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

Yasuhiro Sato[1,2], Eiji Yamamoto[1,3], Kentaro K. Shimizu[4,5*], Atsushi J. Nagano[6*]

[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, 244-0813 Totsuka-ward, Yokohama, Japan

[6]Faculty of Agriculture, Ryukoku University, Yokotani 1-5, Seta Oe-cho, Otsu, Shiga 520-2194, Japan

*Co-correspondence:

Atsushi J. Nagano

anagano@agr.ryukoku.ac.jp

Kentaro K. Shimizu

kentaro.shimizu@ieu.uzh.ch

Running head: Neighbor GWAS of herbivory

ABSTRACT

An increasing number of field studies show that the phenotype of an individual plant depends not only on its genotype but also on that 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 phenotypic variation explained by neighbor effects approached a plateau when an effective spatial scale became narrow. Thus, the effective range of neighbor effects could be estimated based on the scale at which an increase in the explained variation peaked. The power to detect causal variants of neighbor effects was moderate to strong when a trait was governed by tens of variants. In contrast, the power decreased when hundreds of variants underlay a single trait. We applied neighbor GWAS to field herbivory data from 199 accessions of *Arabidopsis thaliana* and found a significant contribution of the neighbor effects to the observed damage variation.

Interestingly, a locus responsive to methyl jasmonate (AT2G34810 called *BBE16*) was located near the second top-scoring single nucleotide polymorphism (SNP) of the neighbor effects on the herbivory, while self-genotype effects had the third top-scoring SNP near a locus encoding flavin-monooxygenase glucosinolate S-oxygenase 2 (FMO GS-OX2). Overall, neighbor GWAS highlight the overlooked role of plant neighborhood effects in shaping phenotypic variation. The neighbor GWAS method is available as an R package at CRAN (https://cran.r-project.org/package=rNeighborGWAS), providing a novel tool to analyze complex traits in spatially structured environments.

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), these effects are often not considered in quantitative genetics 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). *GLABRA1* gene is known to determine hairy or glabrous phenotypes in *Arabidopsis* plants (Hauser et al. 2001), where a flightless leaf beetle (*Phaedon brassicae*) prefers 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 to seek 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; 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 such complex traits 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 GWAS of neighbor effects.

To incorporate neighbor effects into GWAS, we 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 a model trend to generate 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 real 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 & 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 either 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 of *i*-th focal individual plant y_i as:

$$y_i = \beta_1 x_i + \beta_2 \sum_{<i,j>} x_i x_j \qquad (1)$$

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)×(-1)=+1 or (+1)×(+1)=+1. If two neighbors had different alleles, the product $x_i x_j$ became (-1)×(+1)=-1 or (+1)×(-1)=-1. Accordingly, the effects of neighbor genotypic identity on a phenotype depended on the coefficient β_2 and the number of 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 $\sum_{<i,j>} x_i x_j$ 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 as $\epsilon_I = -E \sum_{<i,j>} x_i x_j - H \sum x_i$, which defined the interaction energy of a two-dimensional ferromagnetic Ising model (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 *A. thaliana*. 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 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:

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

where β_0 indicated the intercept, and the term $\beta_1 x_i$ represented fixed self-genotype effects as tested in standard GWAS; β_2 was the coefficient of fixed neighbor effects. The neighbor covariate $\sum_{<i,j>}^{L} x_i x_j^{(s)}$ 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 a sample structure in self and neighbor effects were modeled by a random effect $u_i \sim \mathrm{Norm}(0, \sigma_1^2 \boldsymbol{K}_1 + \sigma_2^2 \boldsymbol{K}_2)$. The residual was

expressed as $e_i \sim \mathrm{Norm}(0, \sigma_e^2)$.

Variation partitioning. To estimate the proportion of phenotypic variation explained (PVE) by 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 $\boldsymbol{K}_1 = \frac{1}{q-1}\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{X}_1$ and $\boldsymbol{K}_2 = \frac{1}{q-1}\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{X}_2$, where q indicated the number of markers. As we defined $x_{i(j)} \in \{+1, -1\}$, the elements of the kinship matrix $\boldsymbol{K}_1$ were scaled to represent the proportion of marker loci shared among $n \times n$ plants such that $\boldsymbol{K}_1 = \left(\frac{k_{ij}+1}{2}\right)$; σ_1^2 and σ_2^2 indicated variance component parameters for the self and neighbor effects.

The elements of n plants $\times$ q markers matrix $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ consisted of explanatory variables for the self and neighbor effects as $\boldsymbol{X}_1 = (x_i)$ and $\boldsymbol{X}_2 = (\frac{\sum_{<i,j>}^{L} x_i x_j^{(s)}}{L})$. The individual-level formula eq. 2 could also be converted into a conventional matrix form as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \qquad (3)$$

where $\boldsymbol{y}$ was $n \times 1$ vector of phenotypes; $\boldsymbol{X}$ was a matrix of fixed effects, including mean, self-genotype x_i, neighbor covariate $\frac{\sum_{<i,j>}^{L} x_i x_j^{(s)}}{L}$, and other confounding covariates for n plants; $\boldsymbol{\beta}$ was a vector that represents coefficients of the fixed effects; $\boldsymbol{Z}$ was a design matrix allocating individuals

to a genotype, and became an identity matrix if all plants were different accessions; $\boldsymbol{u}$ was the random effect with Var($\boldsymbol{u}$) = $\sigma_1^2\boldsymbol{K}_1 + \sigma_2^2\boldsymbol{K}_2$; and $\boldsymbol{e}$ was residual as Var($\boldsymbol{e}$) = $\sigma_e^2\boldsymbol{I}$. The proportion of phenotypic variation explained (PVE) by self and neighbor effects is calculated using a single-random effect model defined by '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 in eq. 3. Furthermore, we could partial out either of the two variance components using a two-random effect model and define these metrics as 'partial' $\text{PVE}_{\text{self}} = \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$ and 'partial' $\text{PVE}_{\text{nei}} = \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$. The partial PVE_{self} was equivalent to the single PVE_{self} when s was set at 0.

We should be cautious when applying the model structure underlying neighbor GWAS to data. 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 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 $\sigma_2^2 \neq 0$. Furthermore, if the neighbor effects range was broad enough to encompass an entire field (i.e., $s \to \infty$), the neighbor covariate and self-genotype x_i would become colinear according to the following equation: $(\sum_{<i,j>}^{L} x_i x_j^{(s)})/L = x_i\left(\sum_{j=1}^{L} x_j^{(s)}\right)/L = x_i\bar{x}_j$,

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 $s \to \infty$. Such a collinearity makes distinguishing self and neighbor effects for neighbor GWAS challenging, when the effective range is broad. Therefore, the maximum range of neighbor effects was restricted during simulation as described below.

Power simulation

To examine the power to detect neighbor effects, we applied the neighbor GWAS to simulated phenotypes. Phenotypes were simulated using a part of the real genotypes of *A. thaliana*. To evaluate the power of the simple linear model, we assumed a complex ecological form of neighbor effects with multiple variance components controlled. The power was evaluated by the receiver operating characteristic (ROC) curve on the association score of -$\log_{10}$(p-value) (e.g., Gage et al. 2018). 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 on 1,307 accessions were downloaded from the Joy Bergelson laboratory website (http://bergelson.uchicago.edu/?page_id=790 accessed on February 9, 2017). We extracted the chromosome 1 and 2 data with the minor allele

239 frequency (MAF) at >0.1, providing a matrix of 1,307 plants with 65,226
240 single nucleotide polymorphisms (SNPs). Pairwise linkage disequilibrium
241 (LD) among the loci was $r^2 = 0.003$ [0.00-0.06: median with upper and
242 lower 95 percentiles]. Before generating a phenotype, each locus was
243 standardized to have a mean of zero and a variance of 1. Subsequently, we
244 randomly selected 1,296 accessions (= 36 × 36 accessions) without any
245 replacements for each iteration and placed them in a 36 × 72 checkered
246 space following *Arabidopsis* experimental settings (see Fig. 2b).
247 ***Phenotype simulation***. To address ecological issues specific to
248 plant neighborhood effects, we considered two extensions, namely
249 asymmetric neighbor effects and spatial scales. Firstly, previous studies
250 have shown that such plant–plant interactions are sometimes asymmetric
251 between two accessions in herbivory (e.g., Bergvall et al. 2006; Verschut et
252 al. 2016; Sato and Kudoh 2017) and height competition (Weiner 1990).
253 Such asymmetric neighbor effects can be tested by statistical interaction
254 terms in a linear model (Bergvall et al. 2006; Sato and Kudoh 2017).
255 Secondly, several studies have shown that the strength of neighbor effects
256 depends on spatial scales (Hambäck et al. 2014), and that the scale of
257 neighbors to be analyzed relies on the dispersal ability of causative
258 organisms (see Hambäck et al. 2009; Sato and Kudoh 2015; Verschut et al.
259 2016; Ida et al. 2018 for insect and mammal herbivores; Rieux et al. 2014
260 for pathogen dispersal) or the size of competing plants (Weiner 1990). We

assumed the distance decay at s-th sites from a focal individual i with the decay coefficient α as $w(s,\alpha) = \mathrm{e}^{-\alpha(s-1)}$, 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. Total variance components resulting from three background effects (i.e., the self, neighbor, and self-by-neighbor effects) was defined as $u_i \sim \mathrm{Norm}(0, \sigma_1^2 \boldsymbol{K}_1 + \sigma_2^2 \boldsymbol{K}_2 + \sigma_{12}^2 \boldsymbol{K}_{12})$. The three variance component parameters σ_1^2, σ_2^2, and σ_{12}^2, determined the relative importance of self-genotype, neighbor, and asymmetric neighbor effects in u_i. Given the elements of the n plants × q marker explanatory matrix with $\boldsymbol{X}_{12} = (\frac{x_i}{L}\sum_{<i,j>}^{L} w(s,\alpha) x_i x_j^{(s)})$, the similarity in asymmetric neighbor effects was calculated as $\boldsymbol{K}_{12} = \frac{1}{q-1}\boldsymbol{X}_{12}^T \boldsymbol{X}_{12}$. To control phenotypic variations, we further partitioned the proportion of phenotypic variation into those explained by major-effect genes and variance components PVE$_{\boldsymbol{\beta}}$ + PVE$_{\boldsymbol{u}}$,

major-effect genes alone PVE$_{\boldsymbol{\beta}}$, and residual error PVE$_{\boldsymbol{e}}$, where PVE$_{\boldsymbol{\beta}}$ + PVE$_{\boldsymbol{u}}$ + PVE$_{\boldsymbol{e}}$ = 1. The *optimize* function in R was used to adjust simulated phenotypes to the given amounts of PVE.

Parameter setting. Ten phenotypes were simulated for each combination of the distance decay coefficient α, the proportion of phenotypic variation explained by major-effect genes PVE$_{\boldsymbol{\beta}}$, the proportion of phenotypic variation explained by major-effect genes and variance components PVE$_{\boldsymbol{\beta}}$ + PVE$_{\boldsymbol{u}}$, and the relative importance of multiple variance components $\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$ as: α = 0.25, 1.0 or 3.0, $\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$ = 6:3:1, 4:4:1, or 3:6:1, PVE$_{\boldsymbol{\beta}}$ and PVE$_{\boldsymbol{\beta}}$ + PVE$_{\boldsymbol{u}}$ = 0.1 and 0.4, 0.3 and 0.4, 0.3 and 0.8, or 0.6 and 0.8. The maximum reference scale was fixed at s = 3. The line of simulations was repeated for the number of causal SNPs at 10, 50, or 300 to examine cases of an oligogenic and polygenic control for a trait. The non-zero coefficients 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 all 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 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.

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}. The stepwise likelihood ratio tests were performed first between the null model and model with a self-genotype effect alone, and then between the self-genotype effect model and model with both self and neighbor effects. The likelihood ratio was calculated as the difference in deviance, i.e., -2 × log-likelihood, which is asymptotically χ^2 distributed with one degree of freedom. The variance components, σ_1^2 and σ_2^2, were estimated using a linear mixed model without any fixed effects. To solve the mixed model with the two random effects, we used the average information restricted maximum likelihood (AI-REML) algorithm implemented in the *lmm.aireml* function in the *gaston* package of R (Perdry and Dandine-Roulland 2018). Subsequently, we replaced the two variance parameters σ_1^2 and σ_2^2 in eq. 2 into their estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ by the AI-REML, and performed association tests by solving a linear mixed model with a fast approximation by an eigen value decomposition (implemented in the *lmm.diago* function: Perdry and Dandine-Roulland 2018). The model

326 likelihood was computed using the *lmm.diago.profile.likelihood* function.
327 We evaluated self and neighbor effects for association mapping based on
328 forward selection of the two fixed effects β_1 and β_2 as described below:
329 1. Computed the null likelihood with $\sigma_1^2 \neq 0$ and $\sigma_2^2 = 0$ in eq. 2
330 2. Tested the self-effect, β_1, by comparing with the null likelihood
331 3. Computed the self-likelihood with $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and β_1 using eq. 2
332 4. Tested the neighbor effects, β_2, by comparing with the self-likelihood
333 We also calculated PVE using the mixed models eq. 3 without β_1 and β_2 as
334 follows:
335 1. Calculated single PVE_{self} or single PVE_{nei} by setting either σ_1^2 or σ_2^2 at 0.
336 2. Tested the single PVE_{self} or single PVE_{nei} by likelihood ratio between the
337 null and one-random effect model
338 3. Calculated partial PVE_{self} and partial PVE_{nei} by
339 estimating σ_1^2 and σ_2^2 simultaneously
340 4. Tested the partial PVE_{self} and partial PVE_{nei} by likelihood ratio between
341 the two- and one-random effect model
342 True or false positive rates were evaluated by ROC curves and area
343 under the ROC curves (AUC) (Gage et al. 2018). An AUC of 0.5 would
344 indicate that GWAS has no power to detect true signals, while an AUC of
345 1.0 would indicate that all the top signals predicted by GWAS agree with
346 true signals. The roc function in the pROC package (Robin et al. 2011) was
347 used to calculate AUC from $-\log_{10}(p\text{-value})$. Factors affecting the AUC were

348 tested by analysis-of-variance (ANOVA) for the self or neighbor effects
349 (AUC_{self} or AUC_{nei}). The AUC had proportional values, and thus was
350 arcsine-square-root transformed to improve the Shapiro–Wilk normality
351 statistic before ANOVA. The AUC and PVE were calculated from $s = 1$ (the
352 first nearest neighbors) to $s = 3$ (up to the third nearest neighbors) cases. The
353 AUC was also calculated using standard linear models without any random
354 effects to examine whether the linear mixed models were superior to the
355 linear models. We also tested the neighbor GWAS model incorporating
356 neighbor phenotype $y_j^{(s)}$ instead of $x_j^{(s)}$.
357
358 ***Arabidopsis* herbivory data**
359 We applied the neighbor GWAS to field data of *Arabidopsis* herbivory. The
360 procedure of the present field experiment followed our previous experiment
361 on summer herbivory on field-grown *A. thaliana* (Sato et al. 2019). We
362 selected 199 worldwide accessions from 2029 accessions sequenced by the
363 RegMap (Horton et al. 2012) and 1001 Genomes project (Alonso-Blanco et
364 al. 2016). Of the 199 accessions, most were overlapped with a previous
365 GWAS of biotic interactions (Horton et al. 2014) and half were included by
366 a GWAS of glucosinolates (Chan et al. 2010). Eight replicates of each of the
367 199 accessions were first prepared in a laboratory and then transferred to the
368 outdoor garden at the Center for Ecological Research, Kyoto University,
369 Japan (Otsu, Japan: 35°06′N, 134°56′E, alt. ca. 200 m: Fig. 2a). Seeds were

370 sown on Jiffy-seven pots (33-mm diameter), and stratified at a temperature
371 of 4 °C for a week. Seedlings were cultivated for 1.5 months under a short-
372 day condition (8 h light: 16 h dark, 20 °C). Plants were then separately
373 potted in plastic pots (6 cm in diameter) filled with mixed soil of
374 agricultural compost (Metro-mix 350, SunGro Co., USA) and perlite at a
375 3:1 ratio. Potted plants were set in plastic trays (10 × 40 cells) in a
376 checkered pattern (Fig. 2b). In the field setting, a set of 199 accessions and
377 an additional Col-0 accession were randomly assigned to each block without
378 replacement (Fig. 2b). Eight replicates of these blocks were set >2 m apart
379 from each other (Fig. 2c). Potted plants were exposed to the field
380 environment for 3 wk in June 2017. At the end of experiment, we scored the
381 percentage of foliage eaten as: 0 for no visible damage, 1 for ≤10%, 2 for
382 >10% and ≤ 25%, 3 for > 25% and ≤ 50%, 4 for >50% and ≤ 75%, and 5 for
383 >75% of the foliage area eaten. All plants were scored by a single person to
384 avoid observer bias. The most predominant herbivore in this field trial was
385 the diamond back moth (*Plutella xylostella*), followed by the small white
386 butterfly (*Pieris rapae*). We also recorded the initial plant size and the
387 presence of inflorescence to incorporate them as covariates. Initial plant size
388 was evaluated by the length of the largest rosette leaf (mm) at the beginning
389 of the field experiment and the presence of inflorescence was recorded 2 wk
390 after transplanting.
391 We estimated the variance components and performed the

392 association tests for the leaf damage score with the neighbor covariate at s =
393 1 and 2. These two scales corresponded to L = 4 (the nearest four neighbors)
394 and L = 12 (up to the second nearest neighbors), respectively, in the
395 *Arabidopsis* dataset. The variation partitioning and association tests were
396 performed using the *gaston* package, as mentioned above. To determine the
397 significance of variance component parameters, we compared the likelihood
398 between mixed models with one or two random effects. For the genotype
399 data, we used an imputed SNP matrix of all the 2029 accessions studied by
400 the RegMap (Horton et al. 2012) and 1001 Genomes project (Alonso-
401 Blanco et al. 2016). Missing genotypes were imputed using BEAGLE
402 (Browning and Browning 2009), as described by Togninalli et al. (2018) and
403 updated on the AraGWAS Catalog (https://aragwas.1001genomes.org). Of
404 the 10,709,466 SNPs from the full imputed matrix, we used 1,242,128 SNPs
405 with MAF at > 0.05 and LD of adjacent SNPs at r^2 <0.8. We considered the
406 initial plant size, presence of inflorescence, experimental blocks, and the
407 edge or center within a block as fixed covariates; these factors explained
408 12.5% of the leaf damage variation (1.2% by initial plant size, Wald test, Z =
409 3.53, p-value<0.001; 2.4% by the presence of inflorescence, Z = -5.69, p-
410 value<10^{-8}; 8.3% by the experimental blocks, likelihood ratio test, χ^2 =
411 152.8, df = 7, p-value<10^{-28}; 0.5% by the edge or center, Z = 3.11, p-value =
412 0.002). After the association mapping, we searched candidate genes within
413 ~10 kb around target SNPs, based on the Araport11 gene model with the

414 latest annotation of The Arabidopsis Information Resource (TAIR)
415 (accessed on 7 September 2019). Gene-set enrichment analysis was
416 performed using the Gowinda algorithm that enables unbiased analysis of
417 GWAS results (Kofler and Schlotterer 2012). We tested SNPS with the top
418 0.1% -$\log_{10}$(*p*-value) score with the option "--gene-definition
419 undownstream10000," "--min-genes 20," and "--mode gene." The GO.db
420 package (Carlson et al. 2018) and the latest TAIR AGI code annotation were
421 used to build input files. To assess implementation, we performed standard
422 GWAS using GEMMA version 0.98 (Zhou and Stephens 2012) and
423 confirmed that the neighbor GWAS yielded the same -$\log_{10}$(*p*-values) scores
424 for self-effects as the GEMMA program (r = 0.9999 among all the SNPs).
425 The R source codes, accession list, and phenotype data are available at the
426 GitHub repository (https://github.com/naganolab/NeighborGWAS).
427
428 **R package, "rNeighborGWAS"**
429 To increase the availability of the new method, we have developed the
430 neighbor GWAS into an R package, which is referred to as
431 "rNeighborGWAS". In addition to the genotype and phenotype data, the
432 package requires a spatial map indicating positions of individuals across a
433 space. In this package, we generalized the discrete space example into a
434 continuous two-dimensional space, allowing the handling of any spatial
435 distribution along the x- and y-axes. Based on the three input files, the

436 rNeighborGWAS package estimates the effective range of neighbor effects
437 by calculating partial PVE_{nei} and performs association mapping of neighbor
438 effects using the linear mixed models described earlier. Details and usage
439 are described in the help files and vignette of the rNeighborGWAS package
440 available via CRAN at https://cran.r-project.org/package=rNeighborGWAS.
441
442
443 **RESULTS**
444
445 **Power simulation**
446 A set of phenotypes were simulated from real genotype data following a
447 complex model (eq. 4), and then fitted by a simplified model (eq. 2).
448 Analyzing the factors affecting the AUC, we found that the number of
449 causal SNPs, the proportion of phenotypic variation (PVE) explained by
450 major-effect genes $PVE_{\boldsymbol{\beta}}$, and distance decay coefficient α were the most
451 influential factors affecting the power of the proposed method to detect
452 neighbor signals (Table 1). In addition to $PVE_{\boldsymbol{\beta}}$, the amount of variance
453 components $PVE_{\boldsymbol{u}}$ also significantly affected the AUC of the neighbor
454 effects, but these additional effects were less significant compared to those
455 of $PVE_{\boldsymbol{\beta}}$ alone (Table 1). The power of the standard GWAS to detect self-
456 genotype effects depended on the number of causal SNPs and $PVE_{\boldsymbol{\beta}}$, but
457 was not significantly influenced by the distance decay coefficient of

neighbor effects (Table 1). Neither the power to detect self nor neighbor effects were significantly affected by the ratio of three variance components $\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$ (Table 1).

The power to detect self-genotype signals changed from strong (AUC_{self}>0.9) to weak (AUC_{self}<0.6) as PVE_{β} decreased and the number of causal SNPs increased (Fig. 3a; Fig. S1a). Compared to the self-genotype effects, it was relatively difficult to detect neighbor effects (Fig. 3b; Fig. S1b). In the case of 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 distance decay coefficient (Fig. 3b; Fig. S1b). The power to detect neighbor effects was very weak (AUC_{nei}>0.5 and AUC_{nei}<0.6) in the case of the number of causal SNPs = 300 (Fig. 3b; Fig. S1b). The result of simulations indicated that 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 shown by 9.7% and 1.7% increases in the power to detect self and neighbor signals, respectively (AUC_{self}, paired *t*-test, mean of the difference = 0.097, *p*-value<10^{-16}; AUC_{nei}, mean of the difference = 0.017, *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 AUC_{nei} decreased to near 0.5.

Notably, there was a clear relationship between the distance decay coefficient α and the proportion of phenotypic variation explained by neighbor effects partial PVE_{nei} or AUC_{nei} at different spatial scales (Fig. 4). If the distance decay was weak (small value of decay coefficient α) and the effective range of neighbor effects was broad, partial PVE_{nei} and AUC_{nei} increased linearly as the reference spatial scale was broadened (Fig. 4a). On the other hand, if the distance decay was strong (large value of decay coefficient α) and the effective scale of neighbor effects was narrow, partial PVE_{nei} saturated at the scale of the first nearest neighbors (Fig. 4c) and AUC_{nei} did not increase (Fig. 4b, c). The simulation results indicated that the effective spatial scales could be estimated by the scale s at which the partial PVE_{nei} peaked, and that the first nearest scale was enough to detect neighbor signals unless the distance decay was very weak.

Examination of variance components provided further insights into PVE_{nei} estimation (Fig. S2). Comparing the true and estimated PVE, we found that the sum of partial PVE_{self} and partial PVE_{nei} matched the total PVE attributable to the model at the nearest neighbor scale (Fig. S2a). Single PVE_{self} or single PVE_{nei} overestimated the respective true PVE_{self} or PVE_{nei} (Fig. S2b, c). In terms of total PVE, a joint mixed model including the partial PVE_{self} and partial PVE_{nei} outperformed that including the single PVE_{self} alone and compensated for the remaining total PVE unexplained by the single PVE_{self} (Fig. S2a). The partial PVE_{nei} underestimated the true

502 PVE_{nei} depending on the distance decay coefficient (Fig. S2e); if the
503 distance decay was strong (α = 3), the underestimation of true PVE_{nei} was
504 moderate even at the nearest scale (s = 1: Fig. S2e), but if the distance decay
505 was weak (α = 0.25), the bias was large at s = 1 and the broader scale (s = 3)
506 should be referred (Fig. S2e). Once the effective range is determined by
507 increasing patterns of the partial PVE_{nei} across s, the partial PVE_{nei} provides
508 a conservative proxy to true PVE_{nei}.
509
510 ***Arabidopsis* herbivory data**
511 To estimate PVE_{self} and PVE_{nei}, we applied a linear mixed model (eq. 3) for
512 the leaf damage score of field-grown *A. thaliana*. Variation partitioning
513 showed a significant contribution of neighbor effects to phenotypic variation
514 in the leaf damage at the nearest scale (single PVE_{nei} = 0.248, χ_1^2 = 17.8, *p*-
515 value = 10^{-5}; partial PVE_{nei} = 0.214, χ_1^2 = 7.23, *p*-value = 0.004 at s = 1: Fig.
516 5). The proportion of phenotypic variation explained by neighbor effects did
517 not increase when the neighbor scale was referred up to the nearest and
518 second nearest individuals (single PVE_{nei} = 0.243, χ_1^2 = 11.1, *p*-value =
519 0.0005; partial PVE_{nei} = 0.14, χ_1^2 = 1.41, *p*-value = 0.166 at s = 2: Fig. 5);
520 therefore, the effective scale of neighbor effects was estimated at s = 1 and
521 variation partitioning was stopped at s = 2. Notably, these neighbor effects
522 explained the leaf damage variation better than self-genotype effects, since
523 single PVE_{nei} was larger than single PVE_{self} at s = 0 (single PVE_{self} = 0.173,

524 χ_1^2 = 10.1, *p*-value = 0.005) and partial PVE_{nei} was larger than the partial
525 PVE_{self} (partial PVE_{self} = 0.041, χ_1^2 = 0.36, *p*-value = 0.54 at s = 1; partial
526 PVE_{self} = 0.091, χ_1^2 = 1.19, *p*-value = 0.200 at s = 2: Fig. 5). We additionally
527 incorporated the neighbor phenotype $y_j^{(s)}$ instead of the neighbor
528 genotype $x_j^{(s)}$ in eq. 2, but partial PVE_{nei} did not increase (partial PVE_{nei} =
529 0.066 and 0.068 at s = 1 and 2, respectively). These results indicated a
530 narrow effective scale and 15-20% variation, which was explained by
531 neighbor effects on the leaf damage.
532 Standard GWAS of the self-genotype effects on the leaf damage
533 detected SNPs with the second and third largest -$\log_{10}$(*p*-values) score on
534 the first chromosome (chr1), though they were not above a threshold of
535 Bonferroni correction (Fig. 6a; Table S2). The second SNP at chr1-
536 21694386 was located within ~10 kb of three loci encoding a disease
537 resistance protein (CC-NBS-LRR class) family. The third SNP at chr1-
538 23149476 was located within ~10 kb of the AT1G62540 locus that encoded
539 flavin-monooxygenase glucosinolate S-oxygenase 2 (FMO GS-OX2). No
540 GOs were significantly enriched for the self-effects on herbivory (false
541 discovery rate > 0.08). A QQ-plot did not exhibit an inflation of *p*-values for
542 the self-genotype effects (Fig. S3).
543 Regarding the neighbor effects on the leaf damage, we found non-
544 significant but weak peaks on the second and third chromosomes (Fig. 6b;
545 Table S2). The second chromosomal region had higher association scores

546 than predicted by the QQ-plot (Fig. S3). A locus encoding FAD-binding
547 Berberine family protein (AT2G34810 named *BBE16*) was located within
548 the ~10 kb window near SNPs with the top eleven -$\log_{10}$(p-values) score at
549 the second chromosome, which are known to be up-regulated by methyl
550 jasmonate (Devoto et al. 2005). Three transposable elements and a
551 pseudogene of lysyl-tRNA synthetase 1 were located near the most
552 significant SNP on the third chromosome. No GOs were significantly
553 enriched for the self-effects on herbivory (false discovery rate > 0.9). We
554 additionally tested asymmetric neighbor effects β_{12} in the real dataset on
555 field herbivory, but the top 0.1% SNPs of neighbor effects β_2 did not
556 overlap with those of asymmetric neighbor effects β_{12} (Table S2).
557 Based on the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$, we ran a post hoc
558 simulation to infer a spatial arrangement that minimizes a population sum of
559 the leaf damage $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{<i,j>} x_i x_j$. The constant intercept β_0,
560 the variance component u_i, and residual e_i were not considered because they
561 were not involved in deterministic dynamics of the model. Figure 7 shows
562 three representatives and a neutral expectation. For example, a mixture of a
563 dimorphism was expected to decrease the total leaf damage for a SNP at
564 chr2-14679190 near the *BBE16* locus ($\hat{\beta}_2 > 0$: Fig. 7a). On the other hand, a
565 clustered distribution of a dimorphism was expected to decrease the total
566 damage for a SNP at chr2-9422409 near the AT2G22170 locus encoding a
567 lipase/lipooxygenase PLAT/LH2 family protein ($\hat{\beta}_1 \approx 0, \hat{\beta}_2 < 0$: Fig. 7b).

Furthermore, near monomorphism was expected to decrease the leaf damage for a SNP at 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 self and neighbor coefficients had no effects, we observed a random distribution and no mitigation of damage i.e., $\sum y_i \approx 0$ (Fig. 7d). These post hoc simulations suggested a potential applicability of neighbor GWAS in optimizing spatial arrangements in field cultivation.

DISCUSSION

Spatial and genetic factors underlying simulated phenotypes

Benchmark tests using simulated phenotypes revealed that appropriate spatial scales could be estimated by variation partitioning of observed phenotypes. When the scale of neighbor effects was narrow or moderate (α = 1.0 or 3.0), the scale of the first nearest neighbors would be optimum for the power to detect neighbor signals. In terms of neighbor effects in the context of plant defense, mobile animals (e.g., mammalian browsers and flying insects) often select a cluster of plant individuals (e.g., Bergvall et al. 2006; Hambäck et al. 2009; Sato and Kudoh 2015; Verschut et al. 2016). In this case, neighbor effects could not be observed among individual plants

590 within a cluster (Sato and Kudoh 2015). The exponential distance decay at α
591 = 0.25 represented situations in which the effective range of neighbor effects
592 was too broad to be detected; only in such situations should more than the
593 nearest neighbors be referred to gain the power to detect neighbor effects.
594 Neighbor GWAS could retain its power if neighbor effects were
595 spatially limited, and several major-effect genes governed a trait. In contrast,
596 when hundreds of causal variants involved a single trait and less than half of
597 phenotypic variation was attributable to neighbor effects, we observed a
598 decrease in the power to detect neighbor signals. In GWAS, false positive
599 rates can be reduced using linear mixed models that deal with kinship
600 structure as a random effect (Korte and Farlow 2013). Indeed, mixed models
601 were superior to standard linear models in this simulation. Our simulation
602 also adjusted the three variance components σ_1^2, σ_2^2, and σ_{12}^2, but their
603 relative contribution did not have significant effects on the power to detect
604 neighbor signals. In addition, the power to detect self-genotype signals
605 largely depended on the number of causal SNPs and PVE by major-effect
606 genes. This result suggests that the power for the standard GWAS to detect
607 the self-genotype effects was independent from neighbor effects.
608
609 **Neighbor GWAS of the field herbivory on *Arabidopsis***
610 Our genetic analysis of neighbor effects is of ecological interest because
611 how host plant genotypes shape variation in plant–herbivore interactions is a

612 long-standing question in population ecology (e.g., Karban 1992;
613 Underwood and Rausher 2000; Utsumi et al. 2011). Despite the low PVE
614 and several confounding factors under field conditions, the present study
615 illustrated the significant contribution of neighbor genotypic identity to
616 spatial variation in herbivory on *A. thaliana*. Surprisingly, the degree of leaf
617 damage variation explained by neighbor effects was greater than that by
618 self-genotype effects, although joint contribution of self and neighbor
619 effects was not doubled compared to the self-genotype effects. As we did
620 not find any significant peaks above the threshold of genome-wide multiple
621 testing, such neighbor effects on herbivory could be under polygenic control
622 and is exemplified by the simulation scenario where a number of variants
623 and strong distance decay caused moderate total phenotypic variation.
624 Regarding the self-genotype effects, we detected *GS-OX2*, a key
625 gene involved in glucosinolate biosynthesis, near the third top-scoring SNP
626 on the first chromosome. Aliphatic glucosinolates are well known as a
627 chemical defense against insect herbivory (van Poecke 2007; Brachi et al.
628 2015), where GS-OX2 catalyzes the conversion from methylthioalkyl to
629 corresponding methylsulfinylalkyl glucosinolates (Li et al. 2008).
630 Specifically, Geiselhardt et al. (2013) reported that *GS-OX2* was up-
631 regulated in response to feeding by larvae of the large white butterfly (*Pieris*
632 *brassicae*) and such an inducible defense was suppressed by prior
633 oviposition by adult butterflies. The present GWAS supported the relevance

634 of *GS-OX2* to herbivory by *Pieris* butterflies, because we observed larvae of
635 *P. rapae* largely feeding on *A. thaliana*.
636 Remarkably, the second top-scoring SNP of the neighbor effects
637 was located near the *BBE16* locus responsive to methyl jasmonate. Methyl
638 jasmonate is well known as a volatile organic chemical that is emitted from
639 damaged tissue and elicits defense response of other plants (Reymond and
640 Farmer 1998; van Poecke 2007). Volatile-mediated plant–plant
641 communication is a key driver of neighbor effects in herbivory (e.g.,
642 Schuman et al. 2015; Erb 2018; Dahlin et al 2018), while a molecular
643 function of *BBE16* remains unexplored. More recent studies have shown
644 that berberine bridge enzyme-like proteins are involved in damage-
645 associated molecular patterns (Benedetti et al. 2018), which might be a
646 subject of future studies on inducible plant defense and neighbor effects.
647
648 **Potential applicability and limitation**
649 GWAS has often been applied to crop plants (Jannink et al. 2010; Hamblin
650 et al. 2011), where genotypes are known, and individuals are evenly
651 transplanted in space. Considering this outlook for agriculture, we provided
652 an example of neighbor GWAS across a lattice space in this study. However,
653 wild plant populations sometimes exhibit more complex spatial patterns
654 than expected by the Ising model (e.g., Kizaki and Katori 1999; Schlicht and
655 Iwasa 2004). In the rNeighborGWAS package, we allowed neighbor GWAS

656 a continuous two-dimensional space. While its application has now been
657 limited to experimental populations, neighbor GWAS has potential
658 compatibility with an emerging discipline of landscape genomics (Bragg et
659 al. 2015). In this context, the additional R package could help future studies
660 to analyze neighbor effects using a wide variety of plant species.

661 Neighbor effects do not necessarily explain phenotypic variation
662 twice as much as conventional self-genotype effects, since some fractions of
663 PVE are not independent of self and neighbor effects, which is due to scale-
664 dependent collinearity between the elements $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ (e.g., r = 0.60 and
665 0.78 at s = 1 and 2 in the herbivory data analysis). Nevertheless, for each
666 explanatory variable, there was no strong correlation and the magnitude of
667 collinearity varied between the self-genotypic variable x_i and the neighbor
668 variable $\sum x_i x_j / L$ (mean ± SD among all SNPs, r = -0.49 ± 0.54 and -0.56 ±
669 0.63 at s = 1 and 2 in the herbivory data analysis). For an appropriate test of
670 the neighbor effects β_2 and its variance component σ_2^2, it is necessary to
671 compare the likelihood ratio between the standard and neighbor GWAS. In
672 such a case, PVE_{nei} is estimated by partial PVE_{nei} using a mixed model
673 including the two variance components into random effects. Given that the
674 collinearity becomes strong when s is large (see subsection "Neighbor
675 GWAS" in the method), we should be cautious when applying neighbor
676 GWAS to small sampling areas or long-range neighbor effects.

677 The present neighbor GWAS focused on single-locus effects and

did not incorporate locus-by-locus interactions. Although it is difficult to consider such an epistasis into GWAS (Gondro et al. 2013), 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. To date, regional-scale interactions among plants have been analyzed using a genome–environment association study of plant community composition (Frachon et al. 2019), but fine-scale neighbor effects have yet to be examined. Recent studies have shown that a single genomic region or a single gene drives neighbor effects in plant growth, and such a genetic effect shapes a positive relationship between plant genotype diversity and yield

(Wuest and Niklaus 2018; McGale et al. 2020). Our newly proposed methodology of neighbor GWAS could be an efficient tool to identify such a key genetic variant responsible for neighbor effects and resulting biodiversity effects.

Neighbor GWAS may also potentially help determine an optimal spatial arrangement in plant cultivation, as suggested by the post hoc simulation. The Ising model is well established in statistical physics (McCoy and Maillard 2012) and is now applied to a machine-learning pipeline that deals with high-dimensional genome data (Fisher and Mehta 2015). Genome-wide polymorphism data are useful not only to seek causal genes in GWAS, but also to predict breeding values of crop plants in 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 implemented as a genomic selection at a population level. Thus, our study provides an avenue for future studies to predict population-level phenotypes in spatially structured environments.

ACKNOWLEDGEMENTS

The authors would like to thank Ü. Seren, A. Korte, and M. Nordborg for kindly providing the full imputed SNP data prior to being publicly available;

T. Tsuchimatsu, K. Iwayama, and J. Bascompte for discussions; 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 Y.S. and E.Y.; Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship (16J30005) to Y.S.; MEXT KAKENHI (18H04785) and the Swiss National Science Foundation to K.K.S.; URPP Global Change and Biodiversity of the University of Zurich to Y.S. and K.S.S.; and JST CREST (JPMJCR15O2 and JPMJCR16O3) to A.J.N. and K.K.S. The field experiment was supported by the Joint Usage/Research Grant of Center for Ecological Research, Kyoto University, Japan.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

The leaf damage data on *A. thaliana* are included in the supporting information (Table S1). The source codes 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).

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. doi:10.1016/j.cell.2016.05.063

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. doi:10.1038/nature08800

Azaele S, Muneepeerakul R, Rinaldo A, Rodriguez-Iturbe I (2010) Inferring plant ecosystem organization from species occurrences. *J Theor Biol* **262**: 323–329. doi:10.1016/j.jtbi.2009.09.026

Barbosa P, Hines J, Kaplan I, Martinson H, Szczepaniec A, Szendrei Z (2009) Associational resistance and associational susceptibility: Having right or wrong neighbors. *Ann Rev Ecol Evol Sys* **40**: 1–20. doi:10.1146/annurev.ecolsys.110308.120242

Benedetti M, Verrascina I, Pontiggia D, Locci F, Mattei B, De Lorenzo G, Cervone F (2018) Four *Arabidopsis* berberine bridge enzyme-like proteins are specific oxidases that inactivate the elicitor-active oligogalacturonides. *Plant J* **94**: 260–273. doi:10.1111/tpj.13852

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. doi:10.1007/s00442-005-0260-8

Brachi B, Meyer CG, Villoutreix R, Platt A, Morton TC, Roux F, Bergelson J (2015) Coselected genes determine adaptive variation in herbivore resistance throughout the native range of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **112**: 4032–4037. doi:10.1073/pnas.1421416112

Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015) Genomic variation across landscapes: insights and applications. *New 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. doi: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. doi: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. doi: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.

788 doi:10.1534/genetics.109.108522

789 Dahlin I, Rubene D, Glinwood R, Ninkovic V (2018) Pest suppression in
790 cultivar mixtures is influenced by neighbor-specific plant-plant
791 communication. *Ecol Appl* **28**: 2187–2196. doi:10.1002/eap.1807

792 Devaux C, Lavigne C, Austerlitz F, Klein EK (2007) Modelling and
793 estimating pollen movement in oilseed rape (*Brassica napus*) at the
794 landscape scale using genetic markers. *Mol Ecol* **16**: 487–499.
795 doi:10.1111/j.1365-294X.2006.03155.x

796 Devoto A, Ellis C, Magusin A, Chang HS, Chilcott C, Zhu T, Turner JG
797 (2005) Expression profiling reveals *COI1* to be a key regulator of genes
798 involved in wound- and methyl jasmonate-induced secondary metabolism,
799 defence, and hormone interactions. *Plant Mol Biol* **58**: 497–513.
800 doi:10.1007/s11103-005-7306-5

801 Dicke M, Baldwin IT (2010) The evolutionary context for herbivore-
802 induced plant volatiles: beyond the 'cry for help'. *Trends Plant Sci* **15**: 167–
803 175. doi:10.1016/j.tplants.2009.12.002

804 Erb M (2018) Volatiles as inducers and suppressors of plant defense and
805 immunity - origins, specificity, perception and signaling. *Curr Opin Plant*
806 *Biol* **44**: 117–121. doi:10.1016/j.pbi.2018.03.008

807 Frachon L, Mayjonade B, Bartoli C, Hautekèete NC, Roux F (2019)
808 Adaptation to plant communities across the genome of *Arabidopsis*
809 *thaliana*. *Mol Biol Evol* **36**: 1442–1456. doi:10.1093/molbev/msz078

Fisher CK, Mehta P (2015) Bayesian feature selection for high-dimensional linear regression via the Ising approximation with applications to genomics. *Bioinformatics* **31**: 1754–1761. doi:×

Gage JL, De Leon N, Clayton MK (2018) Comparing genome-wide association study results from different measurements of an underlying phenotype. *G3: Genes, Genomes, Genetics* **8**: 3715–3722. doi:10.1534/g3.118.200700

Geiselhardt S, Yoneya K, Blenn B, Drechsler N, Gershenzon J, Kunze R, Hilker M (2013) Egg laying of cabbage white butterfly (*Pieris brassicae*) on *Arabidopsis thaliana* affects subsequent performance of the larvae. *PLoS ONE* **8**: e59661. doi:10.1371/journal.pone.0059661

Gondro C, van der Werf J, Hayes B (eds.) (2013) Genome-wide association studies and genomic prediction. *Methods Mol Biol*. Humana Press, New York. doi:10.1007/978-1-62703-447-0

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. doi: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. doi:10.1890/13-0793.1

Hamblin MT, Buckler ES, Jannink JL (2011) Population genetics of

genomics-based crop improvement methods. *Trends Genet* **27**: 98–106. doi:10.1016/j.tig.2010.12.003

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. doi:10.1093/oxfordjournals.molbev.a003963

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. doi:10.1038/ng.1042

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. doi:10.1038/ncomms6320

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. doi:10.1002/ece3.4750

Ising E (1925) Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik* **31**: 253–258.

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomic* **9**: 166–177.

854 doi:10.1093/bfgp/elq001

855 Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin
856 E (2008) Efficient control of population structure in model organism
857 association mapping. *Genetics* **178**: 1709–1723.
858 doi:10.1534/genetics.107.080101

859 Karban R (1992) Plant variation: its effects on populations of herbivorous
860 insects. In: Fritz RS, Simms EL (eds) *Plant resistance to herbivores and*
861 *pathogens: ecology, evolution, and genetics,* University of Chicago Press:
862 Chicago, pp. 195–215.

863 Kizaki S, Katori M (1999) Analysis of canopy-gap structures of forests by
864 Ising-Gibbs states-equilibrium and scaling property of real forests. *J Phys*
865 *Soc Japan* **68**: 2553-2560. doi:10.1143/JPSJ.68.2553

866 Kofler R, Schlotterer C (2012) Gowinda: unbiased analysis of gene set
867 enrichment for genome-wide association studies. *Bioinformatics* **28**: 2084–
868 2085. doi:10.1093/bioinformatics/bts315

869 Korte A, Farlow A (2013) The advantages and limitations of trait analysis
870 with GWAS: a review. *Plant Methods* **9**: 29. doi:10.1186/1746-4811-9-29

871 Li J, Hansen BG, Ober JA, Kliebenstein DJ, Halkier BA (2008) Subclade of
872 flavin-monooxygenases involved in aliphatic glucosinolate biosynthesis.
873 *Plant Physiol* **148**: 1721–1733. doi:10.1104/pp.108.125757

874 McCoy BM, Maillard JM (2012) The importance of the Ising model.
875 Progress *Theor Phys* **127**: 791-817. doi:10.1143/PTP.127.791

McGale E, Valim H, Mittal D, Jimenez JM, Halitschke R, Schuman MC, Baldwin IT (2020) Determining the scale at which variation in a single gene changes population yields. *eLife* **9**. doi:10.7554/eLife.53517

Mundt CC (2002) Use of multiline cultivars and cultivar mixtures for disease management. *Ann Rev Phytopathol* **40**: 381-410. doi: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. doi: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. doi: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. doi:10.1371/journal.pone.0103225

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**: 77. doi: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. doi:10.1111/1442-1984.12201

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. doi: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. doi:10.1086/692603

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. doi:10.1111/1365-2435.12937

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. doi:10.1186/s12870-019-1705-2

Schlicht R, Iwasa Y. (2004) Forest gap dynamics and the Ising model. *J Theoret Biol* **230**: 65–75. doi:10.1016/j.jtbi.2004.04.027

Schuman MC, Allmann S, Baldwin IT (2015) Plant defense phenotypes

determine the consequences of volatile emission for individuals and neighbors. *eLife* **4**: e04490. doi:10.7554/eLife.04490

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. doi: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. doi:10.1093/nar/gkx954

Underwood N, Rausher MD (2000) The effects of host-plant genotype on herbivore population dynamics. *Ecology* **81**: 1565-1576. doi:10.1890/0012-9658(2000)081[1565:TEOHPG]2.0.CO;2

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. doi:10.1086/674991

Underwood N, Hambäck PA, Inouye BD (2020) Pollinators, herbivores, and plant neighborhood effects. *Quart Review Biol* **95**: 37–57. doi:10.1086/707863

Utsumi S, Ando Y, Craig TP, Ohgushi T (2011) Plant genotypic diversity increases population size of a herbivorous insect. *Proc Royal Soc B* **278**: 3108–3115. doi:10.1098/rspb.2011.0239

van Poecke RM (2007) Arabidopsis-insect interactions. *The Arabidopsis Book* **5**: e0107. doi: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. doi: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. doi:10.1073/pnas.1710980115

Weiner J (1990) Asymmetric competition in plant populations. *Trends Ecol Evol* **5**: 360-364. doi:10.1016/0169-5347(90)90095-U

Wuest SE, Niklaus PA (2018) A plant biodiversity effect resolved to a single chromosomal region. *Nat Ecol Evol* **2**: 1933–1939. doi:10.1038/s41559-018-0708-y

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. doi:10.1038/hdy.2016.84

Zeller SL, Kalinina O, Flynn DFB, Schmid B (2012) Mixtures of genetically modified wheat lines outperform monocultures. *Ecol Appl* **22**: 1817–1826. doi:10.1890/11-0876.1

964 Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis
965 for association studies. *Nat Genet* **44**: 821–824.
966
967

TABLES & FIGURES

Table 1. Factors affecting the power to detect major-effect genes in simulated phenotypes. ANOVA tables show the degree of freedom (df), sum of squares (SS), *F*-statistics, and *p*-values. Explanatory variables are the number of causal SNPs, proportion of phenotypic variation explained (PVE) by major-effect genes ($PVE_{\boldsymbol{\beta}}$), total PVE by major-effect genes and variance components ($PVE_{\boldsymbol{\beta}} + PVE_{\boldsymbol{u}}$), ratio of variance component parameters ($\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$), and distance decay coefficient α. For the neighbor effects, the response variable was the maximum Area Under the ROC Curve (AUC) of the spatial scales from $s = 1$ to $s = 3$. The AUC was arcsine square-root transformed to improve normality. NA means not available.

Response	**Factors**	**df**	**SS**	***F***	***p*-value**
AUC_{self}	No. of causal SNPs	1	37.70	2840.2	< 2e-16
	$PVE_{\boldsymbol{\beta}}$	1	2.23	168.2	< 2e-16
	$PVE_{\boldsymbol{\beta}} + PVE_{\boldsymbol{u}}$	1	0.04	2.64	0.104
	$\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$	2	0.01	0.30	0.738
	α	1	0.00	0.01	0.942
	Residuals	1073	14.24	*NA*	*NA*
AUC_{nei}	No. of causal SNPs	1	30.70	1500.4	< 2e-16
	$PVE_{\boldsymbol{\beta}}$	1	3.01	147.3	< 2e-16
	$PVE_{\boldsymbol{\beta}} + PVE_{\boldsymbol{u}}$	1	0.12	5.65	0.01764
	$\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$	2	0.00	0.08	0.925
	α	1	1.87	91.25	< 2e-16
	Residuals	1073	21.96	*NA*	*NA*

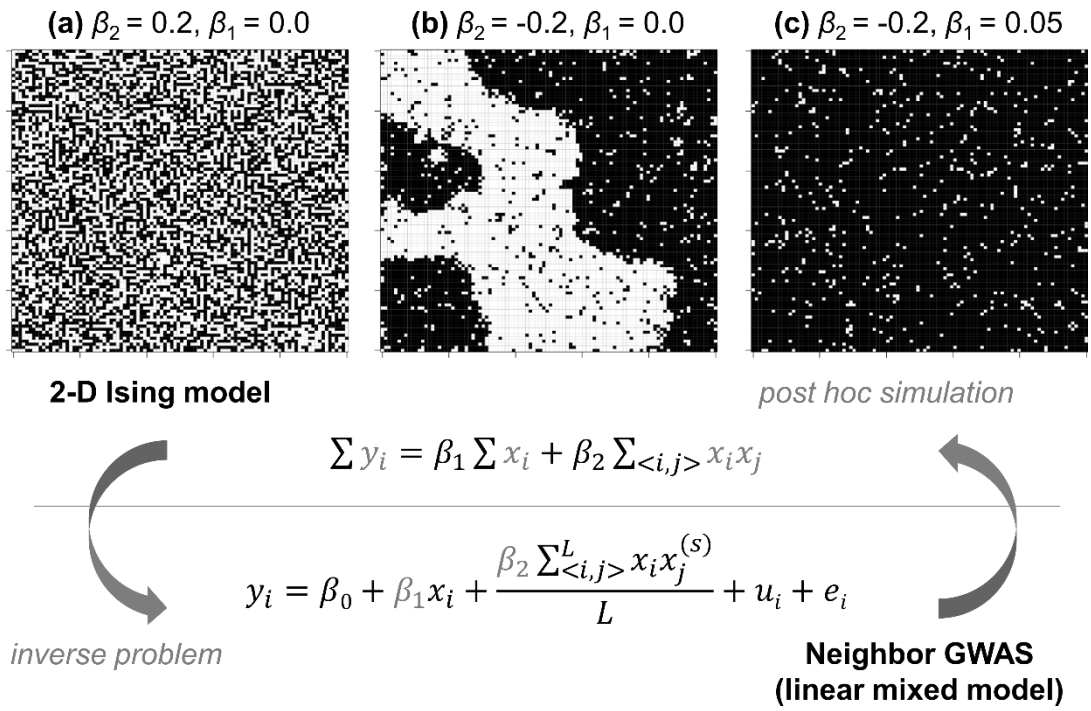


Figure 1. Relationship between Neighbor GWAS and Ising model. Upper panels show spatial arrangements expected by a 2-D Ising model $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{<i,j>} x_i x_j$. (a) If β_2>0, mixed patterns give the argument of the minimum for a population sum of phenotype values $\sum y_i$. (b) If β_2<0, clustered patterns give the argument of the minimum for $\sum y_i$. (c) In addition, β_1 determines overall patterns favoring -1 or +1 states. The figure shows outcomes from a random 100 × 100 lattice after 1000 iterations of simulated annealing. Conversely, the neighbor GWAS was implemented as an inverse problem 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 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 $u_i \sim \mathrm{Norm}(0, \sigma_1^2 \boldsymbol{K}_1 + \sigma_2^2 \boldsymbol{K}_2)$. Once β_1 and β_2 were determined, we could simulate a genotype distribution that maximizes or minimizes $\sum y_i$.

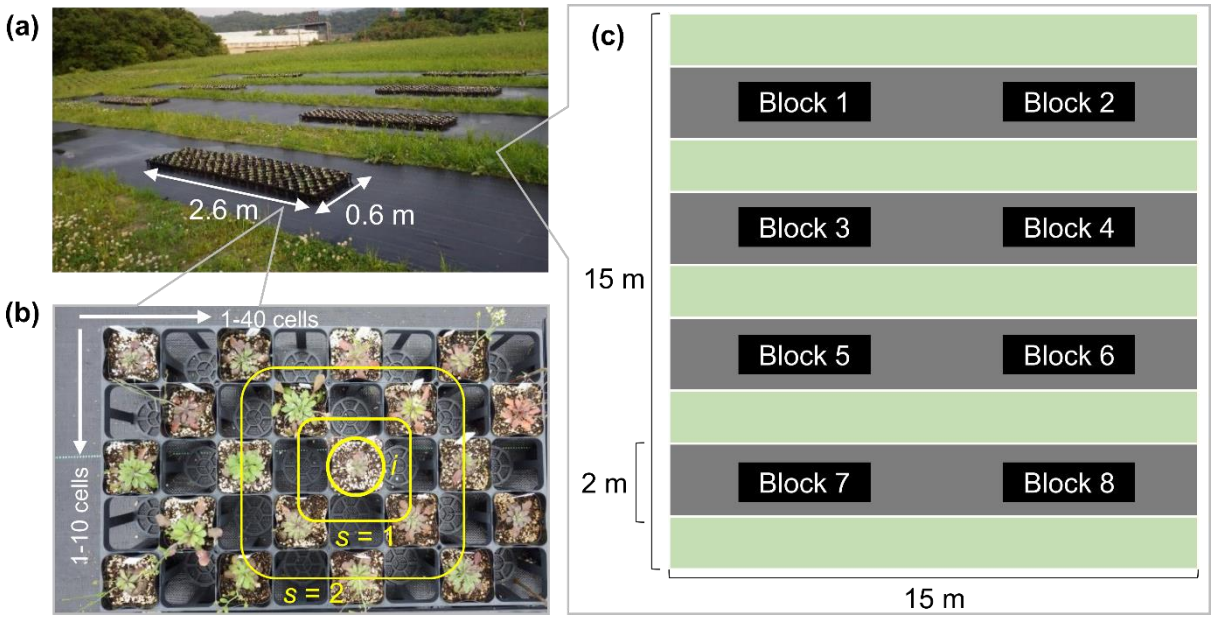


Figure 2. Experimental setting in the *Arabidopsis* herbivory data. (a) Photograph of the field site. Each 0.6 × 2.6 m block included a replicate of 199 accessions and an additional Col-0. (b) *Arabidopsis thaliana* plants were arranged in a 10 × 40 cell plastic tray. The plants were placed in a checkered pattern so that alternate cells were left empty; therefore, each of the 40 rows contained 5 plants and each of the 10 columns contained 20 plants. Yellow lines represent *s*-th neighbor scales from a focal *i*-th plant. (c) A graphical explanation of the experimental area. A meadow (green) was separately covered with weed-masking sheets (grey).

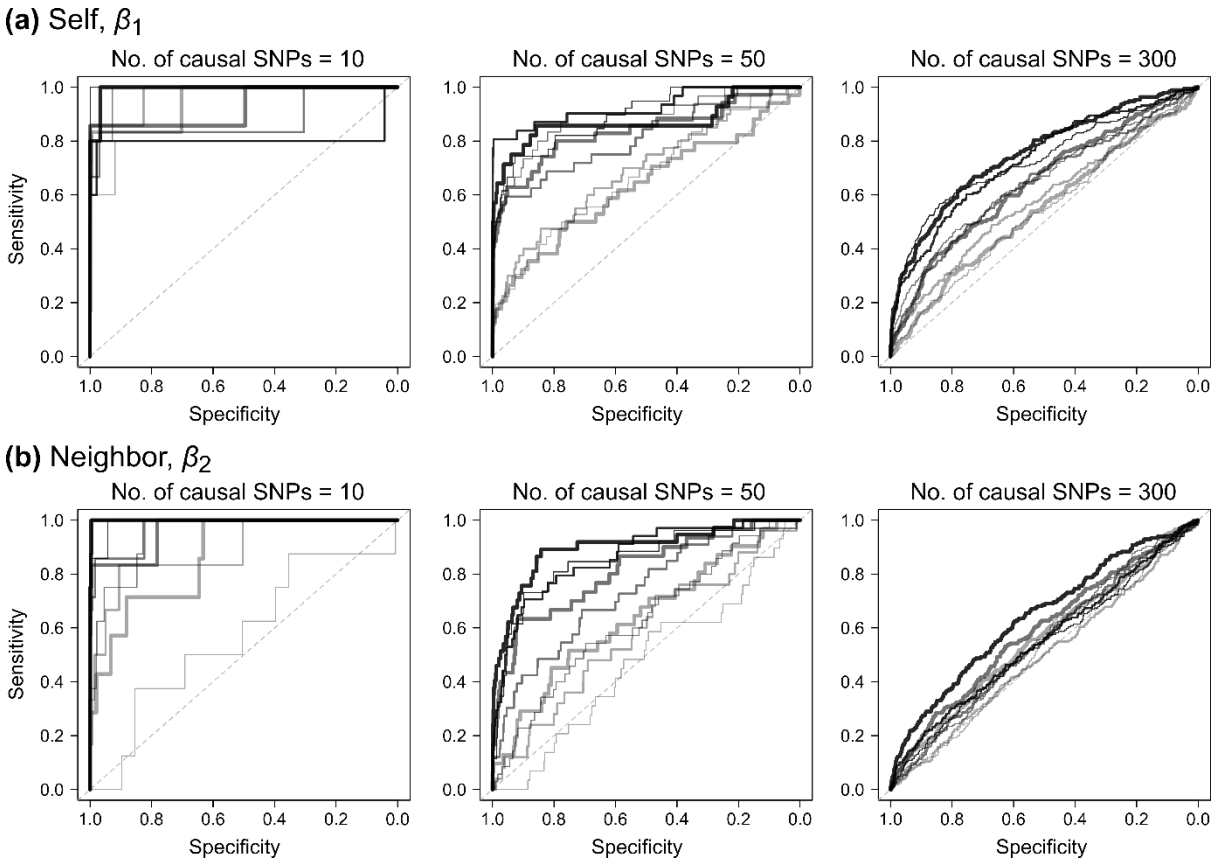


Figure 3. Receiver operating characteristic (ROC) curves representing the sensitivity (true positive rate) and the specificity (1 - false positive rate) in simulated phenotypes. The upper and lower rows give the results of self and neighbor effects with the number of causal SNPs at 10, 50, and 300. The shade of the lines indicates the proportion of phenotypic variation explained by major-effect genes as $PVE_{\boldsymbol{\beta}} = 0.1$ (light gray), 0.3 (dark gray), or 0.6 (black). The thickness of the lines indicates distance decay coefficients as $\alpha = 0.25$ (thin), 1 (medium), or 3 (thick). Entire results are summarized using the area under the ROC curve (AUC) in Figure S1.

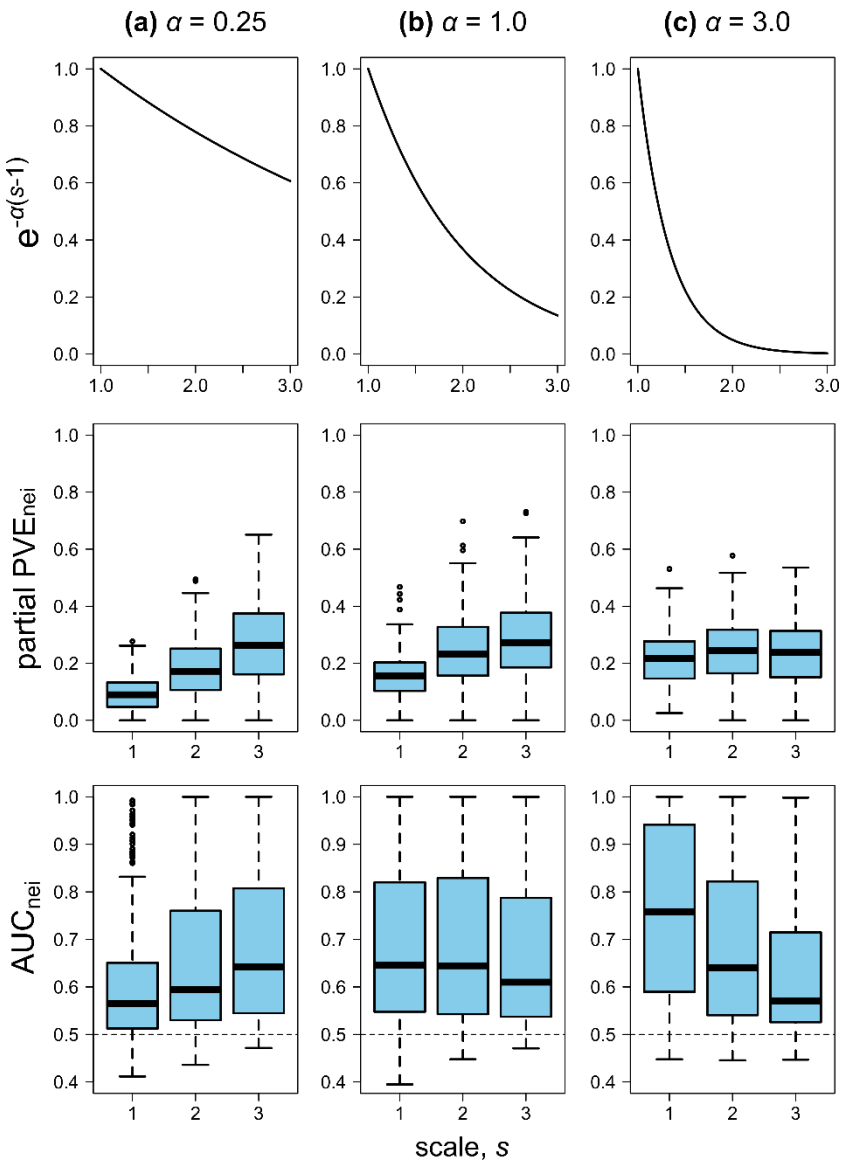
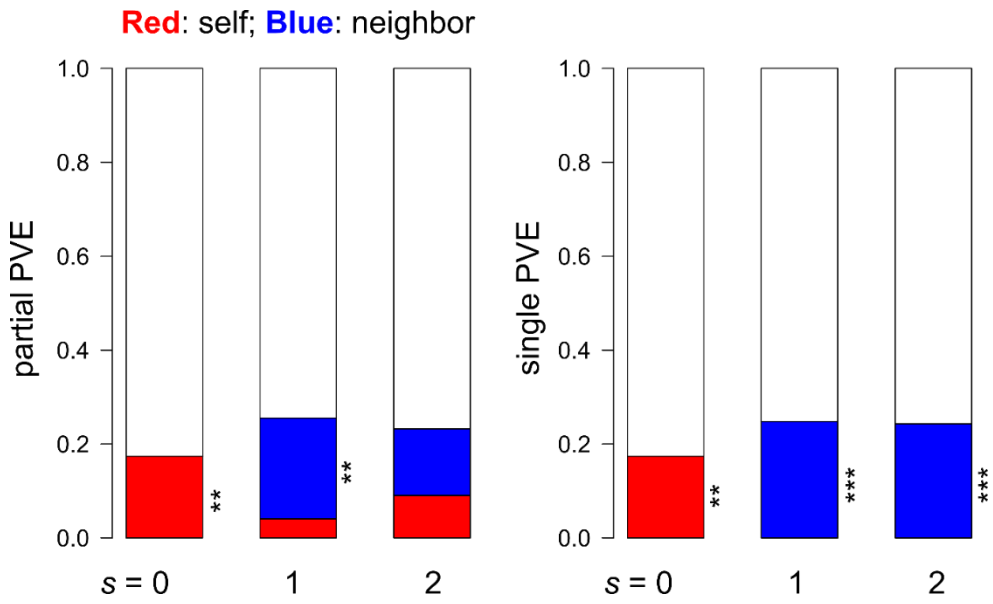


Figure 4. Scale dependence of neighbor effects on simulated phenotypes among all iterations. The broad (a), intermediate (b), and narrow (c) effective range of neighbor effects are represented by weak, moderate, and strong distance decay, respectively. The proportion of phenotypic variation explained by neighbor effects (partial PVE_{nei}) and the area under the ROC curve (AUC) of neighbor effects are shown along the spatial scale from the

first nearest (s = 1) to the third nearest (s = 3) neighbors. Boxplots show center line: median, box limits: upper and lower quartiles, whiskers: 1.5× interquartile range, and points: outliers.

1030



1031

1032 **Figure 5.** Proportion of phenotypic variation explained (PVE) by self-

1033 genotype or neighbor effects on the leaf damage score of field-grown

1034 *Arabidopsis thaliana*. The partial or single PVE is shown at the spatial scale

1035 of $s = 0$ (without any neighbor effects), $s = 1$ and $s = 2$, as defined by the

1036 partial $\text{PVE}_{\text{self}} = \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$; partial $\text{PVE}_{\text{nei}} = \sigma_2^2/(\sigma_1^2 + \sigma_2^2 +$

1037 $\sigma_e^2)$; single $\text{PVE}_{\text{self}} = \sigma_1^2/(\sigma_1^2 + \sigma_e^2)$; and single $\text{PVE}_{\text{nei}} = \sigma_2^2/(\sigma_2^2 + \sigma_e^2)$.

1038 The red or blue fraction indicates PVE by self or neighbor effects,

1039 respectively. The white fraction indicates residuals. Asterisks highlight a

1040 significant fraction with stepwise likelihood ratio tests from simpler to

1041 complex models: $^{**}p$-value < 0.01: $^{***}p$-value < 0.001.

1042

1043

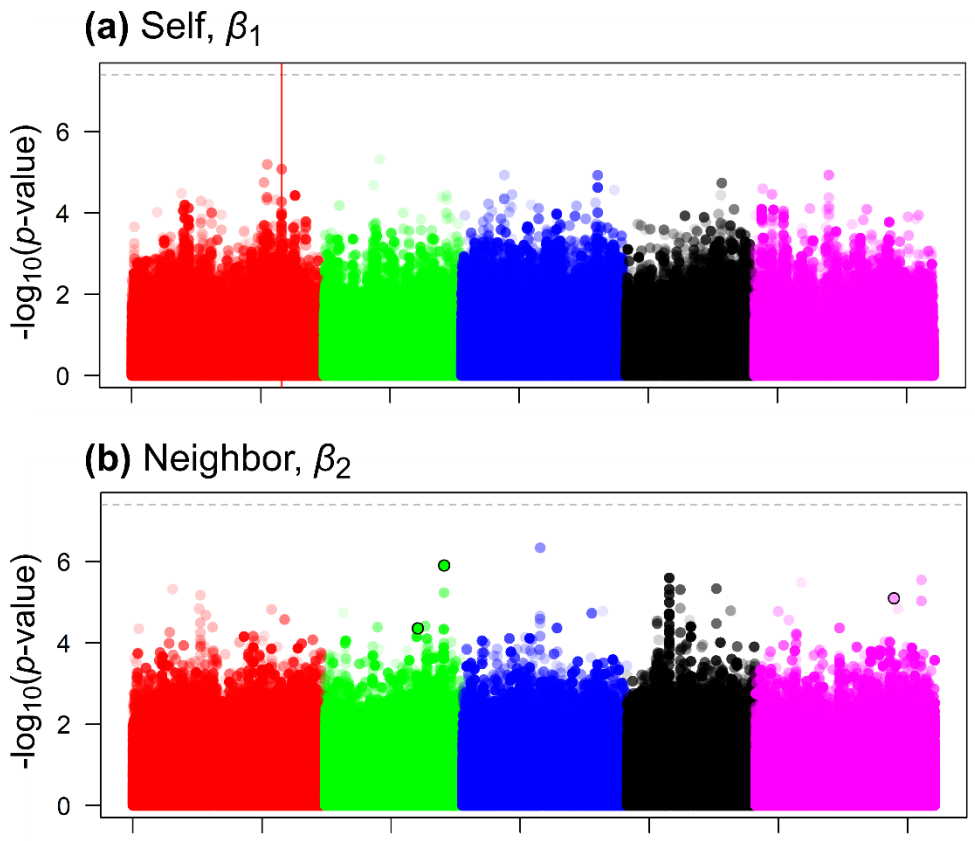


Figure 6. Manhattan plots for the self (a) and neighbor (b) effects on the leaf damage score of field-grown *Arabidopsis thaliana*. The first to fifth chromosomes of *A. thaliana* are differently colored, where lighter plots indicate smaller MAF. Horizontal dashed lines indicate the threshold after Bonferroni correction at p-value < 0.05. The red vertical line in panel (a) indicates a SNP position near the *GS-OX2* locus, while three circles highlighted by a black outline in panel (b) 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.

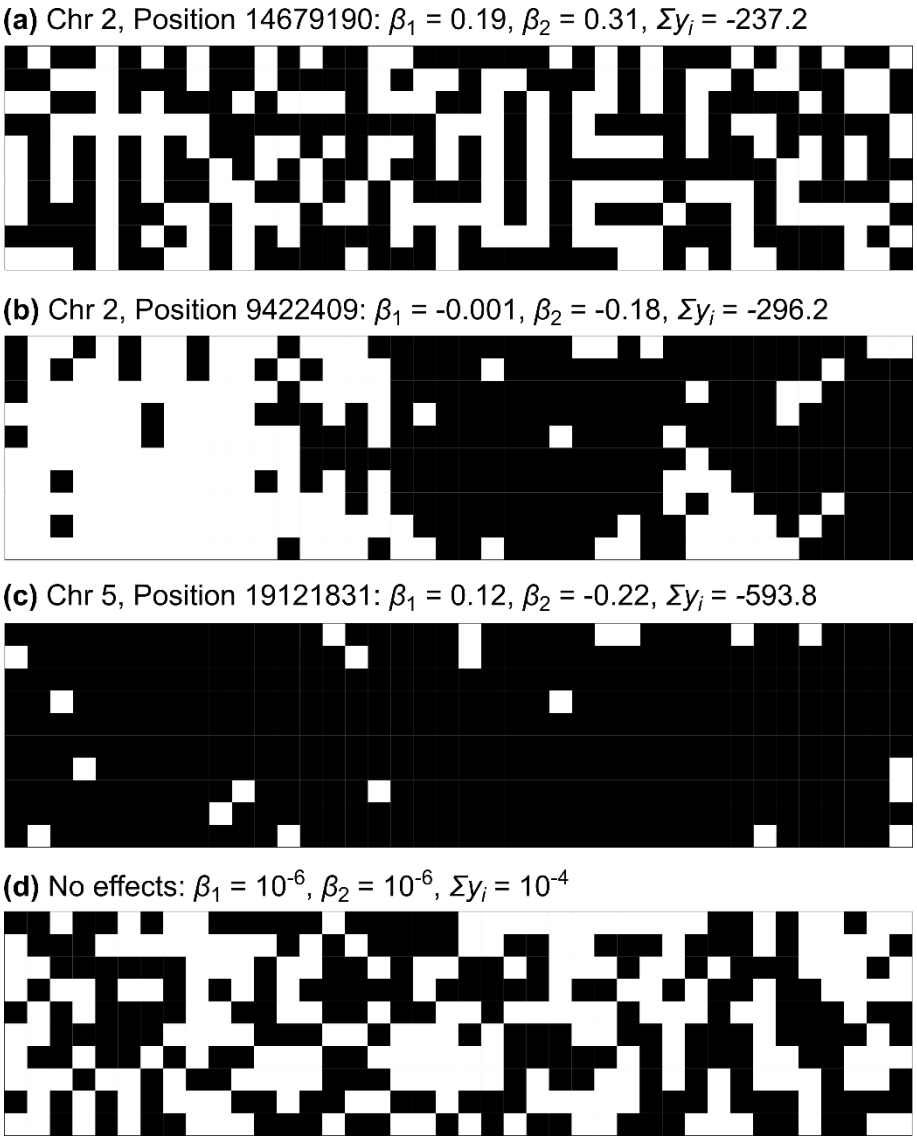


Figure 7. Post hoc simulations exemplifying a spatial arrangement of two alleles expected by the estimated self and neighbor effects, $\hat{\beta}_1$ and $\hat{\beta}_2$, on the leaf damage score of *Arabidopsis thaliana*. Population sum of the leaf damage $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{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