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

Running head: Neighbor GWAS of herbivory

Word count for main text (excluding refences, tables and figures): 6589

ABSTRACT

An increasing number of field studies show that the phenotype of an individual plant depends not only on its genotype but also on 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. 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, the power decreased when hundreds of variants underlay a single trait. We applied the neighbor GWAS to field herbivory data on 199 accessions of *A. 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, the neighbor GWAS highlights the overlooked role of plant neighborhood effects in shaping phenotypic variation, thereby providing a novel and powerful 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 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 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 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; 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 that induce defenses of neighboring plants in a wild tobacco (*Nicotiana attenuata*) (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 (*Phaedon brassicae*) to avoid hairy plants when adult beetles encountered hairy plants at a low frequency (Sato

and Kudoh 2017; Sato et al. 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 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 (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

121 depended on the coefficient β_2 and the number of two alleles in a neighborhood. If the
122 numbers of identical and different alleles were the same near a focal plant, these neighbors
123 offset the sum of the products between the focal plant *i* and all *j* neighbors $\sum_{<i,j>} x_i x_j$ and
124 exerted no effects on a phenotype. When we summed up the phenotype values for the total
125 number of plants *n* and replaced it as $E = -\beta_2,\ H = -\beta_1$ and $\epsilon_I = \sum y_i$, eq. 1 could be
126 transformed as $\epsilon_I = -E \sum_{<i,j>} x_i x_j - H \sum x_i$, which defined the interaction energy of a two-
127 dimensional ferromagnetic Ising model (McCoy and Maillard 2012). The neighbor
128 effect β_2 and self-genotype effect β_1 were interpreted as the energy coefficient E and external
129 magnetic effects H, respectively. An individual plant represented a spin and the two allelic
130 states of each locus corresponded to a north or south dipole. The positive or negative value
131 of $\sum x_i x_j$ indicated a ferromagnetism or paramagnetism, respectively. In this study, we did not
132 consider the effects of allele dominance because this model was applied to inbred *A. thaliana*.
133 However, heterozygotes could be processed if the neighbor covariate $x_i x_j$ was weighted by
134 an estimated degree of dominance in the self-genotypic effects on a phenotype.

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

139 $$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{L} \sum_{<i,j>}^{s} x_i x_{j,s} + u_i + e_i \qquad (2)$$

140 where β_0 indicated the intercept, and the term $\beta_1 x_i$ represented fixed self-genotype effects as
141 tested in conventional GWAS; β_2 was the coefficient of fixed neighbor effects. The neighbor
142 covariate $\sum_{<i,j>}^{s} x_i x_{j,s}$ indicated a sum of products for all combinations between the *i*-th focal
143 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. Variance components due to a sample structure in self and neighbor effects was 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 $e_i \sim \mathrm{Norm}(0, \sigma_e^2)$.

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

$$\boldsymbol{K}_1 = \frac{1}{q-1} \boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{X}_1 \quad \text{and}$$

$$\boldsymbol{K}_2 = \frac{1}{q-1} \boldsymbol{X}_2^{\mathrm{T}} \boldsymbol{X}_2$$

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 $\boldsymbol{K}_1 = \left(\frac{k_{ij}+1}{2}\right)$ and $k_{ij} = [0, 1]$; 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 $\boldsymbol{X}_2 = (\frac{\sum_{<i,j>}^{s} x_i x_{j,s}}{L})$. The individual-level formula eq. 2 could also be converted into a conventional matrix form as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

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 $\frac{\sum_{<i,j>}^{s} x_i x_{j,s}}{L}$, 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 $\mathrm{Var}(\boldsymbol{u}) = \sigma_1^2 \boldsymbol{K}_1 + \sigma_2^2 \boldsymbol{K}_2$; and $\boldsymbol{e}$ was residual as $\mathrm{Var}(\boldsymbol{e}) = \sigma_e^2 \boldsymbol{I}$. In such a mixed model, the proportion of phenotypic variation explained (PVE) by the self and neighbor effects could be calculated as $\mathrm{PVE}_{\mathrm{self}} = \sigma_1^2 / (\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$ and

166 $PVE_{nei} = \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$, respectively. The line of extensions to incorporate neighbor
167 effects into GWAS was hereafter referred to as "neighbor GWAS." Note also that the standard
168 GWAS was a subset of the neighbor GWAS. These two models became equivalent at $s = 0$
169 and $\sigma_2^2 = 0$. When testing the self-genotype effect β_1, we recommended that neighbor effects
170 and its variance component σ_2^2 should be excluded; otherwise, the standard GWAS failed to
171 correct a sample structure because of the additional variance component at $\sigma_2^2 \neq 0$.
172

173 **Power simulation**

174 To examine the power to detect neighbor effects, we applied the neighbor GWAS to
175 simulated phenotypes. Phenotypes were simulated using a part of the real genotypes of *A.*
176 *thaliana*. To evaluate the power of the simple linear model, we assumed a complex ecological
177 form of neighbor effects with multiple variance components controlled. The power was
178 evaluated by the receiver operating characteristic (ROC) curve on the association score of -
179 $\log_{10}$(*p*-value) (e.g., Gage et al. 2018). All analyses were performed using R version 3.6.0 (R
180 Core Team 2019).

181 ***Genotype data.*** To consider a realistic genetic structure in the simulation, we used
182 part of the *A. thaliana* RegMap panel (Horton et al. 2012). The genotype data on 1307
183 accessions were downloaded from the Joy Bergelson laboratory website
184 (http://bergelson.uchicago.edu/?page_id=790 accessed on February 9, 2017). We extracted
185 the chromosome 1 and 2 data with the minor allele frequency (MAF) at >0.1, providing a
186 matrix of 1307 plants with 65 226 single nucleotide polymorphisms (SNPs). Pairwise linkage
187 disequilibrium (LD) among the loci was $r^2 = 0.003$ [0.00-0.06: median with upper and lower
188 95 percentiles]. Before generating a phenotype, each locus was centered by its mean and
189 scaled by its standard deviation. Subsequently, we randomly selected 1296 accessions (= 36 ×

36 accessions) without any replacements for each iteration, and placed them in a 36 × 72 checkered space following *Arabidopsis* experimental settings (see Fig. 2b).

Phenotype simulation. To address ecological issues specific to plant neighborhood effects, we considered two extensions, namely asymmetric neighbor effects and spatial scales. Firstly, previous studies have shown that such plant–plant interactions are sometimes asymmetric between two accessions in herbivory (e.g., Bergvall et al. 2006; Verschut et al. 2016; Sato and Kudoh 2017) and height competition (Weiner 1990). Such asymmetric neighbor effects can be tested by statistical interaction terms in a linear model (Bergvall et al. 2006; Sato and Kudoh 2017). Secondly, several studies have shown that the strength of neighbor effects depends on spatial scales (Hambäck et al. 2014), and that the scale of neighbors to be analyzed relies on the dispersal ability of causative organisms (see Hambäck et al. 2009; Sato and Kudoh 2015; Verschut et al. 2016; Ida et al. 2018 for insect and mammal herbivores; Rieux et al. 2014 for pathogen dispersal) or the size of 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:

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{L}\sum_{<i,j>}^{s} w(s,\alpha) x_i x_{j,s} + \beta_{12}\frac{x_i}{L}\sum_{<i,j>}^{s} w(s,\alpha) x_i x_{j,s} + u_i + e_i \quad (3)$$

where β_{12} was the coefficient for asymmetry in neighbor effects. 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

214 explanatory matrix with $\boldsymbol{X}_{12} = (\frac{x_i}{L}\sum_{<i,j>}^{s} w(s,\alpha)x_i x_{j,s})$, the similarity in asymmetric

215 neighbor effects was calculated as $\boldsymbol{K}_{12} = \frac{1}{q-1}\boldsymbol{X}_{12}^T\boldsymbol{X}_{12}$. To control phenotypic variations, we

216 further partitioned the proportion of phenotypic variation into those explained by major-effect

217 genes and variance components PVE$_{\boldsymbol{\beta}}$ + PVE$_{\boldsymbol{u}}$, major-effect genes alone PVE$_{\boldsymbol{\beta}}$, and residual

218 error PVE$_{\boldsymbol{e}}$, where PVE$_{\boldsymbol{\beta}}$ + PVE$_{\boldsymbol{u}}$ + PVE$_{\boldsymbol{e}}$ = 1. The *optimize* function in R was used to adjust

219 simulated phenotypes to the given amounts of PVE.

220 ***Parameter setting***. Ten phenotypes were simulated for each combination of the

221 distance decay α, the proportion of phenotypic variance explained by major-effect genes

222 PVE$_{\boldsymbol{\beta}}$, variance components PVE$_{\boldsymbol{u}}$, and the relative importance of multiple variance

223 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

224 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

225 scale was fixed at s = 3. The line of simulations was repeated for the number of causal SNPs

226 at 10, 50, or 300 to examine cases of an oligogenic and polygenic control for a trait. The non-

227 zero coefficients for the causal SNPs were randomly sampled from a uniform distribution,

228 Unif(|0.5|, |2.0|), and assigned as some causal SNPs were responsible for both the self and

229 neighbor effects. Of the total number of causal SNPs, 15% had all self, neighbor, and

230 asymmetric neighbor effects (i.e., $\beta_1 \neq 0$ and $\beta_2 \neq 0$ and $\beta_{12} \neq 0$); another 15% had both the

231 self and neighbor effects, but no asymmetry in the neighbor effects ($\beta_1 \neq 0$ and $\beta_2 \neq$

232 0 and $\beta_{12} = 0$); another 35% had self-genotypic effects only ($\beta_1 \neq 0$); and the remaining

233 35% had neighbor effects alone ($\beta_2 \neq 0$). Given its biological significance, we assumed that

234 some loci having neighbor signals possessed asymmetric interactions between neighbors

235 ($\beta_2 \neq 0$ and $\beta_{12} \neq 0$) while the others had symmetric interactions ($\beta_2 \neq 0$ and $\beta_{12} = 0$).

236 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 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, 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). We computed the stepwise likelihood using the *lmm.diago.profile.likelihood* function to first test β_1 in the standard GWAS, and then to test β_2 in the neighbor GWAS. 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 $-\log_{10}(p\text{-value})$. Factors affecting the AUCs were tested by analysis-of-variance (ANOVA). The AUCs had

261 proportional values, and thus were arcsine-square-root transformed to improve the Shapiro–
262 Wilk normality statistic before ANOVA. The AUC and variance components were calculated
263 from s = 1 (the first nearest neighbors) to s = 3 (up to the third nearest neighbors) cases. The
264 AUCs were also calculated using standard linear models without any random effects to
265 examine whether the linear mixed models were superior to the linear models. We also tested
266 the neighbor GWAS model incorporating neighbor phenotype $y_{j,s}$ instead of $x_{j,s}$.
267
268 ***Arabidopsis* herbivory data**
269 We applied the neighbor GWAS to field data of *Arabidopsis* herbivory. The procedure of the
270 present field experiment followed our previous experiment on summer herbivory on field-
271 grown *A. thaliana* (Sato et al. 2019). We selected 199 worldwide accessions from 2029
272 accessions sequenced by the RegMap (Horton et al. 2012) and 1001 Genomes project
273 (Alonso-Blanco et al. 2016). Of the 199 accessions, most were overlapped with a previous
274 GWAS of biotic interactions (Horton et al. 2014) and half were included by a GWAS of
275 glucosinolates (Chan et al. 2010). Eight replicates of each of the 199 accessions were first
276 prepared in a laboratory and then transferred to the outdoor garden at the Center for
277 Ecological Research, Kyoto University, Japan (Otsu, Japan: 35°06′N, 134°56′E, alt. ca. 200
278 m: Fig. 2a). Seeds were sown on Jiffy-seven pots (33-mm diameter), and stratified at a
279 temperature of 4 °C for a week. Seedlings were cultivated for 1.5 months under a short-day
280 condition (8 h light: 16 h dark, 20 °C). Plants were then separately potted in plastic pots (6
281 cm in diameter) filled with mixed soil of agricultural compost (Metro-mix 350, SunGro Co.,
282 USA) and perlite at a 3:1 ratio. Potted plants were set in plastic trays (10 × 40 cells) in a
283 checkered pattern (Fig. 2b). In the field setting, a set of 199 accessions and an additional Col-
284 0 accession were randomly assigned to each block without replacement (Fig. 2b). Eight

285 replicates of these blocks were set >2 m apart from each other (Fig. 2c). Potted plants were
286 exposed to the field environment for 3 wk in June 2017. At the end of experiment, we scored
287 the percentage of foliage eaten as: 0 for no visible damage, 1 for ≤10%, 2 for >10% and ≤
288 25%, 3 for > 25% and ≤ 50%, 4 for >50% and ≤ 75%, and 5 for >75% of the foliage area
289 eaten. All plants were scored by a single person to avoid observer bias. The most
290 predominant herbivore in this field trial was the diamond back moth (*Plutella xylostella*),
291 followed by the small white butterfly (*Pieris rapae*). We also recorded the initial plant size
292 and the presence of inflorescence to incorporate them as covariates. Initial plant size was
293 evaluated by the length of the largest rosette leaf (mm) at the beginning of the field
294 experiment and the presence of inflorescence was recorded 2 wk after transplanting.
295 We estimated the variance components and performed the association tests for the
296 leaf damage score with the neighbor covariate at $s = 1$ and 2. These two scales corresponded
297 to $L = 4$ (the nearest four neighbors) and $L = 12$ (up to the second nearest neighbors),
298 respectively, in the *Arabidopsis* dataset. The variation partitioning and association tests were
299 performed using the *gaston* package, as mentioned above. To determine the significance of
300 variance component parameters, we compared the likelihood between mixed models with one
301 or two random effects. For the genotype data, we used an imputed SNP matrix of all the 2029
302 accessions studied by the RegMap (Horton et al. 2012) and 1001 Genomes project (Alonso-
303 Blanco et al. 2016). Missing genotypes were imputed using BEAGLE (Browning and
304 Browning 2009), as described by Togninalli et al. (2018) and updated on the AraGWAS
305 Catalog (https://aragwas.1001genomes.org). Of the 10 709 466 SNPs from the full imputed
306 matrix, we used 1 242 128 SNPs with MAF at > 0.05 and LD of adjacent SNPs at r^2 <0.8. We
307 considered the initial plant size, presence of inflorescence, experimental blocks, and the edge
308 or center within a block as fixed covariates; these factors explained 12.5% of the leaf damage

309 variation (1.2% by initial plant size, Wald test, Z = 3.53, *p*-value<0.001; 2.4% by the
310 presence of inflorescence, Z = -5.69, *p*-value<10^{-8}; 8.3% by the experimental blocks,
311 likelihood ratio test, χ^2 = 152.8, df = 7, *p*-value<10^{-28}; 0.5% by the edge or center, Z = 3.11,
312 *p*-value = 0.002). After the association mapping, we searched candidate genes within ~10 kb
313 around target SNPs, based on the Araport11 gene model with the latest annotation of The
314 Arabidopsis Information Resource (TAIR) (accessed on 7 September 2019). Gene-set
315 enrichment analysis was performed using the Gowinda algorithm that enables unbiased
316 analysis of GWAS results (Kofler and Schlotterer 2012). We tested SNPS with the top 0.1% -
317 $\log_{10}$(*p*-value) score with the option "--gene-definition undownstream10000," "--min-genes
318 20," and "--mode snp." The GO.db package (Carlson et al. 2018) and the latest TAIR AGI
319 code annotation were used to build input files. The R source codes, accession list, and
320 phenotype data are available at the GitHub repository
321 (https://github.com/naganolab/NeighborGWAS).
322
323 **RESULTS**
324
325 **Power simulation**
326 A set of phenotypes were simulated from real genotype data following a complex model (eq.
327 3), and then fitted by a simplified model (eq. 2). Analyzing the factors affecting AUCs, we
328 found that the number of causal SNPs, the proportion of phenotypic variation (PVE)
329 explained by major-effect genes PVE$_{\boldsymbol{\beta}}$, and distance decay of neighbor effects were the most
330 influential factors affecting the power of the proposed method to detect neighbor signals
331 (Table 1). In addition to PVE$_{\boldsymbol{\beta}}$, the amount of variance components PVE$_{\boldsymbol{u}}$ also significantly
332 affected the AUCs of the self and neighbor effects, but these additional effects were less

333 significant compared to those of PVEβ alone (Table 1). The power of the standard GWAS to
334 detect self-genotype effects depended on the number of causal SNPs and PVEβ, and was not
335 significantly influenced by the distance decay of neighbor effects (Table 1). Neither the
336 power to detect self nor neighbor effects were significantly affected by the ratio of three
337 variance components $\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$ (Table 1).

338 The power to detect self-genotype signals changed from strong (AUC>0.9) to weak
339 (AUC<0.6) as PVEβ decreased and the number of causal SNPs increased (Fig. 3a; Fig. S1a).
340 Compared to the self-genotype effects, it was relatively difficult to detect neighbor effects
341 (Fig. 3b; Fig. S1b). In the case of the number of causal SNPs = 10, the power to detect
342 neighbor signals decreased from high (AUC>0.9) to moderate (AUC>0.7) with the
343 decreasing PVEβ and distance decay (Fig. 3b; Fig. S1b). The power to detect neighbor effects
344 was very weak (AUC> 0.5 and AUC<0.6) in the case of the number of causal SNPs = 300
345 (Fig. 3b; Fig. S1b). The result of simulations indicated that neighbor effects were detectable
346 when a target trait was governed by several major genes and the range of neighbor effects
347 was spatially limited. Additionally, linear mixed models outperformed standard linear models
348 as shown by 9.3% and 1.7% increases in the power to detect self and neighbor signals,
349 respectively (AUC_{self}, paired *t*-test, mean of the difference = 0.093, *p*-value<10^{-16}; AUC_{nei},
350 mean of the difference = 0.017, *p*-value<10^{-16}). When the neighbor phenotype $y_{j,s}$ was
351 incorporated instead of the genotype $x_{j,s}$, the power to detect neighbor effects was very weak,
352 such that AUC decreased to near 0.5.

353 Notably, there was a clear relationship between the distance decay α and the
354 proportion of phenotypic variation explained by neighbor effects PVE_{nei} or AUCs at different
355 spatial scales (Fig. 4). If the distance decay was weak and the effective range of neighbor
356 effects was broad, PVE_{nei} and AUCs increased linearly as the reference spatial scale was

broadened (Fig. 4a). On the other hand, if the distance decay was strong and the effective scale of neighbor effects was narrow, PVE_{nei} saturated at the scale of the first nearest neighbors (Fig. 4c) and AUCs did not increase (Fig. 4b, c). The simulation results indicated that the effective spatial scales could be estimated by calculating PVE_{nei} across different spatial scales, 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 the PVE estimates (Fig. S2; Table S2). Comparing the true and estimated PVE, we found that the estimated PVE by self and neighbor effects was biased as a consequence of the ratio of variance components $\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$, the total amount of PVE, and the magnitude of distance decay (Table S2). The standard heritability namely, PVE_{self} at $s = 0$, was overestimated when neighbor effects involved a simulated phenotype. The bias of estimated PVE_{self} decreased as the reference spatial scale increased (Fig. S2). In contrast, PVE_{nei} was underestimated and its bias decreased as the reference scale increased (Fig. S2). These results indicated that neighbor effects and its variance component should be tested by comparing the standard GWAS and neighbor GWAS, because some fraction of PVE was mutually non-exclusive between the two factors.

***Arabidopsis* herbivory data**

The variation partitioning of leaf damage showed that the PVE by neighbor effects was larger than PVE by self-genotypic effects ($PVE_{self} = 0.041$, $\chi_1^2 = 0.36$, p-value = 0.54; $PVE_{nei} = 0.214$, $\chi_1^2 = 7.23$, p-value = 0.004: Fig. 5). Phenotypic variation explained by neighbor effects on leaf damage did not increase when the neighbor scale was referred up to the nearest and second nearest individuals ($PVE_{self} = 0.091$, $\chi_1^2 = 1.19$, p-value = 0.200; $PVE_{nei} = 0.14$, $\chi_1^2 =$

381 1.41, p-value = 0.166: Fig. 5); therefore, the variation partitioning was stopped at s = 2.
382 Heritability, namely PVE_{self} at s = 0, was 0.173 and significant (χ_1^2 = 10.1, p-value = 0.005:
383 Fig. 5). This range of heritability was overlapped with PVE by neighbor effects alone (PVE_{nei}
384 without self-effects = 0.248 and 0.243 at s = 1 and 2), indicating that some fraction of PVE_{self}
385 and PVE_{nei} was mutually non-exclusive. We additionally incorporated the neighbor
386 phenotype $y_{j,s}$ instead of the neighbor genotype $x_{j,s}$ in eq. 2, but PVE_{nei} did not increase
387 (PVE_{nei} = 0.066 and 0.068 at s = 1 and 2, respectively). These results indicated a narrow
388 effective scale and significant contribution of neighbor effects to the leaf damage score,
389 where the neighbor genotype rather than the neighbor phenotype was likely a source of
390 phenotypic variation.
391 Standard GWAS of the self-genotype effects on the leaf damage detected SNPs with
392 the second and third largest -$\log_{10}$(p-values) score on the first chromosome (chr1), though
393 they were not above a threshold of Bonferroni correction (Fig. 6a; Table S3). The second SNP
394 at chr1-21694386 was located within ~10 kb of three loci encoding a disease resistance
395 protein (CC-NBS-LRR class) family. The third SNP at chr1-23149476 was located within
396 ~10 kb of the AT1G62540 locus that encoded flavin-monooxygenase glucosinolate S-
397 oxygenase 2 (FMO GS-OX2), This gene's GO annotations, "glucosinolate biosynthetic
398 process" and "oxidation-reduction process," were not significantly enriched among genes
399 within ~10 kb around SNPs with the top 0.1% -$\log_{10}$(p-values) score. In contrast, the
400 "sesquiterpene metabolic process" was significantly enriched with four unique genes (false
401 discovery rate = 0.003; Table S4). A QQ-plot did not exhibit an inflation of p-values for the
402 self-genotype effects (Fig. S3).
403 Regarding the neighbor effects on the leaf damage, we found non-significant but
404 weak peaks on the second and third chromosomes (Fig. 6b; Table S3). The second

405 chromosomal region had higher association scores than predicted by the QQ-plot (Fig. S4). A
406 locus encoding FAD-binding Berberine family protein (AT2G34810 named *BBE16*) was
407 located within the ~10 kb window near SNPs with the top eleven -$\log_{10}$(*p*-values) score at the
408 second chromosome, which are known to be up-regulated by methyl jasmonate (Devoto et al.
409 2005). Three transposable elements and a pseudogene of lysyl-tRNA synthetase 1 were
410 located near the most significant SNP on the third chromosome. Gene ontology categories
411 involved in wounding responses were not significantly enriched among genes within ~10 kb
412 around SNPs with the top 0.1% score of -$\log_{10}$(*p*-values); however, "glucosinolate catabolic
413 process" was significantly enriched with seven unique genes (false discovery rate = 0.014:
414 Table S4). We additionally tested asymmetric neighbor effects β_{12} in the real dataset on field
415 herbivory, but the top 0.1% SNPs of neighbor effects β_2 did not overlap with those of
416 asymmetric neighbor effects β_{12} (Table S3).

417 Based on the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$, we ran a post hoc simulation to infer a
418 spatial arrangement that minimizes a population sum of the leaf damage $\sum y_i = \beta_1 \sum x_i +$
419 $\beta_2 \sum_{<i,j>} x_i x_j$. The constant intercept β_0, the variance component u_i, and residual e_i were not
420 considered because they were not involved in deterministic dynamics of the model. Figure 7
421 shows three representatives and a neutral expectation. For example, a mixture of a
422 dimorphism was expected to decrease the total leaf damage for a SNP at chr2-14679190 near
423 the *BBE16* locus ($\hat{\beta}_2 > 0$: Fig. 7a). On the other hand, a clustered distribution of a
424 dimorphism was expected to decrease the total damage for a SNP at chr2-9422409 near the
425 AT2G22170 locus encoding a lipase/lipooxygenase PLAT/LH2 family protein ($\hat{\beta}_1 \approx 0, \hat{\beta}_2 <$
426 0: Fig. 7b). Furthermore, near monomorphism was expected to decrease the leaf damage for a
427 SNP at chr5-19121831 near the AT5G47075 and AT5G47077 loci encoding low-molecular
428 cysteine-rich proteins, LCR20 and LCR6 ($\hat{\beta}_1 > 0, \hat{\beta}_2 < 0$: Fig. 7c). If self and neighbor

429 coefficients had no effects, we observed a random distribution and no mitigation of damage
430 i.e., $\sum y_i \approx 0$ (Fig. 7d). These post hoc simulations suggested a potential applicability of
431 neighbor GWAS in optimizing spatial arrangements in field cultivation.
432
433 **DISCUSSION**
434
435 **Spatial and genetic factors underlying simulated phenotypes**
436 Benchmark tests using simulated phenotypes revealed that appropriate spatial scales could be
437 estimated by variation partitioning of observed phenotypes. When the scale of neighbor
438 effects was narrow or moderate (α = 1.0 or 3.0), the scale of the first nearest neighbors would
439 be optimum for the power to detect neighbor signals. In terms of neighbor effects in the
440 context of plant defense, mobile animals (e.g., mammalian browsers and flying insects) often
441 select a cluster of plant individuals (e.g., Bergvall et al. 2006; Hambäck et al. 2009; Vershut
442 et al. 2016); however, previous studies have not detected neighbor effects among individual
443 plants within a cluster (Hambäck et al. 2014; Sato and Kudoh 2015). The exponential
444 distance decay of α = 0.25 represented situations in which the effective range of neighbor
445 effects was too broad to be detected; only in such situations should more than the nearest
446 neighbors be referred to gain the power to detect neighbor effects.
447 Neighbor GWAS could retain its power as long as neighbor effects were spatially
448 limited, and several major-effect genes governed a trait. In contrast, when hundreds of causal
449 variants involved a single trait and less than half of phenotypic variation was attributable to
450 neighbor effects, we observed a decrease in the power to detect neighbor signals. In GWAS,
451 false positive rates can be reduced using linear mixed models that deal with kinship structure
452 as a random effect (Korte and Farlow 2013). Indeed, mixed models were superior to standard

453 linear models in this simulation. Our simulation also adjusted the three variance
454 components σ_1^2, σ_2^2, and σ_{12}^2, but their relative contribution did not have significant effects on
455 the power to detect neighbor signals. In addition, the power to detect self-genotype signals
456 largely depended on the number of causal SNPs and PVE by major-effect genes. This result
457 suggests that the power for the standard GWAS to detect the self-genotype effects was
458 independent from neighbor effects, though its heritability may confound PVE by neighbor
459 effects.
460
461 **Neighbor GWAS of the field herbivory on *Arabidopsis***
462 Our genetic analysis of neighbor effects is of ecological interest because how host plant
463 genotypes shape variation in plant–herbivore interactions is a long-standing question in
464 population ecology (e.g., Karban 1992; Underwood and Rausher 2000; Utsumi et al. 2011).
465 Despite the low heritability and several confounding factors under field conditions, the
466 present study illustrated the significant contribution of neighbor genotypic identity to spatial
467 variation in herbivory on *A. thaliana*. As we did not find any significant peaks above the
468 threshold of genome-wide multiple testing, such neighbor effects on herbivory could be
469 under polygenic control. Glucosinolates and sesquiterpene-mediated defense might jointly
470 contribute to variation in herbivory among individual plants, as suggested by a gene-set
471 enrichment analysis.
472 Regarding the self-genotype effects, we detected *GS-OX2*, a key gene involved in
473 glucosinolate biosynthesis, near the third top-scoring SNP on the first chromosome. Aliphatic
474 glucosinolates are well known as a chemical defense against insect herbivory (van Poecke
475 2007; Brachi et al. 2015), where GS-OX2 catalyzes the conversion from methylthioalkyl to
476 corresponding methylsulfinylalkyl glucosinolates (Li et al. 2008). Specifically, Geiselhardt et

477 al. (2013) reported that *GS-OX2* was up-regulated in response to feeding by larvae of the
478 large white butterfly (*Pieris brassicae*) and such an inducible defense was suppressed by
479 prior oviposition by adult butterflies. The present GWAS supported the relevance of *GS-OX2*
480 to herbivory by *Pieris* butterflies, because we observed larvae of *P. rapae* largely feeding on
481 *A. thaliana*. Furthermore, our experiment was conducted under the field condition where
482 insect movements and microclimates were not controlled. Thus, the present results provide
483 ecological insights into the gene function of *GS-OX2* from the laboratory to the field
484 environment.

485 Remarkably, the second top-scoring SNP of the neighbor effects was located near the
486 *BBE16* locus responsive to methyl jasmonate. Methyl jasmonate is well known as a volatile
487 organic chemical that is emitted from damaged tissue and elicits defense response of other
488 plants (Reymond and Farmer 1998; van Poecke 2007). Volatile-mediated plant–plant
489 communication is a key driver of neighbor effects in herbivory (e.g., Schuman et al. 2015;
490 Erb 2018; Dahlin et al 2018), while a molecular function of *BBE16* remains unexplored.
491 More recent studies have shown that berberine bridge enzyme-like proteins are involved in
492 damage-associated molecular patterns (Benedetti et al. 2018), which might be a subject of
493 future studies on inducible plant defense and neighbor effects.
494

495 **Potential applicability and limitation**

496 GWAS has often been applied to crop plants (Jannink et al. 2010; Hamblin et al. 2011), where
497 genotypes are known and individuals are evenly transplanted in space. Considering this
498 outlook for agriculture, we provided an example of neighbor GWAS across a lattice space in
499 this study. However, wild plant populations sometimes exhibit more complex spatial patterns
500 than expected by the Ising model (e.g., Kizaki and Katori 1999; Schlicht and Iwasa 2004). A

variation partitioning approach could also be applied to uneven distribution if the neighbor GWAS method will be elaborated to handle distance decay across a complex spatial structure. Now that a vast number of plant genomes are being sequenced (Exposito-Alonso et al. 2020), further studies on spatial statistics are encouraged to enable the neighbor GWAS in wild plants.

We should also note that some fraction of phenotypic variation cannot be separated into self or neighbor effects, as shown by the shift in PVE from $s = 0$ to $s = 1$ in the simulated and real data. Such an overlapped fraction might arise because of a strong correlation between $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ (e.g., $r = 0.60$ and 0.78 at $s = 1$ and 2 in the real data analysis). In eq. 2, the self-genotypic variable x_i was incorporated into the neighbor variable $\sum x_i x_j / L$, and thus the kinship matrix $\boldsymbol{K}_1$ was partially redundant with the similarity matrix of neighbor effects $\boldsymbol{K}_2$. For a proper test of neighbor effects β_2 and its variance component σ_2^2, it is necessary to compare the likelihood ratio between the standard and neighbor GWAS. Once neighbor effects are detected by the stepwise likelihood ratio test, the heritability and PVE should be carefully interpreted regarding the potential bias along the reference spatial scale s.

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 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.

549 **CONFLICT OF INTEREST**

550 The authors declare no conflict of interest.

551

552 **DATA ARCHIVING**

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

554 The R source codes used in this study are available at the GitHub repository

555 (https://github.com/naganolab/NeighborGWAS).

556

557 **REFERENCES**

558 Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KM et al. (2016)

559 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**:

560 481–491. doi:10.1016/j.cell.2016.05.063

561 Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y et al. (2010) Genome-

562 wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**:

563 627–631. doi:10.1038/nature08800

564 Azaele S, Muneepeerakul R, Rinaldo A, Rodriguez-Iturbe I (2010) Inferring plant ecosystem

565 organization from species occurrences. *J Theor Biol* **262**: 323–329.

566 doi:10.1016/j.jtbi.2009.09.026

567 Barbosa P, Hines J, Kaplan I, Martinson H, Szczepaniec A, Szendrei Z (2009) Associational

568 resistance and associational susceptibility: Having right or wrong neighbors. *Ann Rev Ecol*

569 *Evol Sys* **40**: 1–20. doi:10.1146/annurev.ecolsys.110308.120242

570 Benedetti M, Verrascina I, Pontiggia D, Locci F, Mattei B, De Lorenzo G, Cervone F (2018)

571 Four *Arabidopsis* berberine bridge enzyme-like proteins are specific oxidases that inactivate

572 the elicitor-active oligogalacturonides. *Plant J* **94**: 260–273. doi:10.1111/tpj.13852

573 Bergvall UA, Rautio P, Kesti K, Tuomi J, Leimar O (2006) Associational effects of plant
574 defences in relation to within- and between-patch food choice by a mammalian herbivore:
575 neighbour contrast susceptibility and defence. *Oecologia* **147**: 253–260. doi:10.1007/s00442-
576 005-0260-8
577 Brachi B, Meyer CG, Villoutreix R, Platt A, Morton TC, Roux F, Bergelson J (2015)
578 Coselected genes determine adaptive variation in herbivore resistance throughout the native
579 range of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **112**: 4032–4037.
580 doi:10.1073/pnas.1421416112
581 Browning BL, Browning SR (2009) A unified approach to genotype imputation and
582 haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum*
583 *Genet* **84**: 210-223. doi:10.1016/j.ajhg.2009.01.005
584 Carlson M (2018) GO.db: A set of annotation maps describing the entire Gene Ontology. R
585 package version 3.7.0. doi:10.18129/B9.bioc.GO.db
586 Carrasco LR, Harwood TD, Toepfer S, MacLeod A, Levay N, Kiss J et al. (2010) Dispersal
587 kernels of the invasive alien western corn rootworm and the effectiveness of buffer zones in
588 eradication programmes in Europe. *Ann Appl Biol* **156**: 63–77. doi:10.1111/j.1744-
589 7348.2009.00363.x
590 Chan EKF, Rowe HC, Kliebenstein DJ (2010) Understanding the evolution of defense
591 metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* **185**:
592 991–1007. doi:10.1534/genetics.109.108522
593 Dahlin I, Rubene D, Glinwood R, Ninkovic V (2018) Pest suppression in cultivar mixtures is
594 influenced by neighbor-specific plant-plant communication. *Ecol Appl* **28**: 2187–2196.
595 doi:10.1002/eap.1807
596 Devaux C, Lavigne C, Austerlitz F, Klein EK (2007) Modelling and estimating pollen

movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. *Mol Ecol* **16**: 487–499. doi:10.1111/j.1365-294X.2006.03155.x

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

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

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

Exposito-Alonso M, Drost H, Burbano HA, Weigel D (2020) The Earth BioGenome project: opportunities and challenges for plant genomics and conservation. *Plant J* doi:10.1111/tpj.14631

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. 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

645 chemicals of neighboring plants limit visits of herbivorous insects: Associational resistance
646 within a plant population. *Ecol Evol* **8**: 12981–12990. doi:10.1002/ece3.4750
647 Ising E (1925) Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik* **31**: 253–258.
648 Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to
649 practice. *Brief Funct Genomic* **9**: 166–177. doi:10.1093/bfgp/elq001
650 Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008)
651 Efficient control of population structure in model organism association mapping. *Genetics*
652 **178**: 1709–1723. doi:10.1534/genetics.107.080101
653 Karban R (1992) Plant variation: its effects on populations of herbivorous insects. In: Fritz
654 RS, Simms EL (eds) *Plant resistance to herbivores and pathogens: ecology, evolution, and*
655 *genetics,* University of Chicago Press: Chicago, pp. 195–215.
656 Kizaki S, Katori M (1999) Analysis of canopy-gap structures of forests by Ising-Gibbs states-
657 equilibrium and scaling property of real forests. *J Phys Soc Japan* **68**: 2553-2560.
658 doi:10.1143/JPSJ.68.2553
659 Kofler R, Schlotterer C (2012) Gowinda: unbiased analysis of gene set enrichment for
660 genome-wide association studies. *Bioinformatics* **28**: 2084–2085.
661 doi:10.1093/bioinformatics/bts315
662 Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a
663 review. *Plant Methods* **9**: 29. doi:10.1186/1746-4811-9-29
664 Li J, Hansen BG, Ober JA, Kliebenstein DJ, Halkier BA (2008) Subclade of flavin-
665 monooxygenases involved in aliphatic glucosinolate biosynthesis. *Plant Physiol* **148**: 1721–
666 1733. doi:10.1104/pp.108.125757
667 McCoy BM, Maillard JM (2012) The importance of the Ising model. Progress *Theor Phys*
668 **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

741 tomato. *Heredity* **118**: 202–209. doi:10.1038/hdy.2016.84

742 Zeller SL, Kalinina O, Flynn DFB, Schmid B (2012) Mixtures of genetically modified wheat

743 lines outperform monocultures. *Ecol Appl* **22**: 1817–1826. doi:10.1890/11-0876.1

744

745

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. For the neighbor effects, the response variable was the maximum Area Under the ROC Curve (AUC) of the spatial scales from $s = 1$ to $s = 3$. AUCs were 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	30.2	2172.5	< 2e-16
	$PVE_{\boldsymbol{\beta}}$	1	2.24	160.7	< 2e-16
	$PVE_{\boldsymbol{\beta}} + PVE_{\boldsymbol{u}}$	1	0.08	5.40	0.020
	$\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$	2	0.00	0.14	0.869
	α	1	0.02	1.71	0.192
	Residuals	1073	14.93	NA	NA
AUC_{nei}	No. of causal SNPs	1	24.0	1343.4	< 2e-16
	$PVE_{\boldsymbol{\beta}}$	1	1.88	105.4	< 2e-16
	$PVE_{\boldsymbol{\beta}} + PVE_{\boldsymbol{u}}$	1	0.11	6.10	0.014
	$\sigma_1^2 : \sigma_2^2 : \sigma_{12}^2$	2	0.00	0.12	0.886
	α	1	1.46	81.8	< 2e-16
	Residuals	1073	19.20	NA	NA

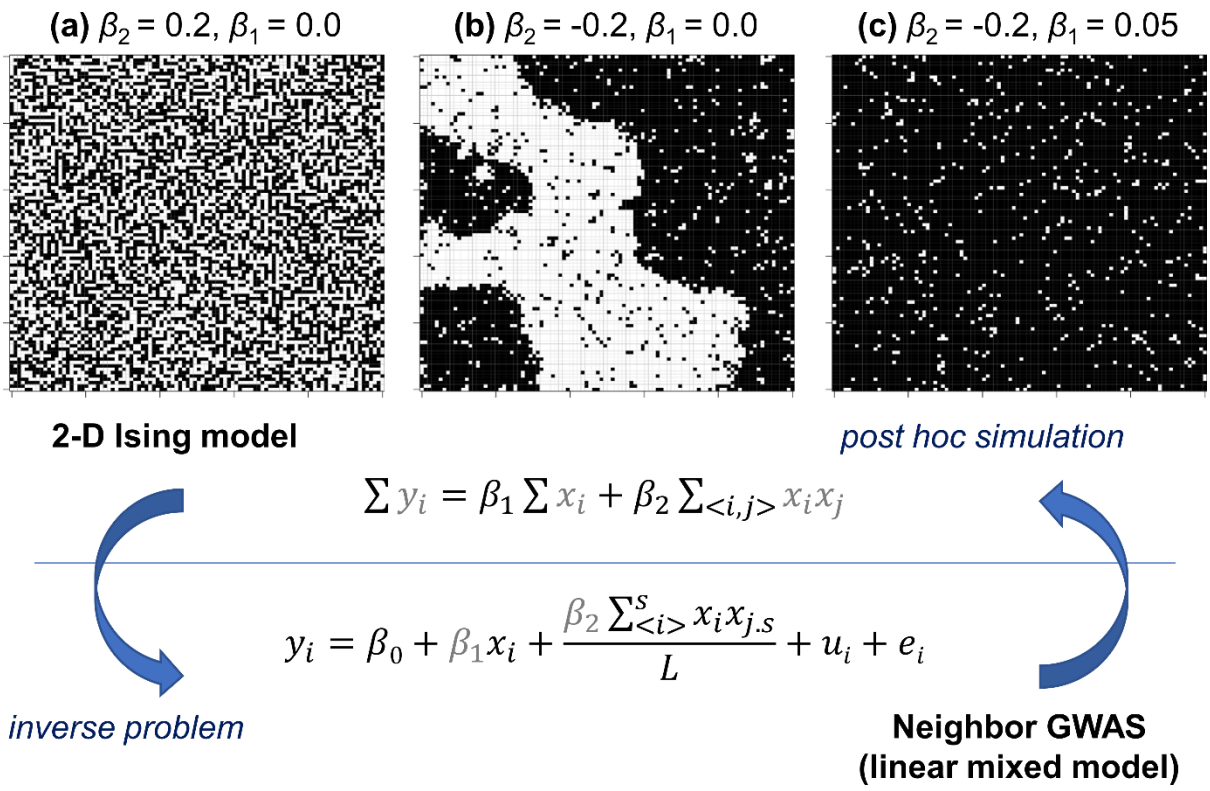

Figure 1. Relationship between Neighbor GWAS and Ising model. Upper panels show spatial arrangements expected by a 2-D Ising model $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{<i,j>} x_i x_j$. (a) If β_2>0, mixed patterns give the argument of the minimum for a population sum of phenotype values $\sum y_i$. (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 $\sum y_i$.

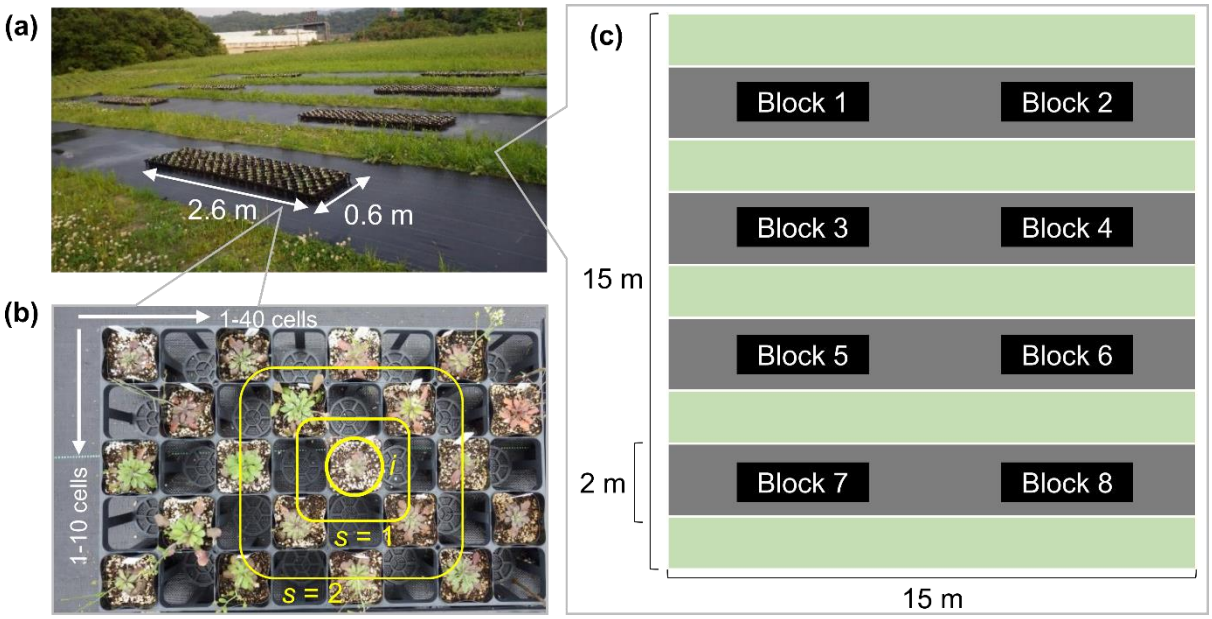

Figure 2. Experimental setting in the *Arabidopsis* herbivory data. (a) Photograph of the field site. Each 0.6 × 2.6 m block included a replicate of 199 accessions and an additional Col-0. (b) *Arabidopsis thaliana* plants were arranged in a 10 × 40 cell plastic tray. The plants were placed in a checkered 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).

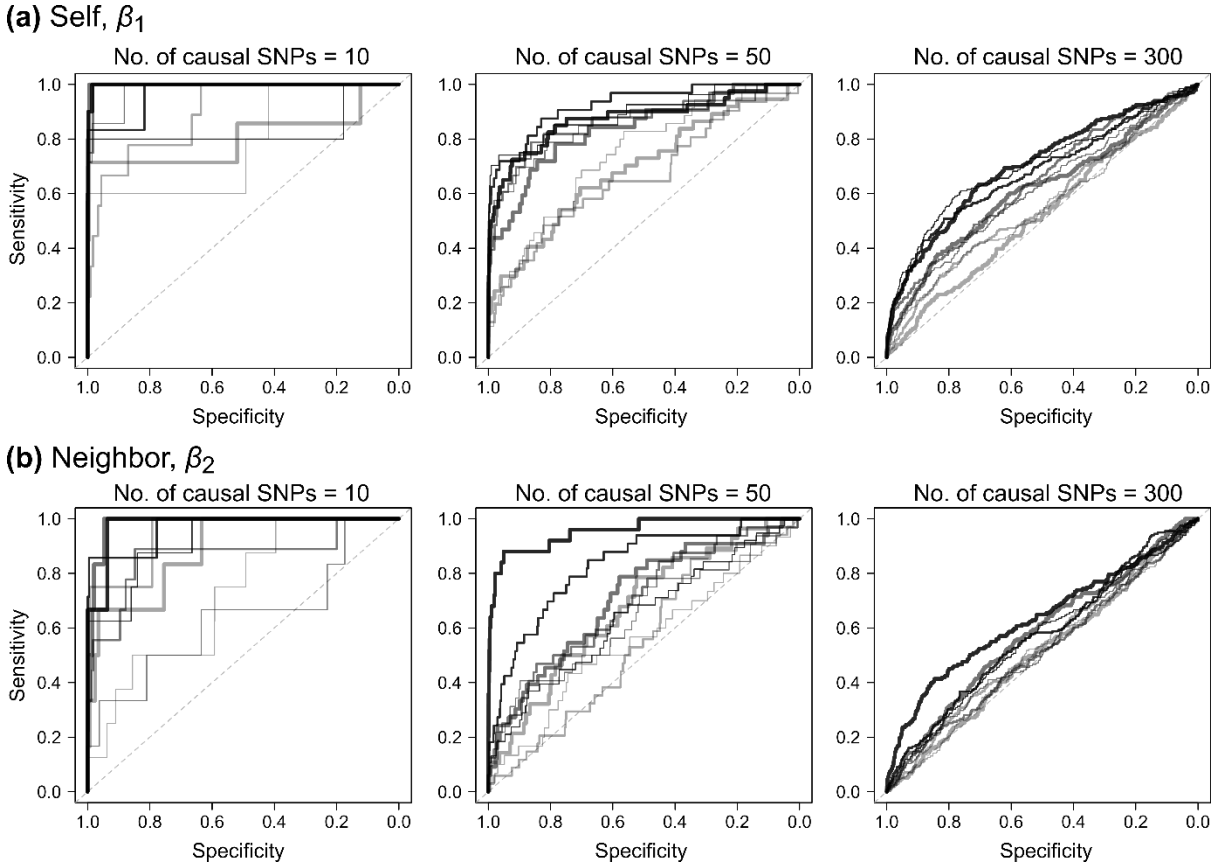

Figure 3. Receiver operating characteristic (ROC) curves representing the sensitivity (true positive rate) and the specificity (1 - false positive rate) in simulated phenotypes. The upper and lower rows give the results of self and neighbor effects with the number of causal SNPs at 10, 50, and 300. The shade of the lines indicates the degree of PVE$_{\boldsymbol{\beta}}$: PVE$_{\boldsymbol{\beta}}$ = 0.1 (light gray), 0.3 (dark gray), 0.6 (black). The thickness of the lines indicates distance decay: α = 0.25 (thin), 1 (medium), 3 (thick).

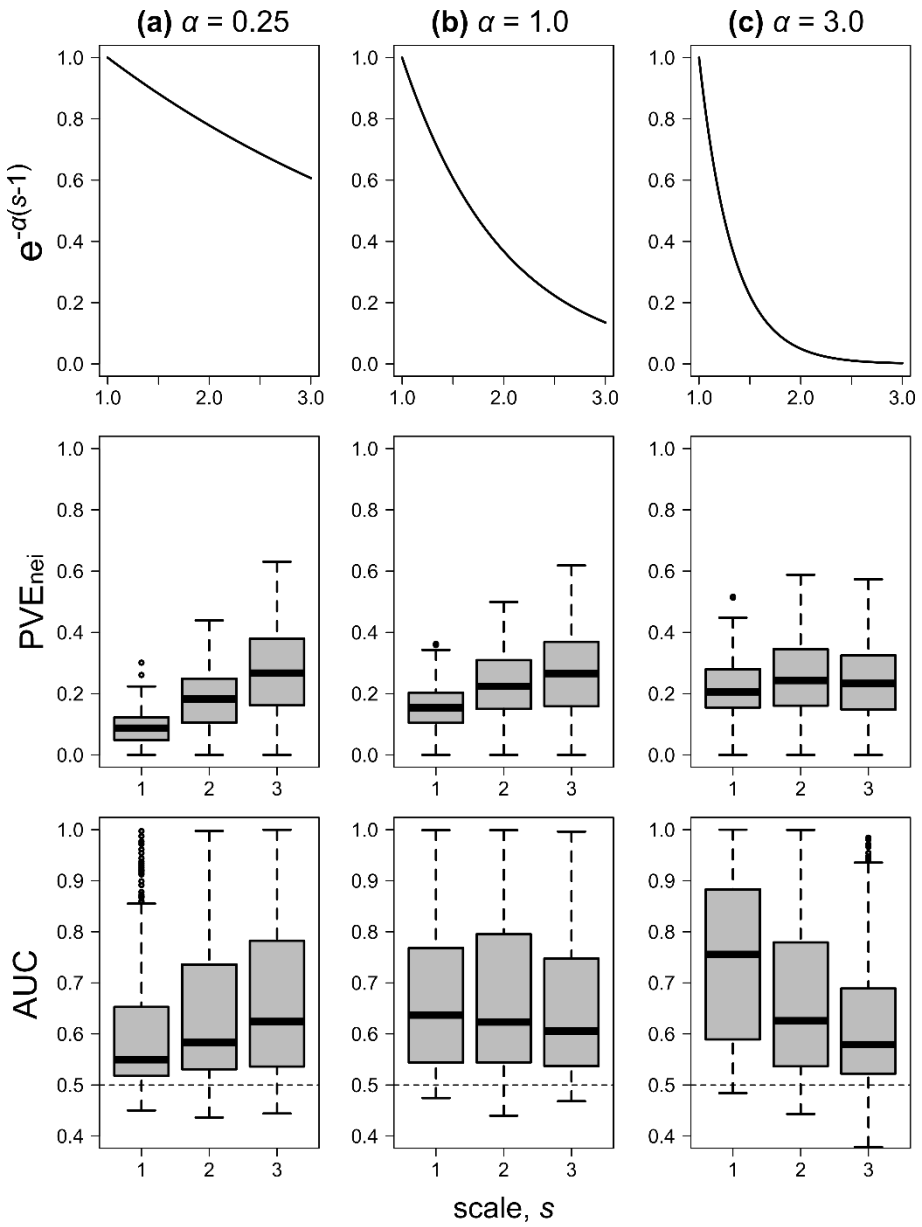

Figure 4. Scale dependence of neighbor effects on simulated phenotypes among all iterations. The broad (a), intermediate (b), and narrow (c) effective range of neighbor effects are represented by weak, moderate, and strong distance decay, respectively. The proportion of phenotypic variation explained by neighbor effects (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.

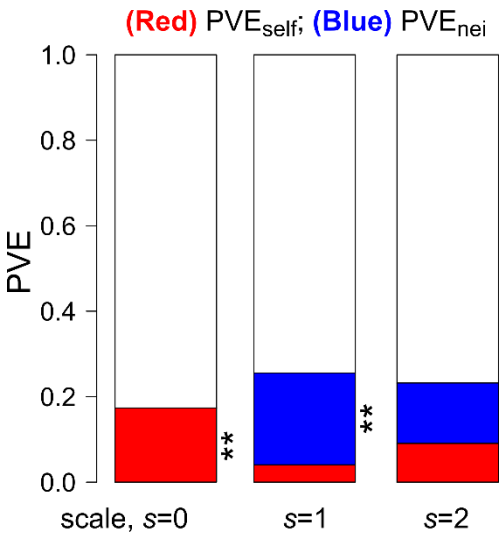

Figure 5. Proportion of phenotypic variation explained (PVE) by self-genotype or neighbor effects on the leaf damage score of field-grown *Arabidopsis thaliana*. The PVEs at the spatial scale of $s = 0$ (without any neighbor effects), $s = 1$ and $s = 2$ are given with different colors. The red fraction indicates PVE by self effects as $\text{PVE}_{\text{self}} = \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$, while the blue fraction indicates PVE by neighbor effects as $\text{PVE}_{\text{nei}} (= \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$. The white fraction indicates residuals. Asterisks highlight a significant fraction with stepwise likelihood ratio tests from simpler to complex models: $^{**}p$-value < 0.01.

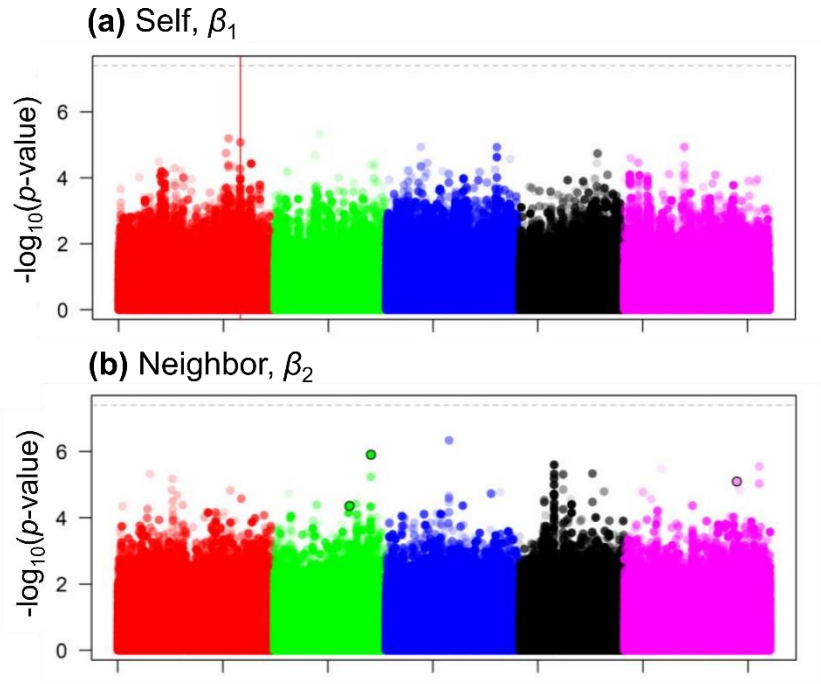

Figure 6. Manhattan plots for the self (a) and neighbor (b) effects on the leaf damage score of field-grown *Arabidopsis thaliana*. The first to fifth chromosomes of *A. thaliana* are differently colored, where lighter plots indicate smaller MAF. Horizontal dashed lines indicate the threshold after Bonferroni 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.

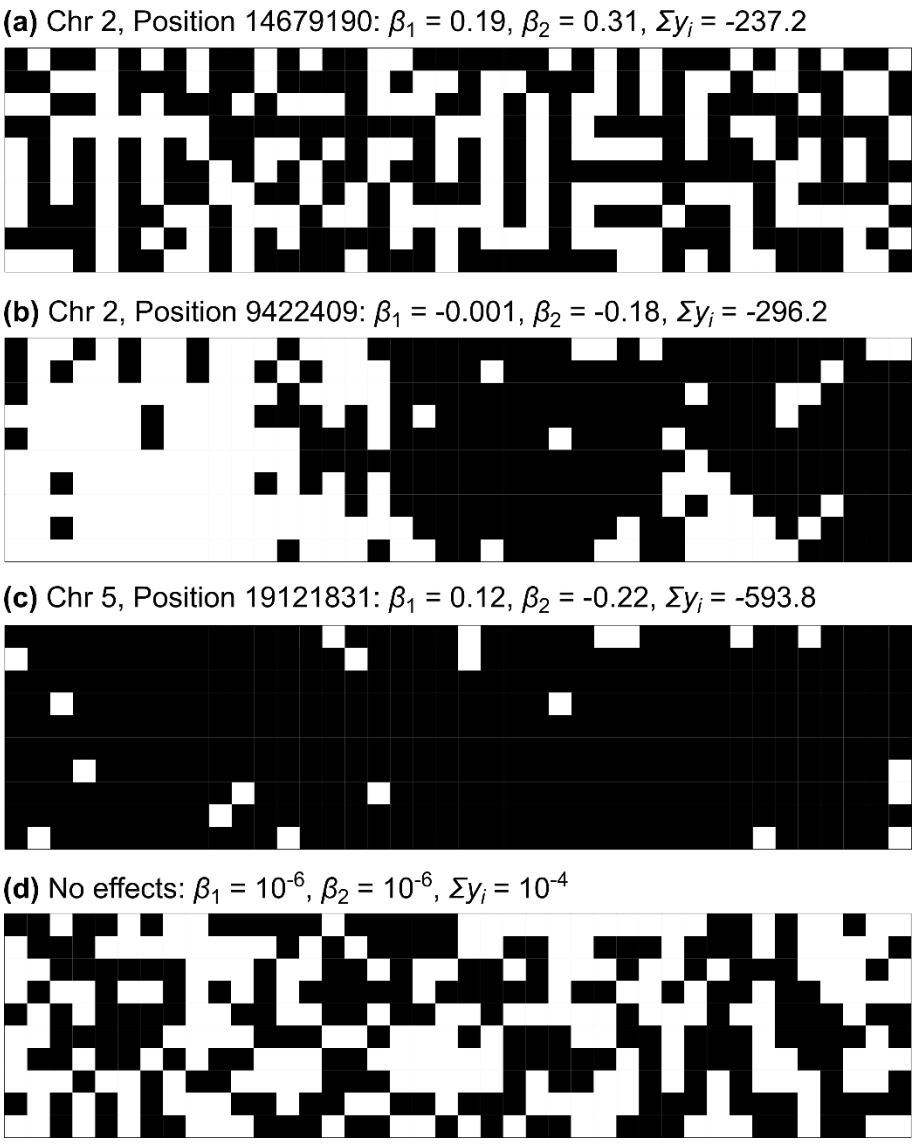

Figure 7. Post hoc simulations exemplifying a spatial arrangement of two alleles expected by the estimated self and neighbor effects, $\hat{\beta}_1$ and $\hat{\beta}_2$, on the leaf damage score of *Arabidopsis thaliana*. Population sum of the leaf damage $\sum y_i = \beta_1 \sum x_i + \beta_2 \sum_{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