Neighbor QTL: an interval mapping method for quantitative trait loci underlying plant neighborhood effects

Yasuhiro Sato*[1], Kazuya Takeda†[2], and Atsushi J. Nagano‡[3]

[1]PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan; Research Institute for Food and Agriculture, Ryukoku University, Yokotani 1-5, Seta Oe-cho, Otsu, Shiga 520-2194, Japan

[2]Center for Ecological Research, Kyoto University, Hirano 2-509-3, Otsu, Shiga 520-2113, Japan

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

September 19, 2020

*Co-correspondence: sato.yasuhiro.36c@kyoto-u.jp

†Y. Sato and K. Takeda equally contributed to the study.

‡Co-correspondence: anagano@agr.ryukoku.ac.jp

Abstract

Phenotypes of sessile organisms, such as plants, rely not only on their own genotype but also on the genotypes of neighboring individuals. Previously, we incorporated such neighbor effects into a single-marker regression using the Ising model of ferromagnetism. However, little is known about how to incorporate neighbor effects in quantitative trait locus (QTL) mapping. In this study, we propose a new method for interval QTL mapping of neighbor effects, named "Neighbor QTL". The algorithm of neighbor QTL involves the following: (i) obtaining conditional self-genotype probabilities with recombination fraction between flanking markers, (ii) calculating neighbor genotypic identity using the self-genotype probabilities, and (iii) estimating additive and dominance deviation for neighbor effects. Our simulation using F2 and backcross lines showed that the power to detect neighbor effects increased as the effective range became smaller. The neighbor QTL was applied to insect herbivory on Col × Kas recombinant inbred lines of *Arabidopsis thaliana*. Consistent with previous evidence, the pilot experiment detected a self QTL effect on the herbivory at *GLABRA1* locus. We also observed a weak QTL on chromosome 4 regarding neighbor effects on the herbivory. The neighbor QTL method is available as an R package (https://cran.r-project.org/package=rNeighborQTL), providing a novel tool to investigate neighbor effects in QTL studies.

1 Introduction

Sessile organisms, such as land plants, have no active mobility to escape neighboring individuals. Field studies have shown that the phenotypes of an individual plant depend not only on their own genotype but also on those of neighboring plants (Barbosa et al., 2009). Such neighbor effects are mediated by direct interactions (e.g., competition and volatile communication) or indirect interactions (e.g., herbivore and pollinator movements), and thus modulate complex traits throughout a plant life cycle, including growth (Subrahmaniam et al., 2018), defense (Schuman et al., 2015; Sato, 2018; Tamura et al., 2020), and reproduction (Underwood et al., 2020). There is increasing appreciation that plant-plant interactions within a species may result in increased yield and population-wide pest resistance (Zeller et al., 2012; Wuest and Niklaus, 2018; Yang et al., 2019). However, knowledge remains limited about how to analyze the quantitative trait locus (QTL) underlying plant neighborhood effects.

QTL mapping is a well-established approach to analyze the loci responsible for complex traits (Broman et al., 2003; Broman and Sen, 2009; Broman et al., 2019). Although genome-wide association studies (GWAS) have now been developed, there are several limitations of this approach such as false positive signals due to the population structure (Hayes, 2013) and small-effect variants being overlooked if they are rare in the sample population (Korte and Farlow, 2013). While recombination events are limited in experimental crosses, the experimental approaches would overcome the problem of population structure and rare variants. In plant genetics, once GWAS leads us to find a pair of target accessions, its biparental population is then subject to QTL mapping (Sonah et al., 2015; Han et al., 2018). Therefore, QTL mapping provides a complementary analysis for GWAS to further dissect complex traits in plant genetics and breeding (Sonah et al., 2015; Rishmawi et al., 2017; Han et al., 2018; Marchadier et al., 2019).

Using the Ising model of statistical physics, our previous study proposed "Neighbor GWAS" that combined neighbor effects and a linear mixed model (Sato et al., 2019b). The core idea of neighbor GWAS was to consider the Ising model as an inverse problem of single-marker regression and, thereby, estimate the effects of neighbor genotypic identity on a trait. However, QTL mapping of neighbor effects is more complicated than single-marker analysis because QTL studies employ the maximum likelihood method for interval mapping between flanking markers (Haley and Knott, 1992; Jansen, 1993; Broman and Sen, 2009). Such an interval mapping requires a stepwise inference from genotype imputation to phenotype prediction. First, conditional genotype probabilities are obtained from the observed marker genotypes and recombination fractions between flanking markers. Then, phenotypes are inferred using the conditional genotype probabilities and marker effects (Haley and Knott, 1992). To adopt interval mapping for neighbor effects, it is necessary to define the effects of neighbor genotypic identity on a quantitative trait.

In this study, we developed an interval mapping method for testing neighbor effects in QTL studies. The proposed method, "Neighbor QTL", was applied to simulated data and recombinant inbred lines (RILs) of *Arabidopsis thaliana*. Furthermore, the new QTL method was built into an R package.

2 Materials and Methods

2.1 Model

We first developed a basic regression model and then defined QTL effects for interval mapping. To combine neighbor effects and a linear model, we focused on the well-known model of statistical physics, Ising model (McCoy and Maillard, 2012). The Ising model defines magnetic energy arising from physical interactions among neighboring magnets. By analogy, we regarded an individual as a magnet, genotypes as dipoles, and a trait as energy. Given

the observed traits (or energy), we estimated interaction coefficients of the Ising model to infer neighbor effects.

2.1.1 Joint regression for self and neighbor effects

To incorporate neighbor effects into a linear regression, we developed a joint model following the single-marker regression of neighbor GWAS (Sato et al., 2019b). We considered a situation where a number of inbred lines occupied finite sites in a two-dimensional space and assumed that an individual is represented by a magnet, whereby two homozygotes at each marker, AA or BB, correspond to north or south dipole (Fig. 1). We defined x_i or x_j as the genotype at a focal marker respectively for i-th focal individual or j-th neighbor, where $x_{i(j)} \in \{\text{AA, BB}\} = \{1, -1\}$. We then used multiple regression to model the effects of self-genotype and neighbor genotypic identity on a trait of i-th individual y_i as

$$y_i = \beta_0 + \beta_1 x_i + \frac{\beta_2}{L} \sum_{<i,j>}^{L} x_i x_j^{(s)} + e_i \qquad (1),$$

where β_0, β_1, and β_2 indicated intercept, self-genotype, and neighbor effects, respectively. The residual for a trait value of the focal individual i was denoted as e_i. The neighbor covariate $\sum_{<i,j>}^{L} x_i x_j^{(s)}$ was the sum of products for all combinations between the i-th focal individual and the j-th neighbor at the s-th scale of spatial distance from the focal individual i (Fig. 1). The total number of neighbors L varied in response to the spatial scale s to be referred. The coefficient of neighbor effects β_2 was scaled by L. If two individuals shared the same genotype at a given locus, the product $x_i x_j$ became positive; the product $x_i x_j$ became negative if two individuals had different genotypes. Thus, the effects of neighbor genotypic identity on a trait y_i was dependent on the coefficient β_2 and the number of two genotypes in a neighborhood.

Notably, the multiple regression model eq. 1 was posed as an inverse problem of the Ising

model. When summing up y_i for all individuals and substituting coefficients as $E = -\beta_2/L$, $H = -\beta_1$ and $\epsilon_I = \sum (y_i - \beta_0)$, eq. 1 could be transformed into the total magnetic energy of a two-dimensional Ising model as $\epsilon_I = -E \sum_{<i,j>}^{L} x_i x_j^{(s)} - H \sum x_i$ (McCoy and Maillard, 2012). In such a case, the neighbor effect β_2 and self-genotype effect β_1 could be interpreted as the interaction coefficient E and external magnetic force H, respectively.

2.1.2 QTL effects of neighbor genotypic identity

To exchange a linear regression into a QTL model, we defined QTL effects for self and neighbor effects. With heterozygosity incorporated, we redefined x_i and x_j by a marker genotype for an i-th focal individual and j-th neighbor as g_i and g_j, respectively. Self QTL effects expected from those genotypes were denoted as $g_{i(j)} \in \{\text{AA, AB, BB}\} = \{a, d, -a\}$, where a and d indicated additive and dominance deviation, respectively. Assuming two possible directions, we then defined QTL effects by neighbor genotypic identity between the individual i and j (Table 1). Given the QTL effects of self and neighbor effects, we decomposed a trait of i-th individual y_i as

$$y_i = \bar{y} + g_i + \frac{\sum_{<i,j>}^{L} g_i g_j^{(s)}}{L} + e_i \qquad (2),$$

where $\bar{y}$ and e_i indicated a population mean of traits and a residual for the focal individual i, respectively. Assuming that two marker effects a and d were unlikely to be equivalent between self and neighbor effects, we introduced a_1 and d_1 to the self QTL effects; and a_2 and d_2 to the neighbor QTL effects. If QTL effects were completely additive (i.e., $a_1 = a_2 = 1$ and $d_1 = d_2 = 0$), the QTL model eq. 2 had the same structure as the linear regression eq. 1. In such an additive model, the coefficients β_1 and $\pm\sqrt{\beta_2}$ represented additive QTL effects. It was also worth noting that the sign of $\pm\sqrt{\beta_2}$ determined the direction of the effects of genotypic identity on a trait (Table 1).

2.1.3 Interval mapping for neighbor genotypic identity

To enable interval mapping, we extended the single-marker QTL model eq. 2 to multiple pseudo-markers. In particular, we modified Haley-Knott regression that approximated the maximum likelihood method by a simple regression (Haley and Knott, 1992; Broman and Sen, 2009). The proposed algorithm consisted of three steps: (i) obtaining conditional self-genotype probabilities, (ii) calculating neighbor genotypic identity from the conditional self-genotype probabilities, and (iii) regressing trait values on the conditional self-genotype probabilities and neighbor genotypic identity.

The first step to obtain conditional self-genotype probabilities was the same as that of standard QTL mapping. Let $p_{i(j)}$ be the probability for the focal individual i or neighbor j to have a certain genotype at an interval pseudo-marker. We defined the conditional self-genotype probability for the individual i as $p_i = \Pr(g_i = \{\text{AA, AB, BB}\}|\mathbf{M})$, and obtained p_i from the number of observed markers $\times$ n individuals matrix $\mathbf{M}$ and its recombination fraction following hidden Markov models (Lander and Green, 1987; Broman et al., 2003). Based on the products of the conditional self-genotype probabilities, we further calculated the conditional probabilities for neighbor genotypic identity p_ip_j. We then defined g_ig_j as the QTL effects by neighbor genotypic identity; and p_ip_j as the expected probability for two genotypes to interact, whereby the expected neighbor QTL effects was $p_ip_jg_ig_j$. These probabilities were summed up for all possible combinations of the genotypes as $\sum_v^3 \sum_w^3 [(p_{i,v}p_{j,w}) \otimes (g_{i,v}g_{j,w})]$, where the subscript v and w indicated the three genotype states AA, AB, and BB.

Similar to Haley-Knott regression, we finally estimated the QTL effects g_i and g_ig_j by regressing the trait values y_i on p_i and $\sum_{<i,j>}^{L} p_ip_j^{(s)}/L$, respectively. The additive and dominance deviation for the self QTL effects a_1 and d_1 were considered as average differences in trait values among AA, AB, or BB genotypes, such that $a_1 = (\bar{y}_{\text{AA}} - \bar{y}_{\text{BB}})/2$ and $d_1 = \bar{y}_{\text{AB}} - (\bar{y}_{\text{AA}} + \bar{y}_{\text{BB}})/2$ (Broman and Sen, 2009). In such a case, the regression coefficient β_1 gave $2\hat{a}_1$ when -1, 0, and 1 dummy groups were assigned for the AA, AB, and BB genotypes,

respectively, or gave $\hat{d}_1$ when 0, 1, and 0 were assigned for the AA, AB, and BB genotypes, respectively (Broman et al., 2003).

For neighbor QTL effects, the additive and dominance deviation a_2 and d_2 were also considered as the average differences in trait values among the nine possible combinations (Table 1) as $a_2 = [(\bar{y}_{\mathrm{AA/AA}} + \bar{y}_{\mathrm{BB/BB}}) - (\bar{y}_{\mathrm{AA/BB}} + \bar{y}_{\mathrm{BB/AA}})]/4$ and $d_2 = \bar{y}_{\mathrm{AB/AB}} - (\bar{y}_{\mathrm{AB/AA}} + \bar{y}_{\mathrm{AB/BB}} + \bar{y}_{\mathrm{AA/AB}} + \bar{y}_{\mathrm{BB/AB}})/4 - (\bar{y}_{\mathrm{AA/AA}} + \bar{y}_{\mathrm{BB/BB}} + \bar{y}_{\mathrm{AA/BB}} + \bar{y}_{\mathrm{BB/AA}})/4$. In this case, trait values y_i could be fitted by a quadratic regression on the group of nine genotype combinations (Fig. 2). Suppose that $y_i = \beta_0 + \beta_1 p_i + \beta_2 (\sum_{<i,j>}^{L} p_i p_j^{(s)}/L) + \beta_3 (\sum_{<i,j>}^{L} p_i p_j^{(s)}/L)^2$ represents such a quadratic regression, where the linear or quadratic coefficient β_2 or β_3 provides estimates for the additive or dominance deviation $\pm 2\hat{a}_2^2$ or $\hat{d}_2^2$, respectively. Practically, we could estimate a_2 and d_2 by the quadratic regression of the trait values y_i on the neighbor genotypic identity $\sum_{<i,j>}^{L} p_i p_j^{(s)}/L$, with nine genotype combinations encoded as AA/AA, BB/BB, AA/AB, AB/AA, AB/AB, AB/BB, BB/AB, AA/BB, BB/AA = {1, 1, 0.25, 0.25, 0.0, -0.25, -0.25, -1, -1}.

Based on the linear and quadratic regression, we decomposed a trait y_i into self and neighbor QTL effects. To distinguish the two effects, we estimated a_1, d_1, a_2, and d_2 by following the six-step iterations.

1. Estimate a_1 by a linear regression on self-genotype probabilities, with -1, 0, and 1 encoded for the AA, AB and BB genotypes, respectively.

2. Estimate d_1 by a linear on self-genotype probabilities regression, with 0, 1, and 0 encoded for the AA, AB and BB genotypes, respectively.

3. Calculate self QTL effects with $\hat{a}_1$ and $\hat{d}_1$.

4. Include the self QTL effect as a covariate at a focal marker.

5. Estimate $\pm a_2$ and d_2 by a quadratic regression on neighbor genotypic identity, with

[-1, 1] dummy groups assigned for nine genotype combinations.

6. Calculate joint QTL effects with $\hat{a}_1$, $\hat{d}_1$, $\hat{a}_2$ and $\hat{d}_2$.

Based on $\hat{a}_1$, $\hat{a}_2$, $\hat{d}_1$ and $\hat{d}_2$, we inferred $\hat{y}_i$ and derived $\log_e$-likelihood (LL) from model deviance. LOD score for the self or neighbor effects were designated as $\mathrm{LOD}_{\mathrm{self}} = \log_{10}[\exp(\mathrm{LL}_{\mathrm{self}} - \mathrm{LL}_{\mathrm{null}})]$ or $\mathrm{LOD}_{\mathrm{nei}} = \log_{10}[\exp(\mathrm{LL}_{\mathrm{nei}} - \mathrm{LL}_{\mathrm{self}})]$, which could be obtained in steps 3 and 6, respectively.

When there were only two genotypes, the quadratic regression was replaced by a linear regression to estimate the additive neighbor effects. For the case of inbred lines lacking AB heterozygotes, we estimated the additive deviation a_2 by a linear regression of trait values y_i on the neighbor genotypic identity $\sum_{<i,j>}^{L} p_i p_j^{(s)}/L$, with 1 and -1 dummy groups assigned for the AA and BB genotypes, respectively. In case of backcross lines lacking BB homozygotes, the additive deviation corresponded to the dominance deviation so that $d_2 = -a_2$. The additive deviation a_2 could be estimated by a linear regression with the AA and AB genotypes encoded into -1 and 0, respectively. These two linear models were equivalent in the sense that both inbred and backcross lines had two genotypes with additive effects.

2.1.4 Variation partitioning with the QTL model

Prior to the genome scan, we estimated the effective spatial scale s by calculating the proportion of phenotypic variation explained (PVE) by neighbor effects. Incorporating two random effects into a linear mixed model, we were able to partition phenotypic variation into PVE by self effects, neighbor effects, and residuals (Sato et al., 2019b). According to previous studies (Henderson et al., 1959; Kang et al., 2008), the linear mixed model was expressed as

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

where $\mathbf{y}$ indicated a phenotype vector as $y_i \in \mathbf{y}$; $\mathbf{X}\boldsymbol{\beta}$ indicated fixed effects with a matrix including a unit vector and all covariates $\mathbf{X}$ and a coefficient vector $\boldsymbol{\beta}$; $\mathbf{Zu}$ indicated random effects with $u_i \in \mathbf{u}$ and a design matrix $\mathbf{Z}$; and $\mathbf{e}$ indicates residuals where $e_i \in \mathbf{e}$. The random effects and residuals were further decomposed as $\mathrm{Var}(\mathbf{u}) = \sigma_1^2\mathbf{K}_1 + \sigma_2^2\mathbf{K}_2$ and $\mathrm{Var}(\mathbf{e}) = \sigma_e^2\mathbf{I}$, where the $n \times n$ individuals similarity matrix for self-genotype or neighbor identity was scaled by the number of markers q as $\mathbf{K}_1 = \mathbf{P}_1^\mathsf{T}\mathbf{P}_1/(q-1)$ or $\mathbf{K}_2 = \mathbf{P}_2^\mathsf{T}\mathbf{P}_2/(q-1)$, respectively. Given that one of two alleles is similar between heterozygotes and homozygotes, here we defined additive polygenic effects for self QTLs as $g_i \in$\{AA, AB, BB\} = \{-1, 0, 1\}; and for neighbor QTLs as $g_ig_j \in$\{AA/AA, BB/BB, AA/AB, AB/AA, AB/AB, AB/BB, BB/AB, AA/BB, BB/AA\} = \{1, 1, 0.5, 0.5, 0.0, -0.5, -0.5, -1, -1\}. In these cases, the $q \times n$ matrix $\mathbf{P}_1$ included expected self-genotype values as elements $\mathbf{P}_1 = (\sum_v^3 p_i g_i)$ and $\mathbf{K}_1$ represented a kinship matrix that was calculated from all the pseudo-markers (Broman et al., 2019). Similarly, the $q \times n$ matrix $\mathbf{P}_2$ included the neighbor genotypic identities as elements $\mathbf{P}_2 = (\sum_{<i,j>}^{L} \sum_v^3 \sum_w^3 [(p_{i,v} p_{j,w}^{(s)}) \otimes (g_i g_j^{(s)})])$ and $\mathbf{K}_2$ represented a genome-wide structure of neighbor genotypic identity. Based on the three variance component parameters, we calculated PVE by self or neighbor effects as $\mathrm{PVE}_{\mathrm{self}} = \sigma_1^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$ or $\mathrm{PVE}_{\mathrm{nei}} = \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_e^2)$. Additionally, the heritability was designated as $h^2 = \sigma_1^2/(\sigma_1^2 + \sigma_e^2)$ when σ_2^2 was set at 0.

Using the linear mixed model eq. 3, our previous simulations revealed that the effective spatial scale of neighbor effects could be determined by increasing the patterns of $\mathrm{PVE}_{\mathrm{nei}}$ from $s = 0$ to a large s (Sato et al., 2019b). If the effective range was narrow, $\mathrm{PVE}_{\mathrm{nei}}$ approached to a plateau at a small value of s. In contrast, $\mathrm{PVE}_{\mathrm{nei}}$ linearly increased with s if the effective range was broad. To generalize these results for a continuous two-dimensional space, here we introduced ΔPVE metric as differences in PVE from s to $s+1$ such that $\Delta\mathrm{PVE} = \mathrm{PVE}_{\mathrm{nei},s+1}$ - $\mathrm{PVE}_{\mathrm{nei},s}$. Using such differential metrics, we quantified how $\mathrm{PVE}_{\mathrm{nei}}$ approached to a plateau across s as follows:

1. Categorize spatial scales as $s \in S$ based on the percentiles for pairwise Euclidean distance between individuals.

2. Calculate PVE_{nei} from $s = 1$ to the maximum elements of S.

3. Calculate $\Delta\text{PVE}_{\text{nei}}$ and determine $s = \arg\max \Delta\text{PVE}_{\text{nei}}$

The proposed algorithm using a differential PVE was called "ΔPVE method" hereafter.

2.1.5 An R package, "rNeighborQTL"

In addition, the neighbor QTL method was built into an R package, "rNeighborQTL". The rNeighborQTL took as input objects from the R/qtl package (Broman et al., 2003), allowing us to save phenotypes and genotypes as common "cross" objects. Because of the stepwise testing, the self QTL effects yielded the same results as standard QTL mapping. For the ΔPVE method, the mixed models eq. 3 were solved using the algorithm of average information restricted maximum likelihood (AI-REML) (Gilmour et al., 1995) implemented in the gaston package (Perdry and Dandine-Roulland, 2020). An additional, but necessary, input file was a spatial map describing the positions of individuals at the x- and y-axes. The rNeighborQTL package is available via CRAN at https://cran.r-project.org/package=rNeighborQTL.

The rNeighborQTL package included several options to analyze a variety of QTL data. Alternative to linear (mixed) models (eq. 1 and eq. 3), logistic (mixed) models could also be selected to handle a binary phenotype (Faraway, 2016; Chen et al., 2016). Because the logistic mixed model did not provide $\hat{\sigma}_e^2$ (Chen et al., 2016; Perdry and Dandine-Roulland, 2020), PVE_{nei} was substituted by the ratio of phenotypic variation explained (RVE) by neighbor effects as $\text{RVE}_{\text{nei}} = \hat{\sigma}_2^2/\hat{\sigma}_1^2$, when a binary trait was subject to the ΔPVE method. The neighbor QTL also allowed additional covariates when conducting a genome scan. This option enabled composite interval mapping (Jansen, 1993), if genetic markers other than a focal

246 locus were considered covariates. When a significant marker was detected by the single-QTL
247 analysis, it was also possible to test two-way interactions, such as namely epistasis, between
248 the neighbor QTL effects across a genome. Details are documented in the rNeighborQTL
249 package.

250 2.2 Simulation

251 We performed a benchmark test using simulated data on F2 and backcross lines. With a ran-
252 dom spatial map generated, we simulated neighbor effects based on "fake.f2" and "fake.bc"
253 autosome genotypes implemented in the R/qtl package (Broman et al., 2003). The spatial
254 positions were sampled from a uniform distribution Unif(1, 100) across a continuous two-
255 dimensional space. We estimated a_1 for self-phenotypes of "fake.f2" and "fake.bc" data after
256 the trait values were scaled to have a mean of zero and variance of 1, and assigned $\max \hat{a}_1$ to
257 a randomly selected marker. In contrast to the major-effect marker, small coefficients, i.e.,
258 $10^{-3} \times \max \hat{a}_1$, were assigned to the other markers to simulate polygenic effects. Additive
259 ($a_2 = \max \hat{a}_1$ and $d_2 = 0.25 \times \max \hat{a}_1$), dominant ($a_2 = d_2 = \max \hat{a}_1$), and overdominant
260 ($a_2 = \max \hat{a}_1$ and $d_2 = 1.25 \times \max \hat{a}_1$) scenarios were analyzed for the F2 lines, while only
261 additive scenario ($a_2 = \max \hat{a}_1$ and $d_2 = -\max \hat{a}_1$) was applicable for the backcross lines.
262 Thirty traits were simulated for true effective distances given at ten to fifty percentiles of
263 pairwise Euclidean distance among individuals. The simulated neighbor effects were added
264 to the self-phenotypes of “fake.f2” or “fake.bc” dataset, with 75% of phenotypic variation be-
265 ing attributable to the neighbor effects. Then we applied the ΔPVE method and a genome
266 scan for the joint traits, and calculated $\mathrm{LOD}_{\mathrm{nei}}$ at $s = \arg\max \Delta\mathrm{PVE}_{\mathrm{nei}}$ to evaluate the
267 power to detect neighbor effects.

2.3 Data

To apply the neighbor QTL on real data, we conducted a pilot QTL experiment using the yellow-striped flea beetle *Phyllotreta striolata* and RILs of *Arabidopsis thaliana* (Fig. S1A). Adults of flea beetles access host plants by jumping, and leaf holes made by these beetles are easily countable. These flea beetles are known to prefer glabrous *A. thaliana* to hairy accessions (Sato et al., 2019a). To observe large phenotypic variation in leaf holes, we selected RILs derived from hairy and glabrous accessions in this study.

2.3.1 Plants and insects

We used 130 accessions, including parental and recombinant inbred lines between Col(*gl1*) and Kas-1 accession (Wilson et al., 2001). Col(*gl1*) plants produce no trichomes, while Kas has sparse trichomes on leaves and stems. The RILs are known to vary in the trichome production, disease resistance (Wilson et al., 2001), and flowering time (Li et al., 2006). The genotype data were available in Wilson et al. (2001). The set of RILs was obtained through the Arabidopsis Biological Resource Center (ABRC) (Stock ID, CS84999: https://abrc.osu.edu/).

Flea beetles were maintained under a long-day condition (16:8 hours light:dark cycles with a 22 °C constant air temperature) in an environmental chamber (Biotron LH-241PFD-SP, NK system, Osaka, Japan). To establish the experimental population, we collected ca. 200 adults from *Brassica* cultivars grown in the field within Otsu City, Shiga Prefecture, Japan (35°01′N 135°51′E) during November 2018 and May 2019. Adults of *P. striolata* consume shoots and especially prefer to young glabrous leaves, whereas larvae consume below-ground tissue of *Brassica* plants; therefore, we reared adults and larvae on leaves and swollen hypocotyls, respectively. Young leaves of Boc choy *Brassica rapa* var. *chinensis* or Chinese cabbage *B. rapa* var. *pekinensis* were supplied for the adults. The larvae were allowed to feed on swollen hypocotyls of the radish *Raphanus sativus* var. *longipinnatus* or

the turnip *B. rapa* subsp. *rapa* buried in moisten vermiculite. Adult females laid eggs in the moisten vermiculite, and it took a month (28 to 32 days) for eggs to become adults.

2.3.2 Experimental procedure

To investigate neighbor effects in herbivory, we allowed adult beetles to feed on RIL seedlings grown in a plastic cell tray. Three seeds for each accession were sown on each compartment of the cell tray (13 $\times$ 10 cells composed of 20 $\times$ 20 mm^2 compartment) with the accessions randomized. The seeds were acclimated under a constant dark condition with 4 °C for seven days, and then allowed to germinate under a long-day condition (16:8 hours light:dark cycles with a 20 °C constant air temperature). The seedlings were grown under the long-day condition for 24 days, with 2000-fold diluted liquid fertilizer (N:P:K = 6:10:5; Hyponex, Hyponex Japan, Osaka) supplied once. On day 14 after the germination, the seedlings were thinned out to leave one seedling per compartment. Prior to the feeding experiment, we recorded the presence or absence of leaf trichomes and the occurrence of bolting by direct observation and determined the rosette diameter (mm) by analyzing seedling images using Image J software (Abràmoff et al., 2004). The cell tray was enclosed by a white mesh cage (length 29.2 cm $\times$ width 41.0 cm $\times$ height 27.0 cm: Fig. S1B). Thirty adult beetles were released into the cage and allowed to feed on plants for 72 hours. We counted leaf holes as a measure of herbivory for each plant as flea beetles left small holes when they fed on leaves (Fig. S1C). The final sample size was 126 individuals; out of 130 accessions, 4 accessions (CS84877, CS84873, CS84950, and CS84894) were not germinated, CS84898 lacked genotype data, and CS84958 had two replicates of individuals. The data are included in the rNeighborQTL package.

2.3.3 Data analysis

We used R version 3.6.0 (R Core Team, 2019) for all statistical analyses. A genetic map for the Col × Kas RILs was estimated using the est.map() function in the R/qtl package (Broman et al., 2003). Self-genotype probabilities were calculated using the calc.genoprob() function implemented in the R/qtl package (Broman et al., 2003). The number of leaf holes was log-transformed and analyzed using linear models. The presence of trichomes and bolting was analyzed using logistic models. When analyzing the number of leaf holes, we incorporated the presence or absence of bolting, the rosette diameter, and the edge (or not) of the cell tray into covariates. The neighbor QTL was performed using the rNeighborQTL package developed above. Examples using the Col × Kas dataset are available in the vignette of rNeighborQTL package, where the usage of each function is also documented. A genome-wide significance level was determined by empirical percentiles of the maximum LOD score among 999 permuted traits. We considered $p < 0.1$ and $p < 0.05$ a suggestive and significant level, respectively. We also set an arbitrary threshold at LOD score of 1.5 when discussing the results.

3 Results and Discussion

3.1 Simulation using F2 and backcross lines

We simulated neighbor effects based on "fake.f2" and "fake.bc" data implemented in the R/qtl package (Broman et al., 2003). The maximum additive deviation of self QTL effects, $\max \hat{a}_1$, was 0.56 and 0.28 for F2 and backcross lines, respectively. These values were assigned for neighbor QTL effects to achieve similar a signal strength between self and neighbor effects, while minor effects were allocated to other loci. Considering the polygenic variation as random effects, we applied the ΔPVE method for simulated traits. The estimated distance

given by $s = \arg\max \Delta\text{PVE}$ increased as the true distance increased (Fig. 3), indicating that the ΔPVE method was effective.

When the efficient distance of neighbor effects was limited, such short-range neighbor effects were well detected using ΔPVE method and the quadratic approximation (median $\text{LOD}_{\text{nei}} > 4$ at the ten percentile of pairwise Euclidean distance: Fig. 3). Although the power to detect long-range neighbor effects was lowered, LOD score was still larger than the Bonferroni threshold (median $\text{LOD}_{\text{nei}} > 4$: Fig. 3). These results indicated that short-range neighbor effects could be detected in any scenario, although it was relatively difficult to detect long-range effects.

For backcross lines, both short- and long-range neighbor effects were well detected (median $\text{LOD}_{\text{nei}} > 4$ for all s: Fig. 3D). The backcross lines had two genotypes with the additive deviation alone and were well fitted using linear approximation (Fig. 3D), whereas the additive traits for F2 lines were less likely fitted using the quadratic model assuming three genotypes with the additive and dominance deviation (Fig. 3A). Given the model structure underlying F2 lines, it is plausible that the quadratic term was unnecessary for the additive F2 traits, indeed, it decreased the power to detect neighbor effects.

3.2 Self QTL effects in Col $\times$ Kas RILs

The observed number of leaf holes ranged from 0 to 38 with a median of 4 (Fig. S1D). The total variation in the number of leaf holes was explained at 5% by the trichome production; 2% by bolting; 10% by the rosette diameter; and 22% by the edge effects (Analysis-of-Variance, $F = 9.1$, 3.7, 20.7, and 43.4; $p = 0.003$, 0.06, 10^{-4}, and 10^{-8}, respectively). With the kinship matrix $\mathbf{K}_1$ considered a random effect, a linear mixed model estimated the heritability as 5.6% for the leaf holes, though it was not significant (Likelihood ratio test, $\chi^2_1 = 1.82, p = 0.18$).

To scan self QTL effects, we conducted standard QTL mapping of the trichome produc-

tion, the number of leaf holes, and bolting (Fig. 4; Table 2). For self QTL effects on the trichome production, we detected a strong peak near the *GLABRA1* locus (>20 LOD_{self} score: Fig. 4B). Considering the rosette diameter and bolting covariates, we observed a suggestive but the largest self QTL effect on the leaf holes at the *GLABRA1* locus (LOD_{self} = 1.97: Fig. 4C). For the bolting, we observed the largest significant peak on the bottom of chromosome 1 (>4 LOD_{self}), and the second largest and suggestive peak on the top of chromosome 4 (LOD_{self} = 1.92: Fig. 4D).

Several studies reported the same QTLs or a particular gene function for the self effects on trichomes, defense, and flowering. Remarkably, *GLABRA1* gene on the chromosome 3 is known to encode a myb transcription factor regulating leaf trichome developments (Ishida et al., 2008) and deter feeding by flea beetles (Sato et al., 2019a). The result that *GLABRA1* possessed self QTL effects on the leaf holes adds a biological value to the insect herbivory data. Furthermore, two self-bolting QTLs on the chromosome 1 and 4 were located near flowering time QTLs in Col × Kas RILs (Li et al., 2006). Thus, our pilot experiment supports previous evidence for the loci responsible for plant development and defense.

3.3 Neighbor QTL effects in Col × Kas RILs

To estimate the effective distance of neighbor effects, we applied ΔPVE method with every ten percentile categories for pairwise Euclidean distance (Fig. 5). For the number of leaf holes, the $\Delta\text{PVE}_{\text{nei}}$ was peaked at 7 distance scale from a focal individual (Fig. 5A), covering almost all the experimental arena from the center plant. At this estimated distance, the neighbor effects explained 6% of total variation in the leaf holes, with the both self and neighbor similarity $\mathbf{K}_1$ and $\mathbf{K}_2$ considered random effects in linear mixed models. At 7.8 distance scale, the neighbor effects explained 8.7% of total variation in leaf holes at the maximum, though it was not significant compared to its heritability by self QTL effects (Likelihood ratio test, $\chi^2_1 = 1.03, p = 0.31$). It seemed plausible that the effective distance

was relatively large for insect herbivory, because the adult beetles were likely free to move within the small experimental cage. On the other hand, the ΔPVE became the largest at the nearest scale for the bolting and explained over a half variation compared to the self QTL effects ($RVE_{nei} = 0.68$ at $s = 2.24$: Fig. 5B), suggesting that the bolting was unlikely affected by distant neighbors. For the trichome production, ΔPVE method revealed that there were little variation explained by neighbor effects ($RVE_{nei} \approx 0$ otherwise models failed to converge).

A genome scan for neighbor effects was performed using the estimated spatial distance (Fig. 5; Table 2). Regarding the neighbor QTL effects on the leaf holes, we observed, although weak, the largest QTL on the top of chromosome 4 at the nga8 marker ($LOD_{nei} = 1.86$), which was also the position the second largest self-bolting QTL occurred. This neighbor QTL had no significant epistasis as shown by < 1.1 LOD score for all the two-way interactions between the nga8 and other markers (Fig. S2A). Neither the neighbor QTL nor *GLABRA1* locus detected above was overlapped with known self-QTLs of powdery mildew resistance (Wilson et al., 2001), suggesting independence of the herbivory QTLs on the disease resistance loci. At the nearest scale for bolting, we found a weak neighbor QTL at the R30025 marker on the chromosome 3 ($LOD_{nei} = 1.8$: Fig. 5; Table 2). This QTL did not have any significant epistasis with the other markers (< 1.1 LOD score: Fig. S2B).

Ecological studies have shown that how easy an individual plant is to find, termed plant apparency, drives neighbor effects through visual crypsis against herbivores (Hambäck et al., 2000; Castagneyrol et al., 2013; Strauss et al., 2015). In the present study, the neighbor QTL involved in the leaf holes was located near a self-bolting QTL at the top of chromosome 4, suggesting the potential importance of plant apparency in neighbor effects in anti-herbivore defense. In addition, the positive sign of the additive neighbor effects a_2 at that marker indicated that the number of leaf holes decreased when neighbors had different genotypes (Table 2). This implies that the mixture of flowering and vegetative plants may acquire

414 population-wide resistance to flea beetles since the effective distance of neighbor effect was
415 sufficiently large to encompass almost the entire experimental arena. These results led us to
416 hypothesize that the self QTL underlying plant apparency might facilitate population-wide
417 anti-herbivore defense, called associational resistance (Hambäck et al., 2000), through its
418 pleiotropy on neighbor effects.

419 3.4 Further applicability and limitation

420 Theoretical advantage of the Ising model lies in its inference of spatial arrangements that
421 optimize total magnetic energy. Once the self and neighbor coefficients are estimated by the
422 marker-based regression, these two coefficients may infer which genotype distributions can
423 minimize or maximize the population-sum of trait values (Sato et al., 2019b). In the context
424 of neighbor QTL, additive effects suggest that positive and negative a_2 favors clustered
425 or mixed patterns for maximizing the sum of trait values, respectively. However, in cases
426 where dominance effects and epistasis are involved, how such a complex genetic basis affects
427 the optimal spatial arrangement remains unexplored. These potential effects of genetic
428 architecture on a population-level outcome of neighbor effects would be of theoretical as well
429 as empirical interest for future studies.

430 Superior to the previous neighbor GWAS, the present neighbor QTL has a flexibility
431 to deal with heterozygosity. However, the use of neighbor QTL is still restricted to auto-
432 somes because sex-dependent inheritance of neighbor effects remains unknown. Standard
433 QTL mapping on sex chromosomes is known to require from one to three degree-of-freedom
434 (Broman et al., 2006), and thus its extension to neighbor effects may be more complex than
435 the self QTL effects. In addition, the neighbor QTL approximated the maximum likelihood
436 method by a quadratic regression, in which phenotype variance was assumed to be equal
437 among the nine combinations among three QTL genotypes. Our simulation revealed that
438 the quadratic approximation could handle the overdominance, but became inferior to linear

approximation if additive effects alone governed a trait. We should thus be aware of statistical models behind the neighbor QTL. Practically, both the intercross and the inbred models might be utilized if a sample population is partially inbred.

3.5 Conclusion

The present neighbor QTL, together with the previous neighbor GWAS (Sato et al., 2019b), provides a sort of novel tools to incorporate neighbor effects into quantitative genetics. These methods may provide insights into genetic architecture underlying neighbor effects as exemplified by the pilot study of insect herbivory on *A. thaliana*. Once the neighbor GWAS screens candidate accessions, their crossed progeny can be inspected by the neighbor QTL. The line of R packages, "rNeighborQTL" and "rNeighborGWAS", would help investigate neighbor effects using a complementary set of GWAS and QTL data.

Acknowledgements

The authors are grateful to E. Yamamoto for helpful comments on the draft; to Dynacom Co., Ltd. for technical assistance on the R package development; and to L. G. Kawaguchi for collecting insect materials from the field. This study was supported by Japan Science and Technology Agency (JST) PRESTO (Grant number, JPMJPR17Q4) and Japan Society for the Promotion of Science (20K15880) to Y.S., and JST CREST (JPMJCR15O2) to A.J.N. No conflicts of interests concern this study.

4 Tables

Table 1: QTL effects expected by genotypic identity between the individuals i and j with AA, AB, or BB genotypes. The additive and dominance deviation is represented by a and d, respectively. The left table shows a case in which a share of same QTL genotypes exerts positive effects on a trait y_i, whereas the right table shows a case in which a share of same genotypes exerts negative effects on y_i

g_i/g_j	AA	AB	BB
AA	a^2	ad	$-a^2$
AB	ad	d^2	$-ad$
BB	$-a^2$	$-ad$	a^2

g_i/g_j	AA	AB	BB
AA	$-a^2$	$-ad$	a^2
AB	$-ad$	d^2	ad
BB	a^2	ad	$-a^2$

Table 2: Estimated QTL effects in Col × Kas RILs of *Arabidopsis thaliana*. Markers with any >1.5 LOD scores (highlighted by bold letters) are shown. Additive effects $2a_1$ indicate the effect size when Kas alleles are replaced by two Col alleles, while $2a_2^2$ indicates the effect size of identical homozygotes over different ones. The sign of $2a_2^2$ determines the direction of neighbor genotypic identity (Table 1). The LOD_{nei} score is shown on the spatial distance at which ΔPVE peaks.

Trait	Marker	Chr	Position (cM)	$2a_1$	LOD_{self}	$\pm 2a_2^2$	LOD_{nei}	Distance
Trichome	GL1	3	65.24	**-2.83**	**22.8**	3.28	0.13	7.82
Holes	GL1	3	65.24	**0.21**	**1.97**	-0.25	0.05	7
	nga8	4	0	-0.07	0.13	**2.67**	**1.86**	7
Bolting	nga692	1	102.0	**-1.04**	**4.15**	-0.72	0.13	2.24
	R30025	3	126.1	0.06	0.01	**2.82**	**1.80**	2.24
	nga8	4	0	**1.05**	**1.92**	0.41	0.02	2.24

5 Figures

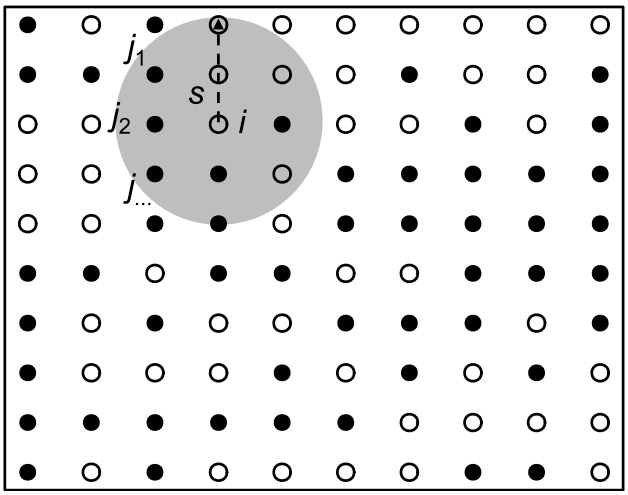

Figure 1: Assumption of neighbor effects in a two-dimensional space. A white or black point indicates an individual having AA or BB genotype, respectively. A grey circle shows an effective area of neighbor effects at the spatial distance s from the focal individual i. Neighbor effects then occur depending on genotype similarity between the focal individual i and all the j-th neighbors within the spatial distance s.

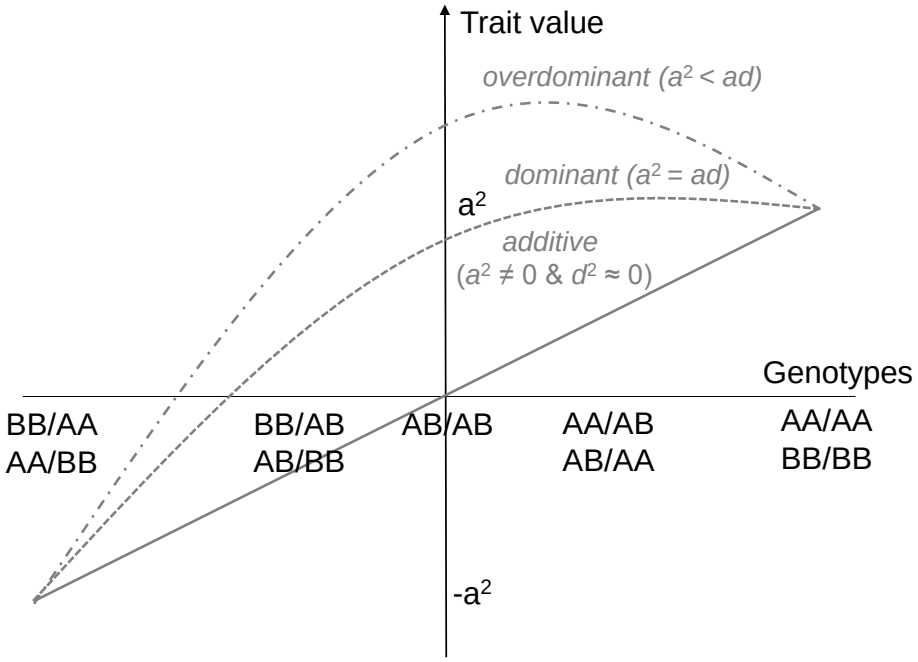

Figure 2: A scheme explaining approximation of neighbor QTL effects by quadratic regression. Trait values y_i are regressed on nine possible combinations of genotype identity between a focal individual i and its neighbor j (Table 1). The additive or dominance deviation a or d is represented by the linear or quadratic term, respectively. If the linear coefficient is negative, it indicates the case in which AA/AA and BB/BB combinations had negative QTL effects on traits (Right of Table 1).

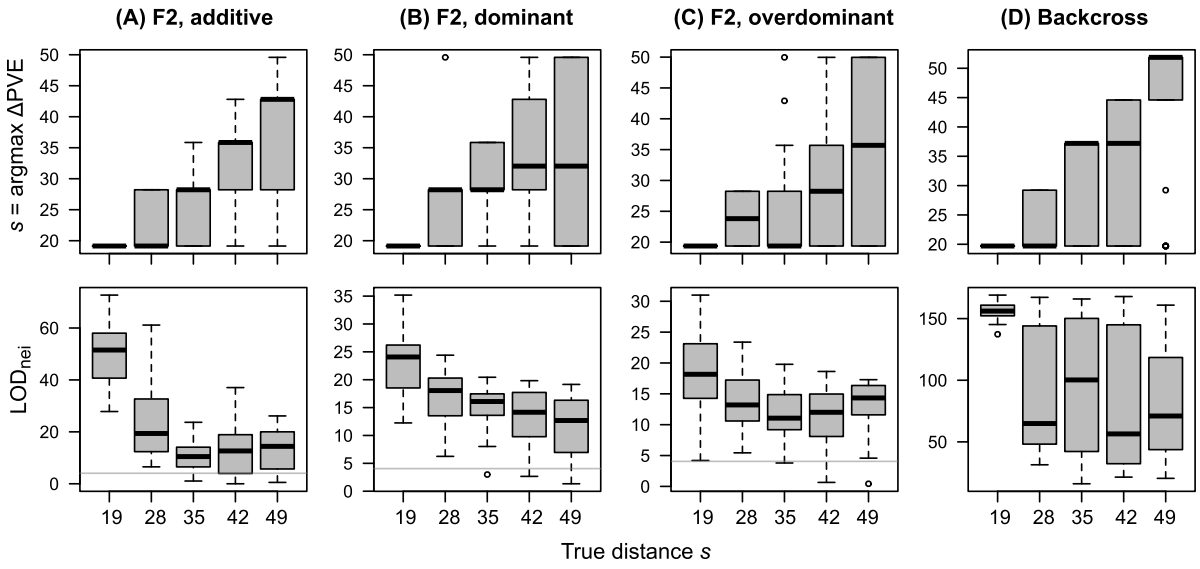

Figure 3: Benchmark test using simulated F2 and backcross datasets. Upper panels show the distance estimated by ΔPVE method, while lower panels show $\mathrm{LOD}_{\mathrm{nei}}$ of a major-effect marker at the estimated distance. The x-axis corresponds to ten to fifty percentiles of pairwise Euclidean distance. Thirty traits were simulated for each distance class. Boxplots represent median by a center line; upper and lower quartiles by box limits; 1.5$\times$ interquartile range by whiskers; and outliers by points. Horizontal lines indicate a LOD threshold at $p = 0.05$ after Bonferroni correction.

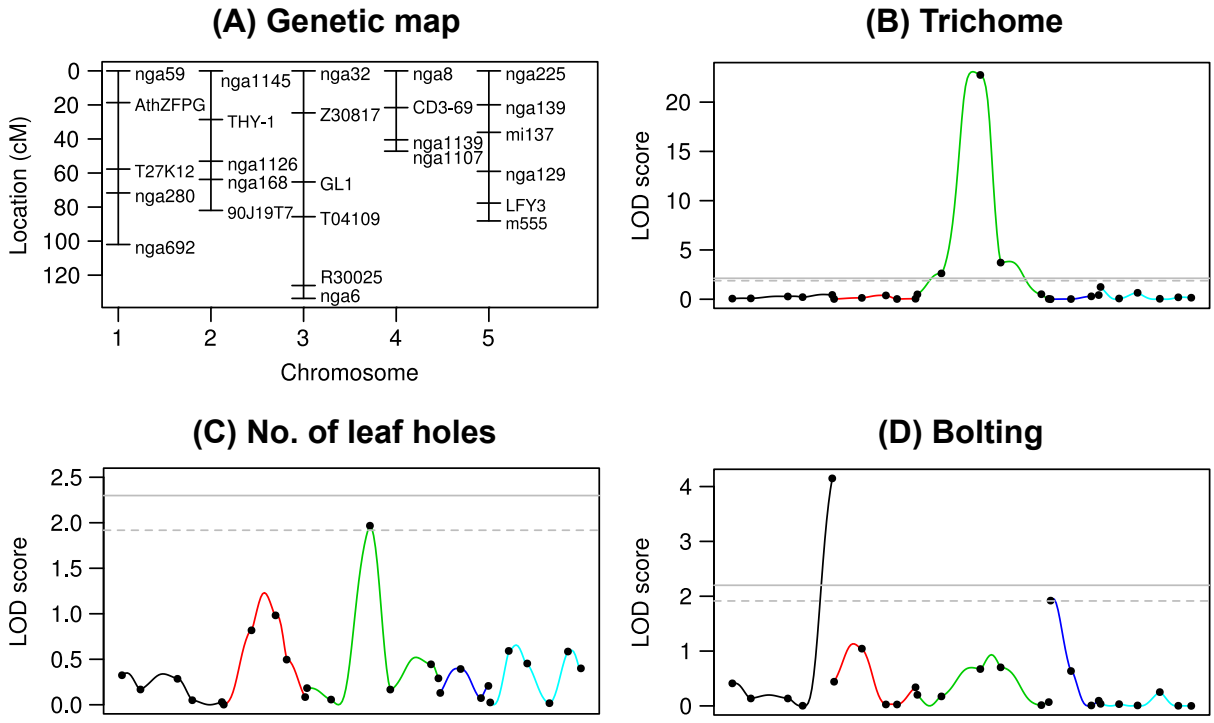

Figure 4: Genetic map and LOD scores for self QTL effects in Col × Kas RILs. (A) Genetic map showing the locations of 26 markers among the five chromosomes of *Arabidopsis thaliana*. LOD_{self} score for the trichome production (B), the number of leaf holes (C), bolting (D). Colors correspond to chromosome numbers, and dots indicate observed markers. A solid and dashed horizontal line indicates a significant ($p < 0.05$) and suggestive ($p < 0.1$) LOD threshold with 999 permutations, respectively.

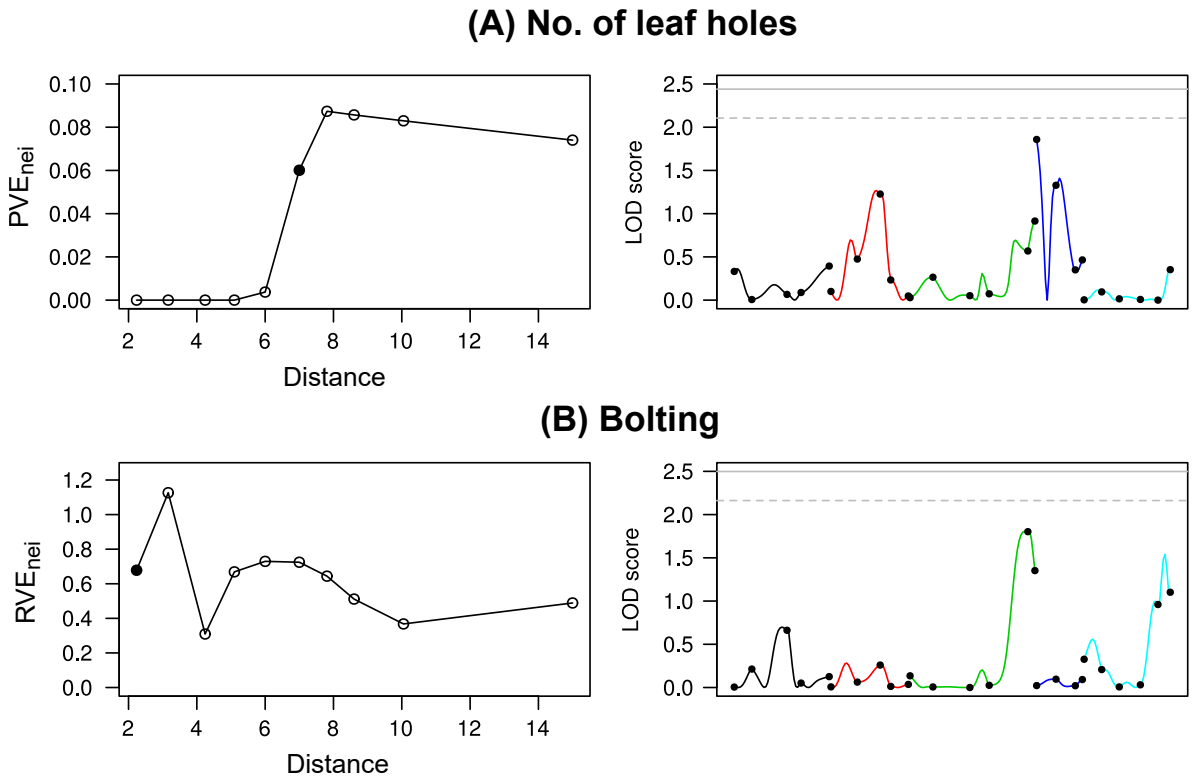

Figure 5: Phenotypic variation explained and LOD score attributed to neighbor effects on the number of leaf holes (A) or the presence of bolting (B) in Col × Kas RILs. Left: Proportion or ratio of phenotypic variation explained by neighbor effects (PVE_{nei} or RVE_{nei}) plotted against the pairwise distance among individuals. A closed point indicates the distance at which ΔPVE peaked. Right: LOD_{nei} score for neighbor QTL effects at the distance at which ΔPVE peaked. Colors correspond to chromosome numbers, and dots indicate observed markers. A solid and dashed horizontal line indicates a significant ($p < 0.05$) and suggestive ($p < 0.1$) LOD threshold with 999 permutations, respectively.

References

Abràmoff, M. D., P. J. Magalhães, and S. J. Ram, 2004: Image processing with imagej. *Biophotonics International*, **11 (7)**, 36–42.

Barbosa, P., J. Hines, I. Kaplan, H. Martinson, A. Szczepaniec, and Z. Szendrei, 2009: Associational resistance and associational susceptibility: Having right or wrong neighbors. *Annual Review of Ecology, Evolution, and Systematics*, **40 (1)**, 1–20, doi:10.1146/annurev.ecolsys.110308.120242.

Broman, K. W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, Ś. Sen, B. S. Yandell, and G. A. Churchill, 2019: R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics*, **211 (2)**, 495–502, doi:10.1534/genetics.118.301595.

Broman, K. W., and Ś. Sen, 2009: Single-QTL analysis. *A Guide to QTL Mapping with R/qtl*, K. W. Broman, and Ś. Sen, Eds., Springer, 75–133.

Broman, K. W., Ś. Sen, S. E. Owens, A. Manichaikul, E. M. Southard-Smith, and G. A. Churchill, 2006: The x chromosome in quantitative trait locus mapping. *Genetics*, **174 (4)**, 2151–2158, doi:10.1534/genetics.106.061176.

Broman, K. W., H. Wu, Ś. Sen, and G. A. Churchill, 2003: R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19 (7)**, 889–890, doi:10.1093/bioinformatics/btg112.

Castagneyrol, B., B. Giffard, C. Péré, and H. Jactel, 2013: Plant apparency, an overlooked driver of associational resistance to insect herbivory. *Journal of Ecology*, **101 (2)**, 418–429.

Chen, H., and Coauthors, 2016: Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics*, **98 (4)**, 653–666, doi:10.1016/j.ajhg.2016.02.012.

Faraway, J. J., 2016: *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press.

Gilmour, A. R., R. Thompson, and B. R. Cullis, 1995: Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 1440–1450.

Haley, C. S., and S. A. Knott, 1992: A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69 (4)**, 315–324, doi:10.1038/hdy.1992.131.

Hambäck, P. A., J. Ågren, and L. Ericson, 2000: Associational resistance: insect damage to purple loosestrife reduced in thickets of sweet gale. *Ecology*, **81 (7)**, 1784–1794.

Han, K., H. Y. Lee, N. Y. Ro, O. S. Hur, J. H. Lee, J. K. Kwon, and B. C. Kang, 2018: QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotechnology Journal*, **16 (9)**, 1546–1558, doi:10.1111/pbi.12894.

Hayes, B., 2013: Overview of statistical methods for genome-wide association studies (GWAS). *Genome-wide association studies and genomic prediction*, C. Gondro, J. Werf, and B. Hayes, Eds., Springer, 149–169.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. Von Krosigk, 1959: The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15 (2)**, 192–218.

Ishida, T., T. Kurata, K. Okada, and T. Wada, 2008: A genetic regulatory network in the development of trichomes and root hairs. *Annual Review of Plant Biology*, **59**, 365–386.

Jansen, R. C., 1993: Interval mapping of multiple quantitative trait loci. *Genetics*, **135 (1)**, 205–211.

505 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin,
506 2008: Efficient control of population structure in model organism association mapping.
507 *Genetics*, **178 (3)**, 1709–1723, doi:10.1534/genetics.107.080101.

508 Korte, A., and A. Farlow, 2013: The advantages and limitations of trait analysis with GWAS:
509 a review. *Plant Methods*, **9 (1)**, 29, doi:10.1186/1746-4811-9-29.

510 Lander, E. S., and P. Green, 1987: Construction of multilocus genetic linkage maps in
511 humans. *Proceedings of the National Academy of Sciences*, **84 (8)**, 2363–2367.

512 Li, Y., P. Roycewicz, E. Smith, and J. O. Borevitz, 2006: Genetics of local adaptation
513 in the laboratory: Flowering time quantitative trait loci under geographic and seasonal
514 conditions in *Arabidopsis*. *PLoS ONE*, **1 (1)**, e105, doi:10.1371/journal.pone.0000105.

515 Marchadier, E., and Coauthors, 2019: The complex genetic architecture of shoot growth
516 natural variation in *Arabidopsis thaliana*. *PLoS Genetics*, **15 (4)**, e1007 954, doi:10.1371/
517 journal.pgen.1007954.

518 McCoy, B. M., and J.-M. Maillard, 2012: The importance of the ising model. *Progress of*
519 *Theoretical Physics*, **127 (5)**, 791–817, doi:10.1143/PTP.127.791.

520 Perdry, H., and C. Dandine-Roulland, 2020: *gaston: Genetic Data Handling (QC, GRM,*
521 *LD, PCA) & Linear Mixed Models*. URL https://CRAN.R-project.org/package=gaston,
522 r package version 1.5.6.

523 R Core Team, 2019: *R: A Language and Environment for Statistical Computing*. Vienna,
524 Austria, R Foundation for Statistical Computing, URL https://www.R-project.org/.

525 Rishmawi, L., J. Bühler, J. Benjamin, M. Hülskamp, and M. Koornneef, 2017: Quantitative
526 trait loci controlling leaf venation in *Arabidopsis*. *Plant, Cell & Environment*, **40 (8)**,
527 1429–1441.

528 Sato, Y., 2018: Associational effects and the maintenance of polymorphism in plant defense
529 against herbivores: review and evidence. *Plant Species Biology*, **33 (2)**, 91–108, doi:10.
530 1111/1442-1984.12201.

531 Sato, Y., R. Shimizu-Inatsugi, M. Yamazaki, K. K. Shimizu, and A. J. Nagano, 2019a:
532 Plant trichomes and a single gene *GLABRA1* contribute to insect community com-
533 position on field-grown *Arabidopsis thaliana*. *BMC Plant Biology*, **19 (1)**, doi:10.1186/
534 s12870-019-1705-2.

535 Sato, Y., E. Yamamoto, K. K. Shimizu, and A. J. Nagano, 2019b: Neighbor GWAS: incorpo-
536 rating neighbor genotypic identity into genome-wide association studies of field herbivory
537 on *Arabidopsis thaliana*. *bioRxiv*, 845735, doi:10.1101/845735.

538 Schuman, M. C., S. Allmann, and I. T. Baldwin, 2015: Plant defense phenotypes determine
539 the consequences of volatile emission for individuals and neighbors. *eLife*, **4**, doi:10.7554/
540 eLife.04490.

541 Sonah, H., L. O'Donoughue, E. Cober, I. Rajcan, and F. Belzile, 2015: Identification of loci
542 governing eight agronomic traits using a GBS-GWAS approach and validation by QTL
543 mapping in soya bean. *Plant Biotechnology Journal*, **13 (2)**, 211–221, doi:10.1111/pbi.
544 12249.

545 Strauss, S. Y., N. I. Cacho, M. W. Schwartz, A. C. Schwartz, and K. C. Burns, 2015:
546 Apparency revisited. *Entomologia Experimentalis et Applicata*, **157 (1)**, 74–85.

547 Subrahmaniam, H. J., C. Libourel, E.-P. Journet, J.-B. Morel, S. Muños, A. Niebel, S. Raf-
548 faele, and F. Roux, 2018: The genetics underlying natural variation of plant–plant inter-
549 actions, a beloved but forgotten member of the family of biotic interactions. *The Plant*
550 *Journal*, **93 (4)**, 747–770.

Tamura, M., T. Ohgushi, and T. Y. Ida, 2020: Intraspecific neighbourhood effect: Population-level consequence of aggregation of highly defended plants. *Functional Ecology*, **34 (3)**, 597–605.

Underwood, N., P. A. Hambäck, and B. D. Inouye, 2020: Pollinators, herbivores, and plant neighborhood effects. *The Quarterly Review of Biology*, **95 (1)**, 37–57.

Wilson, I. W., C. L. Schiff, D. E. Hughes, and S. C. Somerville, 2001: Quantitative trait loci analysis of powdery mildew disease resistance in the *Arabidopsis thaliana* accession kashmir-1. *Genetics*, **158 (3)**, 1301–1309.

Wuest, S. E., and P. A. Niklaus, 2018: A plant biodiversity effect resolved to a single chromosomal region. *Nature Ecology & Evolution*, **2 (12)**, 1933–1939, doi:10.1038/s41559-018-0708-y.

Yang, L.-N., and Coauthors, 2019: Enhanced agricultural sustainability through within-species diversification. *Nature Sustainability*, **2 (1)**, 46–52.

Zeller, S. L., O. Kalinina, D. F. Flynn, and B. Schmid, 2012: Mixtures of genetically modified wheat lines outperform monocultures. *Ecological Applications*, **22 (6)**, 1817–1826.

Supplementary Materials

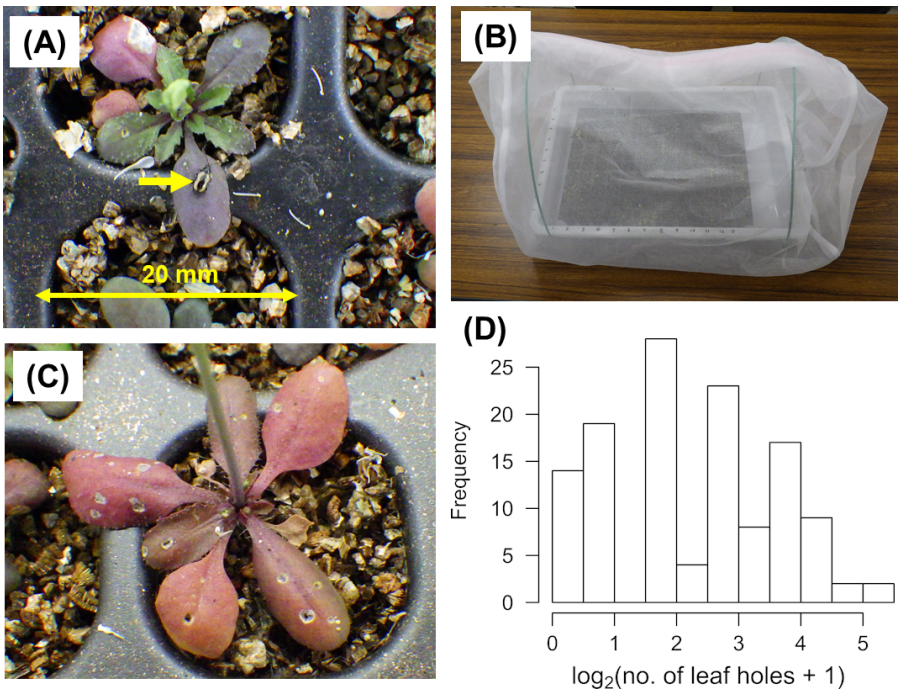

Figure S1: Pilot QTL experiment using *Arabidopsis thaliana* and the yellow-striped flea beetle *Phyllotreta striolata*. (A) An adult beetle on a vegetative plant. (B) An experimental cage including 130 Col × Kas RILs. (C) A plant attacked by *P. striolata*. Leaf holes were made by adult beetles. (D) Histogram for the observed number of leaf holes.

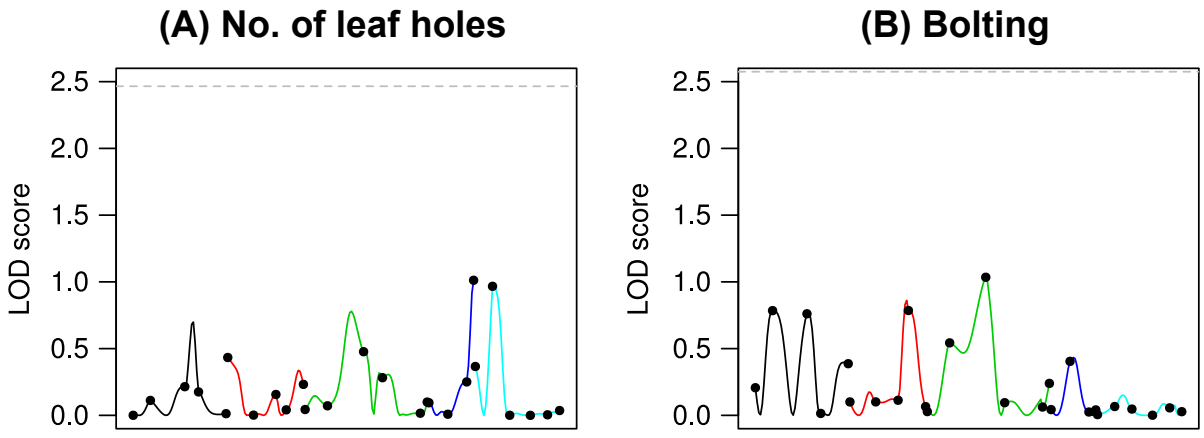

Figure S2: Epistasis in neighbor QTL effects on the number of leaf holes between the nga8 and other markers (A); or on the presence of bolting between the R30025 and other markers (B). Colors correspond to chromosome numbers, and dots indicate observed markers. A dashed horizontal line indicates a suggestive ($p < 0.1$) LOD threshold with 999 permutations.