



Neighbor GWAS: incorporating neighbor genotypic identity into genome-wide association studies of field herbivory

Yasuhiro Sato ^{1,2} · Eiji Yamamoto ^{1,3} · Kentaro K. Shimizu ^{4,5} · Atsushi J. Nagano ⁶

Received: 14 March 2020 / Revised: 25 December 2020 / Accepted: 28 December 2020 / Published online: 29 January 2021
© The Author(s) 2021. This article is published with open access

Abstract

An increasing number of field studies have shown that the phenotype of an individual plant depends not only on its genotype but also on those of neighboring plants; however, this fact is not taken into consideration in genome-wide association studies (GWAS). Based on the Ising model of ferromagnetism, we incorporated neighbor genotypic identity into a regression model, named “Neighbor GWAS”. Our simulations showed that the effective range of neighbor effects could be estimated using an observed phenotype when the proportion of phenotypic variation explained (PVE) by neighbor effects peaked. The spatial scale of the first nearest neighbors gave the maximum power to detect the causal variants responsible for neighbor effects, unless their effective range was too broad. However, if the effective range of the neighbor effects was broad and minor allele frequencies were low, there was collinearity between the self and neighbor effects. To suppress the false positive detection of neighbor effects, the fixed effect and variance components involved in the neighbor effects should be tested in comparison with a standard GWAS model. We applied neighbor GWAS to field herbivory data from 199 accessions of *Arabidopsis thaliana* and found that neighbor effects explained 8% more of the PVE of the observed damage than standard GWAS. The neighbor GWAS method provides a novel tool that could facilitate the analysis of complex traits in spatially structured environments and is available as an R package at CRAN (<https://cran.rproject.org/package=rNeighborGWAS>).

Associate editor Jinliang Wang

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-020-00401-w>.

✉ Kentaro K. Shimizu
kentaro.shimizu@ieu.uzh.ch

✉ Atsushi J. Nagano
anagano@agr.ryukoku.ac.jp

- ¹ PRESTO, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan
- ² Research Institute for Food and Agriculture, Ryukoku University, Yokotani 1-5, Seta Oe-cho, Otsu, Shiga 520-2194, Japan
- ³ Graduate School of Agriculture, Meiji University, Higashi-Mita 1-1-1, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan
- ⁴ Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
- ⁵ Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka, Totsuka-ward, Yokohama 244-0813, Japan
- ⁶ Faculty of Agriculture, Ryukoku University, Yokotani 1-5, Seta Oe-cho, Otsu, Shiga 520-2194, Japan

Introduction

Plants are immobile and thus cannot escape their neighbors. In natural and agricultural systems, individual phenotypes depend not only on the plants’ own genotype but also on the genotypes of other neighboring plants (Tahvanainen and Root 1972; Barbosa et al. 2009; Underwood et al. 2014). This phenomenon has been termed neighbor effects or associational effects in plant ecology (Barbosa et al. 2009; Underwood et al. 2014; Sato 2018). Such neighbor effects were initially reported as a form of interspecific interaction among different plant species (Tahvanainen and Root 1972), but many studies have illustrated that neighbor effects occur among different genotypes within a plant species with respect to: (i) herbivory (Schuman et al. 2015; Sato 2018; Ida et al. 2018), (ii) pathogen infections (Mundt 2002; Zeller et al. 2012), and (iii) pollinator visitations (Underwood et al. 2020). Although neighbor effects are of considerable interest in plant science (Dicke and Baldwin 2010; Erb 2018) and agriculture (Zeller et al. 2012; Dahlin et al. 2018), they are often not considered in quantitative genetic analyses of field-grown plants.

Complex mechanisms underlie neighbor effects through direct competition (Weiner 1990), herbivore and pollinator

movement (Bergvall et al. 2006; Verschut et al. 2016; Underwood et al. 2020), and volatile communication among plants (Schuman et al. 2015; Dahlin et al. 2018). For example, lipoxygenase (*LOX*) genes govern jasmonate-mediated volatile emissions in wild tobacco (*Nicotiana attenuata*) that induce defenses of neighboring plants (Schuman et al. 2015). Even if direct plant–plant communications are absent, herbivores can mediate indirect interactions between plant genotypes (Sato and Kudoh 2017; Ida et al. 2018). For example, the *GLABRA1* gene is known to determine hairy or glabrous phenotypes in *Arabidopsis* plants (Hauser et al. 2001), and the flightless leaf beetle (*Phaedon brassicae*) is known to prefer glabrous plants to hairy ones (Sato et al. 2017). Consequently, hairy plants escape herbivory when surrounded by glabrous plants (Sato and Kudoh 2017). Yet, there are few hypothesis-free approaches currently available for the identification of the key genetic variants responsible for plant neighborhood effects.

Genome-wide association studies (GWAS) have been increasingly adopted to resolve the genetic architecture of complex traits in the model plant, *Arabidopsis thaliana* (Atwell et al. 2010; Seren et al. 2017; Togninalli et al. 2018), and crop species (Hamblin et al. 2011). The interactions of plants with herbivores (Brachi et al. 2015; Nallu et al. 2018), microbes (Horton et al. 2014; Wang et al. 2018), and other plant species (Frachon et al. 2019) are examples of the complex traits that are investigated through the lens of GWAS. To distinguish causal variants from the genome structure, GWAS often employs a linear mixed model with kinship considered as a random effect (Kang et al. 2008; Korte and Farlow 2013). However, because of combinatorial explosion, it is generally impossible to test the full set of inter-genomic locus-by-locus interactions (Gondro et al. 2013); thus, some feasible and reasonable approach should be developed for the GWAS of neighbor effects.

To incorporate neighbor effects into GWAS, we have focused on a theoretical model of neighbor effects in magnetic fields, known as the Ising model (Ising 1925; McCoy and Maillard 2012), which has been applied to forest gap dynamics (Kizaki and Katori 1999; Schlicht and Iwasa 2004) and community assembly (Azaele et al. 2010) in plant ecology. Using the Ising analogy, we compare individual plants to a magnet: the two alleles at each locus correspond to the north and south dipoles, and genome-wide multiple testing across all loci is analogous to a number of parallel two-dimensional layers. The Ising model has a clear advantage in its interpretability, such that: (i) the optimization problem for a population sum of trait values can be regarded as an inverse problem of a simple linear model, (ii) the sign of neighbor effects determines the model's trend with regard to the generation of a clustered or

checkered spatial pattern of the two states, and (iii) the self-genotypic effect determines the general tendency to favor one allele over another (Fig. 1).

In this study, we proposed a new methodology integrating GWAS and the Ising model, named “neighbor GWAS.” The method was applied to simulated phenotypes and actual data of field herbivory on *A. thaliana*. We addressed two specific questions: (i) what spatial and genetic factors influenced the power to detect causal variants? and (ii) were neighbor effects significant sources of leaf damage variation in field-grown *A. thaliana*? Based on the simulation and application, we determined the feasibility of our approach to detect neighbor effects in field-grown plants.

Materials and methods

Neighbor GWAS

Basic model

We analyzed neighbor effects in GWAS as an inverse problem of the two-dimensional Ising model, named “neighbor GWAS” hereafter (Fig. 1). We considered a situation where a plant accession has one of two alleles at each locus, and a number of accessions occupied a finite set of field sites, in a two-dimensional lattice. The allelic status at each locus was represented by x , and so the allelic status at each locus of the i th focal plant and the j th neighboring plants was designated as $x_{i(j)} \in \{-1, +1\}$. Based on a two-dimensional Ising model, we defined a phenotype value for the i th focal individual plant y_i as:

$$y_i = \beta_1 x_i + \beta_2 \sum_{\langle i,j \rangle} x_i x_j \quad (1)$$

where β_1 and β_2 denoted self-genotype and neighbor effects, respectively. If two neighboring plants shared the same allele at a given locus, the product $x_i x_j$ turned into $(-1) \times (-1) = +1$ or $(+1) \times (+1) = +1$. If two neighbors had different alleles, the product $x_i x_j$ became $(-1) \times (+1) = -1$ or $(+1) \times (-1) = -1$. Accordingly, the effects of neighbor genotypic identity on a particular phenotype depended on the coefficient β_2 and the number of the two alleles in a neighborhood. If the numbers of identical and different alleles were the same near a focal plant, these neighbors offset the sum of the products between the focal plant i and all j neighbors $\sum_{\langle i,j \rangle} x_i x_j$ and exerted no effects on a phenotype. When we summed up the phenotype values for the total number of plants n and replaced it as $E = -\beta_2$, $H = -\beta_1$ and $\epsilon_I = \sum y_i$, Eq. 1 could be transformed into $\epsilon_I = -E \sum_{\langle i,j \rangle} x_i x_j - H \sum x_i$, which defined the interaction energy of a two-dimensional ferromagnetic Ising model

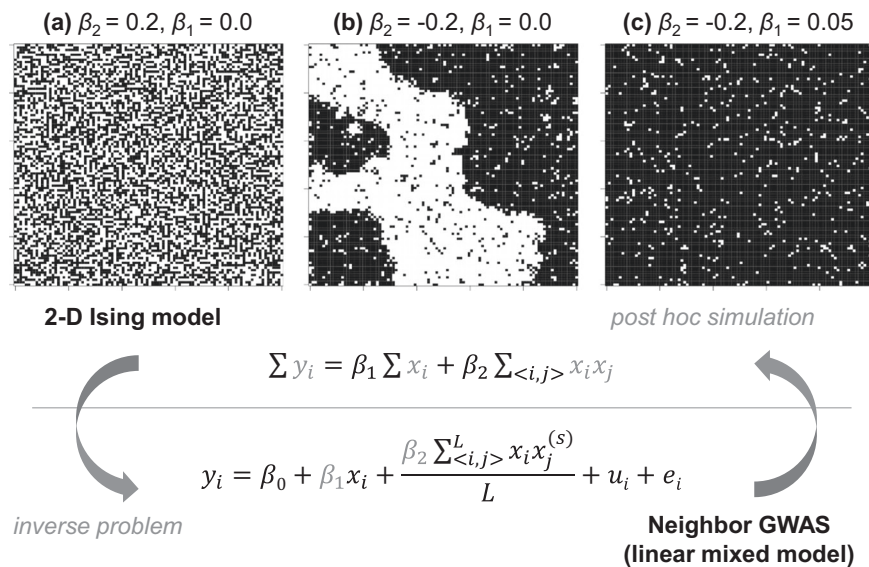


Fig. 1 Relationship between the neighbor GWAS and Ising model. Upper panels show the spatial arrangements expected by a 2-D Ising model $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{\langle i,j \rangle} x_i x_j$. **a** If $\beta_2 > 0$, mixed patterns give the argument of the minimum for a population sum of phenotype values $\sum y_i$. **b** If $\beta_2 < 0$, clustered patterns give the argument of the minimum for $\sum y_i$. **c** In addition, β_1 determines the overall patterns favoring -1 or $+1$ states. The figures show outcomes from a random 100×100 lattice after 1000 iterations of simulated annealing. Conversely, the neighbor GWAS was implemented as an inverse problem

(McCoy and Maillard 2012). The neighbor effect β_2 and self-genotype effect β_1 were interpreted as the energy coefficient E and external magnetic effects H , respectively. An individual plant represented a spin and the two allelic states of each locus corresponded to a north or south dipole. The positive or negative value of $\sum x_i x_j$ indicated a ferromagnetism or paramagnetism, respectively. In this study, we did not consider the effects of allele dominance because this model was applied to inbred plants. However, heterozygotes could be processed if the neighbor covariate $x_i x_j$ was weighted by an estimated degree of dominance in the self-genotypic effects on a phenotype.

Association tests

For association mapping, we needed to determine β_1 and β_2 from the observed phenotypes and considered a confounding sample structure as advocated by previous GWAS (e.g., Kang et al. 2008; Korte and Farlow 2013). Extending the basic model (Eq. (1)), we described a linear mixed model at an individual level as:

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{L} \sum_{\langle i,j \rangle}^L x_i x_j^{(s)} + u_i + e_i \quad (2)$$

where β_0 indicated the intercept; the term $\beta_1 x_i$ represented fixed self-genotype effects as tested in standard GWAS; and

of the 2-D Ising model, where genotypes and its spatial arrangement, x_i and $x_i x_j$, were given while the coefficients β_1 and β_2 were to be estimated from the observed phenotypes y_i . In addition, the variance component due to self and neighbor effects was considered a random effect in a linear mixed model, such that $u_i \in \mathbf{u}$ and $\mathbf{u} \sim \text{Norm}(0, \sigma_1^2 \mathbf{K}_1 + \sigma_2^2 \mathbf{K}_2)$. Once β_1 and β_2 were determined, we could simulate a genotype distribution that maximizes or minimizes $\sum y_i$.

β_2 was the coefficient of fixed neighbor effects. The neighbor covariate $\sum_{\langle i,j \rangle}^L x_i x_j^{(s)}$ indicated a sum of products for all combinations between the i th focal plant and the j th neighbor at the s th spatial scale from the focal plant i , and was scaled by the number of neighboring plants, L . The number of neighboring plants L was dependent on the spatial scale s to be referred. Variance components due to the sample structure of self and neighbor effects were modeled by a random effect $u_i \in \mathbf{u}$ and $\mathbf{u} \sim \text{Norm}(0, \sigma_1^2 \mathbf{K}_1 + \sigma_2^2 \mathbf{K}_2)$. The residual was expressed as $e_i \in \mathbf{e}$ and $\mathbf{e} \sim \text{Norm}(0, \sigma_e^2 \mathbf{I})$, where \mathbf{I} represented an identity matrix.

Variation partitioning

To estimate the proportion of phenotypic variation explained (PVE) by the self and neighbor effects, we utilized variance component parameters in linear mixed models. The $n \times n$ variance-covariance matrices represented the similarity in self-genotypes (i.e., kinship) and neighbor covariates among n individual plants as $\mathbf{K}_1 = \frac{1}{q-1} \mathbf{X}_1^T \mathbf{X}_1$ and $\mathbf{K}_2 = \frac{1}{q-1} \mathbf{X}_2^T \mathbf{X}_2$, where the elements of n plants $\times q$ markers matrix \mathbf{X}_1 and \mathbf{X}_2 consisted of explanatory variables for the self and neighbor effects as $\mathbf{X}_1 = (x_i)$ and $\mathbf{X}_2 = (\frac{\sum_{\langle i,j \rangle}^L x_i x_j^{(s)}}{L})$. As we defined $x_{i(j)} \in \{+1, -1\}$, the elements of the kinship matrix \mathbf{K}_1 were scaled to

represent the proportion of marker loci shared among $n \times n$ plants such that $\mathbf{K}_1 = \left(\frac{k_{ij}+1}{2}\right)$; σ_1^2 and σ_2^2 indicated variance component parameters for the self and neighbor effects.

The individual-level formula Eq. (2) could also be converted into a conventional matrix form as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (3)$$

where \mathbf{y} was an $n \times 1$ vector of the phenotypes; \mathbf{X} was a matrix of fixed effects, including a unit vector, self-genotype x_i , neighbor covariate $(\sum_{<i,j>}^L x_i x_j^{(s)})/L$, and other confounding covariates for n plants; $\boldsymbol{\beta}$ was a vector that represents the coefficients of the fixed effects; \mathbf{Z} was a design matrix allocating individuals to a genotype, and became an identity matrix if all plants were different accessions; \mathbf{u} was the random effect with $\text{Var}(\mathbf{u}) = \sigma_1^2 \mathbf{K}_1 + \sigma_2^2 \mathbf{K}_2$; and \mathbf{e} was residual as $\text{Var}(\mathbf{e}) = \sigma_e^2 \mathbf{I}$.

Because our objective was to test for neighbor effects, we needed to avoid the detection of false positive neighbor effects. The self-genotype value x_i and neighbor genotypic identity $\sum_{<i,j>}^L x_i x_j^{(s)}$ would become correlated explanatory variables in a single regression model (*sensu* colinear) due to the minor allele frequency (MAF) and the spatial scale of s . When MAF is low, neighbors $x_j^{(s)}$ are unlikely to vary in space and most plants will have similar values for neighbor identity $\sum_{<i,j>}^L x_i x_j^{(s)}$. Furthermore, if the neighbor effects range was broad enough to encompass an entire field (i.e., $s \rightarrow \infty$), the neighbor covariate and self-genotype x_i would become colinear according to the equation: $(\sum_{<i,j>}^L x_i x_j^{(s)})/L = x_i (\sum_{j=1}^L x_j^{(s)})/L = x_i \bar{x}_j$, where \bar{x}_j indicates a population-mean of neighbor genotypes and corresponds to a population-mean of self-genotype values \bar{x}_i , if $s \rightarrow \infty$. The standard GWAS is a subset of the neighbor GWAS and these two models become equivalent at $s=0$ and $\sigma_2^2=0$. When testing the self-genotype effect β_1 , we recommend that the neighbor effects and its variance component σ_2^2 should be excluded; otherwise, the standard GWAS fails to correct a sample structure because of the additional variance component at $\sigma_2^2 \neq 0$. To obtain a conservative conclusion, the significance of β_2 and σ_2^2 should be compared using the standard GWAS model based on self-effects alone.

Given the potential collinearity between the self and neighbor effects, we defined different metrics for the proportion of phenotypic variation explained (PVE) based on self or neighbor effects. Using a single-random effect model, we calculated PVE for either the self or neighbor effects as follows:

‘single’ $\text{PVE}_{\text{self}} = \sigma_1^2 / (\sigma_1^2 + \sigma_e^2)$ when s and σ_2^2 were set at 0, or

‘single’ $\text{PVE}_{\text{nei}} = \sigma_2^2 / (\sigma_2^2 + \sigma_e^2)$ when σ_1^2 was set at 0.

Using a two-random effect model, we could focus on one variable while considering relationships between two variables (*sensu* partial out) for either of the two variance components. We defined such a partial PVE as:

‘partial’ $\text{PVE}_{\text{self}} = \sigma_1^2 / (\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$ and

‘partial’ $\text{PVE}_{\text{nei}} = \sigma_2^2 / (\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$.

As the partial PVE_{self} was equivalent to the single PVE_{self} when s was set at 0, the net contribution of neighbor effects at $s \neq 0$ was given as

‘net’ $\text{PVE}_{\text{nei}} = (\text{partial } \text{PVE}_{\text{self}} + \text{partial } \text{PVE}_{\text{nei}}) - \text{single } \text{PVE}_{\text{self}}$,

which indicated the proportion of phenotypic variation that could be explained by neighbor effects, but not by the self-genotype effects.

Simulation

To examine the model performance, we applied the neighbor GWAS to simulated phenotypes. Phenotypes were simulated using a subset of the actual *A. thaliana* genotypes. To evaluate the performance of the simple linear model, we assumed a complex ecological form of neighbor effects with multiple variance components controlled. The model performance was evaluated in terms of the causal variant detection and accuracy of estimates. All analyses were performed using R version 3.6.0 (R Core Team 2019).

Genotype data

To consider a realistic genetic structure in the simulation, we used part of the *A. thaliana* RegMap panel (Horton et al. 2012). The genotype data for 1307 accessions were downloaded from the Joy Bergelson laboratory website (http://bergelson.uchicago.edu/?page_id=790 accessed on February 9, 2017). We extracted data for chromosomes 1 and 2 with MAF at >0.1 , yielding a matrix of 1307 plants with 65,226 single nucleotide polymorphisms (SNPs). Pairwise linkage disequilibrium (LD) among the loci was $r^2 = 0.003$ [0.00–0.06: median with upper and lower 95 percentiles]. Before generating a phenotype, genotype values at each locus were standardized to a mean of zero and a variance of 1. Subsequently, we randomly selected 1,296 accessions (= 36×36 accessions) without any replacements for each iteration and placed them in a 36×72 checkered space, following the *Arabidopsis* experimental settings (see Fig. S1).

Phenotype simulation

To address ecological issues specific to plant neighborhood effects, we considered two extensions, namely asymmetric neighbor effects and spatial scales. Previous

studies have shown that plant–plant interactions between accessions are sometimes asymmetric under herbivory (e.g., Bergvall et al. 2006; Verschut et al. 2016; Sato and Kudoh 2017) and height competition (Weiner 1990); where one focal genotype is influenced by neighboring genotypes, while another receives no neighbor effects. Such asymmetric neighbor effects can be tested by statistical interaction terms in a linear model (Bergvall et al. 2006; Sato and Kudoh 2017). Several studies have also shown that the strength of neighbor effects depends on spatial scales (Hambäck et al. 2014), and that the scale of neighbors to be analyzed relies on the dispersal ability of the causative organisms (see Hambäck et al. 2009; Sato and Kudoh 2015; Verschut et al. 2016; Ida et al. 2018 for insect and mammal herbivores; Rieux et al. 2014 for pathogen dispersal) or the size of the competing plants (Weiner 1990). We assumed the distance decay at the s th sites from a focal individual i with the decay coefficient α as $w(s, \alpha) = e^{-\alpha(s-1)}$, since such an exponential distance decay has been widely adopted in empirical studies (Devaux et al. 2007; Carrasco et al. 2010; Rieux et al. 2014; Ida et al. 2018). Therefore, we assumed a more complex model for simulated phenotypes than the model for neighbor GWAS as follows:

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{L} \sum_{<i,j>}^L w(s, \alpha) x_i x_j^{(s)} + \beta_{12} \frac{x_i}{L} \sum_{<i,j>}^L w(s, \alpha) x_i x_j^{(s)} + u_i + e_i \quad (4)$$

where β_{12} was the coefficient for asymmetry in neighbor effects. By incorporating an asymmetry coefficient, the model (Eq. (4)) can deal with cases where neighbor effects are one-sided or occur irrespective of a focal genotype (Fig. 2). Total variance components resulting from three background effects (i.e., the self, neighbor, and self-by-neighbor effects) were defined as $u_i \in \mathbf{u}$ and $\mathbf{u} \sim \text{Norm}(0, \sigma_1^2 \mathbf{K}_1 + \sigma_2^2 \mathbf{K}_2 + \sigma_{12}^2 \mathbf{K}_{12})$. The three variance component parameters σ_1^2 , σ_2^2 , and σ_{12}^2 , determined the relative importance of the self-genotype, neighbor, and asymmetric neighbor effects in u_i . Given the elements of n plants \times q marker explanatory matrix with $\mathbf{X}_{12} = (\frac{x_i}{L} \sum_{<i,j>}^L w(s, \alpha) x_i x_j^{(s)})$, the similarity in asymmetric neighbor effects was calculated as $\mathbf{K}_{12} = \frac{1}{q-1} \mathbf{X}_{12}^T \mathbf{X}_{12}$. To control phenotypic variations, we further partitioned the proportion of phenotypic variation into those explained by the major-effect genes and variance components $\text{PVE}_\beta + \text{PVE}_u$, major-effect genes alone PVE_β , and residual error PVE_e , where $\text{PVE}_\beta + \text{PVE}_u + \text{PVE}_e = 1$. The *optimize* function in R was used to adjust the simulated phenotypes to the given amounts of PVE.

Parameter setting

Ten phenotypes were simulated with varying combination of the following parameters, including the distance decay coefficient α , the proportion of phenotypic variation explained by the major-effect genes PVE_β , the proportion of phenotypic variation explained by major-effect genes and variance components $\text{PVE}_\beta + \text{PVE}_u$, and the relative contributions of self, symmetric neighbor, and asymmetric neighbor effects, i.e., $\text{PVE}_{\text{self}}:\text{PVE}_{\text{nei}}:\text{PVE}_{\text{sn}}$. We ran the simulation with different combinations, including $\alpha = 0.01, 1.0, \text{ or } 3.0$; $\text{PVE}_{\text{self}}:\text{PVE}_{\text{nei}}:\text{PVE}_{\text{sn}} = 8:1:1, 5:4:1, \text{ or } 1:8:1$; and PVE_β and $\text{PVE}_\beta + \text{PVE}_u = 0.1 \text{ and } 0.4, 0.3 \text{ and } 0.4, 0.3 \text{ and } 0.8, \text{ or } 0.6 \text{ and } 0.8$. The maximum reference scale was fixed at $s = 3$. The line of simulations was repeated for 10, 50, or 300 causal SNPs to examine cases of oligogenic and polygenic control of a trait. The non-zero coefficients (i.e., signals) for the causal SNPs were randomly sampled from -1 or 1 digit and then assigned, as some causal SNPs were responsible for both the self and neighbor effects. Of the total number of causal SNPs, 15% had self, neighbor, and asymmetric neighbor effects (i.e., $\beta_1 \neq 0$ and $\beta_2 \neq 0$ and $\beta_{12} \neq 0$); another 15% had both the self and neighbor effects, but no asymmetry in the neighbor effects ($\beta_1 \neq 0$ and $\beta_2 \neq 0$ and $\beta_{12} = 0$); another 35% had self-genotypic effects only ($\beta_1 \neq 0$); and the remaining 35% had neighbor effects alone ($\beta_2 \neq 0$). Given its biological significance, we assumed that some loci having neighbor signals possessed asymmetric interactions between the neighbors ($\beta_2 \neq 0$ and $\beta_{12} \neq 0$), while the others had symmetric interactions ($\beta_2 \neq 0$ and $\beta_{12} = 0$). Therefore, the number of causal SNPs in β_{12} was smaller than that in the main neighbor effects β_2 . According to this assumption, the variance component σ_{12}^2 was also assumed to be smaller than σ_2^2 . To examine extreme conditions and strong asymmetry in neighbor effects, we additionally analyzed the cases with $\text{PVE}_{\text{self}}:\text{PVE}_{\text{nei}}:\text{PVE}_{\text{sn}} = 1:0:0, 0:1:0, \text{ or } 1:1:8$.

Summary statistics

The simulated phenotypes were fitted by Eq. (2) to test the significance of coefficients β_1 and β_2 , and to estimate single or partial PVE_{self} and PVE_{nei} . To deal with potential collinearity between x_i and neighbor genotypic identity $\sum_{<i,j>}^L x_i x_j^{(s)}$, we performed likelihood ratio tests between the self-genotype effect model and the model with both self and neighbor effects, which resulted in conservative tests of significance for β_2 and σ_2^2 . The simulated phenotype values were standardized to have a mean of zero and a variance of 1, where true β was expected to match the estimated coefficients $\hat{\beta}$ when multiplied by the standard deviation of non-standardized phenotype values.

using the true positive rate of the ROC curves (i.e., y-axis of the ROC curve) at a stringent specificity level, where the false positive rate (x-axis of the ROC curve) = 0.05. The roc function in the pROC package (Robin et al. 2011) was used to calculate the ROC and AUC from $-\log_{10}(p \text{ value})$. Factors affecting the AUC or sensitivity were tested by analysis-of-variance (ANOVA) for the self or neighbor effects (AUC_{self} or AUC_{nei} ; self or neighbor sensitivity). The AUC and PVE were calculated from $s = 1$ (the first nearest neighbors) to $s = 3$ (up to the third nearest neighbors) cases. The AUC was also calculated using standard linear models without any random effects, to examine whether the linear mixed models were superior to the linear models. We also tested the neighbor GWAS model incorporating the neighbor phenotype $y_j^{(s)}$ instead of $x_j^{(s)}$. The accuracy of the total PVE estimates was defined as $\text{PVE accuracy} = (\text{estimated total PVE} - \text{true total PVE}) / \text{true total PVE}$. The accuracy of the effect size estimates was evaluated using mean absolute errors (MAE) between the true and estimated β_1 or β_2 for the self and neighbor effects (MAE_{self} and MAE_{nei}). Factors affecting the accuracy of PVE and effect size estimates were also tested using ANOVA. Misclassifications between self and neighbor signals were further evaluated by comparing p value scores between zero and non-zero coefficients. If $-\log_{10}(p \text{ value})$ scores of zero β are the same or larger than non-zero β , it infers a risk of misspecification of the true signals.

Arabidopsis herbivory data

We applied the neighbor GWAS to field data of *Arabidopsis* herbivory. The procedure for this field experiment followed that of our previous experiment (Sato et al. 2019). We selected 199 worldwide accessions (Table S1) from 2029 accessions sequenced by the RegMap (Horton et al. 2012) and 1001 Genomes project (Alonso-Blanco et al. 2016). Of the 199 accessions, most were overlapped with a previous GWAS of biotic interactions (Horton et al. 2014) and half were included by a GWAS of glucosinolates (Chan et al. 2010). Eight replicates of each of the 199 accessions were first prepared in a laboratory and then transferred to the outdoor garden at the Center for Ecological Research, Kyoto University, Japan (Otsu, Japan: 35°06'N, 134°56'E, alt. ca. 200 m; Fig. S1a). Seeds were sown on Jiffy-seven pots (33-mm diameter) and stratified at a temperature of 4 °C for a week. Seedlings were cultivated for 1.5 months under a short-day condition (8 h light: 16 h dark, 20 °C). Plants were then separately potted in plastic pots (6 cm in diameter) filled with mixed soil of agricultural compost (Metro-mix 350, SunGro Co., USA) and perlite at a 3:1 ratio. Potted plants were set in plastic trays (10 × 40 cells) in a checkered pattern (Fig. S1b). In the field setting, a set of 199 accessions and an additional Col-0 accession were

randomly assigned to each block without replacement (Fig. S1c). Eight replicates of these blocks were set >2 m apart from each other (Fig. S1c). Potted plants were exposed to the field environment for 3 weeks in June 2017. At the end of the experiment, the percentage of foliage eaten was scored as: 0 for no visible damage, 1 for ≤10%, 2 for >10% and ≤25%, 3 for >25% and ≤50%, 4 for >50% and ≤75%, and 5 for >75%. All plants were scored by a single person to avoid observer bias. The most predominant herbivore in this field trial was the diamond back moth (*Plutella xylostella*), followed by the small white butterfly (*Pieris rapae*). We also recorded the initial plant size and the presence of inflorescence to incorporate them as covariates. Initial plant size was evaluated by the length of the largest rosette leaf (mm) at the beginning of the field experiment and the presence of inflorescence was recorded 2 weeks after transplanting.

We estimated the variance components and performed the association tests for the leaf damage score with the neighbor covariate at $s = 1$ and 2. These two scales corresponded to $L = 4$ (the nearest four neighbors) and $L = 12$ (up to the second nearest neighbors), respectively, in the *Arabidopsis* dataset. The variation partitioning and association tests were performed using the *gaston* package, as mentioned above. To determine the significance of the variance component parameters, we compared the likelihood between mixed models with one or two random effects. For the genotype data, we used an imputed SNP matrix of the 2029 accessions studied by the RegMap (Horton et al. 2012) and 1001 Genomes project (Alonso-Blanco et al. 2016). Missing genotypes were imputed using BEAGLE (Browning and Browning 2009), as described by Togninalli et al. (2018) and updated on the AraGWAS Catalog (<https://aragwas.1001genomes.org>). Of the 10,709,466 SNPs from the full imputed matrix, we used 1,242,128 SNPs with MAF at >0.05 and LD of adjacent SNPs at $r^2 < 0.8$. We considered the initial plant size, presence of inflorescence, experimental blocks, and the edge or center within a block as fixed covariates; these factors explained 12.5% of the leaf damage variation (1.2% by initial plant size, Wald test, $Z = 3.53$, $p \text{ value} < 0.001$; 2.4% by the presence of inflorescence, $Z = -5.69$, $p \text{ value} < 10^{-8}$; 8.3% by the experimental blocks, likelihood ratio test, $\chi^2 = 152.8$, $df = 7$, $p \text{ value} < 10^{-28}$; 0.5% by the edge or center, $Z = 3.11$, $p \text{ value} = 0.002$). After the association mapping, we searched candidate genes within ~10 kb around the target SNPs, based on the Araport11 gene model with the latest annotation of The Arabidopsis Information Resource (TAIR) (accessed on 7 September 2019). Gene-set enrichment analysis was performed using the Gowinda algorithm that enables unbiased analysis of the GWAS results (Kofler and Schlotterer 2012). We tested the SNPs with the top 0.1% $-\log_{10}(p \text{ value})$ scores, with the option

“--gene-definition undownstream10000,” “--min-genes 20,” and “--mode gene.” The GO.db package (Carlson 2018) and the latest TAIR AGI code annotation were used to build input files for gene ontologies (GOs). The R source codes, accession list, and phenotype data are available at the GitHub repository (<https://github.com/naganolab/NeighborGWAS>).

R package, “rNeighborGWAS”

To increase the availability of the new method, we have developed the neighbor GWAS into an R package, which is referred to as “rNeighborGWAS”. In addition to the genotype and phenotype data, the package requires a spatial map indicating the positions of individuals across a space. In this package, we generalized the discrete space example into a continuous two-dimensional space, allowing it to handle any spatial distribution along the x - and y -axes. Based on the three input files, the rNeighborGWAS package estimates the effective range of neighbor effects by calculating partial PVE_{nei} and performs association mapping of the neighbor effects using the linear mixed models described earlier. Details and usage are described in the help files and vignette of the rNeighborGWAS package available via CRAN at <https://cran.r-project.org/package=rNeighborGWAS>.

To assess its implementation, we performed standard GWAS using GEMMA version 0.98 (Zhou and Stephens 2012) and the rNeighborGWAS. The test phenotype data were the leaf damage scores for the 199 accessions described above or flowering time under long-day conditions for 1057 accessions (“FT16” phenotype collected by Atwell et al. 2010 and Alonso-Blanco et al. 2016). The flowering time phenotype was downloaded from the AraPheno database (<https://arapheno.1001genomes.org/>; Seren et al. 2017). The full imputed genotype data were compiled for 1057 accessions, whose genotypes and flowering time phenotype were both available. The cut-off value of the MAF was set at 5%, yielding 1,814,755 SNPs for the 1057 accessions. The same kinship matrix defined by K_1 above was prepared as an input file. We calculated p values using likelihood ratio tests in the GEMMA program, because the rNeighborGWAS adopted likelihood ratio tests.

Results

Simulation

We conducted simulations to test the capability of the neighbor GWAS to estimate PVE and marker-effects. As expected by the model and data structure, correlation was detected between the self-genotypic variable x_i and the

neighbor variable $\sum x_j/L$ in the simulated genotypes (Fig. S2). The level of Pearson’s correlation coefficient r varied from a slight correlation to complete correlation as the MAF became smaller, from 0.5 to 0.1 (Fig. S2). The correlation was also more severe as the scale of s was increased. For example, even at $s = 2$, we could cut off the MAF at >0.4 to keep $|r|$ below 0.6 for all SNPs. The element-wise correlation between K_1 and K_2 indicated that at least 60% of the variation was overlapping between the two genome-wide variance-covariance matrices in the partial genotype data used for this simulation ($R^2 = 0.62$ at $s = 1$; $R^2 = 0.79$ at $s = 2$; $R^2 = 0.84$ at $s = 3$).

A set of phenotypes were then simulated from the real genotype data following a complex model (Eq. (4)), and then fitted using a simplified model (Eq. (2)). The accuracy of the total PVE estimation was the most significantly affected by the spatial scales of s (Table 1). The total PVE was explained relatively well by the single PVE_{self} that represented the additive polygenic effects of the self-genotypes (Fig. 3 left). Inclusion of partial PVE_{nei} accounted for the rest of the true total PVE, which was considered the net contribution of neighbor effects to phenotypic variation (Fig. 3 right). The net PVE_{nei} was largest when the effective range of the neighbor effects was narrow (i.e., strong distance decay at $\alpha = 3$) and the contribution of the partial PVE_{nei} was much larger than that of PVE_{self} (the case of $\alpha = 3$ in Fig. 3 right). However, the sum of the single PVE_{self} (=partial PVE_{self} at $s = 0$) and the partial PVE_{nei} did not match the true total PVE (Fig. 3 left), as expected by the correlation between the self and neighbor effects (Fig. S2). Due to such correlation, the single PVE_{self} or single PVE_{nei} mostly overrepresented the actual amounts of PVE_{self} or PVE_{nei} , respectively (Fig. S3). The overrepresentation of the single PVE_{self} and single PVE_{nei} was observed even when either the self or neighbor effects were absent in the simulation (Fig. S4). These results indicate that (i) single PVE_{nei} should not be used, (ii) partial PVE_{nei} suffered from its collinearity with the single PVE_{self} , and (iii) net PVE_{nei} provides a conservative estimate for the genome-wide contribution of neighbor effects to phenotypic variation.

Although the partial PVE_{nei} could not be used to quantify the net contribution of the neighbor effects, this metric inferred spatial scales at which neighbor effects remained effective. If the distance decay was weak (small value of decay coefficient α) and thus the effective range of the neighbor effects was broad, partial PVE_{nei} increased linearly as the reference spatial scale was broadened (the case of $\alpha = 0.01$ in Fig. 3 left). On the other hand, if the distance decay was strong (large value of decay coefficient α) and thus the effective scale of the neighbor effects was narrow, partial PVE_{nei} decreased as the reference spatial scale was broadened or saturated at the scale of the first nearest neighbors (the case of $\alpha = 3$ in Fig. 3 left). Considering the

Table 1 Factors affecting variance estimation and causal variant detection in the simulated phenotypes.

Response	Factors	df	SS	<i>F</i>	<i>p</i> value
PVE accuracy	No. of causal SNPs	1	0.00	0.0	0.95
	PVE_{β}	1	0.01	0.6	0.44
	$PVE_{\beta} + PVE_u$	1	0.02	0.61	0.43
	$PVE_{self}:PVE_{nei}:PVE_{sxn}$	2	0.25	4.95	0.007
	α	1	0.96	38.46	$6.1e-10$
	<i>s</i>	1	8.49	341.0	$<2.2e-16$
	Residuals	4312	107.34	NA	NA
AUC_{self}	No. of causal SNPs	1	13.12	2998.6	$<2.2e-16$
	PVE_{β}	1	0.77	176.6	$<2.2e-16$
	$PVE_{\beta} + PVE_u$	1	0.02	4.04	0.045
	$PVE_{self}:PVE_{nei}:PVE_{sxn}$	2	8.08	923.54	$<2.2e-16$
	α	1	0.01	2.19	0.14
	Residuals	1073	4.69	NA	NA
	AUC_{nei}	No. of causal SNPs	1	25.82	2225.1
AUC_{nei}	PVE_{β}	1	2.30	198.1	$<2.2e-16$
	$PVE_{\beta} + PVE_u$	1	0.03	2.24	0.14
	$PVE_{self}:PVE_{nei}:PVE_{sxn}$	2	20.97	903.48	$<2.2e-16$
	α	1	0.96	83.00	$<2.2e-16$
	<i>s</i>	1	0.079	6.83	0.0090
	Residuals	3232	37.50	NA	NA
	Self sensitivity	No. of causal SNPs	1	74.204	1317.15
PVE_{β}		1	2.236	39.69	$4.0e-10$
$PVE_{\beta} + PVE_u$		1	0.06	1.06	0.30
$PVE_{self}:PVE_{nei}:PVE_{sxn}$		2	11.955	106.10	$<2.2e-16$
α		1	0.089	1.57	0.21
Residuals		1073	60.449	NA	NA
Neighbor sensitivity		No. of causal SNPs	1	98.052	1153.56
	PVE_{β}	1	4.649	54.70	$2.0e-13$
	$PVE_{\beta} + PVE_u$	1	0.016	0.19	0.67
	$PVE_{self}:PVE_{nei}:PVE_{sxn}$	2	23.196	136.45	$<2.2e-16$
	α	1	1.852	21.79	$3.0e-06$
	<i>s</i>	1	0.096	1.13	0.29
	Residuals	3232	274.72	NA	NA
MAE_{self}	No. of causal SNPs	1	105.32	323.44	$<2.2e-16$
	PVE_{β}	1	1.80	5.54	0.02
	$PVE_{\beta} + PVE_u$	1	0.14	0.44	0.51
	$PVE_{self}:PVE_{nei}:PVE_{sxn}$	2	36.11	55.45	$<2.2e-16$

Table 1 (continued)

Response	Factors	df	SS	<i>F</i>	<i>p</i> value
	α	1	0.73	2.23	0.14
	Residuals	1073	349.41	NA	NA
MAE_{nei}	No. of causal SNPs	1	2.73	15.06	$1.0e-04$
	PVE_{β}	1	16.89	93.17	$<2.2e-16$
	$PVE_{\beta} + PVE_u$	1	3.51	19.34	$1.0e-05$
	$PVE_{self}:PVE_{nei}:PVE_{sxn}$	2	80.22	221.25	$<2.2e-16$
	α	1	0.39	2.15	0.14
	<i>s</i>	1	45.87	253.01	$<2.2e-16$
	Residuals	3232	585.91	NA	NA

The accuracy of the proportion of the phenotypic variation explained (PVE) was defined as the PVE accuracy = (estimated total PVE – true total PVE)/true total PVE. The power was represented by the area under the ROC curve (AUC). The sensitivity to distinguish self or neighbor signals (Self or neighbor sensitivity) was evaluated using the true positive rate of the ROC curve, when the false positive rate = 0.05. The accuracy of the effect size estimates was evaluated using the mean absolute errors (MAE) between the true and estimated fixed effects. ANOVA tables show the degree of freedom (df), sum of squares (SS), *F*-statistics, and *p* values. Explanatory factors are the number of causal SNPs, proportion of phenotypic variation explained (PVE) by major-effect genes (PVE_{β}), total PVE by major-effect genes and variance components ($PVE_{\beta} + PVE_u$), relative contribution of self, symmetric, and asymmetric neighbor effects ($PVE_{self}:PVE_{nei}:PVE_{sxn}$), and distance decay coefficient α . For the neighbor effects, the difference of the reference spatial scales ($s = 1-3$) was also considered an explanatory variable. NA means not available.

spatial dependency of the partial PVE_{nei} , we could estimate the effective spatial scales by $\Delta PVE_{nei} = \text{partial } PVE_{nei,s+1} - \text{partial } PVE_{nei,s}$ and by the scale that resulted in the maximum ΔPVE_{nei} as $s = \arg \max \Delta PVE_{nei}$ (Fig. S5).

The spatial scale that yielded the maximum AUC for neighbor signals (AUC_{nei}) coincided with the patterns of the partial PVE_{nei} across the range of *s*. If the distance decay was weak ($\alpha = 0.01$) and thus the effective range of the neighbor effects was broad, the AUC_{nei} increased linearly as the reference spatial scale was broadened (the case of $\alpha = 0.01$ in Fig. 4). If the distance decay was strong (large value of decay coefficient α) and thus the effective scale of the neighbor effects was narrow, the AUC_{nei} did not increase even when the reference spatial scale was broadened (the case of $\alpha = 3$ in Fig. 4). Thus, the first nearest scale was enough to detect neighbor signals, unless the distance decay was very weak.

In terms of the AUC, we also found that the number of causal SNPs, the amount of PVE by neighbor effects (controlled by the total $PVE = PVE_{\beta} + PVE_u$; and ratio of $PVE_{self}:PVE_{nei}$), and the distance decay coefficient α were significant factors affecting the power to detect neighbor signals (AUC_{nei} ; Table 1). The power to detect self-genotype signals (AUC_{self}) depended on the number of

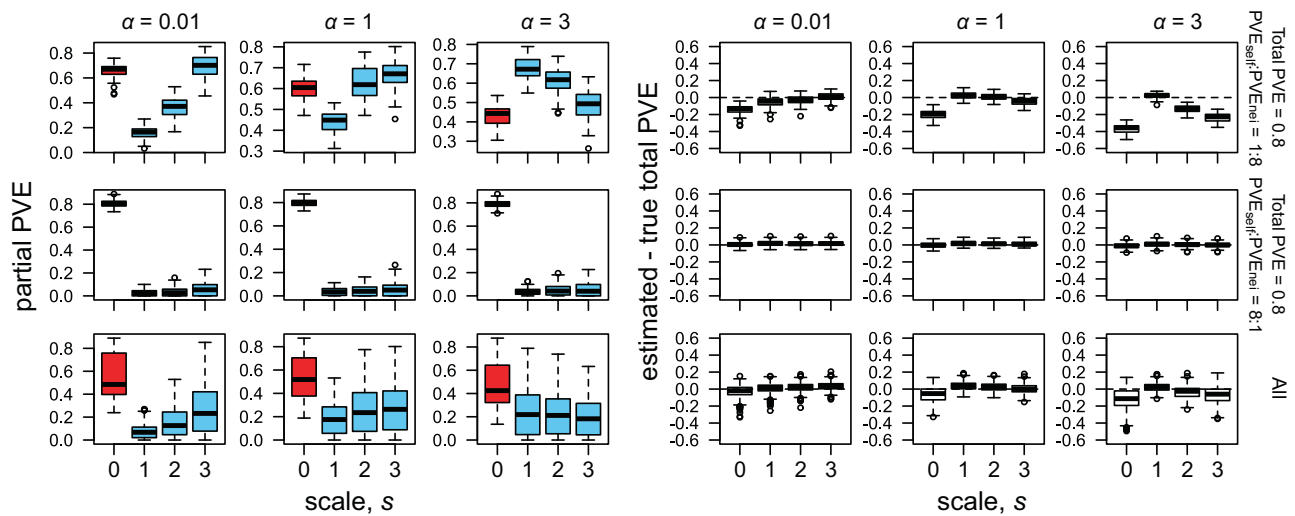


Fig. 3 Spatial scale dependence of PVE estimates in simulated phenotypes. The broad, intermediate, and narrow effective range of neighbor effects are represented by weak ($\alpha = 0.01$), moderate ($\alpha = 1$), and strong ($\alpha = 3$) distance decay coefficients, respectively. Partial PVE (left) and the difference between the true and estimated total PVE (right) are shown along the spatial scale from the first nearest ($s = 1$) to the third nearest ($s = 3$) neighbors, with distinct relative contributions of the self or neighbor effects to a phenotype ($\text{PVE}_{\text{self}}:\text{PVE}_{\text{nei}} = 1:8$ or $8:1$; upper or middle panels) and pooled results among all parameter

conditions (All; lower panels). The difference between the true and estimated total PVE (right) represents the accuracy of total PVE estimates. Boxplots show center line: median, box limits: upper and lower quartiles, whiskers: $1.5 \times$ interquartile range, and points: outliers. In the left panels, red boxes indicate partial PVE_{self} at $s = 0$ (corresponded to single PVE_{self}), while blue boxes indicate partial PVE_{nei} at $s \neq 0$. In the right panels, horizontal dashed lines indicate a perfect match between the estimated and true total PVE.

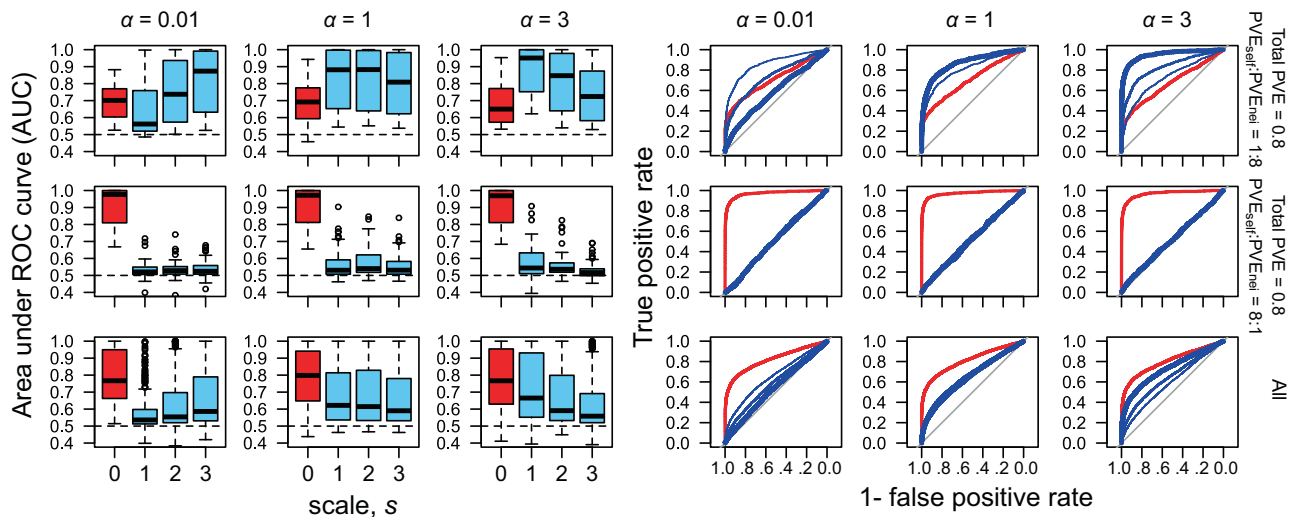


Fig. 4 Spatial scale dependence of the power to detect causal SNPs in simulated phenotypes. The broad, intermediate, and narrow effective range of neighbor effects are represented by weak ($\alpha = 0.01$), moderate ($\alpha = 1$), and strong ($\alpha = 3$) distance decay coefficients, respectively. Receiver operating characteristic (ROC) curves (right) and the area under the ROC curve (AUC) (left) are shown alongside the spatial scales from the first nearest ($s = 1$) to the third nearest ($s = 3$) neighbors, with the distinct relative contributions of the self and neighbor effects to a phenotype ($\text{PVE}_{\text{self}}:\text{PVE}_{\text{nei}} = 1:8$ or $8:1$; upper or

middle panels) and pooled results among all parameter conditions (All; lower panels). Red boxes and curves indicate self-effects, while blue boxes indicate neighbor effects. The thickness of the blue curves indicates reference spatial scales as follows: $s = 1$ (thick), $s = 2$ (medium), or $s = 3$ (thin) for neighbor effects. The horizontal dashed lines in the left panels indicates that the $\text{AUC} = 0.5$, i.e., no detection of causal variants. The ROC curves in the right panels are depicted based on ten iterations with 50 causal SNPs.

causal SNPs and PVE_{β} but was not significantly influenced by the distance decay coefficient of the neighbor effects (α) (Table 1). The power to detect self-genotype signals changed from strong ($\text{AUC}_{\text{self}} > 0.9$) to weak ($\text{AUC}_{\text{self}} < 0.6$),

depending on the number of causal SNPs, the PVE by the major-effect genes, and as the relative contribution from the PVE_{self} increased (Fig. S6 upper). Compared to the self-genotype effects, it was relatively difficult to detect

neighbor signals (Fig. 4 right; Fig. S6 lower), ranging from strong ($AUC_{nei} > 0.9$) to little (AUC_{nei} near to 0.5) power. When the number of causal SNPs = 10, the power to detect neighbor signals decreased from high ($AUC_{nei} > 0.9$) to moderate ($AUC_{nei} > 0.7$) with the decreasing PVE_{β} and the distance decay coefficient (Fig. S6 lower). There was almost no power to detect neighbor signals (AUC_{nei} near to 0.5) when the number of causal SNPs = 50 and PVE_{nei} had low relative contributions (Fig. S6 lower). The result of the simulations indicated that strong neighbor effects were detectable when a target trait was governed by several major genes and the range of neighbor effects was spatially limited. Additionally, linear mixed models outperformed standard linear models as there were 8.8% and 1.4% increases in their power to detect self and neighbor signals, respectively (AUC_{self} , paired t -test, mean of the difference = 0.088, p value $< 10^{-16}$; AUC_{nei} , mean of the difference = 0.014, p value $< 10^{-16}$). When the neighbor phenotype $y_j^{(s)}$ was incorporated instead of the genotype $x_j^{(s)}$, the power to detect neighbor effects was very weak, such that the AUC_{nei} decreased to almost 0.5.

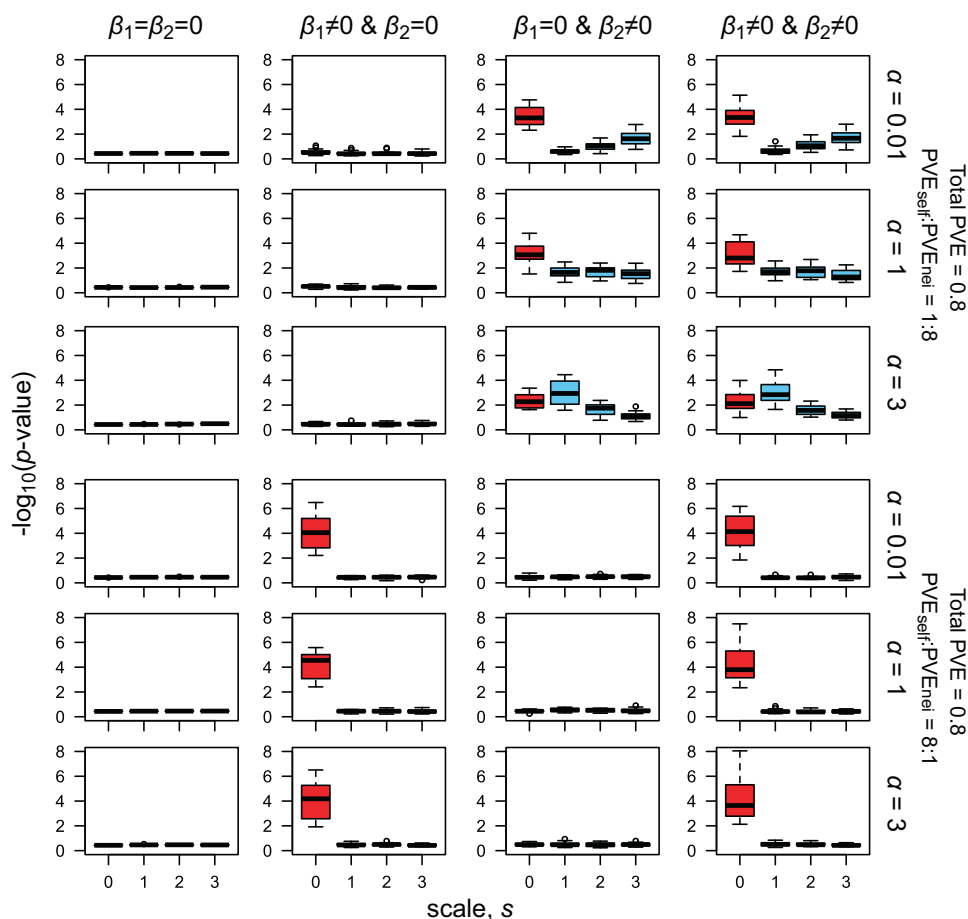
To examine misclassifications between the self and neighbor signals, we compared the sensitivity, effect size estimates, and p value scores among causal SNPs having non-zero coefficients of the true β_1 and β_2 . The sensitivity to distinguish the self or neighbor signals was largely affected by the number of causal SNPs, the amount of PVE by the major-effect genes PVE_{β} , and the relative contribution of the self and neighbor effects (controlled by $PVE_{self}:PVE_{nei}$) (Table 1; Fig. S7). The mean absolute errors of the self-effect estimates for $\hat{\beta}_1$ (MAE_{self}) largely depended on the number of causal SNPs and the relative contribution of the variance components (Table 1; Fig. S8 upper), while those of the neighbor effect estimates for $\hat{\beta}_2$ (MAE_{nei}) were dependent on the relative contribution of the variance components and the spatial scales to be referred (Table 1; Fig. S8 lower). Given that the self and neighbor signals were sufficiently detected when the number of causal SNPs was 50 (Fig. S6 middle column), p values under this condition were compared between the causal and non-causal SNPs. We observed that strong self-signals ($\beta_1 \neq 0$) were unlikely to be detected as neighbor effects (Fig. 5 lower). Causal SNPs responsible for both the self and neighbor effects ($\beta_1 \neq 0$ and $\beta_2 \neq 0$) were better detected than the non-causal SNPs ($\beta_1 = 0$ and $\beta_2 = 0$) (Fig. 5 right column). The sensitivity to distinguish neighbor signals from self-signals was large when the true contribution of the neighbor effects was as large as $PVE_{self}:PVE_{nei} = 1:8$, but decreased when the contribution of the self-effects was as large as $PVE_{self}:PVE_{nei} = 8:1$ (Fig. S7 lower). In contrast, if the contribution of the neighbor effects was relatively large ($PVE_{self}:PVE_{nei} = 1:8$), the SNPs responsible for the neighbor effects alone ($\beta_1 = 0$ and $\beta_2 = 0$) could also be detected as self-effects

(Fig. 5). As expected by the level of correlation (Fig. S2), the false positive detection of the self-signals was more likely when the distance decay coefficient was small and thus the effective range of the neighbor effects was broad (the case of $\alpha = 0.01$ and $\beta_1 = 0$ and $\beta_2 \neq 0$ in Fig. 5). This coincided with the strength of the correlation (Fig. S2), as the false positive detection of self-signals and false negative detection of neighbor signals are more likely if the MAF was small (Fig. S9). Consistent with the false positive detection, the sensitivity to distinguish self-signals from neighbor signals remained large, even when the contribution of the neighbor effects was far larger (the case of $PVE_{self}:PVE_{nei} = 1:8$ in Fig. S7 upper). Strong self-effects (p value $< 10^{-5}$ for $\hat{\beta}_1$) and slight neighbor effects (p value < 0.05 for $\hat{\beta}_2$ at $s = 1$ and $\alpha = 3$) remained when asymmetric neighbor effects were strong ($\beta_1 \neq 0$ and $\beta_2 \neq 0$ and $\beta_{12} \neq 0$ and $PVE_{self}:PVE_{nei}:PVE_{sxn} = 1:1:8$; Fig. S10). These results indicate that (i) the collinearity may lead to the false positive detection of self-effects, yet is unlikely to result in the false positive detection of neighbor effects, and that (ii) smaller MAFs are more likely to cause the false positive detection of self-effects and decrease the power to detect true neighbor effects.

***Arabidopsis* herbivory data**

To estimate PVE_{self} and PVE_{nei} , we applied a linear mixed model (Eq. (3)) to the leaf damage score data for the field-grown *A. thaliana*. The leaf damage variation was significantly explained by the single PVE_{self} that represented additive genetic variation (single $PVE_{self} = 0.173$, $\chi^2_1 = 10.1$, p value = 0.005; Fig. 6a; Fig. S11 right). Variation partitioning showed a significant contribution of neighbor effects to the phenotypic variation in the leaf damage at the nearest scale (partial $PVE_{nei} = 0.214$, $\chi^2_1 = 7.23$, p value = 0.004 at $s = 1$; Fig. S11 left). The proportion of phenotypic variation explained by the neighbor effects did not increase when the neighbor scale was referred up to the nearest and second nearest individuals (partial $PVE_{nei} = 0.14$, $\chi^2_1 = 1.41$, p value = 0.166 at $s = 2$; Fig. S11 left); therefore, the effective scale of the neighbor effects was estimated at $s = 1$ and variation partitioning was stopped at $s = 2$. These results indicated that the effective scale of the neighbor effects on the leaf damage was narrow ($s = 1$) and the net PVE_{nei} at $s = 1$ explained an additional 8% of the PVE compared to the additive genetic variation attributable to the single PVE_{self} (Fig. 6a). The genotype data had moderate to strong element-wise correlation between K_1 and K_2 in these analyses ($r = 0.60$ and 0.78 at $s = 1$ and 2 among 199 accessions with eight replicates). We additionally incorporated the neighbor phenotype $y_j^{(s)}$ instead of the neighbor genotype $x_j^{(s)}$ in Eq. (2), but the partial PVE_{nei} did not increase (partial $PVE_{nei} = 0.066$ and 0.068 at $s = 1$ and 2 , respectively).

Fig. 5 Signals of the self and neighbor effects when 50 causal SNPs governed either the strong self (lower panels) or neighbor (upper panels) effects. The score of $-\log_{10}(p$ value) is averaged within each iteration and is shown for the non-causal SNPs ($\beta_1 = \beta_2 = 0$), SNPs responsible for self-effects alone ($\beta_1 \neq 0$ and $\beta_2 = 0$), SNPs responsible for neighbor effects alone ($\beta_1 = 0$ and $\beta_2 \neq 0$), and SNPs responsible for both self and neighbor effects ($\beta_1 \neq 0$ and $\beta_2 \neq 0$). Red and blue boxes show $-\log_{10}(p$ value) distributions among the iterations for the self and neighbor effects, respectively.



The standard GWAS of the self-genotype effects on the leaf damage detected the SNPs with the second and third largest $-\log_{10}(p$ values) scores, on the first chromosome (chr1), though they were not above the threshold for Bonferroni correction (Fig. 6b; Table S2). The second SNP at chr1-21694386 was located within ~ 10 kb of the three loci encoding a disease resistance protein (CC-NBS-LRR class) family. The third SNP at chr1-23149476 was located within ~ 10 kb of the AT1G62540 locus that encodes a flavin-monoxygenase glucosinolate S-oxygenase 2 (FMO GS-OX2). No GOs were significantly enriched for the self-effects on herbivory (false discovery rate > 0.08). A QQ-plot did not exhibit an inflation of p values for the self-genotype effects (Fig. S12a).

Regarding the neighbor effects on leaf damage, we found non-significant but weak peaks on the second and third chromosomes (Fig. 6c; Table S2). The second chromosomal region had higher association scores than those predicted by the QQ-plot (Fig. S12b). A locus encoding FAD-binding Berberine family protein (AT2G34810 named *BBE16*), which is known to be up-regulated by methyl jasmonate (Devoto et al. 2005), was located within the ~ 10 kb window near SNPs with the top eleven $-\log_{10}(p$ values) scores on the second chromosome. Three transposable elements and a

pseudogene of lysyl-tRNA synthetase 1 were located near the most significant SNP on the third chromosome. No GOs were significantly enriched for the self-effects on herbivory (false discovery rate > 0.9). We additionally tested the asymmetric neighbor effects of β_{12} in the real dataset on field herbivory, but the top 0.1% of the SNPs for the neighbor effects for β_2 , did not overlap with those of the asymmetric neighbor effects β_{12} (Table S2).

Based on the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$, we ran a post hoc simulation to infer a spatial arrangement that minimizes a population sum of the leaf damage $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{\langle i,j \rangle} x_i x_j$. The constant intercept β_0 , the variance component u_i , and residual e_i were not considered because they were not involved in the deterministic dynamics of the model. Figure 7 shows three representatives and a neutral expectation. For example, a mixture of a dimorphism was expected to decrease the total leaf damage for an SNP at chr2-14679190 near the *BBE16* locus ($\hat{\beta}_2 > 0$; Fig. 7a). On the other hand, a clustered distribution of a dimorphism was expected to decrease the total damage for an SNP at chr2-9422409 near the AT2G22170 locus encoding a lipase/lipoxygenase PLAT/LH2 family protein ($\hat{\beta}_1 \approx 0, \hat{\beta}_2 < 0$; Fig. 7b). Furthermore, a near monomorphism was expected to decrease the leaf damage for an SNP at

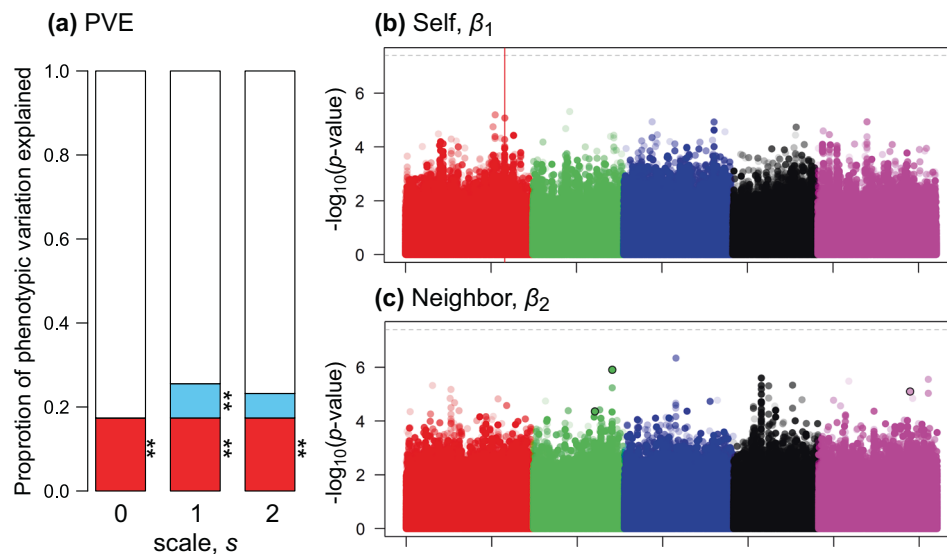


Fig. 6 Pilot GWAS of leaf damage scores on field-grown *Arabidopsis thaliana*. **a** Proportion of phenotypic variation explained (PVE) by the self-genotype (red) or neighbor effects (blue). The PVE_{self} was represented by the single PVE_{self} that represented additive genetic variance, while the net contribution of the neighbor effects was evaluated using the net $PVE_{\text{nei}} = \text{total PVE} - \text{single } PVE_{\text{self}}$. Asterisks highlight a significant fraction with stepwise likelihood ratio tests, from simpler to complex models: ** p value < 0.01. **b, c** Manhattan

plots for the self or neighbor effects. The first to fifth chromosomes are differently colored, where lighter plots indicate smaller MAF. Horizontal dashed lines indicate the threshold after Bonferroni correction at p value < 0.05. The red vertical line in panel **b** indicates a SNP position near the *GS-OX2* locus, while the three circles highlighted by a black outline in panel **c** indicates the variants subject to the post hoc simulation (Fig. 7). Results of the self and neighbor effects are shown at $s = 0$ (i.e., standard GWAS) and $s = 1$, respectively.

chr5-19121831 near the AT5G47075 and AT5G47077 loci encoding low-molecular cysteine-rich proteins, LCR20 and LCR6 ($\hat{\beta}_1 > 0, \hat{\beta}_2 < 0$: Fig. 7c). If the self and neighbor coefficients had no effects, we would observe a random distribution and no mitigation of damage i.e., $\sum y_i \approx 0$ (Fig. 7d). These post hoc simulations suggested a potential application for neighbor GWAS for the optimization of the spatial arrangements in field cultivation.

Comparing self p values between the neighbor GWAS and GEMMA

To ascertain whether the self-genotype effects in the neighbor GWAS agree with those of a standard GWAS, we compared the p value scores between the rNeighborGWAS package and the commonly used GEMMA program (Fig. S13). For the leaf damage score, the neighbor GWAS yielded almost the same $-\log_{10}(p$ values) scores for the self-effects as the GEMMA program ($r = 0.9999$ among all the 1,242,128 SNPs: Fig. S13a). The standard GWAS, using the flowering time phenotype, also yielded the consistent $-\log_{10}(p$ values) scores between the neighbor GWAS and GEMMA ($r = 0.9999$ among all the 1,814,755 SNPs: Fig. S13b). Both the flowering time GWAS using the neighbor GWAS and GEMMA found two significant SNPs above the genome-wide Bonferroni threshold on chromosome 5 (chr5-18590741 and chr5-18590743, MAF = 0.49

and 0.49, $-\log_{10}(p$ value) = 7.797 and 7.797 for the neighbor GWAS; chr5-18590741 and chr5-18590743, MAF = 0.49 and 0.49, $-\log_{10}(p$ value) = 7.798 and 7.798 for GEMMA), which were located within the *Delay of Germination 1 (DOG1)* locus, that was reported previously by Alonso-Blanco et al. (2016). Another significant SNP was observed at the top of chromosome 4 (chr4-317979, MAF = 0.12, $-\log_{10}(p$ value) = 7.787 and 7.933 for the neighbor GWAS and GEMMA), which was previously identified as a quantitative trait locus underlying flowering time in long-day conditions (Aranzana et al. 2005).

Discussion

Spatial and genetic factors underlying simulated phenotypes

Benchmark tests using simulated phenotypes revealed that appropriate spatial scales could be estimated using the partial PVE_{nei} of the observed phenotypes. When the scale of the neighbor effects was narrow or moderate ($\alpha = 1.0$ or 3.0), the scale of the first nearest neighbors would be optimum for increasing the AUC to detect neighbor signals. In terms of the neighbor effects in the context of plant defense, mobile animals (e.g., mammalian browsers and flying insects) often select a cluster of plant individuals

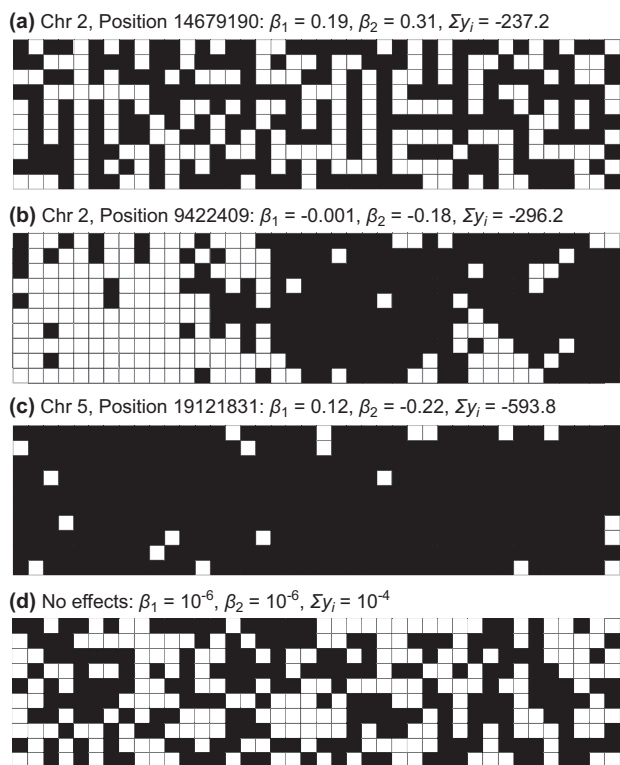


Fig. 7 Post hoc simulations exemplifying a spatial arrangement of the two alleles expected by the estimated self and neighbor effects, $\hat{\beta}_1$ and $\hat{\beta}_2$, on the leaf damage score of *Arabidopsis thaliana*. The population sum of the leaf damage $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{\langle i,j \rangle} x_i x_j$ was minimized using 1000 iterations of simulated annealing from a random distribution of two alleles in a 10×40 space.

(e.g., Bergvall et al. 2006; Hambäck et al. 2009; Sato and Kudoh 2015; Verschut et al. 2016). In this case, the neighbor effects could not be observed among individual plants within a cluster (Sato and Kudoh 2015). The exponential distance decay at $\alpha = 0.01$ represented situations in which the effective range of the neighbor effects was too broad to be detected; only in such situations should more than the nearest neighbors be referred to, to gain the power to detect neighbor effects. We also considered the asymmetric neighbor effects where the neighbor genotype similarity had significant effects on one genotype, but not on another genotype. In this situation, strong self-effects could be observed when the symmetric neighbor effects were weakened. This additional result suggests that asymmetric neighbor effects should be tested if strong self-effects and weak symmetric neighbor effects are both detected at a single locus.

Neighbor effects are more likely to contribute to phenotypic variation when its effective range becomes narrow due to a strong distance decay ($\alpha = 3$), as suggested by the net PVE_{nei} . However, the total phenotypic variation was explained relatively well by the single PVE_{self} that represented additive polygenic effects. Previous studies showed that genetic interactions could lead to an overrepresentation

of narrow-sense heritability in GWAS (e.g., Zuk et al. 2012; Young and Durbin 2014). This occurs because the SNP heritability is represented by the genetic similarity between individuals, and thereby covariance of the kinship matrix helps to fit the phenotypic variance attributable to gene-by-gene interactions (Young and Durbin 2014; Schrauf et al. 2020). This problem is also observed in the neighbor GWAS that models pairwise interactions at a focal locus among neighboring individuals. Given the difficulty in distinguishing the kinship and genetic interactions, we conclude that the non-independence of the self and neighbor effects is an intrinsic feature of the neighbor GWAS, and that the difference of the PVE between a standard and neighbor GWAS i.e., net PVE_{nei} should be used as a conservative estimate of PVE_{nei} .

Neighbor GWAS of the field herbivory on *Arabidopsis*

Our genetic analysis of the neighbor effects is of ecological interest, as the question of how host plant genotypes shape variations in plant–herbivore interactions, is a long-standing question in population ecology (e.g., Karban 1992; Underwood and Rausher 2000; Utsumi et al. 2011). Despite the low PVE and several confounding factors under field conditions, the present study illustrated the significant contribution of neighbor genotypic identity to the spatial variation of the herbivory on *A. thaliana*. Although the additional fraction explained by the neighbor effects was 8%, this amount was plausible in the GWAS of complex traits. For example, the variance components of epistasis explained 10–14% PVE on average for 46 traits in yeast (Young and Durbin 2014). Even when heritability is high, the significant variants have often explained a small fraction of PVE, which is known as the missing heritability problem in plants and animals (Brachi et al. 2011; López-Cortegano and Caballero 2019).

Regarding the self-genotype effects, we detected *GS-OX2* near the third top-scoring SNP on the first chromosome. *GS-OX2* catalyzes the conversion of methylthioalkyl to methylsulfinylalkyl glucosinolates (Li et al. 2008) and is up-regulated in response to feeding by the larvae of the large white butterfly (*Pieris brassicae*) (Geiselhardt et al. 2013). On the other hand, the second top-scoring SNP of the neighbor effects was located near the *BBE16* locus, responsive to methyl jasmonate, a volatile organic chemical that is emitted from damaged tissue and elicits the defense responses of other plants (Reymond and Farmer 1998; van Poecke 2007). However, because none of the associations were significant above a genome-wide Bonferroni threshold, they should be interpreted cautiously. Nearby genes should only be considered candidates, and further work is necessary to confirm that they exert any neighbor effects on herbivory.

Potential limitation

Despite many improvements, it is more difficult for GWAS to capture rare causal variants than common ones (Lee et al. 2014; Auer, Lettre 2015; Bomba et al. 2017). This problem is more severe in neighbor GWAS, because smaller MAFs result in stronger collinearity between the self-genotype effects x_i and the neighbor genotypic identity $\sum_{<i,j>} x_i x_j^{(s)}$. Our simulations showed that the rare variants responsible for the neighbor effects might be misclassified as self-effects, though the opposite was not found, i.e., the misclassification of self-signals into neighbor effects could be suppressed. In GWAS, genotype data usually contain minor alleles and possess kinship structures to some extent, making collinearity unavoidable. To anticipate false positive detection of neighbor effects, the significance of variance components and marker effects involving neighbor effects should always be compared using the standard GWAS model.

The present neighbor GWAS focused on single-locus effects and did not incorporate locus-by-locus interactions. Although it is challenging to integrate all the association tests for epistasis into GWAS (Gondro et al. 2013; Young and Durbin 2014), it is possible that multiple combinations among different variants govern neighbor effects. For example, neighbor effects on insect herbivory may occur due to the joint action of volatile-mediated signaling and the accumulation of secondary metabolites (Dicke and Baldwin 2010; Erb 2018). The linear mixed model could be extended as exemplified by the asymmetric neighbor effects; however, we need to reconcile multiple criteria including the collinearity of explanatory variables, inflation of p values, and computational costs. Further customization is warranted when analyzing more complex forms of neighbor effects.

Conclusion

Based on the newly proposed methodology, we suggest that neighbor effects are an overlooked source of phenotypic variation in field-grown plants. GWAS have often been applied to crop plants (Jannink et al. 2010; Hamblin et al. 2011), where genotypes are known, and individuals are evenly transplanted in space. Considering this outlook for agriculture, we provided an example of neighbor GWAS across a lattice space in this study. However, wild plant populations sometimes exhibit more complex spatial patterns than those expected by the Ising model (e.g., Kizaki and Katori 1999; Schlicht and Iwasa 2004). In the rNeighborGWAS package, we allowed neighbor GWAS for a continuous two-dimensional space. While its application has now been limited to experimental populations, neighbor GWAS has the potential for compatibility with the

emerging discipline of landscape genomics (Bragg et al. 2015). In this context, the additional R package could help future studies to test self and neighbor effects using a wide variety of plant species.

Neighbor GWAS may also have the potential to help determine optimal spatial arrangements for plant cultivation, as suggested by the post hoc simulation. Genome-wide polymorphism data are useful not only for identifying causal variants in GWAS, but also for predicting the breeding values of crop plants for genomic selection (e.g., Jannink et al. 2010; Hamblin et al. 2011; Yamamoto et al. 2017). Given that the neighbor GWAS consists of a marker-based regression, this methodology could also be expanded as a genomic selection tool to help predict population-level phenotypes in spatially structured environments.

Data availability

The leaf damage data on *A. thaliana* are included in the supporting information (Table S1). The simulation code and R script used in this study are available at the GitHub repository (<https://github.com/naganolab/NeighborGWAS>). R package version of the neighbor GWAS method is available at CRAN (<https://cran.r-project.org/package=rNeighborGWAS>).

Acknowledgements The authors would like to thank Ü. Seren, A. Korte, and M. Nordborg for kindly providing the full imputed SNP data prior to it being publicly available; T. Tsuchimatsu, K. Iwayama, J. Bascompte, and anonymous reviewers for discussions and comments; and Dynacom Co., Ltd. for technical assistance with the R package development. This study was supported by the Japan Science and Technology Agency (JST) PRESTO (Grant number, JPMJPR17Q4 and JPMJPR16Q9) to YS and EY; Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship (16J30005) to YS; MEXT KAKENHI (18H04785) and the Swiss National Science Foundation (31003A_182318) to KKS; URPP Global Change and Biodiversity of the University of Zurich to YS and KSS; and JST CREST (JPMJCR15O2 and JPMJCR16O3) to AJN and KKS. The field experiment was supported by the Joint Usage/Research Grant of Center for Ecological Research, Kyoto University, Japan.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not

included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM et al. (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166:481–491. <https://doi.org/10.1016/j.cell.2016.05.063>
- Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K et al. (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet* 1:e60. <https://doi.org/10.1371/journal.pgen.0010060>
- Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631. <https://doi.org/10.1038/nature08800>
- Auer PL, Lettre G (2015) Rare variant association studies: considerations, challenges and opportunities. *Genome Med* 7:16. <https://doi.org/10.1186/s13073-015-0138-2>
- Azaele S, Muneeppeerakul R, Rinaldo A, Rodriguez-Iturbe I (2010) Inferring plant ecosystem organization from species occurrences. *J Theor Biol* 262:323–329. <https://doi.org/10.1016/j.jtbi.2009.09.026>
- Barbosa P, Hines J, Kaplan I, Martinson H, Szczepaniak A, Szendrei Z (2009) Associational resistance and associational susceptibility: having right or wrong neighbors. *Ann Rev Ecol Syst* 40:1–20. <https://doi.org/10.1146/annurev.ecolsys.110308.120242>
- Bergvall UA, Rautio P, Kesti K, Tuomi J, Leimar O (2006) Associational effects of plant defences in relation to within- and between-patch food choice by a mammalian herbivore: neighbour contrast susceptibility and defence. *Oecologia* 147:253–260. <https://doi.org/10.1007/s00442-005-0260-8>
- Bomba L, Walter K, Soranzo N (2017) The impact of rare and low-frequency genetic variants in common disease. *Genome Biol* 18:77. <https://doi.org/10.1186/s13059-017-1212-4>
- Brachi B, Meyer CG, Villoutreix R, Platt A, Morton TC, Roux F et al. (2015) Coselected genes determine adaptive variation in herbivore resistance throughout the native range of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 112:4032–4037. <https://doi.org/10.1073/pnas.1421416112>
- Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol* 12:232. <https://doi.org/10.1186/gb-2011-12-10-232>
- Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015) Genomic variation across landscapes: insights and applications. *N. Phytol* 207:953–967
- Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84:210–223. <https://doi.org/10.1016/j.ajhg.2009.01.005>
- Carlson M (2018) GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.7.0. <https://doi.org/10.18129/B9.bioc.GO.db>
- Carrasco LR, Harwood TD, Toepfer S, MacLeod A, Levay N, Kiss J et al. (2010) Dispersal kernels of the invasive alien western corn rootworm and the effectiveness of buffer zones in eradication programmes in Europe. *Ann Appl Biol* 156:63–77. <https://doi.org/10.1111/j.1744-7348.2009.00363.x>
- Chan EKF, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* 185:991–1007. <https://doi.org/10.1534/genetics.109.108522>
- Dahlin I, Rubene D, Glinwood R, Ninkovic V (2018) Pest suppression in cultivar mixtures is influenced by neighbor-specific plant-plant communication. *Ecol Appl* 28:2187–2196. <https://doi.org/10.1002/eap.1807>
- Devaux C, Lavigne C, Austerlitz F, Klein EK (2007) Modelling and estimating pollen movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. *Mol Ecol* 16:487–499. <https://doi.org/10.1111/j.1365-294X.2006.03155.x>
- Devoto A, Ellis C, Magusin A, Chang HS, Chilcott C, Zhu T et al. (2005) Expression profiling reveals *COI1* to be a key regulator of genes involved in wound- and methyl jasmonate-induced secondary metabolism, defence, and hormone interactions. *Plant Mol Biol* 58:497–513. <https://doi.org/10.1007/s1103-005-7306-5>
- Dicke M, Baldwin IT (2010) The evolutionary context for herbivore-induced plant volatiles: beyond the ‘cry for help’. *Trends Plant Sci* 15:167–175. <https://doi.org/10.1016/j.tplants.2009.12.002>
- Erb M (2018) Volatiles as inducers and suppressors of plant defense and immunity - origins, specificity, perception and signaling. *Curr Opin Plant Biol* 44:117–121. <https://doi.org/10.1016/j.pbi.2018.03.008>
- Frachon L, Mayjonade B, Bartoli C, Hautekèete NC, Roux F (2019) Adaptation to plant communities across the genome of *Arabidopsis thaliana*. *Mol Biol Evol* 36:1442–1456. <https://doi.org/10.1093/molbev/msz078>
- Gage JL, De Leon N, Clayton MK (2018) Comparing genome-wide association study results from different measurements of an underlying phenotype. *G3: Genes/Genomes/Genet* 8:3715–3722. <https://doi.org/10.1534/g3.118.200700>
- Geiselhardt S, Yoneya K, Blenn B, Drechsler N, Gershenson J, Kunze R et al. (2013) Egg laying of cabbage white butterfly (*Pieris brassicae*) on *Arabidopsis thaliana* affects subsequent performance of the larvae. *PLoS ONE* 8:e59661. <https://doi.org/10.1371/journal.pone.0059661>
- Gondro C, van der Werf J, Hayes B (eds) (2013) Genome-wide association studies and genomic prediction. *Methods in Molecular Biology*. Humana Press, New York. <https://doi.org/10.1007/978-1-62703-447-0>
- Hamblin MT, Buckler ES, Jannink JL (2011) Population genetics of genomics-based crop improvement methods. *Trends Genet* 27:98–106. <https://doi.org/10.1016/j.tig.2010.12.003>
- Hambäck PA, Björkman M, Rämert B, Hopkins RJ (2009) Scale-dependent responses in cabbage herbivores affect attack rates in spatially heterogeneous systems. *Basic Appl Ecol* 10:228–236. <https://doi.org/10.1016/j.baae.2008.06.004>
- Hambäck PA, Inouye BD, Andersson P, Underwood N (2014) Effects of plant neighborhoods on plant-herbivore interactions: resource dilution and associational effects. *Ecology* 95:1370–1383. <https://doi.org/10.1890/13-0793.1>
- Hauser MT, Harr B, Schlötterer C (2001) Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene *GLABROUS1*. *Mol Biol Evol* 18:1754–1763. <https://doi.org/10.1093/oxfordjournals.molbev.a003963>
- Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S et al. (2014) Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nat Commun* 5:5320. <https://doi.org/10.1038/ncomms6320>
- Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A et al. (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* 44:212–216. <https://doi.org/10.1038/ng.1042>
- Ida TY, Takanashi K, Tamura M, Ozawa R, Nakashima Y, Ohgushi T (2018) Defensive chemicals of neighboring plants limit visits of herbivorous insects: Associational resistance within a plant

- population. *Ecol Evol* 8:12981–12990. <https://doi.org/10.1002/ece3.4750>
- Ising E (1925) Beitrag zur theorie des ferromagnetismus. *Z für Phys* 31:253–258
- Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genom* 9:166–177. <https://doi.org/10.1093/bfpg/elq001>
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ et al. (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723. <https://doi.org/10.1534/genetics.107.080101>
- Karban R (1992) Plant variation: its effects on populations of herbivorous insects. In: Fritz RS, Simms EL (eds) *Plant resistance to herbivores and pathogens: ecology, evolution, and genetics*, University of Chicago Press, Chicago, pp. 195–215
- Kizaki S, Katori M (1999) Analysis of canopy-gap structures of forests by Ising-Gibbs states-equilibrium and scaling property of real forests. *J Phys Soc Jpn* 68:2553–2560. <https://doi.org/10.1143/JPSJ.68.2553>
- Kofler R, Schlotterer C (2012) Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* 28:2084–2085. <https://doi.org/10.1093/bioinformatics/bts315>
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29. <https://doi.org/10.1186/1746-4811-9-29>
- Lee S, Abecasis GR, Boehnke M, Lin X (2014) Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 95:5–23. <https://doi.org/10.1016/j.ajhg.2014.06.009>
- Li J, Hansen BG, Ober JA, Kliebenstein DJ, Halkier BA (2008) Subclade of flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol* 148:1721–1733. <https://doi.org/10.1104/pp.108.125757>
- López-Cortegano E, Caballero A (2019) Inferring the nature of missing heritability in human traits using data from the GWAS catalog. *Genetics* 212:891–904. <https://doi.org/10.1534/genetics.119.302077>
- McCoy BM, Maillard JM (2012) The importance of the Ising model. *Prog Theor Phys* 127:791–817. <https://doi.org/10.1143/PTP.127.791>
- Mundt CC (2002) Use of multiline cultivars and cultivar mixtures for disease management. *Ann Rev Phytopathol* 40:381–410. <https://doi.org/10.1146/annurev.phyto.40.011402.113723>
- Nallu S, Hill JA, Don K, Sahagun C, Zhang W, Meslin C et al. (2018) The molecular genetic basis of herbivory between butterflies and their host plants. *Nat Ecol Evol* 2:1418–1427. <https://doi.org/10.1038/s41559-018-0629-9>
- Perdry H, Dandine-Roulland C (2018) gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. R package version 1.5.4. <https://CRAN.R-project.org/package=gaston>
- R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Reymond P, Farmer EE (1998) Jasmonate and salicylate as global signals for defense gene expression. *Curr Opin Plant Biol* 1:404–411. [https://doi.org/10.1016/s1369-5266\(98\)80264-1](https://doi.org/10.1016/s1369-5266(98)80264-1)
- Rieux A, Soubeyrand S, Bonnot F, Klein EK, Ngando JE, Mehl A et al. (2014) Long-distance wind-dispersal of spores in a fungal plant pathogen: estimation of anisotropic dispersal kernels from an extensive field experiment. *PLoS One* 9:e103225. <https://doi.org/10.1371/journal.pone.0103225>
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma* 12:77. <https://doi.org/10.1186/1471-2105-12-77>
- Sato Y (2018) Associational effects and the maintenance of polymorphism in plant defense against herbivores: review and evidence. *Plant Species Biol* 33:91–108. <https://doi.org/10.1111/1442-1984.12201>
- Sato Y, Ito K, Kudoh H (2017) Optimal foraging by herbivores maintains polymorphism in defence in a natural plant population. *Funct Ecol* 31:2233–2243. <https://doi.org/10.1111/1365-2435.12937>
- Sato Y, Kudoh H (2015) Tests of associational defence provided by hairy plants for glabrous plants of *Arabidopsis halleri* subsp. *gemmifera* against insect herbivores. *Ecol Entomol* 40:269–279. <https://doi.org/10.1111/een.12179>
- Sato Y, Kudoh H (2017) Herbivore-mediated interaction promotes the maintenance of trichome dimorphism through negative frequency-dependent selection. *Am Nat* 190:E67–E77. <https://doi.org/10.1086/692603>
- Sato Y, Shimizu-Inatsugi R, Yamazaki M, Shimizu KK, Nagano AJ (2019) Plant trichomes and a single gene *GLABRA1* contribute to insect community composition on field-grown *Arabidopsis thaliana*. *BMC Plant Biol* 19:163. <https://doi.org/10.1186/s12870-019-1705-2>
- Schlicht R, Iwasa Y (2004) Forest gap dynamics and the Ising model. *J Theor Biol* 230:65–75. <https://doi.org/10.1016/j.jtbi.2004.04.027>
- Schrauf MF, Martini JWR, Simianer H, De los Campos G, Cantet R, Freudenthal J et al. (2020) Phantom epistasis in genomic selection: on the predictive ability of epistatic models. *G3: Genes Genomes Genet* 10:3137–3145. <https://doi.org/10.1534/g3.120.401300>
- Schuman MC, Allmann S, Baldwin IT (2015) Plant defense phenotypes determine the consequences of volatile emission for individuals and neighbors. *eLife* 4:e04490. <https://doi.org/10.7554/eLife.04490>
- Seren Ü, Grimm D, Fitz J, Weigel D, Nordborg M, Borgwardt K et al. (2017) AraPheno: a public database for *Arabidopsis thaliana* phenotypes. *Nucleic Acids Res* 45:D1054–D1059. <https://doi.org/10.1093/nar/gkw986>
- Tahvanainen JO, Root RB (1972) The influence of vegetational diversity on the population ecology of a specialized herbivore, *Phyllotreta cruciferae* (Coleoptera: Chrysomelidae). *Oecologia* 10:321–346. <https://doi.org/10.1007/BF00345736>
- Togninalli M, Seren Ü, Meng D, Fitz J, Nordborg M, Weigel D et al. (2018) The AraGWAS Catalog: a curated and standardized *Arabidopsis thaliana* GWAS catalog. *Nucleic Acids Res* 46:D1150–D1156. <https://doi.org/10.1093/nar/gkx954>
- Underwood N, Hambäck PA, Inouye BD (2020) Pollinators, herbivores, and plant neighborhood effects. *Quart Rev Biol* 95:37–57. <https://doi.org/10.1086/707863>
- Underwood N, Inouye BD, Hambäck PA (2014) A conceptual framework for associational effects: when do neighbors matter and how would we know? *Quart Rev Biol* 89:1–19. <https://doi.org/10.1086/674991>
- Underwood N, Rausher MD (2000) The effects of host-plant genotype on herbivore population dynamics. *Ecology* 81:1565–1576. [https://doi.org/10.1890/0012-9658\(2000\)081\[1565:TEOHPG\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[1565:TEOHPG]2.0.CO;2)
- Utsumi S, Ando Y, Craig TP, Ohgushi T (2011) Plant genotypic diversity increases population size of a herbivorous insect. *Proc R Soc B* 278:3108–3115. <https://doi.org/10.1098/rspb.2011.0239>
- van Poecke RM (2007) *Arabidopsis*-insect interactions. *Arabidopsis Book* 5:e0107. <https://doi.org/10.1199/tab.0107>
- Verschut TA, Becher PG, Anderson P, Hambäck PA (2016) Disentangling associational effects: both resource density and resource frequency affect search behaviour in complex environments. *Funct Ecol* 30:1826–1833. <https://doi.org/10.1111/1365-2435.12670>
- Wang M, Roux F, Bartoli C, Huard-Chauveau C, Meyer C, Lee H et al. (2018) Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proc Natl*

- Acad Sci USA 115:E5440–E5449. <https://doi.org/10.1073/pnas.1710980115>
- Weiner J (1990) Asymmetric competition in plant populations. *Trends Ecol Evol* 5:360–364. [https://doi.org/10.1016/0169-5347\(90\)90095-U](https://doi.org/10.1016/0169-5347(90)90095-U)
- Yamamoto E, Matsunaga H, Onogi A, Ohyama A, Miyatake K, Yamaguchi H et al. (2017) Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. *Heredity* 118:202–209. <https://doi.org/10.1038/hdy.2016.84>
- Young AI, Durbin R (2014) Estimation of epistatic variance components and heritability in founder populations and crosses. *Genetics* 198:1405–1416. <https://doi.org/10.1534/genetics.114.170795>
- Zeller SL, Kalinina O, Flynn DFB, Schmid B (2012) Mixtures of genetically modified wheat lines outperform monocultures. *Ecol Appl* 22:1817–1826. <https://doi.org/10.1890/11-0876.1>
- Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824
- Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 109:1193–1198. <https://doi.org/10.1073/pnas.1119675109>