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]Kazusa DNA Research Institute, Kazusa-kamatari 2-6-7, Kisarazu, Chiba 292-0818, 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]Department of Plant Life Sciences, 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

ABSTRACT

An increasing number of field studies show 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 in this study. The proposed method, named "neighbor GWAS", was applied to simulated and real phenotypes using *Arabidopsis thaliana* accessions. Our simulations showed that phenotypic variation explained by neighbor effects approached a plateau when an effective spatial scale became narrow. Thus, the effective scale of neighbor effects could be estimated by patterns of the phenotypic variation explained. The power to detect causal variants of neighbor effects was moderate to strong when a trait was governed by tens of variants. In contrast, there was a reasonable power down when hundreds of variants underlay a single trait. We applied the neighbor GWAS to field herbivory data on 200 accessions of *A. thaliana*, and found that the neighbor effects more largely contributed to the observed damage variation than self-genotype effects. Interestingly, several defensin family genes were associated with neighbor effects on the herbivory, while self-genotype effects were related to flavin-monooxygenase glucosinolate S-oxygenase 2 (FMO GS-OX2). Overall, the neighbor GWAS highlights the overlooked but significant role of plant neighborhood effects in shaping phenotypic variation, thereby providing a novel and powerful tool to dissect complex traits in spatially structured environments.

Keywords: *Arabidopsis thaliana*, GWAS, Ising model, Neighbor effects, Plant-insect interaction

INTRODUCTION

Plants are immobile and thus cannot escape their neighbors. In natural and agricultural fields, individual phenotypes depend not only on the plants' own genotype but also on those of other plants in a neighborhood (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 thus far illustrated that neighbor effects occur among different genotypes within a plant species in herbivory (Schuman et al. 2015; Sato 2018; Ida et al. 2018), pathogen infections (Mundt 2002; Zeller et al. 2012), and pollinator visitations (Genung et al. 2012). Although neighbor effects are of considerable interest in plant science (Dicke and Baldwin 2010; Erb 2018) and its potential application to agriculture (Zeller et al. 2012; Dahlin et al. 2018), these effects are not often 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; Genung et al. 2012; Verschut et al. 2016), and volatile communication among plants (Schuman et al. 2015; Dahlin et al. 2018). For example, lipoxygenase (*LOX*) genes govern jasmonate-mediated volatile emissions that induce defenses of neighboring plants in *Nicotiana* (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* (*GL1*) polymorphism determines hairy or glabrous phenotypes in *Arabidopsis* plants (Hauser et al. 2001) and allow a flightless leaf beetle to avoid hairy plants when encountered at a low frequency (Sato and Kudoh 2017; Sato et al. 2017). Yet, there are few hypothesis-free

73 approaches to seek key genetic variants responsible for plant neighborhood effects.

74 Genome-wide association studies (GWAS) have been increasingly adopted to
75 resolve the genetic architecture of complex traits in the model plant, *Arabidopsis thaliana*
76 (Atwell et al. 2010; Togninalli 2018) and crop species (Hamblin et al. 2011). Plant
77 interactions with herbivores (Brachi et al. 2015; Nallu et al. 2018), microbes (Horton et al.
78 2014; Wang et al. 2018), and other plant species (Frachon et al. 2019), are one of such
79 complex traits dissected through the lens of GWAS. To distinguish causal variants from the
80 genome structure, GWAS often employs a linear mixed model with kinship considered in a
81 random effect (Kang et al. 2008; Korte and Farlow 2013). However, it is generally impossible
82 to test all the gene-by-gene interactions due to combinatorial explosion (Gondro et al. 2013);
83 thus, some feasible and reasonable approach should be invented for GWAS of neighbor
84 effects.

85 To incorporate neighbor effects into GWAS, we focused on a theoretical model of
86 neighbor effects in a magnetic field, known as the Ising model (Ising 1925; McCoy and
87 Maillard 2012), which has been applied to forest gap dynamics and community assembly in
88 plant ecology (Kizaki and Katori 1999; Schlicht and Iwasa 2004; Azaele et al. 2010).
89 Assuming that an individual plant is a magnet, two alleles at each locus correspond to the
90 north or south dipole, whereby genome-wide multiple testing across all loci is analogous to a
91 number of parallel two-dimensional layers. The Ising model has a clear advantage in its
92 interpretability, such that (i) the optimization problem for a population sum of trait values can
93 be regarded as an inverse problem of a simple linear model, (ii) the sign of neighbor effects
94 determines a model trend to generate a clustered or checkered spatial pattern of the two
95 states, and (iii) the self-genotypic effect determines the general tendency to favor one allele
96 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? (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 (Fig. 1). We consider a situation where a plant accession has either of two alleles at each locus, and a number of accessions occupy a finite set of field sites in a two-dimensional lattice. Let x represent allelic status at each locus, the allelic status at each locus of the i-th focal plant and j-th neighboring plants can be designated as $x_{i(j)} \in \{-1, +1\}$. Based on a two-dimensional Ising model, we can define a phenotype value of i-th focal individual plant y_i as

$$y_i = \beta_1 x_i + \beta_2 \sum_{j=1}^{L} x_i x_j \qquad \text{[eq. 1]},$$

where β_1 and β_2 denote self-genotype and neighbor effects, respectively and L is the number of neighboring plants to refer. 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 $\sum_{j=1}^{L} x_i x_j$ 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 can be transformed as $\epsilon_I = -E \sum_{<i,j>} x_i x_j - H \sum x_i$, which defines 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 field 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.

Mixed model. For association mapping, we needed to determine β_1 and β_2 from observed phenotypes and consider 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

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{L} \sum_{j=1}^{L} x_i x_j + u_i + e_i \qquad \text{[eq. 2]},$$

where β_0 indicates the intercept, and the term $\beta_1 x_i$ represents fixed self-genotype effects as tested in conventional GWAS; β_2 is the coefficient of fixed neighbor effects, and the neighbor covariate $\sum_{j=1}^{L} x_i x_j$ is scaled by the number of neighboring plants, L. Variance components due to a sample structure in self and neighbor effects was modeled by a random effect $u_i \sim \text{Norm}(0, \sigma_S^2 \mathbf{K}_S + \sigma_N^2 \mathbf{K}_N)$. The residual was expressed as $e_i \sim \text{Norm}(0, \sigma_e^2)$.

144 The n × n variance-covariance matrices represented the similarity in self-genotypes
145 (i.e., kinship) and neighbor covariates among n individual plants as:

146 $$\mathbf{K}_S = \frac{1}{(q-1)} \mathbf{X}_S^T \mathbf{X}_S \text{ and}$$

147 $$\mathbf{K}_N = \frac{1}{q-1} \mathbf{X}_N^T \mathbf{X}_N,$$

148 where q indicates the number of markers. As we defined $x_{i(j)} \in \{-1, +1\}$, the elements of the
149 kinship matrix $\mathbf{K}_S$ are scaled to represent the proportion of marker loci shared among n × n
150 plants such that $\mathbf{K}_S = (0.5k_{S,ij} + 0.5)$ and $k_{S,ij} = [0, 1]$; σ_S^2 and σ_N^2 indicates variance
151 component parameters for the self and neighbor effects.

152 The n plants × q markers matrix $\mathbf{X}_S$ and $\mathbf{X}_N$ are explanatory variables for the self
153 and neighbor effects as $\mathbf{X}_S = (x_i)$ and $\mathbf{X}_N = \left(\frac{\sum_{j=1}^{L} x_i x_j}{L}\right)$. The individual-level formula eq. 2 could
154 also be converted into a conventional matrix form as:

155 $$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e} \qquad \text{[eq. 3]},$$

156 where **y** is n × 1 vector of phenotypes. **X** is a matrix of fixed effects, including mean, self-
157 genotype x_i, neighbor covariate $(\sum_{j=1}^{L} x_i x_j)/L$, and other confounding covariates for n plants;
158 **β** is a vector that represents coefficients of the fixed effects; **Z** is a design matrix allocating
159 individuals to a genotype, and becomes an identity matrix if all plants are different
160 accessions; **u** is the random effect with $\text{Var}(\mathbf{u}) = \sigma_S^2 \mathbf{K}_S + \sigma_N^2 \mathbf{K}_N$, and **e** is residual as $\text{Var}(\mathbf{e}) =$
161 $\sigma_e^2 \mathbf{I}$. In such a mixed model, the proportion of phenotypic variation explained (PVE) by the
162 self and neighbor effects could be calculated as $\text{PVE}_{self} = \sigma_S^2/(\sigma_S^2+\sigma_N^2+\sigma_e^2)$ and $\text{PVE}_{nei} =$
163 $\sigma_N^2/(\sigma_S^2+\sigma_N^2+\sigma_e^2)$, respectively. The line of extensions to incorporate neighbor effects into
164 GWAS is referred to as "neighbor GWAS" hereafter.
165

166 **Power simulation**

167 To examine the power to detect neighbor effects, we applied the neighbor GWAS to
168 simulated phenotypes. Phenotypes were simulated using a part of the real genotypes of *A.*
169 *thaliana*. To evaluate the power of the simple linear model, we assumed a complex ecological
170 form of neighbor effects with multiple variance components controlled. The power was
171 evaluated by the receiver operating characteristic (ROC) curve on the association score of -
172 $\log_{10}$(p-value) (e.g., Gage et al. 2018). All analyses were performed using R version 3.4.0 (R
173 Core Team 2017).

174 ***Genotype data.*** To consider a realistic genetic structure in the simulation, we used
175 part of the *A. thaliana* RegMap panel (Horton et al. 2012). The genotype data on 1307
176 accessions were downloaded from the Joy Bergelson laboratory website
177 (http://bergelson.uchicago.edu/?page_id=790 accessed on 9 February 2017). We extracted the
178 chromosome 1 and 2 data with the minor allele frequency (MAF) at >0.1, providing a matrix
179 of 1307 plants with 65,226 single nucleotide polymorphisms (SNPs). Pairwise linkage
180 disequilibrium (LD) among the loci was $r^2 = 0.003$ [0.00-0.06: median with upper and lower
181 95 percentiles]. Before generating a phenotype, each locus was centered by its mean and
182 scaled by its standard deviation. Subsequently, we randomly selected 1296 accessions (= 36 ×
183 36 accessions) without any replacements for each iteration, and placed them in a 36 × 72
184 checkered space following *Arabidopsis* experimental settings (see Fig. 2b).

185 ***Phenotype simulation***. To address ecological issues specific to plant neighborhood
186 effects, we considered two extensions, namely asymmetric neighbor effects and spatial
187 scales. Firstly, previous studies showed that such plant-plant interactions are sometimes
188 asymmetric between two accessions in herbivory (e.g., Bergvall et al. 2006; Verschut et al.
189 2016; Sato and Kudoh 2017) and height competition (Weiner 1990). Such asymmetric

neighbor effects can be tested by statistical interactions terms in a linear model (Bergvall et al. 2006; Sato and Kudoh 2017). Secondly, several studies showed that the strength of neighbor effects depended on spatial scales (Hambäck et al. 2014) and the scale of neighbors to be analyzed relied on the dispersal ability of causative organisms (see Hambäck et al. 2009; Sato and Kudoh 2015; Vershut et al. 2018; Ida et al. 2018 for insect and mammal herbivores; Rieux et al. 2014 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 $w(s, \alpha) = e^{-\alpha(s-1)}$, 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:

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{L}\sum_{j=1}^{L} w(s, \alpha)x_i x_j + \beta_{12}\frac{x_i}{L}\sum_{j=1}^{L} w(s, \alpha)x_i x_j + u_i + e_i \quad \text{[eq. 4]},$$

where β_{12} is the coefficient for asymmetry in neighbor effects. Total variance components due to the three background effects i.e., the self, neighbor, and self-by-neighbor effects is defined as $u_i \sim \text{Norm}\left(0, \sigma_S^2\mathbf{K}_S + \sigma_N^2\mathbf{K}_N + \sigma_{SxN}^2\mathbf{K}_{SxN}\right)$. The three variance component parameters σ_S^2, σ_N^2, and σ_{SxN}^2, determined the relative importance of self-genotype, neighbor, and asymmetric neighbor effects in u_i. Given the n plants × q marker explanatory matrix with $\mathbf{X}_{S\times N} = \left(\frac{x_i}{L}\sum_{j=1}^{L} w(s, \alpha)x_i x_j\right)$, the similarity in asymmetric neighbor effects was calculated as $\mathbf{K}_{S\times N} = \frac{1}{q-1}\mathbf{X}_{S\times N}^{T}\mathbf{X}_{S\times N}$. To control phenotypic variations, we further partitioned the proportion of phenotypic variation into those explained by major-effect genes and variance components $PVE_{\beta} + PVE_{u}$, major-effect genes alone PVE_{β}, and residual error PVE_{e}, where $PVE_{\beta} + PVE_{u} + PVE_{e} = 1$. The *optimize* function in R was used to adjust simulated phenotypes to given amounts of PVE.

Parameter setting. Fifteen phenotypes were simulated for each combination of the distance decay α, the proportion of phenotypic variance explained by major-effect genes $PVE_{\boldsymbol{\beta}}$, variance components $PVE_{\boldsymbol{u}}$, and the relative importance of multiple variance components $\sigma^2_S:\sigma^2_N:\sigma^2_{SxN}$ as: α = 0.25, 1.0 or 3.0, $\sigma^2_S:\sigma^2_N:\sigma^2_{SxN}$= 6:3:1, 4:4:1, or 3:6:1, $PVE_{\boldsymbol{\beta}}$ = 0.1, 0.3 or 0.6, and $PVE_{\boldsymbol{\beta}} + PVE_{\boldsymbol{u}}$ = 0.4 or 0.8. The maximum reference scale was fixed at s = 3. The line of simulations was repeated for the number of causal SNPs at 20 or 200 to examine cases of an oligogenic and polygenic control for a trait. The non-zero coefficients for the causal SNPs were randomly sampled from a uniform distribution, Unif(|0.5|, |2.0|), and 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 ones ($\beta_2 \neq 0$ 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^2_{SxN} was also assumed to be smaller than σ^2_N.

***Summary statistics*.** The simulated phenotypes were fitted by eq. 2 to test the significance of coefficients β_1 and β_2, and to estimate the variance component due to self and neighbor effects 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, PVE_{self} and PVE_{nei}, were estimated using 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, the two variance parameters σ_S^2 and σ_N^2 were replaced with their estimates $\hat{\sigma}_S^2$ and $\hat{\sigma}_N^2$ by the AI-REML, and association tests were performed by solving linear mixed models with a fast approximation by an eigen value decomposition (implemented in the *lmm.diago* function: Perdry and Dandine-Roulland 2018). The likelihood was computed using the *lmm.diago.profile.likelihood* function to test β_1 or β_2. True or false positive rates were evaluated by ROC curves and area under the ROC curves (AUC) (Gage et al. 2018). An AUC of 0.5 would indicate that GWAS has no power to detect true signals, while an AUC of 1.0 would indicate that all the top signals predicted by GWAS agree with true signals. The roc function in the pROC package (Robin et al. 2011) was used to calculate AUC from $-\log_{10}$(p-value). The AUC and variance components were calculated from s = 1 (the first nearest neighbors) to s = 3 (up to the third nearest neighbors) cases. The AUCs were also calculated using standard linear models without any random effects to examine whether the linear mixed models were superior to the linear models.

Arabidopsis herbivory data

We applied the neighbor GWAS to field data of *Arabidopsis* herbivory. This field experiment followed our previous publication on a summer herbivory on field-grown *A. thaliana* (Sato et al. 2019). We used 200 worldwide accessions comprising the RegMap (Horton et al. 2012) and 1001 Genomes project (Alonso-Blanco et al. 2016), of which 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 the 200 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: 35º 06' N, 134º 56' E, alt. ca. 200 m: Fig. 2a). Seeds were sown on Jiffy-seven pots (33-mm diameter), and stratified under 4 ℃ for a week. Seedlings were cultivated for 1.5 months under a short-day condition (8 h light: 16 h dark, 20 ℃). Plants were then separately potted in plastic pots (6 cm in diameter) filled with mixed soil of agricultural composts (Metro-mix 350, SunGro Co., USA) and perlites at a 3:1 L ratio. In the field setting, 200 accessions were randomly assigned in a checkered manner within a block (Fig. 2b). Eight replicates of these blocks were set >2 m apart from each other (Fig. 2c). Potted plants were exposed to the field environment for 3 wk in June 2017. At the end of experiment, we scored leaves eaten 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% of the leaf area eaten. 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 wk 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 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 all 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, and experimental blocks as fixed covariates. After the association mapping, we searched candidate genes within ~10 kb around target SNPs, based on the Araport11 gene model with the latest TAIR annotation (accessed on 7 September 2019). Gene ontology (GO) enrichment analysis was applied to the candidate genes near the top 0.1% SNP score. GO categories including >20 and <200 genes were tested by Fisher's exact probability tests and adjusted by false discovery rate (FDR: Benjamini and Hochberg 1995). The GO.db package (Carlson et al. 2018) and the latest TAIR AGI code annotation were used for the GO enrichment analysis. The R source codes, accession list, and phenotype data are available at the GitHub repository (https://github.com/naganolab/NeighborGWAS).

RESULTS

Power simulation

A set of phenotypes were simulated from real genotype data following a complex model eq. 4, and then fitted by a simplified model eq. 2. Analyzing the factors affecting AUCs, we found that the proportion of phenotypic variation (PVE) explained by major-effect genes PVE$_\beta$ and distance decay of neighbor effects were the most influential on the power to detect neighbor signals (Table 1b, d). In addition to PVE$_\beta$, the amount of variance components

309 $PVE_{\mathbf{u}}$ also significantly affected the AUCs of the self and neighbor effects, but these
310 additional effects were less significant compared to those of $PVE_{\boldsymbol{\beta}}$ alone (Table 1). In
311 contrast, the AUCs of neither self nor neighbor effects were significantly affected by the ratio
312 of three variance components $\sigma^2_S:\sigma^2_N:\sigma^2_{SxN}$ (Table 1).

313 Notably, there was a clear relationship between the distance decay α and the
314 proportion of phenotypic variation explained by neighbor effects PVE_{nei} or AUCs at different
315 spatial scales (Fig. 3). If the distance decay was weak and the effective range of neighbor
316 effects was broad, PVE_{nei} and AUCs increased linearly as the reference spatial scale was
317 broadened (Fig. 3a). On the other hand, if the distance decay was strong and the effective
318 scale of neighbor effects was narrow, PVE_{nei} saturated at the scale of the first nearest
319 neighbors (Fig. 3c) or AUCs did not increase (Fig. 3b, c). These results remained the same
320 between the number of causal SNPs = 20 and 200 (Fig. 3 and Fig. S1). The line of simulation
321 results indicated that the effective spatial scales could be estimated by calculating PVE_{nei}
322 across different spatial scales.

323 In the case of the number of causal SNPs = 20, signals of major-effect genes were
324 well detected as AUC ranged from moderate (>0.7) to high (>0.9) (Fig. S2). For the case of
325 the number of causal SNPs = 200, it became relatively difficult to detect the major-effect
326 genes as AUCs were ≤ 0.75 (Fig. S2). The line of simulations indicated that neighbor effects
327 were detectable when a target trait was governed by several major genes and the range of
328 neighbor effects was spatially limited. Additionally, linear mixed models outperformed
329 standard linear models in the detection of self and neighbor signals (ΔAUC_{self} = 0.105 [0.101
330 – 0.109], ΔAUC_{nei} = 0.024 [0.021 – 0.026]: 10,000-times bootstrap mean with 95%
331 confidence intervals). This indicated that the mixed models were more appropriate for the
332 neighbor GWAS to deal with spurious associations due to a sample structure.

333

334 ***Arabidopsis* herbivory data**

335 The variation partitioning of leaf damage showed that the PVE by neighbor effects were
336 larger than PVE by self-genotypic effects (PVE_{self} = 0.026, χ^2_1 = 0.151, p-value = 0.70;
337 PVE_{nei} = 0.218, χ^2_1 = 7.17, p-value = 0.0074: Fig. 4a). Heritability, namely PVE_{self} without
338 neighbor effects, was 0.159 (χ^2_1 = 8.73, p-value = 0.003: Fig. S3). This range of heritability
339 was overlapped with PVE by neighbor effects alone (PVE_{nei} without self-effects = 0.24 at
340 scale s = 1, χ^2_1 = 15.7, p-value < 0.0001: Fig. S3), indicating that there was an intersection
341 between PVE by self and neighbor effects on the leaf damage variation. Phenotypic variation
342 explained by neighbor effects on leaf damage did not increase when the neighbor scale was
343 referred up to the nearest and second nearest individuals (PVE_{self} = 0.083, χ^2_1 = 1.03, p-value
344 = 0.311; PVE_{nei} = 0.13, χ^2_1 = 1.29, p-value = 0.256: Fig. 4a); therefore, the variation
345 partitioning was stopped at s = 2. These results indicated a narrow effective scale and
346 significant contribution of neighbor effects to the leaf damage score.

347 Association mapping of the self-genotype effects on the leaf damage found a SNP
348 with the largest -$\log_{10}$(p-values) score at "chr1-23149476". This SNP was located within ~10
349 kb of the AT1G62540 locus that encoded flavin-monooxygenase glucosinolate S-oxygenase 2
350 (FMO GS-OX2), though this was not above a threshold of Bonferroni correction. Gene
351 ontology annotation of "cellular response to extracellular stimulus" was marginally enriched
352 among genes within ~10 kb around SNPs with the top 0.1% -$\log_{10}$(p-values) score which
353 corresponded to p-values at < 0.00096 (FDR<0.1: Table 2a). A QQ-plot did not exhibit an
354 inflation of p-values for the self-genotype effects (Fig. S4).

355 We found a marginally significant SNP for neighbor effects at the second and third
356 chromosome (Fig. 4c), of which the second chromosomal region had higher association

scores than expected by the QQ-plot (Fig. S4). A locus encoding FAD-binding Berberine family protein (AT2G34810 named *BBE16*) were located within the ~10 kb window near the SNP with the largest -$\log_{10}$(p-values) at the second chromosome, which are known to be up-regulated by methyl jasmonate (Devoto et al. 2005). Three transposable elements and a pseudogene of lysyl-tRNA synthetase 1 were located near the most significant SNP at the third chromosome. These top ten SNPs significantly related to the neighbor effects exhibited positive estimates of β_1 and β_2. Three defense-related GO annotations of "killing cells of other organisms" and "disruption of cells of other organism" were significantly enriched among genes within ~10 kb around SNPs with the top 0.1% score of -$\log_{10}$(p-values) (FDR<0.05: Table 2b). Of the genes with these GO annotations, we found 22 low-molecular weight cysteine-rich proteins or plant defensin family proteins (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 $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{<i,j>} x_i x_j$. The constant intercept β_0, the variance component u_i, and residual e_i were not considered because they were not involved in deterministic dynamics of the model. Figure 5 shows three representatives and a neutral expectation. For example, a mixture of a dimorphism was expected to decrease the total leaf damage for a SNP at "chr2-14679190" near the *BBE16* locus ($\hat{\beta}_2$>0: Fig. 5a). On the other hand, a clustered distribution of a dimorphism was expected to decrease the total damage for a SNP at "chr2-9422409" near the AT2G22170 locus encoding a lipase/lipooxygenase PLAT/LH2 family protein ($\hat{\beta}_1 \approx 0$, $\hat{\beta}_2$<0: Fig. 5b). 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. 5c). No self and neighbor effects

led to a random distribution and no mitigation of damage i.e., $\sum y_i \approx 0$ (Fig. 5d). These post hoc simulations suggested a potential applicability of neighbor GWAS in optimizing spatial arrangements in field cultivation.

DISCUSSION

Spatial and genetic factors affecting the power to detect signals

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 ($\alpha = 1.0$ 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 plant defense, mobile animals, such as mammalian browsers and flying insects, often select a cluster of plant individuals (e.g., Bergvall et al. 2006; Hambäck et al. 2009; Vershut et al. 2016); however, neighbor effects could not be detected among individual plants within a cluster (Hambäck et al. 2014; Sato and Kudoh 2015). This case was represented by the exponential distance decay of $\alpha = 0.25$; only in such a special case should more than the first nearest be referred to gain the power.

Neighbor GWAS could retain its power as long as neighbor effects were spatially limited and several major-effect genes governed a trait. In contrast, when hundreds of causal variants involved a single trait and less than half of phenotypic variation was attributable to neighbor effects, we observed a reasonable power down of neighbor GWAS. In GWAS, false positive rates can be reduced using linear mixed models that deal with kinship structure as a random effect (Korte and Farlow 2013). Indeed, mixed models were superior to standard linear models in this simulation. Our simulation also adjusted the three variance components

404 σ^2_S, σ^2_N, and σ^2_{SxN}, but their relative contribution did not have significant effects on the power.
405 This was likely due to the fact that the self-genotypic variable x_i was encompassed into the
406 neighbor variable $\sum x_i x_j/L$, and thus the kinship matrix $\mathbf{K}_S$ was partially redundant with the
407 similarity matrix of neighbor effects $\mathbf{K}_N$. Indeed, elemental-wise correlations between $\mathbf{K}_S$
408 and $\mathbf{K}_N$ were strong in our simulations ($R^2 > 0.7$). Thus, linear mixed models gain the power
409 to detect neighbor effects if signals are strong, but likelihood ratio tests are reliable enough to
410 deal with the correlated variables.
411
412 **Candidate genes related to field herbivory on *Arabidopsis***
413 Our *Arabidopsis* data successfully detected known defense-related genes involved in the self-
414 genotype effects on leaf damage. Aliphatic glucosinolates are a major chemical defense
415 against insect herbivory (Brachi et al. 2015; Kerwin et al. 2017). Specifically, FMO GS-OX2
416 is involved in aliphatic glucosinolate biosynthesis by catalyzing the conversion from
417 methylthioalkyl to corresponding methylsulfinylalkyl glucosinolates (Li et al. 2008).
418 Furthermore, previous GWAS reported methionine synthase 2 (AT3G03780), disease
419 resistance protein (TIR-NBS-LRR class) family (AT4G16950), and monodehydroascorbate
420 reductase 4 (AT3G27820) as candidate genes involved in self-resistance to the white butterfly
421 *Pieris rapae* (Davila-Olivas et al. 2017; Nallu et al. 2018). In this field experiment, we
422 observed larvae of *P. rapae* largely feeding on *A. thaliana*, and the GWAS of self-genotype
423 effects on leaf damage detected the above three candidate genes near SNPs with the top 0.1%
424 association score. Thus, our GWAS results seemed convincing in terms of the detection of
425 known defense-related genes in the self-genotypic effects on herbivory.
426 Notably, the neighbor effects in herbivory were relevant to candidate genes
427 disrupting cells of other organisms. Plant defensins are stable and cysteine-rich peptides that

428 confer plant resistance by killing cells of other organisms (Stotz 2009). While anti-fungal
429 resistance is a well-known function of plant defensins (Stotz 2009), they can also act as
430 protease inhibitors against insect herbivores (Bloch and Richardson 1991; Pelegrini et al.
431 2008; Choi et al. 2009). Typical examples of neighbor effects may be a direct induction of
432 plant defense via volatile organic chemicals (e.g., Schuman et al. 2015; Dahlin et al. 2018),
433 but ecological studies have shown that herbivore host choice is one of the most important
434 processes leading indirect neighbor effects to anti-herbivore defenses (Bergvall et al. 2006;
435 Verschut et al. 2016; Sato et al. 2017). The findings of our neighbor GWAS suggest a putative
436 role of plant defensins in modulating insect feeding behaviors and thus neighbor effects in
437 herbivory.
438
439 **Conclusion and applicability**
440 Based on the newly proposed methodology, we suggest that neighbor effects are a more
441 important source of phenotypic variation in field-grown plants than currently appreciated. To
442 date, regional-scale interactions among plants have been analyzed using a genome-
443 environment association study of plant community composition (Frachon et al. 2019), but
444 fine-scale neighbor effects have yet to be examined. Using tens of *A. thaliana* accessions and
445 their experimentally mixed populations, Wuest and Niklaus (2018) recently showed that a
446 single genomic region drives neighbor effects in plant growth via soil improvements, and this
447 genetic effect shapes a positive relationship between plant genotype diversity and
448 productivity. Our newly proposed methodology of neighbor GWAS could be a powerful tool
449 to identify such a key genetic variant responsible for neighbor effects and resulting
450 biodiversity effects.
451 Neighbor GWAS may also potentially help determine an optimal spatial arrangement

452 in plant cultivation, as suggested by the post hoc simulation. The Ising model is well
453 established in statistical physics (McCoy and Maillard 2012) and is now applied to a
454 machine-learning pipeline that deals with high-dimensionality in genomics data (Fisher and
455 Mehta 2015). Genome-wide polymorphism data are useful not only to seek causal genes in
456 GWAS, but also to predict breeding values of crop plants in genomic selection (e.g., Jannink
457 et al. 2010; Hamblin et al. 2011; Yamamoto et al. 2017). Although it is still challenging to
458 determine β_1 and β_2 for all SNPs efficiently, the linear model of neighbor GWAS could also
459 be implemented as a genomic selection at a population level. Thus, our study provides an
460 avenue for future studies to predict population-level phenotypes in spatially structured
461 environments.
462

ACKNOWLEDGEMENTS

The authors thank Ü. Seren, A. Korte, and M. Nordborg for kindly providing the full imputed SNP data; and T. Tsuchimatsu and K. Iwayama for discussions. 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.; 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 that there are no conflicts of interests in this study.

DATA ARCHIVING

The leaf damage data on *A. thaliana* are included in the supporting information (Table S1). The R source codes used in this study are available at the GitHub repository (https://github.com/naganolab/NeighborGWAS).

REFERENCES

1. 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
2. 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.

3. 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

4. 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.

5. Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc B 57: 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

6. 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

7. Bloch C, Richardson M. (1991). A new family of small (5 kDa) protein inhibitors of insect α-amylases from seeds or sorghum (*Sorghum bicolor* (L) Moench) have sequence homologies with wheat γ-purothionins. FEBS letters 279: 101-104.

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.

9. 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.

10. Carlson M. (2018). GO.db: A set of annotation maps describing the entire Gene

511 Ontology. R package version 3.7.0.

512 11. Carrasco LR, Harwood TD, Toepfer S, MacLeod A, Levay N, Kiss J, Baker RHA,
513 Mumford JD, Knight JD. (2010). Dispersal kernels of the invasive alien western corn
514 rootworm and the effectiveness of buffer zones in eradication programmes in Europe.
515 Ann Appl Biol 156: 63–77. doi:10.1111/j.1744-7348.2009.00363.x

516 12. Chan EKF, Rowe HC, Kliebenstein DJ. (2010). Understanding the evolution of defense
517 metabolites in *Arabidopsis thaliana* using genome-wide association mapping. Genetics
518 185: 991–1007.

519 13. Choi MS, Kim YH, Park HM, Seo BY, Jung JK, Kim ST. et al. (2009). Expression of
520 BrD1, a plant defensin from *Brassica rapa*, confers resistance against brown planthopper
521 (*Nilaparvata lugens*) in transgenic rices. Mol Cells 28: 131–137. doi:10.1007/s10059-
522 009-0117-9.

523 14. Dahlin I, Rubene D, Glinwood R, Ninkovic V. (2018). Pest suppression in cultivar
524 mixtures is influenced by neighbor-specific plant-plant communication. Ecol Appl 28:
525 2187–2196. doi:10.1002/eap.1807

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

529 16. Davila-Olivas NH, Kruijer W, Gort G, Wijnen CL, van Loon JJA, Dicke M. (2017).
530 Genome-wide association analysis reveals distinct genetic architectures for single and
531 combined stress responses in *Arabidopsis thaliana*. New Phytol 213: 838–851.
532 doi:10.1111/nph.14165

533 17. Devoto A, Ellis C, Magusin A, Chang HS, Chilcott C, Zhu T, Turner JG. (2005).
534 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.

18. Dicke M, Baldwin IT. (2010). The evolutionary context for herbivore-induced plant volatiles: beyond the 'cry for help'. Trends Plant Sci 15: 167–175. doi:10.1016/j.tplants.2009.12.002
19. 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. doi:10.1016/j.pbi.2018.03.008
20. 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.
21. 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.
22. 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.
23. Genung MA, Bailey JK, Schweitzer JA. (2012). Welcome to the neighbourhood: interspecific genotype by genotype interactions in *Solidago* influence above- and belowground biomass and associated communities. Ecol Lett 15: 65–73.
24. Gondro C, van der Werf J, Hayes B. (eds.) (2013). Genome-wide association studies and genomic prediction. Methods Mol Biol. Humana Press, New York.
25. 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

559 26. Hambäck PA, Inouye BD, Andersson P, Underwood N. (2014). Effects of plant
560 neighborhoods on plant-herbivore interactions: resource dilution and associational
561 effects. Ecology 95: 1370–1383. doi:10.1890/13-0793.1
562 27. Hamblin MT, Buckler ES, Jannink JL. (2011). Population genetics of genomics-based
563 crop improvement methods. Trends Genet 27: 98–106.
564 28. Hauser MT, Harr B, Schlötterer C. (2001). Trichome distribution in *Arabidopsis thaliana*
565 and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene
566 *GLABROUS1*. Mol Biol Evol 18: 1754–1763.
567 29. Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, et al. (2012).
568 Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions
569 from the RegMap panel. Nat Genet 44: 212–216. doi:10.1038/ng.1042
570 30. Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, et al.
571 (2014). Genome-wide association study of *Arabidopsis thaliana* leaf microbial
572 community. Nat Commun 5: 5320. doi:10.1038/ncomms6320
573 31. Ida TY, Takanashi K, Tamura M, Ozawa R, Nakashima Y, Ohgushi T. (2018). Defensive
574 chemicals of neighboring plants limit visits of herbivorous insects: Associational
575 resistance within a plant population. Ecol Evol 8: 12981–12990. doi:10.1002/ece3.4750
576 32. Ising E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik* 31: 253–
577 258.
578 33. Jannink JL, Lorenz AJ, Iwata H. (2010). Genomic selection in plant breeding: from
579 theory to practice. Brief Funct Genomic 9: 166–177.
580 34. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. (2008).
581 Efficient control of population structure in model organism association mapping.
582 Genetics 178: 1709–1723.

583 35. Kerwin RE, Feusier J, Muok A, Lin C, Larson B, Copeland D. et al. (2017). Epistasis ×
584 environment interactions among *Arabidopsis thaliana* glucosinolate genes impact
585 complex traits and fitness in the field. New Phytol 215: 1249–1263.
586 36. Kizaki S, Katori M. (1999). Analysis of canopy-gap structures of forests by Ising-Gibbs
587 states-equilibrium and scaling property of real forests. J Phys Soc Japan **68**: 2553-2560.
588 37. Korte A, Farlow A. (2013). The advantages and limitations of trait analysis with GWAS:
589 a review. Plant Methods 9: 29.
590 38. Li J, Hansen BG, Ober JA, Kliebenstein DJ, Halkier BA. (2008). Subclade of flavin-
591 monooxygenases involved in aliphatic glucosinolate biosynthesis. Plant Physiol 148:
592 1721–1733.
593 39. McCoy BM, Maillard JM. (2012). The importance of the Ising model. Progress Theor
594 Phys 127: 791-817.
595 40. Mundt CC. (2002). Use of multiline cultivars and cultivar mixtures for disease
596 management. Ann Rev Phytopathol 40: 381-410.
597 41. Nallu S, Hill JA, Don K, Sahagun C, Zhang W, Meslin C. et al. (2018). The molecular
598 genetic basis of herbivory between butterflies and their host plants. Nat Ecol Evol 2:
599 1418-1427. doi:10.1038/s41559-018-0629-9
600 42. Pelegrini PB, Lay FT, Murad AM, Anderson MA, Franco OL. (2008). Novel insights on
601 the mechanism of action of A-amylase inhibitors from the plant defensin family.
602 Proteins: Struct Funct Bioinformatics 73: 719–729. doi:10.1002/prot.22086.
603 43. Perdry H, Dandine-Roulland C. (2018). gaston: Genetic Data Handling (QC, GRM, LD,
604 PCA) & Linear Mixed Models. R package version 1.5.4. https://CRAN.R-
605 project.org/package=gaston
606 44. R Core Team. (2017). R: A language and environment for statistical computing. R

607 Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

608 45. Rieux A, Soubeyrand S, Bonnot F, Klein EK, Ngando JE, Mehl A. et al. (2014). Long-
609 distance wind-dispersal of spores in a fungal plant pathogen: estimation of anisotropic
610 dispersal kernels from an extensive field experiment. PLoS ONE 9: e103225.
611 doi:10.1371/journal.pone.0103225

612 46. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. (2011).
613 pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC*
614 *Bioinformatics* **12**: 77. doi:10.1186/1471-2105-12-77

615 47. Schlicht R, Iwasa Y. (2004). Forest gap dynamics and the Ising model. J Theoret Biol
616 230: 65–75. doi:10.1016/j.jtbi.2004.04.027

617 48. Sato Y. (2018). Associational effects and the maintenance of polymorphism in plant
618 defense against herbivores: review and evidence. Plant Species Biol 33: 91–108.
619 doi:10.1111/1442-1984.12201

620 49. Sato Y, Kudoh H. (2015). Tests of associational defence provided by hairy plants for
621 glabrous plants of *Arabidopsis halleri* subsp. *gemmifera* against insect herbivores. Ecol
622 Entomol 40: 269–279. doi:10.1111/een.12179

623 50. Sato Y, Kudoh H. (2017). Herbivore-mediated interaction promotes the maintenance of
624 trichome dimorphism through negative frequency-dependent selection. Am Nat 190:
625 E67–E77. doi:10.1086/692603

626 51. Sato Y, Ito K, Kudoh H. (2017). Optimal foraging by herbivores maintains
627 polymorphism in defence in a natural plant population. Funct Ecol 31: 2233–2243.
628 doi:10.1111/1365-2435.12937

629 52. Sato Y, Shimizu-Inatsugi R, Yamazaki M, Shimizu KK, Nagano AJ. (2019). Plant
630 trichomes and a single gene *GLABRA1* contribute to insect community composition on

631 field-grown *Arabidopsis thaliana*. BMC Plant Biol 19:163. doi:10.1186/s12870-019-
632 1705-2
633 53. Schuman MC, Allmann S, Baldwin IT. (2015). Plant defense phenotypes determine the
634 consequences of volatile emission for individuals and neighbors. eLife 4: e04490
635 54. Stotz HU, Thomson J, Wang Y. (2009). Plant defensins: defense, development and
636 application. Plant Signal Behav 4: 1010–1012.
637 55. Tahvanainen JO, Root RB. (1972). The influence of vegetational diversity on the
638 population ecology of a specialized herbivore, *Phyllotreta cruciferae* (Coleoptera:
639 Chrysomelidae). Oecologia 10: 321–346.
640 56. Togninalli M, Seren Ü, Meng D, Fitz J, Nordborg M, Weigel D. et al. (2018). The
641 AraGWAS Catalog: a curated and standardized *Arabidopsis thaliana* GWAS catalog.
642 Nucleic Acids Res 46: D1150–D1156.
643 57. Underwood N, Inouye BD, Hambäck PA. (2014). A conceptual framework for
644 associational effects: when do neighbors matter and how would we know? Q Rev Biol
645 89: 1–19.
646 58. Verschut TA, Becher PG, Anderson P, Hambäck PA. (2016). Disentangling associational
647 effects: both resource density and resource frequency affect search behaviour in complex
648 environments. Funct Ecol 30: 1826-1833.
649 59. Wang M, Roux F, Bartoli C, Huard-Chauveau C, Meyer C, Lee H. et al. (2018). Two-
650 way mixed-effects methods for joint association analysis using both host and pathogen
651 genomes. Proc Natl Acad Sci USA 115: E5440–E5449. doi:10.1073/pnas.1710980115
652 60. Weiner J. (1990). Asymmetric competition in plant populations. Trends Ecol Evol 5: 360-
653 364.
654 61. Wuest SE, Niklaus PA. (2018). A plant biodiversity effect resolved to a single

chromosomal region. Nat Ecol Evol 2: 1933–1939.

62. 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.

63. Zeller SL, Kalinina O, Flynn DFB, Schmid B. (2012). Mixtures of genetically modified wheat lines outperform monocultures. Ecol Appl 22: 1817–1826.

663 **TABLES & FIGURES**

664 **Table 1.** Factors affecting the power to detect signals in simulated phenotypes. The response
665 variable was the maximum Area Under the ROC Curve (AUC) of the spatial scales from s =
666 1 to s = 3. ANOVA tables show the degree of freedom (df), sum of squares (SS), F-statistics,
667 and p-values.

(a) AUC_{self}, No. of causal SNPs = 20

Factors	df	SS	F	p-value
σ^2_S:σ^2_N:σ^2_{SxN}	2	0.0110	1.91	0.149
α	1	0.0003	0.10	0.750
PVE_{β}	1	0.410	142.7	< 2.2e-16
PVE_{β} + PVE_{u}	1	0.0196	6.81	0.00933
residuals	534	1.53		

(b) AUC_{nei}, No. of causal SNPs = 20

Factors	df	SS	F	p-value
σ^2_S:σ^2_N:σ^2_{SxN}	2	0.013	1.4	0.259
α	1	1.006	205	< 2.2e-16
PVE_{β}	1	1.101	225	< 2.2e-16
PVE_{β} + PVE_{u}	1	0.060	12.2	0.000520
residuals	174	2.62		

(c) AUC_{self}, No. of causal SNPs = 200

Factors	df	SS	F	p-value
σ^2_S:σ^2_N:σ^2_{SxN}	2	0.0000	0.01	0.9909
α	1	0.0044	6.05	0.0142
PVE_{β}	1	0.798	1106	<2.2e-16
PVE_{β} + PVE_{u}	1	0.0115	16.0	7.29E-05
residuals	534	0.385		

(d) AUC_{nei}, No. of causal SNPs = 200

Factors	df	SS	F	p-value
σ^2_S:σ^2_N:σ^2_{SxN}	2	0.0011	0.81	0.445
α	1	0.195	295	< 2.2e-16
PVE_{β}	1	0.248	375	< 2.2e-16
PVE_{β} + PVE_{u}	1	0.0050	7.60	0.00605
residuals	534	0.352		

668

669

670

671

672

673

674

675

676

677

678 **Table 2.** GO enrichment analysis of the leaf damage score with Fisher's exact probability
679 tests at FDR < 0.1. Candidate genes within ~10 kb around SNPs with the top 0.1%
680 association score -$\log_{10}$(p-values) were subject to the GO analysis.

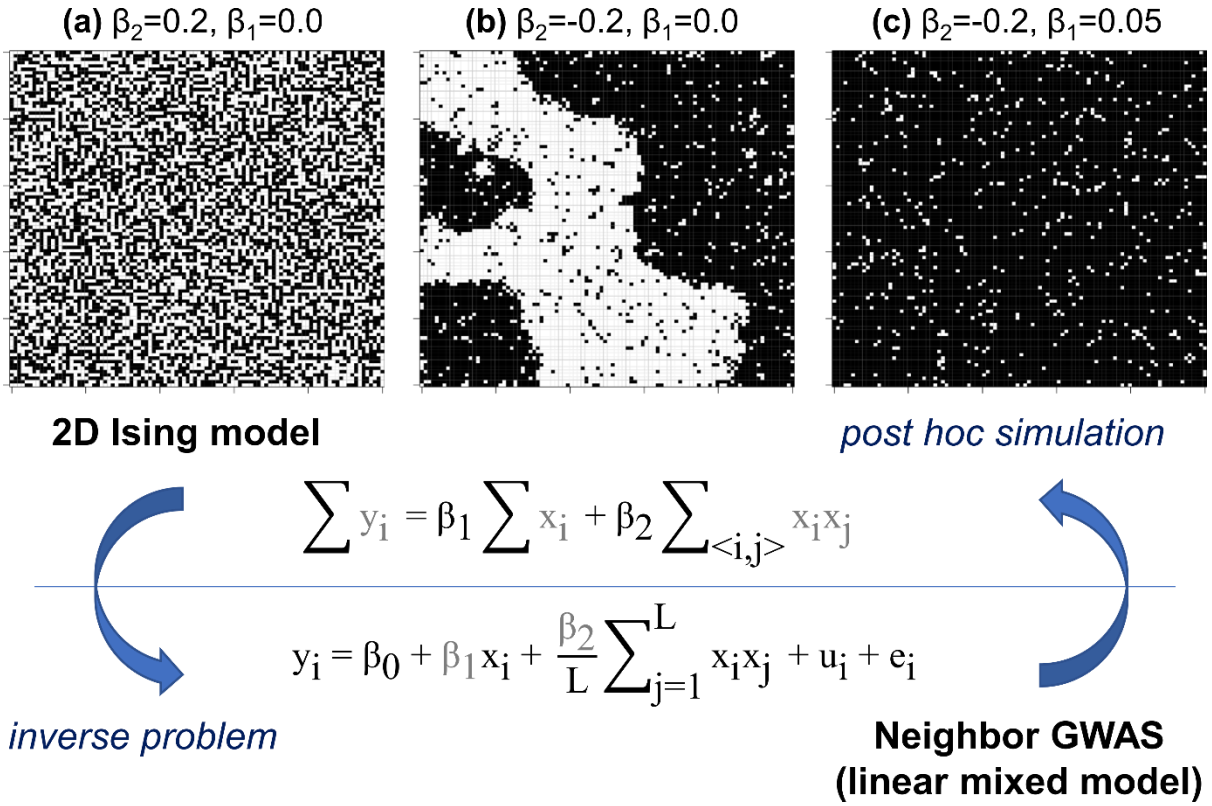
(a) Self, β_1

GO	FDR	Description
GO:0043531	0.0071	ADP binding
GO:0009267	0.0083	cellular response to starvation
GO:0031669	0.0083	cellular response to nutrient levels
GO:0050662	0.0127	coenzyme binding
GO:0031668	0.0546	cellular response to extracellular stimulus
GO:0042594	0.0546	response to starvation
GO:0071496	0.0673	cellular response to external stimulus
GO:0009605	0.0829	response to external stimulus
GO:0004553	0.0829	hydrolase activity, hydrolyzing O-glycosyl compounds
GO:0016798	0.0829	hydrolase activity, acting on glycosyl bonds
GO:0031667	0.0969	response to nutrient levels
GO:0005618	0.0969	cell wall
GO:0030312	0.0969	external encapsulating structure
GO:0000166	0.0982	nucleotide binding
GO:1901265	0.0982	nucleoside phosphate binding

(b) Neighbor, β_2

GO	FDR	Description
GO:0004857	0.0271	enzyme inhibitor activity
GO:0031640	0.0271	killing of cells of other organism
GO:0001906	0.0271	cell killing
GO:0044364	0.0271	disruption of cells of other organism
GO:0043531	0.0271	ADP binding
GO:0035821	0.0307	modification of morphology or physiology of other organism
GO:0010393	0.0307	galacturonan metabolic process
GO:0045488	0.0307	pectin metabolic process
GO:0042545	0.0341	cell wall modification
GO:0044419	0.0595	interspecies interaction between organisms

681
682
683



684

685 **Figure 1**. Relationship between Neighbor GWAS and Ising model. Upper panels show spatial
686 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 $\beta_2>0$, mixed
687 patterns give the argument of the minimum for a population sum of phenotype values Σy_i. (b)
688 If $\beta_2<0$, clustered patterns give the argument of the minimum for Σy_i. (c) In addition, β_1
689 determines overall patterns favoring -1 or +1 states. Shown are outcomes from a random 100
690 × 100 lattice after 1000 iterations of Gibbs sampling. Conversely, the neighbor GWAS was
691 implemented as an inverse problem of the 2-D Ising model, where genotypes and its spatial
692 arrangement, x_i and $x_i x_j$, are given while the coefficients β_1 and β_2 are to be estimated from
693 observed phenotypes y_i. In addition, the variance component due to self and neighbor effects
694 was considered a random effect in a linear mixed model, such that $u_i \sim$ Norm(0,
695 $\sigma_S^2 \mathbf{K}_S + \sigma_N^2 \mathbf{K}_N$). Once β_1 and β_2 are determined, we could simulate a genotype distribution that
696 maximizes or minimizes Σy_i.

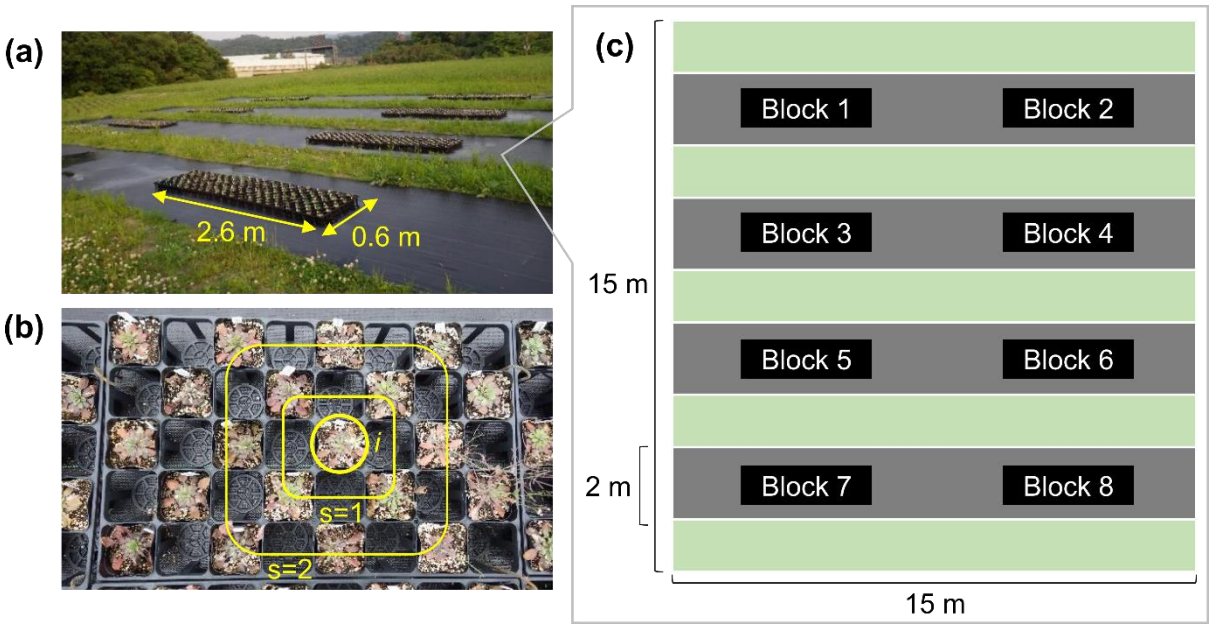


Figure 2. Experimental setting in the *Arabidopsis* herbivory data. (a) Photograph of the field site. Each 0.6 × 2.6 m block included a replicate of 200 accessions, where 5 × 40 plants were assigned to a row and column, respectively. (b) *Arabidopsis thaliana* plants were arranged in a checkered manner. 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).

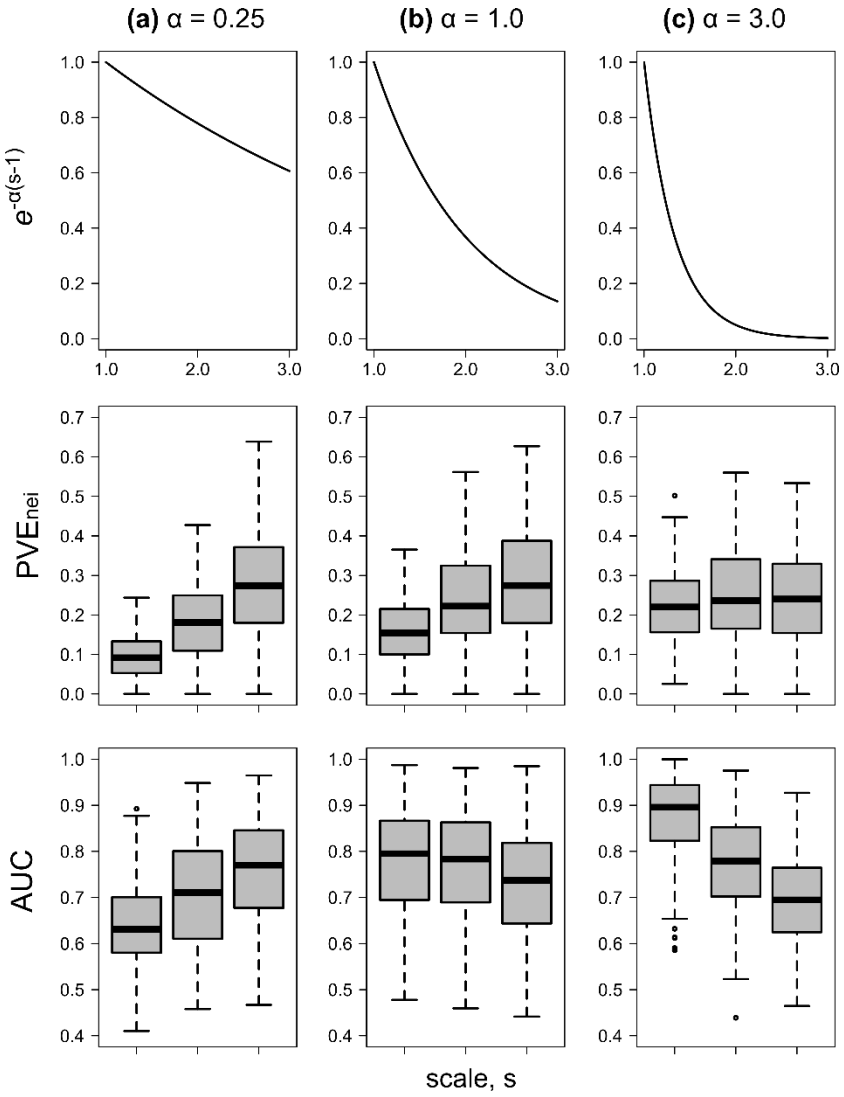


Figure 3. Scale dependence of neighbor effects on simulated phenotypes among all iterations with the number of causal SNPs = 20. The broad (a), intermediate (b), and narrow (c) effective range of neighbor effects are represented by weak, moderate, and strong distance decay, respectively. The proportion of phenotypic variation explained by neighbor effects (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. An AUC at 1.0 indicates a perfect detection of signals. Boxplots show center line: median, box limits: upper and lower quartiles, whiskers: 1.5× interquartile range, and points: outliers. The case for the number of causal SNPs = 200 is given in Figure S1.

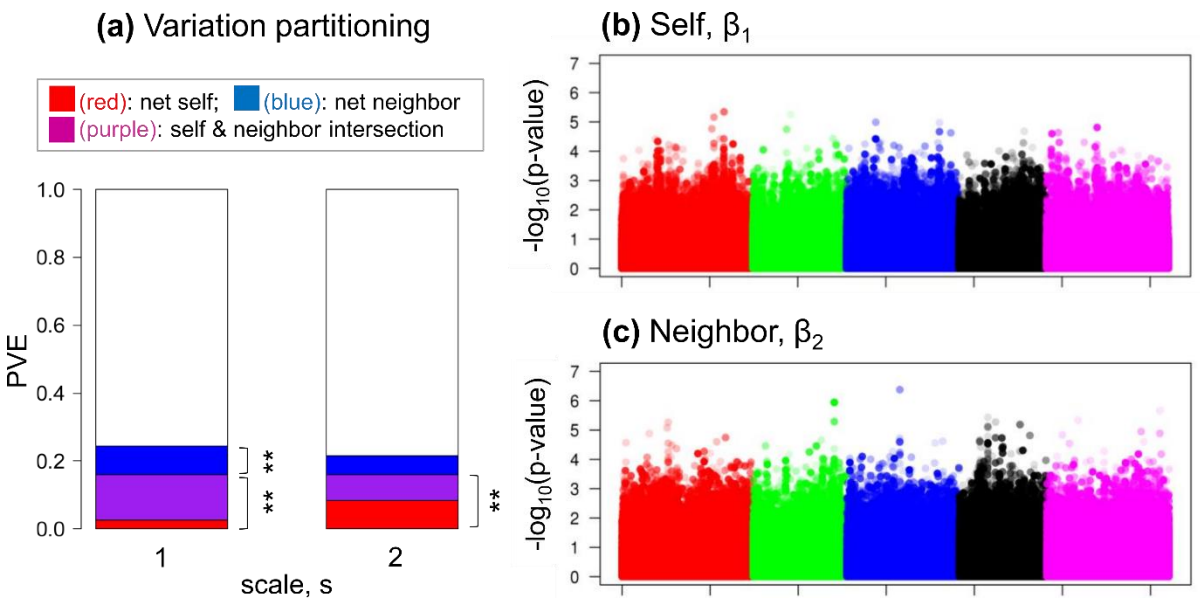
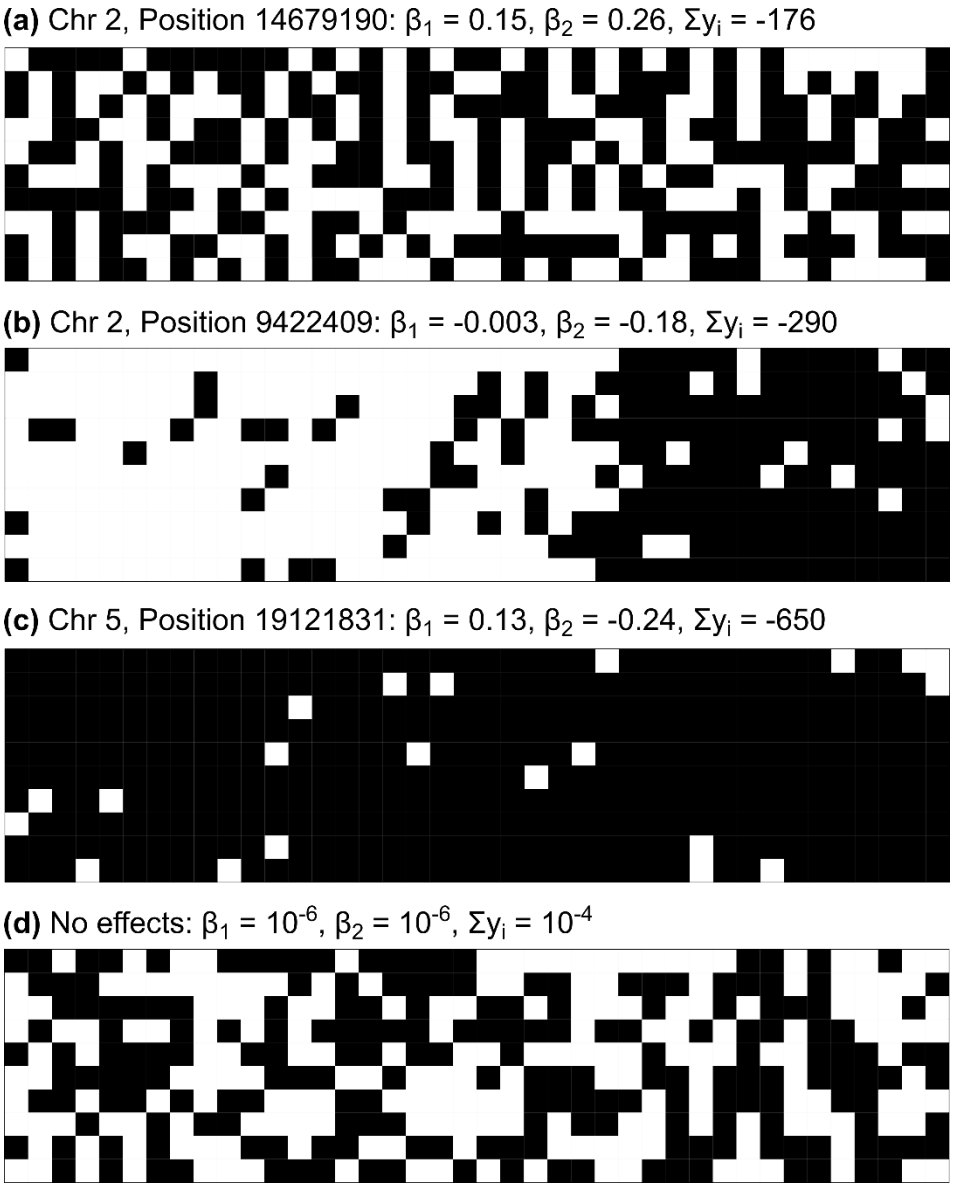


Figure 4. Neighbor GWAS of the leaf damage score on field-grown *Arabidopsis thaliana*. (a) The proportion of leaf damage variation explained by self-genotype effects PVE_{self} (= $\sigma_S^2/(\sigma_S^2+\sigma_N^2+\sigma_e^2)$: blue fraction), neighbor effects PVE_{nei} (= $\sigma_N^2/(\sigma_S^2+\sigma_N^2+\sigma_e^2)$: red fraction), and residuals at the spatial scale of s = 1 and s = 2. Asterisks highlight a significant fraction with likelihood ratio tests: **p-value < 0.01. (b, c) Manhattan plots for the self and neighbor effects on the leaf damage score. Different colors highlight the first to fifth chromosomes of *A. thaliana*. Lighter plots indicate smaller MAF. Results of neighbor effects are shown at s = 1.



728

729 **Figure 5.** Post hoc simulations exemplifying a spatial arrangement of two alleles expected by
730 the estimated self and neighbor effects, β_1 and β_2, on the leaf damage score of *Arabidopsis*
731 *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
732 using 1000 iterations of Gibbs sampling from a random distribution of two alleles in a 10 ×
733 40 space.

734

735