitocchi E, Pieri A, Bellucci E, Di Vittori V, et al. (2025) Agroecological genomics and participatory science: Optimizing crop mixtures for agricultural diversification. *Trends in Plant Science* <https://doi.org/10.1016/j.tplants.2025.04.008> | [PubMed](#)
23. Sato Y, Wuest SE (2025) The Genetics of Plant-Plant Interactions and Their Cascading Effects on Agroecosystems—from Model Plants to Applications. *Plant and Cell Physiology* **66**:477-489 <https://doi.org/10.1093/pcp/pcae092> | [PubMed](#)
24. Sato Y, Yamamoto E, Shimizu KK, Nagano AJ (2021) Neighbor GWAS: Incorporating neighbor genotypic identity into genome-wide association studies of field herbivory. *Heredity* **126**:597-614 <https://doi.org/10.1038/s41437-020-00401-w> | [PubMed](#)
25. Sato Y, Takahashi Y, Xu C, Shimizu KK (2023) Detecting frequency-dependent selection through the effects of genotype similarity on fitness components. *Evolution* **77**:1145-1157 <https://doi.org/10.1093/evolut/qpad028> | [PubMed](#)
26. Hamazaki K, Iwata H (2020) RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method. Perteau M, editor. *PLOS Computational Biology* **16**:e1007663 <https://doi.org/10.1371/journal.pcbi.1007663> | [PubMed](#)
27. Cibra BA (2018) An Introduction to the Ising Model. *The American Mathematical Monthly* **94**:937-959 <https://doi.org/10.1080/00029890.1987.12000742>
28. Schneider KA (2008) Maximization principles for frequency-dependent selection I: The one-locus two-allele case. *Theoretical Population Biology* **74**:251-262 <https://doi.org/10.1016/j.tpb.2008.07.006> | [PubMed](#)
29. Barker HL, Holeski LM, Lindroth RL (2018) Genotypic variation in plant traits shapes herbivorous insect and ant communities on a foundation tree species. *PLOS One* **13**:e0200954 <https://doi.org/10.1371/journal.pone.0200954> | [PubMed](#)
30. Barker HL, Riehl JF, Bernhardsson C, Rubert-Nason KF, Holeski LM, Ingvarsson PK, et al. (2019) Linking plant genes to insect communities: Identifying the genetic bases of plant traits and community composition. *Molecular Ecology* **28**:4404-4421 <https://doi.org/10.1111/mec.15158> | [PubMed](#)
31. Riehl JFL, Cole CT, Morrow CJ, Barker HL, Bernhardsson C, Rubert-Nason K, et al. (2023) Genomic and transcriptomic analyses reveal polygenic architecture for ecologically important traits in aspen (*Populus tremuloides* Michx.). *Ecology and Evolution* **13**:e10541 <https://doi.org/10.1002/ece3.10541> | [PubMed](#)
32. Jung M, Roth M, Aranzana MJ, Auwerkerken A, Bink M, Denancé C, et al. (2020) The apple REFPOP—a reference population for genomics-assisted breeding in apple. *Horticulture Research* **7**:189 <https://doi.org/10.1038/s41438-020-00408-8> | [PubMed](#)

33. Jung M, Keller B, Roth M, Aranzana MJ, Auwerkerken A, Guerra W, et al. (2022) Genetic architecture and genomic predictive ability of apple quantitative traits across environments. *Horticulture Research* **9**:uhac028 <https://doi.org/10.1093/hr/uhac028> | PubMed
34. Flutre T, Le Cunff L, Fodor A, Launay A, Romieu C, Berger G, et al. (2022) A genome-wide association and prediction study in grapevine deciphers the genetic architecture of multiple traits and identifies genes under many new QTLs. *G3:Genes | Genomes | Genetics* **12**:jkac103 <https://doi.org/10.1093/g3journal/jkac103> | PubMed
35. Muir WM (2005) Incorporation of Competitive Effects in Forest Tree or Animal Breeding Programs. *Genetics* **170**:1247-1259 <https://doi.org/10.1534/genetics.104.035956> | PubMed
36. Castagna A, Csepregi K, Neugart S, Zipoli G, Večeřová K, Jakab G, et al. (2017) Environmental plasticity of Pinot noir grapevine leaves: A trans-European study of morphological and biochemical changes along a 1,500-km latitudinal climatic gradient. *Plant, Cell & Environment* **40**:2790-2805 <https://doi.org/10.1111/pce.13054> | PubMed
37. Liu M, Zhao Y, Fan P, Kong J, Wang Y, Xu X, et al. (2023) Grapevine plantlets respond to different monochromatic lights by tuning photosynthesis and carbon allocation. *Horticulture Research* **10**:uhad160 <https://doi.org/10.1093/hr/uhad160> | PubMed
38. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, et al. (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics* **6**:e1000940 <https://doi.org/10.1371/journal.pgen.1000940> | PubMed
39. Guo T, Mu Q, Wang J, Vanous AE, Onogi A, Iwata H, et al. (2020) Dynamic effects of interacting genes underlying rice flowering-time phenotypic plasticity and global adaptation. *Genome Research* **30**:673-683 <https://doi.org/10.1101/gr.255703.119> | PubMed
40. Barry CS, Giovannoni JJ (2007) Ethylene and Fruit Ripening. *Journal of Plant Growth Regulation* **26**:143 <https://doi.org/10.1007/s00344-007-9002-y>
41. Pech J, Purgatto E, Bouzayen M, Latché A (2012) Ethylene and Fruit Ripening. 1st. In: McManus MT (Ed). *Annual Plant Reviews* **44** Wiley. pp. 275-304 <https://doi.org/10.1002/9781118223086.ch11>
42. Li T, Holopainen JK, Kokko H, Tervahauta AI, Blande JD (2012) Herbivore-induced aspen volatiles temporally regulate two different indirect defences in neighbouring plants. *Functional Ecology* **26**:1176-1185 <https://doi.org/10.1111/j.1365-2435.2012.01984.x>
43. Lindroth RL, St. Clair SB (2013) Adaptations of quaking aspen (*Populus tremuloides* Michx.) for defense against herbivores. *Forest Ecology and Management* **299**:14-21 <https://doi.org/10.1016/j.foreco.2012.11.018>
44. Costa F (2015) MetaQTL analysis provides a compendium of genomic loci controlling fruit quality traits in apple. *Tree Genetics & Genomes* **11**:819 <https://doi.org/10.1007/s11295-014-0819-9>
45. Segura V, Durel C-E, Costes E (2009) Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: QTL mapping. *Tree Genetics & Genomes* **5**:165-179 <https://doi.org/10.1007/s11295-008-0181-x>
46. Guo C, Xu Y, Shi M, Lai Y, Wu X, Wang H, et al. (2017) Repression of miR156 by miR159 regulates the timing of the juvenile-to-Adult transition in *Arabidopsis*. *The Plant Cell* **29**:1293-1304 <https://doi.org/10.1105/tpc.16.00975> | PubMed
47. Morales A, Kaiser E, Yin X, Harbinson J, Molenaar J, Driever SM, et al. (2018) Dynamic modelling of limitations on improving leaf CO₂ assimilation under fluctuating irradiance. *Plant, Cell & Environment* **41**:589-604 <https://doi.org/10.1111/pce.13119> | PubMed
48. Akram I, Rohr L, Shimizu KK, Shimizu-Inatsugi R, Sato Y (2026) Spatial heterogeneity of disease infection attributable to neighbour genotypic identity in barley cultivars. *silico Plants* **8**:diaf023 <https://doi.org/10.1093/insilicoplants/diaf023>
49. Wolf JB, Brodie III ED, Cheverud JM, Moore AJ, Wade MJ (1998) Evolutionary consequences of indirect genetic effects. *Trends in Ecology & Evolution* **13**:64-69 [https://doi.org/10.1016/S0169-5347\(97\)01233-0](https://doi.org/10.1016/S0169-5347(97)01233-0) | PubMed

50. Bourke PM, Evers JB, Bijma P, Van Apeldoorn DF, Smulders MJM, Kuyper TW, et al. (2021) Breeding Beyond Monoculture: Putting the “Intercrop” Into Crops. *Frontiers in Plant Science* **12**:734167 <https://doi.org/10.3389/fpls.2021.734167> | PubMed
 51. Pélissier R, Ballini E, Temple C, Ducasse A, Colombo M, Frouin J, et al. (2023) The genetic identity of neighboring plants in intraspecific mixtures modulates disease susceptibility of both wheat and rice. *PLOS Biology* **21**:e3002287 <https://doi.org/10.1371/journal.pbio.3002287> | PubMed
 52. Perdry H, Dandine-Roulland C (2022) Gaston: Genetic data handling (QC, GRM, LD, PCA) & linear mixed models. CRAN. <https://CRAN.R-project.org/package=gaston>
 53. Haller BC, Messer PW (2019) SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution* **36**:632–637 <https://doi.org/10.1093/molbev/msy228> | PubMed
 54. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. (2011) pROC: An open-source package for r and s+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**:77 <https://doi.org/10.1186/1471-2105-12-77> | PubMed
 55. Daccord N, Celton J-M, Linsmith G, Becker C, Choisne N, Schijlen E, et al. (2017) High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* **49**:1099–1106 <https://doi.org/10.1038/ng.3886> | PubMed
 56. Sato Y (2025) Negative frequency-dependent selection underlies overyielding through neighbor genotypic effects in *Arabidopsis thaliana*. *bioRxiv* <https://doi.org/10.1101/2025.05.14.654149>
 57. Bijma P (2014) The quantitative genetics of indirect genetic effects: A selective review of modelling issues. *Heredity* **112**:61–69 <https://doi.org/10.1038/hdy.2013.15> | PubMed
- Barker HL, et al. (2018) Data from: Genotypic variation in plant traits shapes herbivorous insect and ant communities on a foundation tree species. Dryad Digital Repository. <https://doi.org/10.5061/dryad.fr045hv>
- Barker HL, et al. (2019) Data from: Ingvarsson PK, et al. Linking plant genes to insect communities: Identifying the genetic bases of plant traits and community composition. Dryad Digital Repository. <https://doi.org/10.5061/dryad.st463>
- Sato Y, et al. (2023) Data from: Detecting frequency-dependent selection through the effects of genotype similarity on fitness components. Dryad Digital Repository. <https://doi.org/10.5061/dryad.zs7h44jdv>

Peer reviews

Reviewer #1 (Public review):

This study presents a new model of phenotypic variation incorporating direct and indirect genetic effects, as well as a new implementation (RAINBOWR) for quantification, genomic prediction and GWAS. It includes a simulation study to test the model and implementation, and three applications to plant species.

The abstract describes the main novelty and significance of the study as follows: "Recent studies have utilized high-resolution polymorphism data to enable genomic prediction (GP) and genome-wide association study (GWAS) of IGEs, but unified methods remain limited". I disagree with this statement (e.g., using ASREML: <https://doi.org/10.1186/s12711-018-0409-7>, using LIMIX: <https://doi.org/10.1186/s13059-021-02415-x>; etc.).

The parameterisation of genetic effects in the model is not standard and complex. Hence, the simulation study is key, and the results need to be presented in a very rigorous manner. I have several points to make on this:

(1) L172 says the estimated parameters are "close to" the real parameters. The results of the simulation study need to be quantitative (see

<https://www.biorxiv.org/content/10.64898/2026.03.10.710784v1.supplementary-material> [↗](#) for example).

(2) Figure 2h: the estimates seem to be biased, no?

(3) Figure 2 in general: why isn't there a difference between cov and noncov? Do we not expect the inclusion or non-inclusion of a covariance term to affect the other genetic parameters and the results presented in Figure 2?

(4) Does "total BLUPs were highly correlated between models with and without ρ " really validate the model?

(5) As far as the GWAS is concerned, the results of the simulation study should include a figure showing whether the p-values are inflated (as observed in the grape application), and not just a ROC curve.

The model only includes IID residuals, whereas the importance of including non-genetic social effects (IEE) has been demonstrated in many settings, and other IGE plant studies have used sophisticated spatially structured residuals (e.g. 10.1111/nph.12035). Can the authors justify why they considered only IID residuals? In the three applications presented, wouldn't it be appropriate to include spatially structured residuals and potentially other relevant covariates?

It remains unclear why the authors chose such an unconventional parameterisation of the DGE IGE models for the questions asked in this study. It seemed appropriate to study frequency-dependent selection (previous paper), but for this study, focused on IGE quantification and GWAS, the classical models (e.g. early models by P. Bijma but also more recent models that allow for distance-dependent IGE) seem appropriate, and they are much simpler and easier to interpret, and have been validated in many settings). The Discussion paragraph L274-284 only strengthens my doubts.

<https://doi.org/10.7554/eLife.111650.1.sa2>

Reviewer #2 (Public review):

Summary:

In this study, Sato and Hamazaki have expanded upon previous work, describing quantitative genetic models for direct and indirect genetic effects and applied this to both simulated and real plant datasets of three different tree species. The methods are clearly described and accompanied by a number of R packages freely available to the wider community.

Strengths:

The main strength lies in the joint modelling of DGE, IGE and their covariance while also simultaneously modelling single-SNP fixed effects (including SNP interactions across neighbours) and a polygenic effect that goes beyond a simple kinship correction as found in many traditional GWAS models, to a compound kinship structure that accounts for DGE, IGE and their interaction.

Weaknesses:

There were some aspects that deserved more attention from the authors. For example, the authors found that a very large amount of phenotypic variation in citric acid content in grapes was explained by neighbour identity, along with over 1000 significant SNPs, yet there was little to no discussion of this result and how it could have arisen (apart from some mention of volatiles and ethylene - but without being explicit on the mechanism here). The simulation study also only considered the scenario of equal direct and indirect genetic

variances, while previous studies, as well as the 3 real datasets presented in this study, show that DGE variance is almost always larger than IGE variance. A simulation study cannot be exhaustive, of course, but it seems more likely that in reality and for most traits, IGE will be more difficult to detect than DGE.

<https://doi.org/10.7554/eLife.111650.1.sa1>

Reviewer #3 (Public review):

Summary:

The authors aimed at studying the genetics of interactions between individuals, notably the genetic architecture of indirect genetic effects. For that, they mobilized a technique known as "genome-wide association" study. GWASs are typically formalized as linear mixed models (LMMs) with fixed effects to identify the oligogenic component of the genetic architecture (usually SNPs tested one by one, as done here), and with random effects to quantify the overall contribution of the polygenic component of the genetic architecture (using a kinship matrix). They used an LMM with a few corrections and improvements from one of their already-published model, assessed it on data they had already simulated in a previous work, and applied it to three datasets generated and originally analyzed by others, focusing only on direct genetic effects. The results on simulated data confirmed that it was necessary to adapt their previous model. The results on real data confirmed the presence of negative correlation between direct and indirect genetic effects (for two out of three species), as was already known from other studies. They found a few SNPs with significant, indirect effects, which led them to identify candidate genes, but they did not validate them.

Strengths:

The main strength of the manuscript lies in the question tackled by the authors, i.e., related to indirect genetic effects, with the ambition to go beyond the estimation of overall effects towards the distinction between polygenic and oligogenic components of genetic architecture. They also found, in an apple dataset, a significant IGE SNP that also happens to be in a DGE-associated region.

Weaknesses:

- (1) Overall, the authors do not engage sufficiently with the existing literature, and do not provide strong evidence that their approach is more powerful or more interpretable than others. Hence, this work seems rather incremental.
- (2) The authors used an LMM that corresponds to a previous LMM they already published in 2021, with a few changes that appeared more like corrections than improvements. Their model raised several questions.
- (3) First of all, their previous model included the polygenic component of direct genetic effects (modeled as random with a kinship matrix), but not the polygenic component of indirect genetic effects. As a consequence, the initial model did not allow both direct and indirect genetic effects to be correlated, although this correlation is the hallmark of the topic: a negative correlation can lead to selection on direct effects only to deliver a negative genetic gain (Griffing, 1967). This was corrected in their new model here, so that it is similar in this respect to the other models. They highlighted that, on simulated data, their new model could "infer a trade-off between DGEs and IGEs", but that was the very goal of introducing the correlation parameter, so it was reassuring at least to know that they could estimate it on simulated data. On real data, they found evidence for it being negative, which was already the case in Cappa and Cantet (2008) for a tree species, in Haug et al (2023) for annual crops, in Montazeaud et al (2023) for *A. thaliana*, etc. They tested for significance but did not provide

any confidence interval. They showed the proportion of variance explained by the covariance, but did not discuss the sign or magnitude of this correlation.

(4) Although the authors included a correlation parameter between DGE and IGE in their updated model, they did not specify if the residual errors were correlated, too. In fact, they did not even specify a distribution for them. It is already known that allowing for correlated errors may not change the estimates (Haug et al, 2021), but in some settings it can be important (Bergsma et al, 2008).

(5) In appendix S4, they say that the "ordinal" model (I am not sure of what they meant by this word) "defines polygenic DGE and IGE by random effects without fixed effects for each SNP". However, this is not correct; see Baud et al (2021), for instance. In any LMM, it is straightforward to include a single fixed effect for a given SNP, and to do it one SNP at a time. Moreover, they claimed that "compared to the ordinal model (Equation S4), the proposed model (Equation 1) is more extensible to incorporate SNP-wise fixed effects while distinguishing variance-covariance matrices", without providing more evidence than this statement.

(6) The authors seemed keen to convince us that the fact that their model is analogous to the Ising model of ferromagnetics was an advantage in itself. But why would it be? Beyond the mere analogy, it should be a matter of modelling choice, and thus be clearly motivated. For instance, they chose to assess the strength of the association between the trait in the focal individual (y_{k_i}) and the average (dis)similarity between the focal individual and all its neighbors (in neighborhood k), calling the latter "indirect genetic effect". Moreover, it is not clear if what they called "IGE" is $\beta_{q,2}$, u_2 , both, or also $\beta_{q,12}$, etc? Furthermore, they should have used another term as this is not the same as the "indirect genetic effects" of the other models. In these models, what is called the indirect genetic effects can be modeled as depending on group size (see Hadfield and Wilson, 2007; Bijma, 2010). In which sense would the approach of the authors be better? How does it relate to the other models? Do they have more power? Is their term more interpretable?

(7) Another way in which the authors' model may be different from the other models is in the way it models interactions between direct genetic effects and aggregate (dis)similarity between focal and neighbors. At the level of the polygenic components, other models simply have a (DGExIGE) term capturing the deviations from the additivity of DGE + IGE (e.g., Wright, 1985, in the multispecific context). Here, the authors indeed mentioned "interactions between polygenic DGEs and IGEs" and introduced the K_{12} matrix, but it is not clear how different (or similar) it is from the more classical (DGExIGE) term. At the level of the oligogenic component, the authors introduced $\beta_{q,12}$, but it is not clear, to me at least, how it relates to K_{12} and K_{21} .

(8) The authors checked their model on simulated data for various levels of correlation between u_1 (GE) and u_2 .

(9) It is not clear why they have higher absolute errors with negative covariance than with a positive one.

(10) As a causative IGE SNP, the authors considered one with a $\beta_{q,2}$ significantly different from 0. However, they also have two other coefficients, $\beta_{q,1}$ and $\beta_{q,12}$, for each SNP q . How is the FDR in RAINBOW controlled in such a case? This is not detailed.

(11) In their simulations, the causative IGE SNPs were also causative DGE SNPs. However, this may increase power. From the manuscript title, one could assume that the authors' goal was to distinguish between the SNPs that are both DGE and IGE, versus the ones that are IGEs only.

(12) From what I understood, the authors first estimated the (co)variance components once and for all on the model without any SNP, and they then used the values to fit the GWAS model one SNP at a time. This assumes that the inclusion of SNP effects modeled as fixed would not change anything regarding the (co)variance components, but this is not warranted.

(13) The authors applied their model to three datasets of perennial plants.

(14) They only used their model and did not provide evidence that their model gave a significant improvement compared to other models, such as the one of Baud et al (2021).

(15) In Figures 3, 4 and 5, having an indication of which cases have a significant correlation between u_1 and u_2 would have helped.

(16) Concerning the Aspen dataset, it is not clear why the authors claimed that "the negative effects of neighboring genotypes were amplified as trees matured" as the PVE_cov in Figure 3 in 2015 are not systematically more negative than those of Figure 3 in 2014.

(17) When discussing their results, the authors should engage more with the literature estimating DGE-IGE correlations (see some of the references above).

(18) Concerning the apple dataset, they mentioned that "metabolite accumulation in ripening fruits may be facilitated by volatile chemicals, such as ethylene", but they did not find any evidence for significant IGE SNPs localized close to a gene involved in ethylene production. Claiming that these are testable hypotheses should have been made earlier, in the introduction, than a posteriori in the discussion.

<https://doi.org/10.7554/eLife.111650.1.sa0>