Exploring Efficient Linear Mixed Models to Detect Quantitative Trait Locus-by-Environment Interactions

AUTHORS

Eiji Yamamoto[*,†,1], Hiroshi Matsunaga[‡]

AFFILIATIONS

[*] Graduate School of Agriculture, Meiji University, Kawasaki 214-8571, Japan

[†] PRESTO, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan

[‡] Institute of Vegetable and Floriculture Science, National Agriculture and Food Research Organization, Ano, Tsu 514-2392, Japan

SHORT RUNNING TITLE: QTL-by-environment interactions

KEY WORDS: Genome-wide association study (GWAS), Quantitative trait locus (QTL), Genotype-by-environment interaction (G×E), QTL-by-environment interaction (Q×E), Linear mixed model (LMM)

CORRESPONDING AUTHOR

Eiji Yamamoto, Graduate School of Agriculture, Meiji University, Kawasaki 214-8571, Japan. Tel/Fax: +81-44-934-7046. Email: yame@meiji.ac.jp

ABSTRACT

Genotype-by-environment interactions (G×E) are important for understanding genotype–phenotype relationships. To date, various statistical models have been proposed to account for G×E effects, especially in genomic selection (GS) studies. Generally, GS does not focus on the detection of each quantitative trait locus (QTL), while the genome-wide association study (GWAS) was designed for QTL detection. G×E modeling methods in GS can be included as covariates in GWAS using unified linear mixed models (LMMs). However, the efficacy of G×E modeling methods in GS studies has not been evaluated for GWAS. In this study, we performed a comprehensive comparison of LMMs that integrate the G×E modeling methods to detect both QTL and QTL-by-environment interaction (Q×E) effects. Model efficacy was evaluated using simulation experiments. For the fixed effect terms representing Q×E effects, simultaneous scoring of specific and non-specific environmental effects was recommended because of the higher recall and improved genomic inflation factor value. For random effects, it was necessary to account for both G×E and genotype-by-trial (G×T) effects to control genomic inflation factor value. Thus, the recommended LMM includes fixed QTL effect terms that simultaneously score specific and non-specific environmental effects and random effects accounting for both G×E and G×T. The LMM

was applied to real tomato phenotype data obtained from two different cropping seasons. We detected not only QTLs with persistent effects across the cropping seasons but also QTLs with Q×E effects. The optimal LMM identified in this study successfully detected more QTLs with Q×E effects.

INTRODUCTION

Phenotypes are determined not only by genetic potential but also by environmental growth conditions. More specifically, the rank of phenotypic values often changes when the same genotype set is phenotyped under different environmental conditions (Cooper and DeLacy 1994). The phenotypic response to the environment is explained by reaction norms that describe the pattern of phenotypic expression of a genotype across different environments. Genotype-by-environment interactions (G×E) occur when the slopes of the reaction norms of two different genotypes are not parallel across environments (Malosetti *et al.* 2013). G×E analysis is important for a precise understanding of genotype–phenotype relationships as well as for the design of crop varieties that fit a given environment.

To date, various statistical approaches have been proposed for G×E analysis (reviewed in Malosetti *et al.* 2013). The simplest approach is analysis of variance, which compares the mean and variance of the phenotypic values of genotypes in multiple environments. However, more flexible methods were eventually developed, beginning with the Additive Main effects and Multiplicative Interaction (AMMI) model, which divides the genetic contribution for a phenotypic value into additive main effects (i.e., genetic effects not specific to the environment) and G×E effects (Gauch 1988).

Thus, G×E effects in the AMMI model are expressed as multiplicative interaction terms consisting of the products of genotypic and environmental scores. Next, principal component (PC) analysis of these scores is conducted (Gauch 1988). This approach allows graphical representation of the interactivity between genotypes and environments (Gauch 1988). In particular, a genotype will show the highest performance in the closest environment in the biplot. The Genotype main effects and G×E (GGE) model resembles the AMMI model, but is focused on the total genetic effect in each environment (Yan *et al.* 2000). In the GGE model, both additive main effects and environment-specific genetic effects are scored together; thus, genotypes in the GGE biplot represent the overall genetic effect, while the AMMI biplot represents only the G×E effect (Yan *et al.* 2000). The AMMI and the GGE models provided an important perspective on G×E analysis, and facilitated discussion about whether genetic effects for a phenotypic value within a given environment should be separated into specific and non-specific environmental effects.

In the AMMI and GGE models, genotypic differences are treated as categorical variables (Gauch 1988; Yan *et al.* 2000). However, the availability of molecular markers has allowed the quantitative description of genotypic differences using a genomic relationship matrix (GRM) (Endelman and Jannink 2012). GRM applications

in G×E analysis have advanced considerably in genomic selection (GS) studies (reviewed in Crossa *et al.* 2017). In GS, a training population that have been phenotyped and genotyped is used to construct a model that predicts the genetic potential of unphenotyped individuals (Meuwissen *et al.* 2001). Most GS models that account for G×E effects are constructed as linear mixed models (LMMs) (Crossa *et al.* 2017). LMMs treat genetic effects as random effects using a variance–covariance matrix designed using GRMs. The most straightforward method is to consider G×E random effects as a lack of genetic correlation between environments (Lopez-Cruz *et al.* 2015). If information on the similarity among environments is available as a variance–covariance matrix, then G×E random effects can be designed as the product of a GRM and the variance–covariance matrix (Cuevas *et al.* 2017). If there are multiple options to model random effects, then it may be preferable to integrate all random effect terms into a single model and estimate their relative contributions to phenotypic values (Jarquín *et al.* 2014). In all of these methods, random effects with G×E have common variance across environments (Lopez-Cruz *et al.* 2015; Cuevas *et al.* 2017). Sousa *et al.* (2017) extended the models to allow different variances across environments, which resulted in increased GS prediction accuracy in some experiments.

Generally, GS models do not test the significance of each marker effect because their objective is phenotype prediction and not the detection of quantitative trait loci (QTLs) (Meuwissen *et al.* 2001; Crossa *et al.* 2017). For QTL detection, genetic mapping approaches such as the genome-wide association study (GWAS) are necessary (Cortes *et al.* 2020). Currently, the most frequently used GWAS method is based on unified LMMs (Yu *et al.* 2006). In the unified LMMs, total genetic effects are divided into a fixed effect for each marker genotype and random effects that are modeled for all marker genotypes (Yu *et al.* 2006). Then, the statistical significance of the fixed effect is analyzed to estimate the QTL in linkage disequilibrium (LD) with the marker.

Because the methods used to model G×E in GS studies were designed for random effects (Jarquín *et al.* 2014; Lopez-Cruz *et al.* 2015; Cuevas *et al.* 2017; Sousa *et al.* 2017), they can be applied to random effects in unified LMMs for GWAS. Although several studies have detected QTL-by-environment interaction (Q×E) effects using LMMs (Boer *et al.* 2007; Mathews *et al.* 2008), the efficacy of G×E random effects methods in GS studies for Q×E analysis has not been surveyed, because this is a recent application. Therefore, the advantages and disadvantages of the various methods developed to model G×E random effects in Q×E analysis are poorly understood. In the present study, we performed a comprehensive comparison of the methods used to model

G×E random effects in GWAS. The objective of this study was to identify the most effective LMMs for detecting QTLs with Q×E effects. For fixed QTL effects, we compared two methods that were designed based on ideas represented in the AMMI (Gauch 1988) and the GGE (Yan *et al.* 2000) models. The first method divides the total genetic effects of a QTL in LD with a marker into two effects: the additive main effect, which is not specific to the environment, and the Q×E effect. The second method simultaneously scores the additive main effect and the Q×E effect. All combinations of the modeling methods for fixed effect terms and random effect terms were compared using simulation experiments. Finally, LMMs selected based on the simulation experiments were applied to GWAS using tomato (*Solanum lycopersicum* L.) phenotype data.

MATERIALS AND METHODS

Assumptions

To clarify the aim of the present study and its differences from similar recent studies (e.g., Moore *et al.* 2019; Dahl *et al.* 2020), we describe our assumptions as follows:

Experimental design:

- The environment in this study can be geographic location, weather condition, and other artificial experimental conditions. In the real phenotype data used in this study, cropping season was the environmental differences (Figure 1).
- Differences among environments and/or trials are represented as categorical variables.
- Phenotypic values of a genotype in an environment were obtained for several trials (Figure 1).

Genetic effects:

- The target traits are controlled by several major QTLs and numerous minor QTLs, and the major QTLs were the detection target.
- Genetic effects in the same environment are affected by genotype-by-trial (G×T) effects.

LMMs

We modeled the effects of Q×E and G×E using unified LMMs, in which the total genetic effects were divided into fixed and random effect terms:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\mathrm{Fixed}}(\boldsymbol{x}) + f_{\mathrm{Random}}(\boldsymbol{G}) + \boldsymbol{\varepsilon}, \quad (1)$$

where $\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ indicate $n \times 1$ vectors for phenotypic values and residuals, respectively; $\boldsymbol{T}$ is an $n \times TR$ design matrix that assigns phenotypic values to trials, where TR is the number of trials; $\boldsymbol{t}$ is a $TR \times 1$ vector of population-wide mean for each trial; $\boldsymbol{S}$ is a $n \times c$ matrix whose column elements are the eigenvectors from principal component analysis of genotype data from all markers. $\boldsymbol{s}$ indicates a $c \times 1$ vector of fixed effects for $\boldsymbol{S}$. $\boldsymbol{S}$ and $\boldsymbol{s}$ are used to decrease the rate of false signals generated by the population structure (Price *et al.* 2006; Yang *et al.* 2011; Li *et al.* 2014). In our analyses of simulations and real data in this study, $\boldsymbol{S}$ consisted of the first and second eigenvectors, which explained 30.9% and 7.8% of the total genetic variation, respectively (Yamamoto *et al.* 2016). $f_{\mathrm{Fixed}}(\boldsymbol{x})$ represents the fixed effect terms for a major QTL in LD with a single-nucleotide polymorphism (SNP) marker. $f_{\mathrm{Random}}(\boldsymbol{G})$ represents the random effect terms for the background genetic effects. $\boldsymbol{x}$ is an $n \times 1$ vector of SNP genotype values coded as {0, 1, 2} = {aa, Aa, AA}; $\boldsymbol{G}$ is an $m \times m$

GRM containing *m* genotypes. In the GRM, the genetic relationship between individuals *j* and *k* (G_{jk}) is defined as:

$$G_{jk} = \sum_{i=1}^{I} (x_{ij} - 2p_i)(x_{ik} - 2p_i)/2p_i(1 - p_i), \quad (2)$$

where x_{ij} is the coded genotype value for the *i*-th SNP of the *j*-th individual, p_i is the minor allele frequency for the *i*-th SNP, and I is the total number of markers. The GRM was calculated using the *A.mat* function in the R package "rrBLUP" (Endelman and Jannink 2012). Because the effects of major QTL and the Q×E effects were included in $f_{\text{Fixed}}(\boldsymbol{x})$, the proportion of variance explained by $f_{\text{Fixed}}(\boldsymbol{x})$ was subjected to a statistical test, whereas $f_{\text{Random}}(\boldsymbol{G})$ and the other terms were covariates. Detailed descriptions of $f_{\text{Fixed}}(\boldsymbol{x})$ and $f_{\text{Random}}(\boldsymbol{G})$ are provided in the sections "Modeling fixed effects" and "Modeling random effects", respectively. In a standard unified LMM (Yu *et al.* 2006), Eq. 1 is described as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + \boldsymbol{x}\alpha + \boldsymbol{u_G} + \boldsymbol{\varepsilon}, \quad (3)$$

where α models the additive main effect of a QTL in LD with the SNP. We refer to the fixed QTL effect term in Eq. 3 (i.e., $\boldsymbol{x}\alpha$) as the additive main effect term (Table 1). $\boldsymbol{u_G}$ models the random effects as follows:

$$\boldsymbol{u_G} \sim MVN(\boldsymbol{0}, [\boldsymbol{Z_G G Z_G'}]\sigma_G^2), \quad (4)$$

where MVN is the multivariate normal distribution; $\boldsymbol{Z_G}$ represents an $n \times m$ incidence matrix for the phenotype and random effects; and σ_G^2 is the variance for $\boldsymbol{u_G}$.

Modeling fixed effects: To model Q×E as fixed effects in the LMMs, we used two formulae. The first is as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + \boldsymbol{x}\beta + \sum_l^L \{(\boldsymbol{\pi}_l \circ \boldsymbol{x})\gamma_l\} + f_{\mathrm{Random}}(\boldsymbol{G}) + \boldsymbol{\varepsilon}, \quad (5)$$

where β is the QTL effect not specific to the environment, and γ_l is the QTL effect specific to the *l*-th environment. $\boldsymbol{\pi}_l$ is an $n \times 1$ vector containing indicator variables that determine whether the phenotypic value is obtained from *l*-th environment $\{1\}$ or not $\{0\}$. The symbol $\circ$ indicates the Hadamard product for the left and right vectors or matrices. In this equation, $\boldsymbol{x}\beta$ is analogous to the additive main genetic effect in the AMMI model (Gauch 1988). Therefore, we refer to the fixed QTL effect terms for an SNP in Eq. 5 (i.e., $\boldsymbol{x}\beta + \sum_l^L \{(\boldsymbol{\pi}_l \circ \boldsymbol{x})\gamma_l\}$) as AMMI-type Q×E effect terms (Table 1). The second formula used to model Q×E is as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + \sum_l^L \{(\boldsymbol{\pi}_l \circ \boldsymbol{x})\zeta_l\} + f_{\mathrm{Random}}(\boldsymbol{G}) + \boldsymbol{\varepsilon}, \quad (6)$$

where ζ_l is the QTL effect in the *l*-th environment. Unlike Eq. 5, Eq. 6 does not include the QTL effect not specific to the environment (i.e., $\boldsymbol{x}\beta$). Therefore, the Q×E

effect terms are analogous to the GGE model (Yan *et al.* 2000). We refer to the Q×E effect terms in Eq. 6 (i.e., $\sum_l^L\{(\boldsymbol{\pi}_l \circ \boldsymbol{x})\zeta_l\}$) as GGE-type Q×E effect terms (Table 1).

Modeling random effects: Appropriate statistical modeling for random effects has been the key to efficient GWAS using LMMs (Yu *et al.* 2006). In a standard GWAS, random effects are modeled as in Eq. 4. Therefore,

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\text{Fixed}}(\boldsymbol{x}) + \boldsymbol{u_G} + \boldsymbol{\varepsilon}. \quad (7)$$

Notably, $\boldsymbol{u_G}$ does not account for the G×E effects (Yu *et al.* 2006). Therefore, we extended Eq. 7 as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\text{Fixed}}(\boldsymbol{x}) + \boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{\varepsilon}, \quad (8)$$

where $\boldsymbol{u_{GE}}$ models the G×E effects as follows:

$$\boldsymbol{u_{GE}} \sim MVN(\boldsymbol{0}, [\boldsymbol{Z_G G Z_G'}] \circ [\boldsymbol{Z_E Z_E'}]\sigma_{GE}^2), \quad (9)$$

where $\boldsymbol{Z_E}$ is the incidence matrix for the phenotypic values and environmental differences (e.g., cropping season in Figure 1). Thus, Eq. 9 allows for different random effects among environments. The third random effect model is as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\text{Fixed}}(\boldsymbol{x}) + \boldsymbol{u_G} + \boldsymbol{u_{GT}} + \boldsymbol{\varepsilon}, \quad (10)$$

where $\boldsymbol{u_{GT}}$ models the reaction norm for G×T effects as follows:

$$\boldsymbol{u_{GT}} \sim MVN(\boldsymbol{0}, [\boldsymbol{Z_G G Z_G'}] \circ [\boldsymbol{Z_T Z_T'}]\sigma_{GT}^2), \quad (11)$$

where $\boldsymbol{Z_T}$ is the incidence matrix for the phenotype and trials. Thus, Eq. 11 allows for independent random effects between trials and has the potential to capture the unexpected G×T effects mentioned in the second assumption for genetic effects. Eqs. 8 and 10 can be combined as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\text{Fixed}}(\boldsymbol{x}) + \boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}} + \boldsymbol{\varepsilon}. \quad (12)$$

Eq. 12 can model both environment-specific and trial-specific random effects. The above random effect models are used in Jarquín *et al.* (2014) and Lopez-Cruz *et al.* (2015). The last random effect model is as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\text{Fixed}}(\boldsymbol{x}) + \boldsymbol{u_G} + \sum_{tr}^{TR} \boldsymbol{u_{tr}} + \boldsymbol{\varepsilon}, \quad (13)$$

where *TR* is the total number of trials included in the phenotype data. $\sum_{tr}^{TR} \boldsymbol{u_{tr}}$ is represented as follows:

$$\sum_{tr}^{TR} \boldsymbol{u_{tr}} \sim MVN(\boldsymbol{0}, \sigma_1^2 \begin{bmatrix} \boldsymbol{Z_G G Z_G'} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{bmatrix} + \sigma_2^2 \begin{bmatrix} \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z_G G Z_G'} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{bmatrix}$$

$$+ \cdots + \sigma_{TR}^2 \begin{bmatrix} \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{Z_G G Z_G'} \end{bmatrix}). \quad (14)$$

This random effect design is the same as that of Sousa *et al.* (2017) and Dahl *et al.* (2020), which allow for different heritability between trials. Therefore, the random

effect model in Eq. 14 is more flexible and can capture G×T effects better than Eqs. 9 and 11. Information on the random effect terms is summarized in Table 1.

Fitting the LMMs: The LMMs described above include different numbers of random effect terms. For example, Eq. 7 includes only one random effect term (i.e., $\boldsymbol{u_G}$) whereas Eq. 8 includes two (i.e., $\boldsymbol{u_G}$ and $\boldsymbol{u_{GE}}$). However, the model fitting in this study was performed using the same procedure. To explain the procedure, we generalized, tentatively, the random effects as follows:

$$\boldsymbol{u_r} \sim MVN(\boldsymbol{0}, [\boldsymbol{Z_G G Z_G'}] \circ [\boldsymbol{Z_r Z_r'}]\sigma_r^2), \quad (15)$$

where $\boldsymbol{u_r}$ is the *r*-th random effect, $\boldsymbol{Z_r}$ is the incident matrix for phenotype and *r*-th random effects, and σ_r^2 is the variance for $\boldsymbol{u_r}$. To solve the LMMs with *R* random effects, we estimated the genetic variances without fixed QTL effect terms, as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + \sum_r^R \boldsymbol{u_r} + \boldsymbol{\varepsilon}. \quad (16)$$

The solution of Eq. 16 was used to estimate a weight (w_r) for each random effect term. For example, the weight of the first random effect (w_1) is calculated as follows:

$$w_1 = \hat{\sigma}_1^2 \,/\, \sum_r^R \hat{\sigma}_r^2. \quad (17)$$

The variance–covariance matrix for the integrated random effects ($\boldsymbol{K}$) was calculated as:

$$\boldsymbol{K} = \sum_r^R \{[\boldsymbol{Z_G G Z_G'}] \circ [\boldsymbol{Z_r Z_r'}] w_r\}. \quad (18)$$

Next, the integrated random effects $\boldsymbol{u_K}$ were derived as follows:

$$\boldsymbol{u_K} \sim MVN(\boldsymbol{0}, \boldsymbol{K}\sigma_K^2), \quad (19)$$

where σ_K^2 is the variance for $\boldsymbol{u_K}$. Thus, the LMMs used to calculate the test statistics in the present study can be expressed as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\text{Fixed}}(\boldsymbol{x}) + \boldsymbol{u_K} + \boldsymbol{\varepsilon}. \quad (20)$$

This method made GWAS computationally feasible for the present study. All calculations and parameter estimations for the LMMs were performed using the R package “gaston” (Perdry and Dandine-Roulland 2018).

Statistical tests for the fixed QTL effect terms: The statistical significance of the fixed QTL effect terms in each LMM (i.e., $f_{\text{Fixed}}(\boldsymbol{x})$) was evaluated using the log-likelihood (LL) ratio test (LRT). The tests and formulae used to calculate the deviance and degrees of freedom are described in Table 2. The LRT used in this study can be divided into two categories (Table 2). The first category is a test for all QTL-effect terms (Table 2); it tests for the existence of a QTL in LD with the SNP. Thus, signals detected by this test can represent QTLs with or without Q×E effects. The second category is a test only for interaction terms (Table 2), which will be significant only when the QTL has Q×E effects (Malosetti *et al.* 2013). The LL of the maximum likelihood estimates was

calculated using the *lmm.diago.profile.likelihood* function in the R package "gaston" (Perdry and Dandine-Roulland 2018). The *p* value of each test was calculated using the chi-square test based on the deviance and degrees of freedom (df) (Table 2).

Although the LRT can be used to obtain more accurate statistics in small- and moderate-sized samples, its drawback is the absence of test statistics for each coefficient included in the fixed effect terms that account for Q×E effects. As a complement to the LRT, we performed the Wald test for each coefficient for the Q×E effects. The Wald test scores were calculated as follows:

$$W_{il} = \hat{\theta}_{il}^2 / var(\hat{\theta}_{il}), \quad (21)$$

where $\hat{\theta}_{il}$ and W_{il} indicate the estimated effect size of the *i*-th SNP in the *l*-th environment and the Wald score, respectively. Next, the *p* value of W_{il} was calculated based on the chi-square test (df = 1).

We determined the genome-wide significant thresholds based on the false discovery rate (FDR), which is commonly applied in GWAS (Benjamini and Hochberg 1995; Storey and Tibshirani 2003). The FDR was calculated as described by Storey and Tibshirani (2003).

Simulation experiments

We designed simulation experiments to evaluate the QTL detection power of the LMMs, under conditions described in detail in Table 3. The simulated major QTLs were randomly selected from the 16,782 SNP markers in a real dataset for tomatoes (Yamamoto *et al.* 2016). The simulated phenotypic values were generated using the following equation:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_{tr}} \\ \vdots \\ \boldsymbol{y_{TR}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q_a} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{Q_a} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{Q_a} \end{bmatrix} \begin{bmatrix} \boldsymbol{q_1} \\ \vdots \\ \boldsymbol{q_{tr}} \\ \vdots \\ \boldsymbol{q_{TR}} \end{bmatrix} + \{(\boldsymbol{u_G} \times w_G) + (\boldsymbol{u_{GE}} \times w_{GE}) + (\boldsymbol{u_{GT}} \times w_{GT})\}/\omega + \boldsymbol{\varepsilon}/\varphi, \quad (22)$$

where $\boldsymbol{y_{tr}}$ is a $m \times 1$ vector for the phenotypic values in the *tr*-th simulated trial; $\boldsymbol{Q_a}$ is an $m \times 3$ matrix whose column elements consist of SNP genotypes selected for simulated QTLs; $\boldsymbol{q_{tr}}$ is a 3×1 vector for the QTL effect in the *tr*-th trial. The elements of the $\boldsymbol{q_{tr}}$ vectors become the same for trials under the same environmental conditions. $(\boldsymbol{u_G} \times w_G)$, $(\boldsymbol{u_{GE}} \times w_{GE})$, and $(\boldsymbol{u_{GT}} \times w_{GT})$ are random effects that follow Eqs. 4, 9, and 11, respectively. ω and φ were scalars required to adjust the phenotypic values to satisfy the given proportion of variance explained by each major QTL (PVE_{QTL}) and heritability. In this study, the random effect values (i.e., $\boldsymbol{u_G}$, $\boldsymbol{u_{GE}}$, and $\boldsymbol{u_{GT}}$) were generated using the *mvrnorm* function in R using a corresponding variance–covariance matrix (R Core Team 2019). The relative contributions of the three random effects were

adjusted by multiplying the values generated by *mvrnorm* and the specified weight parameters (w_G, w_{GE}, and w_{GT} in Table 3). The residual values ($\boldsymbol{\varepsilon}$) were generated using the *rnorm* function in R. In the present study, PVE_{QTL} and heritability for phenotypic values from all environments and trials were set to 0.1 and 0.5, respectively. To satisfy these settings, the optimal ω and φ were determined using the *optimize* function in R.

We assumed four Q×E effects: persistence, divergence, convergence, and crossover (Malosetti *et al.* 2013) (Table 3). Persistence means that the QTL has a persistent effect across environments and, therefore, has no Q×E effects. Divergence means that the QTL has different degrees of effect size between environments. Convergence means that the QTL shows an effect in a particular environment, but not in other environments. Crossover means that the direction of the QTL effect differs among environments.

GWAS efficiency evaluation

We calculated recall, precision, and F-measure to evaluate GWAS power as in the previous studies (Hamazaki and Iwata 2020; Saber and Shapiro 2020; Shafquat *et al.* 2020). Recall is the proportion of true positives that are correctly identified; precision is

the proportion of true positives among the retrieved positive signals; and F-measure represents the harmonic mean of recall and precision, calculated as follows:

$$\text{F measure} = 2/(\text{Recall}^{-1} + \text{Precision}^{-1}). \quad (23)$$

Thus, a high F-measure can only be achieved by balancing high precision and high recall. Generally, recall, precision, and F-measure are affected by the genome-wide significance threshold (Gage *et al.* 2018). Therefore, we also calculated the receiver operating characteristic (ROC) curve, and evaluated GWAS efficiency based on the area under the ROC curve (AUC) (Gage *et al.* 2018). An AUC of 0.5 indicates that the GWAS has no power to detect true signals, whereas an AUC of 1.0 indicates that all top signals of the GWAS agree with the true signals. We used the *roc* function in the "pROC" package (Robin *et al.* 2011) to calculate the AUC from $-\log_{10}(p)$ values.

Genomic inflation factor

Although the LMMs applied in GWAS are designed to avoid inflation of $-\log_{10}(p)$ values due to population structure (Yu *et al.* 2008), the inclusion of too many covariates in an LMM often results in inflation or deflation of *p* values from the expected distribution (Moore *et al.* 2019). Therefore, the genomic inflation factor (λ_{GC}; Devlin

and Roeder 1999) is often used to evaluate the degree of *p* inflation or deflation (Voorman *et al.* 2011; Moore *et al.* 2019). In this study, λ_{GC} was calculated as follows:

$$\lambda_{\mathrm{GC}} = \frac{log_{10}(p_{obs}[0.5])}{log_{10}(0.5)}, \quad (24)$$

where $p_{obs}[0.5]$ indicates the 0.5 quantile of the observed *p* values. $\lambda_{\mathrm{GC}} > 1$ indicates inflation of $-\log_{10}(p)$ values and an increase in false positive signals, whereas $\lambda_{\mathrm{GC}} < 1$ indicates deflation of $-\log_{10}(p)$ values and an increase in the rate of false negatives (Devlin and Roeder 1999).

Tomato genotype and phenotype data

We applied the LMMs to real tomato genotype and phenotype data (Yamamoto *et al.* 2016, 2017). We used 96 big-fruited tomato F_1 varieties intended for the fresh market. These varieties were developed by various organizations such as seed companies and the public sector (Yamamoto *et al.* 2016). The parental combinations of the F_1 varieties are unknown. Genotype data consisting of 16,782 SNP markers were obtained using Axiom myDesign genotyping arrays (Affymetrix Co., Ltd., Santa Clara, CA, USA). All SNP markers had a minor allele frequency of > 0.05, and a missing value rate of 0. Phenotyping was performed in the winter and spring cropping seasons in four and two trials, respectively. All plants were grown hydroponically using a high-wire system in a

greenhouse at the National Agriculture and Food Research Organization at the Institute of Vegetable and Tea Science in Tsu, Japan. One plant per variety was grown in each trial. Among the phenotypes obtained (Yamamoto *et al.* 2016, 2017), we focused on the average fruit weight and fruit set ratio. The fruit set ratio indicates the ratio of flowers that reached fruit set. Values were transformed using the empirical logit transformation.

Data availability

The phenotype and genotype data, R scripts, and R package developed for this study are available at https://github.com/yame-repos/gwasQxE.

RESULTS

Evaluation of power to detect Q×E effects

Common results among all simulation conditions: In LMMs including only the additive main effect term, there was little difference among the methods used to model random effects (Figure 2). The tests including Q×E effect terms showed different modes of action depending on the random effects. The random effects '$\boldsymbol{u_G}$' and '$\boldsymbol{u_G} + \sum_{tr}^{TR} \boldsymbol{u_{tr}}$' (Table 1) showed higher recall than other random effect models (Figure 2). Conversely, the precision, F-measure, and AUC of tests using '$\boldsymbol{u_G}$' and '$\boldsymbol{u_G} + \sum_{tr}^{TR} \boldsymbol{u_{tr}}$' were lower than those of the other random effect models (Figure 2). These results indicate that signals detected in the tests that used '$\boldsymbol{u_G}$' and '$\boldsymbol{u_G} + \sum_{tr}^{TR} \boldsymbol{u_{tr}}$' included more false discoveries than those in the other random effect models.

Persistence: In this simulation, tests for interaction terms showed no power (i.e., recall ≈ 0 and AUC ≈ 0.5) (Figure 2A). This result is reasonable because no Q×E effects were included in the simulated QTLs (Table 3). LMMs including only the additive main effect were the fittest models for persistence QTLs. LMMs that included only the additive main effect showed the highest AUC, without obvious disadvantages in terms

of recall, precision, and F-measure (Figure 2A). These results indicate that LMMs that include only the additive main effect are recommended for detecting persistence QTLs.

Divergence: The difference between the simulation conditions of ‘persistence’ and ‘divergence’ is that the latter QTLs have smaller PVE_{QTL} in environment A (Table 3). Because of the difference in realized PVE_{QTL} between environments, we expected LMMs that included Q×E effect terms to have higher power than those that included only the additive main effect. However, the results for ‘divergence’ resembled those for ‘persistence’ (Figure 2A and B). These results indicate that it is difficult to identify the ‘divergence’ effect using the LMMs in this study.

Convergence: Under this simulation condition, a QTL has an effect in environment B, but no effect in environment A (Table 3). Unlike persistence and divergence QTLs, LMMs that included only the additive main effect showed the lowest recall and AUC values (Figure 2C). This result indicates that Q×E terms are necessary to detect convergence QTLs. Another difference between ‘convergence’ compared to ‘persistence’ and ‘divergence’ is that the tests for interaction terms showed recognizable

degrees of recall, precision, and F-measure (Figure 2C). These results indicate that the convergence Q×E effect can be detected using LMMs including Q×E fixed effect terms.

Crossover: Under this simulation condition, the QTL effect takes opposite directions in environments A and B (Table 3). LMMs with only an additive main effect term showed near-zero recall, precision, and F-measure (Figure 2D). These results indicate that Q×E terms are necessary to detect crossover QTLs. Recall, precision, F-measure, and AUC were equivalent between all QTL-effect terms and interaction terms (Figure 2D). These results indicate that crossover QTLs were detected as QTLs with Q×E effects and were not misrecognized as QTLs with only an additive main effect.

Genomic inflation factor

The genomic inflation factor (λ_{GC}) was calculated to assess the deviation of test statistics from the expected null distribution (Figure 3). Because we simulated only three major QTLs in the experiments, λ_{GC} should be equal to 1. Thus, the condition $\lambda_{\mathrm{GC}} \approx 1$ is necessary for precise calculation of FDR for a genome-wide significant threshold. In the tests for LMMs with only an additive main effect term (Table 2), $-\log_{10}(p)$ values were close to the theoretically expected distribution (i.e., $\lambda_{\mathrm{GC}} \approx 1$)

(Figure 3A). The tests for LMMs including Q×E effect terms yielded different distributions depending on the random effects. The λ_{GC} values were inflated and fluctuated when the random effects were '$\boldsymbol{u_G}$' or '$\boldsymbol{u_G} + \sum_{tr}^{TR} \boldsymbol{u_{tr}}$' (Figure 3A). The use of the random effects '$\boldsymbol{u_G} + \boldsymbol{u_{GT}}$' or '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' restrained λ_{GC} fluctuation, but the mean values differed depending on the fixed effect terms (Figure 3A). Figure 3B compares quantile–quantile (Q–Q) plots of mean $-\log_{10}(p)$ values between LMMs with the random effects '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' (Table 2). The $-\log_{10}(p)$ values of the AMMI-type Q×E effect terms showed deflation compared with the theoretically expected distribution (Figure 3B). These results indicate that GGE-type Q×E effect terms are more appropriate than the AMMI-type. Next, we focused on the relationship between GGE-type Q×E effect terms and the method used to model random effects. The random effects '$\boldsymbol{u_G}$', '$\boldsymbol{u_G} + \boldsymbol{u_{GT}}$,' and '$\boldsymbol{u_G} + \sum_{tr}^{TR} \boldsymbol{u_{tr}}$' showed inflation of λ_{GC}, whereas '$\boldsymbol{u_G} + \boldsymbol{u_{GE}}$' and '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' showed $\lambda_{\mathrm{GC}} \approx 1$ (Figure 3C). These results indicate that an LMM that includes GGE-type Q×E effect terms and random effects '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' is recommended for calculating FDR as a genome-wide significant threshold.

GWAS for tomato agronomic traits

In the evaluation of the power to detect Q×E effects, LMMs that included only the additive main effect term were more efficient for persistence and divergence QTLs (Figure 2A and B). Conversely, LMMs including Q×E effects were necessary to detect convergence and crossover QTLs (Figure 2C and D). In the analysis using the genomic inflation factor, an LMM including GGE-type Q×E effect terms and the random effect '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' was recommended (Figure 3). Therefore, we used two LMMs for the tomato agronomic trait data. The first was an LMM that included only the additive main effect term and the random effect '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' (Table 1), and the second was an LMM that included GGE-type Q×E effect terms and the random effect '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' (Table 1). We focused on QTLs with FDR > 0.05, which are often used as genome-wide significant thresholds in GWAS (Cortes *et al.* 2020).

For average fruit weight, a significant signal on chromosome 9 was detected for the additive main effect and all QTL effect terms, including Q×E (Figure 4A and Table 4). The signal disappeared when the test was performed only for interaction terms (Figure 4A). These results indicate that the QTL detected on chromosome 9 is a persistence or divergence QTL. In the tests for interaction terms, we detected a significant signal on chromosome 10. The estimated effect size and the Wald score

suggest that the signal is a convergence QTL that shows the effect only in the spring cropping season (Table 4).

For fruit set ratio, a significant signal was detected on chromosome 12 in tests for the additive main effect (Figure 4B; Table 4). This signal disappeared in the other tests (Figure 4B). The effect size and Wald score of the signal on chromosome 12 were similar for the winter and spring cropping seasons (Table 4). These results indicate that the QTL on chromosome 12 is a persistence QTL. Alternatively, tests for all QTL effect terms, and tests for only interaction terms, detected significant signals on chromosome 1 (Figure 4B). The effect size of the signal was opposite in the winter and spring cropping seasons (Table 4). These results suggest that the QTL on chromosome 1 is a crossover QTL.

DISCUSSION

In this study, we explored efficient LMMs to detect QTLs with Q×E effects. For this objective, we compared the efficacy of LMMs with various combinations of fixed QTL effect terms and random effect modeling methods (Tables 1 and 2). Efficacy was evaluated using recall, precision, F-measure, and AUC (Figure 2). None of the tested LMMs showed high values for all parameters (Figure 2). Notably, LMMs with higher recall showed lower precision and F-measure (Figure 2). This result is reasonable because recall and false positives are highly correlated in GWAS (Bian and Holland 2017). Generally, recall, precision, and F-measure change depending on the level of the genome-wide significance threshold (Gage *et al.* 2018). These problems make it difficult to select the most efficient LMM. Recently, AUC has been used to evaluate GWAS efficiency because it is independent of the genome-wide significance threshold (Gage *et al.* 2018; Moore *et al.* 2019; Hamazaki and Iwata 2020; Shafquat *et al.* 2020). In the present study, AUC values were affected by random effect modeling methods to a smaller degree than recall, precision, and F-measure (Figure 2). Therefore, the AUC was useful for selecting the most effective LLM in this study.

Inflation or deflation of λ_{GC} causes various problems in GWAS (Devlin and Roeder 1999). For example, although FDR is commonly applied to determine

genome-wide significant thresholds in GWAS (Benjamini and Hochberg 1995; Storey and Tibshirani 2003), its calculation requires a theoretically expected *p* value distribution (i.e., $\lambda_{GC} \approx 1$). Genomic control (GC) corrects inflated or deflated *p* values using the λ_{GC} value (Devlin and Roeder 1999). However, the applicability of a given inflation factor (i.e., λ_{GC}) to correct marker *p* values differs according to allele frequency and correlation with other covariates, and therefore use of a uniform overall inflation factor (i.e., λ_{GC}) may results in a loss of power (Price *et al.* 2006; Wang *et al.* 2012; Moore *et al.* 2019). Notably, $\lambda_{\mathrm{GC}} \approx 1$ is theoretically correct for the present study because we simulated only three major QTLs (Table 3) and, therefore, most *p* values should follow the expected distribution. In this study, the use of AMMI-type Q×E effect terms resulted in deflated λ_{GC} (Figure 3), although the AUC values were equivalent to those of the GGE-type Q×E effect (Figure 2). This result is attributed to the df of the LRT (Table 2). Because of the additive main effect term ($\boldsymbol{x}\beta$ in Eq. 5), the LRT that used AMMI-type Q×E effect terms had one more df than that using GGE-type Q×E effect terms (Table 3). Therefore, we concluded that using GGE-type Q×E effect terms is preferable to GWAS for Q×E effects. However, a desirable feature of the AMMI-type Q×E effect terms is the separation of a QTL effect into additive main and Q×E effects. To perform exhaustive Q×E effect analysis with minimal oversight, our

findings suggest that (1) genome-wide analysis with LMMs should be performed using GGE-type Q×E effect terms, and (2) Q×E effects should be assessed using AMMI-type Q×E effect terms only for significant signals detected in (1).

In this study, the random effect '$\boldsymbol{u_G} + \sum_{tr}^{TR} \boldsymbol{u_{tr}}$' showed low precision, F-measure, and AUC values (Figure 2) and was less effective in controlling λ_{GC} (Figure 3), although its efficacy has been demonstrated in a previous study (Sousa *et al.* 2017). One possible reason for this discrepancy is that of differences in methods used to estimate the variance component in LMM fitting. In this study, we performed direct estimation using the average information restricted maximum likelihood method without fixed QTL effect terms (Gilmour *et al.* 1995; Perdry and Dandine-Roulland 2018). Conversely the variance component estimation and fitting method used in Sousa *et al.* (2017) were based on a Bayesian approach that updates the parameters using a Gibbs sampler (Pérez and de Los Campos 2014). Because Sousa *et al.* (2017) performed model fitting to construct prediction models for GS, one model fitting process per trait was sufficient. Conversely, GWAS must perform model fitting for every marker in the data, which can be computationally unfeasible under the Bayesian approach. Therefore, we did not use the Bayesian approach for model fitting in this study, even though it has the potential to increase the power of QTL detection.

The findings of the present study can be summarized as follows. First, LMMs including Q×E effect terms are necessary to detect QTLs with Q×E effects (Figure 2). Second, the random effect '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$' is necessary to control λ_{GC} and calculate FDR appropriately (Figure 3). We applied these findings to real phenotypic data for tomatoes to detect QTLs associated with agronomic traits. A fruit weight QTL on chromosome 9 was located on the *fw9.1* region, which was identified in a genetic mapping population derived from a cross between cultivated and wild tomato (Tanksley *et al.* 1996). A fruit set QTL on chromosome 1 was close to *fset1.3*, which was detected in a genetic mapping population derived from eight tomato varieties (Diouf *et al.* 2020). Interestingly, Diouf *et al.* (2020) suggested that *fset1.3* has Q×E effects, which is consistent with the results of this study (Figure 4 and Table 4). Thus, the findings of this study have the potential to contribute to the identification of more QTLs with Q×E effects.

ACKNOWLEDGMENTS

This work was supported by the Japan Science and Technology Agency (JST) PRESTO (Grant number, JPMJPR16Q9), JSPS KAKENHI (Grant number, 19H02951), and Meiji University (Interdisciplinary Research Program in Mathematical Modeling and Life Sciences) to E.Y.

COMPETING INTERESTS

The authors declare no competing interests.

LITERATURE CITED

- Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B 57: 289−300.
- Bian, Y., and J. B. Holland, 2017 Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. Heredity 118: 585−593.
- Boer, M. P., D. Wright, L. Feng, D. W. Podlich, L. Luo *et al.*, 2007 A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177: 1801−1813.
- Cooper, M., and I. H. DeLacy, 1994 Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theor. Appl. Genet. 88: 561−572.
- Cortes, L. T., Z. W. Zhang, and J. M. Yu, 2020 Status and prospects of genome-wide association studies in plants. Plant Genome doi:10.1002/tpg2.20077.
- Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín *et al.*, 2017 Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22: 961–975.
- Cuevas, J., J. Crossa, O. A. Montesinos-López, J. Burgueño, P. Pérez-Rodríguez, and G. de los Campos, 2017 Bayesian genomic prediction with genotype × environment interaction kernel

models. G3 (Bethesda) 7: 41−53.

- Dahl, A., K. Nguyen, N. Cai, M. J. Gandal, J. Flint, and N. Zaitlen, 2020 A Robust Method Uncovers Significant Context-Specific Heritability in Diverse Complex Traits. Am. J. Hum. Genet. 106: 71−91.
- Devlin, B., and K. Roeder, 1999 Genomic control for association studies. Biometrics 55: 997−1004.
- Diouf, I., L. Derivot, S. Koussevitzky, Y. Carretero, F. Bitton *et al.*, 2020 Genetic basis of phenotypic plasticity and genotype x environment interaction in a multi-parental tomato population. J. Exp. Bot. 71: 5365–5376.
- Endelman, J. B., and J. L. Jannink, 2012 Shrinkage estimation of the realized relationship matrix. G3 (Bethesda) 2: 1405–1413.
- Gage, J. L., N. De Leon, and M. K. Clayton, 2018 Comparing genome-wide association study results from different measurements of an underlying phenotype. G3 (Bethesda) 8: 3715–3722.
- Gauch Jr, H. G., 1988 Model selection and validation for yield trials with interaction. Biometrics 44: 705−715.
- 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 51, 1440–1450.
- Hamazaki, K., and H. Iwata, 2020 RAINBOW: Haplotype-based genome-wide association

study using a novel SNP-set method. PLoS Comput. Biol. 16: e1007663.

- Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt *et al.*, 2014 A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127: 595−607.
- Li, F., B. Chen, K. Xu, J. Wu, W. Song *et al*., 2014 Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). DNA Res. 21: 355–367.
- Lopez-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland *et al.*, 2015 Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3 (Bethesda) 5: 569−582.
- Malosetti, M., J. M. Ribaut, and F. A. van Eeuwijk, 2013 The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Front. Physiol. 4: 44.
- Mathews, K. L., M. Malosetti, S. Chapman, L. McIntyre, M. Reynolds *et al.*, 2008 Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor. Appl. Genet. 117: 1077−1091.
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

- Moore, R., F. P. Casale, M. J. Bonder, D. Horta, L. Franke *et al.*, 2019 A linear mixed-model approach to study multivariate gene–environment interactions. Nat. Genet. 51: 180−186.
- Perdry, H., and C. Dandine-Roulland, 2018 gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. R package version 1.5.6. Available at: https://CRAN.R-project.org/package=gaston
- Pérez, P., and G. de Los Campos, 2014 Genome-wide regression and prediction with the BGLR statistical package. Genetics 198: 483−495.
- Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick *et al.*, 2006 Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38: 904–909.
- R Core Team, 2019 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/
- Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek *et al.*, 2011 pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77.
- Saber, M. M., and B. J. Shapiro, 2020 Benchmarking bacterial genome-wide association study methods using simulated genomes and phenotypes. Microb. Genom. 6: e000337.
- Shafquat, A., R. G. Crystal, and J. G. Mezey, 2020 Identifying novel associations in GWAS by hierarchical Bayesian latent variable detection of differentially misclassified phenotypes. BMC

bioinformatics 21: 1−25.

- Sousa, M. B., J. Cuevas, E. G. de Oliveira Couto, P. Pérez-Rodríguez, D. Jarquín *et al.*, 2017 Genomic-enabled prediction in maize using kernel models with genotype× environment interaction. G3 (Bethesda) 7: 1995−2014.
- Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100: 9440−9445.
- Tanksley, S. D., S. Grandillo,T. M. Fulton, D. Zamir,Y. Eshed *et al.*, 1996 Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor. Appl. Genet. 92: 213−224.
- Voorman, A., T. Lumley, B. McKnight, and K. Rice, 2011 Behavior of QQ-plots and genomic control in studies of gene-environment interaction. PloS One 6: e19416.
- Wang, S., W. Chen, X. Chen, F. Hu, K. J. Archer *et al.*, 2012 Double genomic control is not effective to correct for population stratification in meta-analysis for genome-wide association studies. Front. Genet. 3: 300.
- Yamamoto, E., H. Matsunaga, A. Onogi, H. Kajiya-Kanegae, M. Minamikawa *et al.*, 2016 A simulation-based breeding design that uses whole-genome prediction in tomato. Sci. Rep. 6: 19454.
- Yamamoto, E., H. Matsunaga, A. Onogi, A. Ohyama, K. Miyatake, *et al.*, 2017 Efficiency of

genomic selection for breeding population design and phenotype prediction in tomato. Heredity 118: 202−209.

- Yan, W., L. A. Hunt, Q. Sheng, and Z. Szlavnics, 2000 Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci. 40: 597−605.
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011 GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88: 76–82.
- Yu, J., G. Pressoir, W. H. Briggs, I. V. Bi, M. Yamasaki *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203−208.

FIGURE LEGENDS

Figure 1

Environments and trials used in this study. Cropping season (i.e., winter or spring harvest cropping seasons) was the environmental variable in the tomato phenotypic dataset. Each trial represents a period of plant growth and phenotype evaluation.

Figure 2

Bar plots of power to detect (A) persistence, (B) divergence, (C) convergence, and (D) crossover quantitative trait loci (QTLs) in the simulated phenotypes, under the assumption of multiple environments and multiple trials. Labels on the *x*-axes correspond to the tests described in Table 1 and 2. Recall, precision, and F-measure were calculated using a false discovery rate of 0.05 as the genome-wide significance threshold. Values represent means of 100 simulations.

Figure 3

Evaluation of linear mixed models (LMMs) based on deviation of the *p* value distribution from the null hypothesis. (A) Box and whisker plots of genomic inflation factors for *p* values. Red horizontal line indicates the theoretically expected value (i.e.,

λ_{GC} = 1). (B) Quantile–quantile (Q–Q) plots of *p* values obtained from LMMs that include random effect '$\boldsymbol{u_G} + \boldsymbol{u_{GE}} + \boldsymbol{u_{GT}}$.' (C) Q–Q plots of *p* values obtained from LMMs with GGE-type Q×E fixed effect terms. In (B) and (C), *p* values shown in the panel represent means of 1,000 quantiles from 400 experiments (100 simulations × 4 conditions in Table 3), and a black dashed line indicates the *p* value distribution under the null hypothesis.

Figure 4

Genome-wide association study Manhattan plots for tomato agronomic traits. Red horizontal lines indicate the genome-wide significance threshold (false discovery rate < 0.05). (A) Average fruit weight. (B) Fruit set ratio.

Table 1. Terms used for the linear mixed models examined in this study.

Terms	Equation in main text	Definition
		Fixed effects
Additive main	Eq. 3	No Q×E effect assumed.
AMMI-type Q×E	Eq. 5	QTL effect divided into additive main and Q×E effects.
GGE-type Q×E	Eq. 6	Additive main and Q×E effects scored together.
		Random effects
$\boldsymbol{u_G}$	Eq. 7	No G×E random effects assumed.
$\boldsymbol{u_{GE}}$	Eq. 8	G×E random effects not correlated between environments but have common variance.
$\boldsymbol{u_{GT}}$	Eq. 10	G×T random effects not correlated between environments but have common variance.
$\sum_{tr}^{TR} \boldsymbol{u_{tr}}$	Eq. 13	G×T effects not correlated between trials and have different variances.

AMMI, Additive Main effects and Multiplicative Interaction model; G×E, genotype-by-environment interaction; GGE, Genotype main effects and G×E model; G×T, genotype-by-trial interaction; Q×E, QTL-by-environment interaction; QTL, quantitative trait locus.

Table 2. Hypotheses and terms of the chi-square tests performed in this study.

Test	Terms	Deviance formula[a]	Degrees of freedom[b]
All QTL-effect terms	Additive main	$-2 \times (LL_{Eq.3} - LL_{Eq.16})$	1
	AMMI-type Q×E	$-2 \times (LL_{Eq.5} - LL_{Eq.16})$	$L + 1$
	GGE-type Q×E	$-2 \times (LL_{Eq.6} - LL_{Eq.16})$	L
Interaction terms	AMMI-type Q×E	$-2 \times (LL_{Eq.5} - LL_{Eq.3})$	L
	GGE-type Q×E	$-2 \times (LL_{Eq.6} - LL_{Eq.3})$	$L - 1$

[a] $LL_{Eq.n}$ indicates the log-likelihood of Eq. *n* in the main text.

[b] L indicates the number of environments included in the test.

672

Table 3. Parameters used for simulated phenotypes.

Condition	Major QTLs[a]						Random effects[b]
	Trial		1	2	3	4	w_G:w_{GE}:w_{GT}
Persistence	Environment		A	A	B	B	1:1:1
	Effect size		1	1	1	1	
	Realized PVE_{QTL}[c] per trial	Min	0.079	0.082	0.078	0.074	
		Median	0.106	0.106	0.104	0.104	
		Max	0.165	0.147	0.153	0.160	
Divergence	Environment		A	A	B	B	1:1:1
	Effect size		0.5	0.5	1	1	
	Realized PVE_{QTL} per trial	Min	0.025	0.025	0.077	0.065	
		Median	0.043	0.047	0.127	0.130	
		Max	0.076	0.078	0.228	0.207	
Convergence	Environment		A	A	B	B	1:1:1
	Effect size		0	0	1	1	
	Realized PVE_{QTL} per trial	Min	0.000	0.000	0.036	0.030	
		Median	0.000	0.000	0.123	0.127	
		Max	0.000	0.000	0.280	0.255	
Crossover	Environment		A	A	B	B	1:1:1
	Effect size		–1	–1	1	1	
	Realized PVE_{QTL} per trial	Min	0.010	0.011	0.011	0.010	
		Median	0.057	0.064	0.057	0.063	
		Max	0.165	0.174	0.146	0.153	

[a] Effect size and environment indicate effect size of the simulated QTLs for each trial and the environment identifier for the trial, respectively.

[b] Values in this column are relative weights of random effects (see Eq. 22).

[c] PVE_{QTL}: proportion of variance explained by each major QTL.

Table 4. Statistics obtained in the genome-wide association study of tomato phenotypic data.

Trait	SNP	Chr	pos (cM)	$-\log_{10}(p)$				Environment	
				Add[a]	All[b]	Int[c]		Winter	Spring
Fruit weight	AX-95796523	9	53.0	5.03	5.24	1.45	θ_l[d]	1.03	0.59
							Wald[e]	6.29	1.90
	AX-95776890	4	29.7	3.90	3.18	0.00	θ_l	–0.63	–0.62
							Wald	3.67	3.18
	AX-95774753	8	101.4	3.58	2.93	0.17	θ_l	–0.50	–0.56
							Wald	3.11	3.18
	AX-107533066	1	106.4	2.85	3.55	1.88	θ_l	–0.50	–1.02
							Wald	1.58	4.25
	AX-107537279	10	40.5	0.45	3.16	3.66	θ_l	–0.18	1.45
							Wald	0.16	2.28
Fruit set	AX-95767902	12	19.9	3.61	2.93	0.09	θ_l	–0.53	–0.60
							Wald	2.41	2.92
	AX-107525913	1	117.8	0.00	4.15	4.91	θ_l	–0.23	0.41
							Wald	1.10	2.15

[a] Additive main effect

[b] All QTL-effect terms

[c] Interaction terms

[d] Estimated effect size

[e] $-\log_{10}(p)$ value from chi-square test of the Wald score

APPENDIX

Dominant effect QTLs

Because the materials used in this study were F_1 varieties, we extended the analyses in the main text to include dominant effect QTLs. In this appendix, we briefly summarize the methods and results of this analysis.

LMMs: To include dominant effects, Eq. 1 was extended as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + f_{\text{Fixed}}(\boldsymbol{x}, \boldsymbol{h}) + f_{\text{Random}}(\boldsymbol{G}, \boldsymbol{D}) + \boldsymbol{\varepsilon}, \quad (25)$$

where $f_{\text{Fixed}}(\boldsymbol{x}, \boldsymbol{h})$ and $f_{\text{Random}}(\boldsymbol{G}, \boldsymbol{D})$ are the fixed and the random effect terms that include dominant effects, respectively. $\boldsymbol{h}$ is an $n \times 1$ vector of SNP genotype heterozygosity coded as {0, 1, 0} = {aa, Aa, AA}. $\boldsymbol{D}$ is the dominance covariance matrix between individuals *j* and *k* (D_{jk}), and was defined as follows:

$$D_{jk} = \sum_{i=1}^{I} \{h_{ij} - 2p_i(1-p_i)\}\{h_{ik} - 2p_i(1-p_i)\} / \{2p_i(1-p_i)[1 - 2p_i(1-p_i)]\}, \quad (26)$$

where h_{ij} is the coded SNP genotype heterozygosity value for the *i*-th SNP of the *j*-th individual. The derivation of $\boldsymbol{D}$ is explained in detail in Su *et al.* (2012). Without consideration of differences among environments, the random effects for $\boldsymbol{D}$ are modeled as follows:

$$\boldsymbol{d_D} \sim MVN(\boldsymbol{0}, [\boldsymbol{Z_D D Z_D'}]\sigma_D^2). \quad (27)$$

We omitted the derivation of dominant random effects for G×E (i.e., $\boldsymbol{d_D}$, $\boldsymbol{d_{DE}}$, $\boldsymbol{d_{DT}}$, and $\sum_{tr}^{TR} \boldsymbol{d_{tr}}$) because these are analogous to the equations used for additive G×E random effects (i.e., $\boldsymbol{u_G}$, $\boldsymbol{u_{GE}}$, $\boldsymbol{u_{GT}}$, and $\sum_{tr}^{TR} \boldsymbol{u_{tr}}$). When the dominant effect terms were included, Eq. 3 was extended as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + \boldsymbol{x}\alpha + \boldsymbol{h}\tau + f_{\text{Random}}(\boldsymbol{G}, \boldsymbol{D}) + \boldsymbol{\varepsilon}, \quad (28)$$

where τ is the dominant effect for a QTL in LD with the SNP. An LMM including the AMMI-type Q×E effect terms (Eq. 5) was extended as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + \boldsymbol{x}\beta + \sum_l^L \{(\boldsymbol{\pi}_l \circ \boldsymbol{x})\gamma_l\} + \boldsymbol{h}\delta + \sum_l^L \{(\boldsymbol{\pi}_l \circ \boldsymbol{h})\vartheta_l\} + f_{\text{Random}}(\boldsymbol{G}, \boldsymbol{D}) + \boldsymbol{\varepsilon}, \quad (29)$$

where δ is the dominant QTL effect not specific to the environment, and ϑ_l is the dominant QTL effect specific to the *l*-th environment. An LMM including the GGE-type Q×E effect terms (Eq. 6) was extended as follows:

$$\boldsymbol{y} = \boldsymbol{Tt} + \boldsymbol{Ss} + \sum_l^L \{(\boldsymbol{\pi}_l \circ \boldsymbol{x})\zeta_l\} + \sum_l^L \{(\boldsymbol{\pi}_l \circ \boldsymbol{h})\kappa_l\} + f_{\text{Random}}(\boldsymbol{G}, \boldsymbol{D}) + \boldsymbol{\varepsilon}. \quad (30).$$

where κ_l is the dominant QTL effect in the *l*-th environment.

Simulation conditions: The equation used to simulate phenotypic values, including dominant effects, was as follows:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_{tr}} \\ \vdots \\ \boldsymbol{y_{TR}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q_d} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{Q_d} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \cdots & \boldsymbol{Q_d} \end{bmatrix} \begin{bmatrix} \boldsymbol{q_1} \\ \vdots \\ \boldsymbol{q_{tr}} \\ \vdots \\ \boldsymbol{q_{TR}} \end{bmatrix} + \frac{\{(\boldsymbol{u_G} \times w_G) + (\boldsymbol{u_{GE}} \times w_{GE}) + (\boldsymbol{u_{GT}} \times w_{GT})\}}{\omega} + \frac{\boldsymbol{\varepsilon}}{\varphi}. \quad (31)$$

The difference between Eqs. 22 and 31 is that the coding of the SNP genotype values $\boldsymbol{Q_a}$ in Eq. 22 is {0, 1, 2} = {aa, Aa, AA}, whereas that of $\boldsymbol{Q_d}$ in Eq. 31 is {0, 3, 2} = {aa, Aa, AA}. This dominant effect design was based on a real dominant-effect QTL in tomatoes (Krieger *et al.* 2010).

Results and Discussion: The results obtained using the genomic inflation factor (λ_{GC}) were similar to those of LMMs without dominant effects, except that the values were slightly deflated in all tests (Figure S1A), which was likely to be because the additional fixed terms in the dominant effect LMMs resulted in larger df in the subsequent chi-square test. The inclusion of dominant effect terms in the LMMs showed an advantage in terms of both recall and AUC for detecting the dominant-effect QTLs simulated in this study (Figure S1B). This result is reasonable because models including dominant effect terms generally provide a more effective fit to phenotypic values with a dominant effect (Letter *et al.* 2007). Nevertheless, LMMs that included dominant effect terms did not detect dominant-effect QTLs for the real agronomic trait data used in this study (Figure S2). One possible reason for this result is that the deflation of $-\log_{10}(p)$

values, as indicated in λ_{GC} (Figure S1A), resulted in false negative QTLs. Another possibility is that, as suggested in some previous studies, the contribution of dominant effects is less evident for quantitative traits, and the use of an additive effect model is sufficient for genome-wide analysis (Varona *et al.* 2018). In addition, another problem in the inclusion of dominant effect terms have been discussed by Balding (2006). That is, the existence of few homozygotes for an allele results in the neglect of the homozygote genotypes and linear regression fitting between homozygote for the another allele and heterozygote genotypes (Balding 2006). Because the materials used in the present study were F_1 varieties that include many genetic loci where heterozygous genotypes are major and homozygous are minor, this scenario is possible.

FIGURE LEGENDS

Figure S1

Evaluation of linear mixed models (LMMs) including dominant effect terms. (A) Box and whisker plots of genomic inflation factors (λ_{GC}) for *p* values. Red horizontal line indicates the theoretically expected value (i.e., $\lambda_{GC} = 1$). (B) Bar plots of power to detect quantitative trait loci (QTLs) in simulated phenotypes, assuming multiple environments and multiple trials. Recall values were calculated using a false discovery rate of 0.05 as

the genome-wide significance threshold. Values represent means of 100 simulations. The fixed and random effect terms used in this analysis was GGE-type Q×E and '$\boldsymbol{u_G}$ + $\boldsymbol{u_{GE}}$ + $\boldsymbol{u_{GT}}$ + $\boldsymbol{d_D}$ + $\boldsymbol{d_{DE}}$ + $\boldsymbol{d_{DT}}$,' respectively.

Figure S2

Genome-wide association study Manhattan plots of tomato agronomic traits. Red horizontal lines indicate the genome-wide significance threshold (false discovery rate < 0.05). (A) Average fruit weight. (B) Fruit set ratio.

LITERATURE CITED

- Balding, D. J., 2006 A tutorial on statistical methods for population association studies. Nat. Rev. Genet. 7: 781–791.
- Krieger, U., Z. B. Lippman, and D. Zamir, 2010 The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nat. Genet. 42: 459−463.
- Letter, G., C. Lange, and J. N. Hirschhorn, 2007 Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet. Epidemiol. 31: 358–362.
- Varona, L., A. Legarra, M. A. Toro, and Z. G. Vitezica, 2018 Non-additive effects in genomic

selection. Front. Genet. 9: 78.

- Su, G., O. F. Christensen, T. Ostersen, M. Henryon, and M. S. Lund, 2012 Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One 7: e45293.

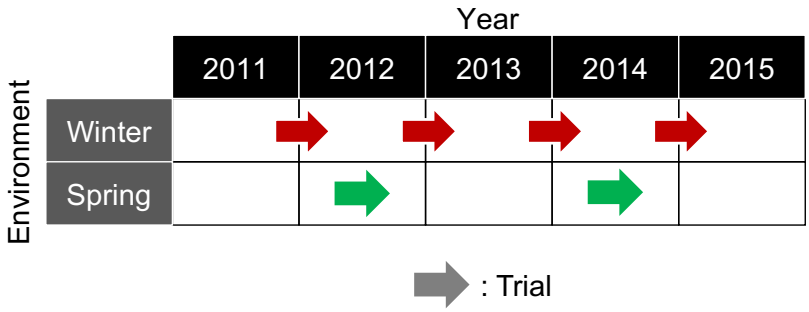

Year
2011
2012
2013
2014
2015
Environment
Winter
Spring
: Trial

Figure 1

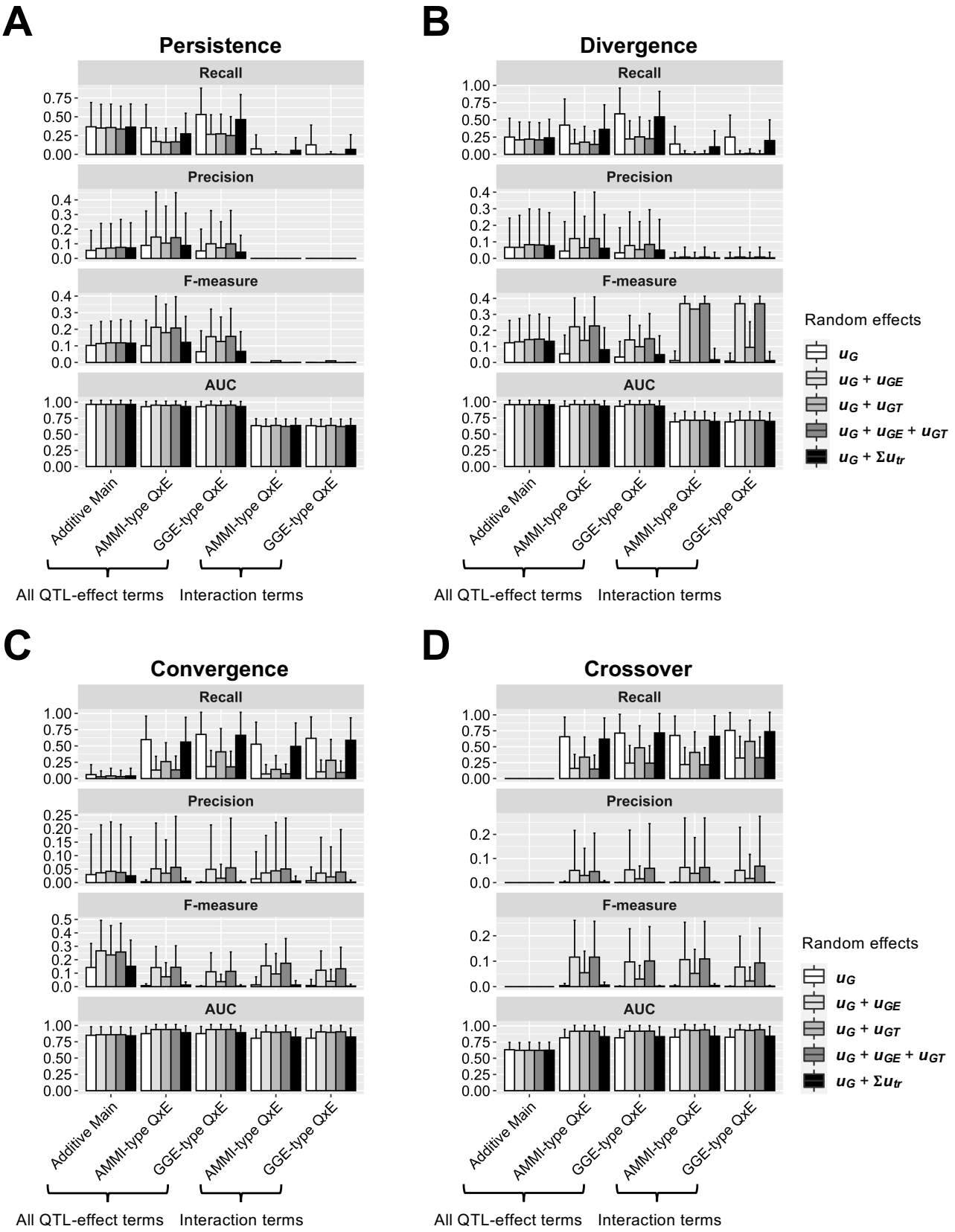

A
Persistence
B
Divergence
C
Convergence
D
Crossover
Recall
Precision
F-measure
AUC
Additive Main
AMMI-type QxE
GGE-type QxE
AMMI-type QxE
GGE-type QxE
All QTL-effect terms
Interaction terms
Random effects
u_G
$u_G + u_{GE}$
$u_G + u_{GT}$
$u_G + u_{GE} + u_{GT}$
$u_G + \Sigma u_{tr}$

Figure 2

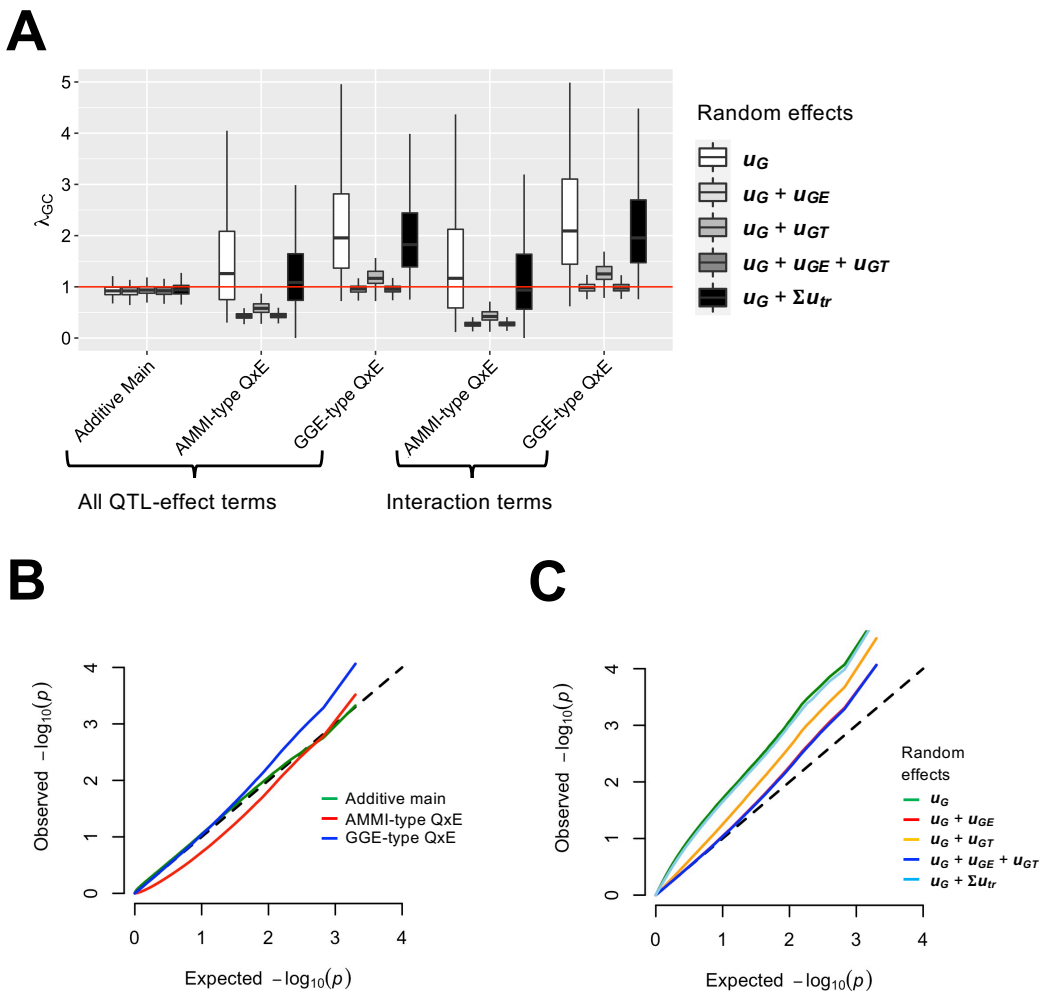

A
λ_{GC}
Random effects
u_G
$u_G + u_{GE}$
$u_G + u_{GT}$
$u_G + u_{GE} + u_{GT}$
$u_G + \Sigma u_{tr}$
Additive Main
AMMI-type QxE
GGE-type QxE
AMMI-type QxE
GGE-type QxE
All QTL-effect terms
Interaction terms
B
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Additive main
AMMI-type QxE
GGE-type QxE
C
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Random effects
u_G
$u_G + u_{GE}$
$u_G + u_{GT}$
$u_G + u_{GE} + u_{GT}$
$u_G + \Sigma u_{tr}$

Figure 3

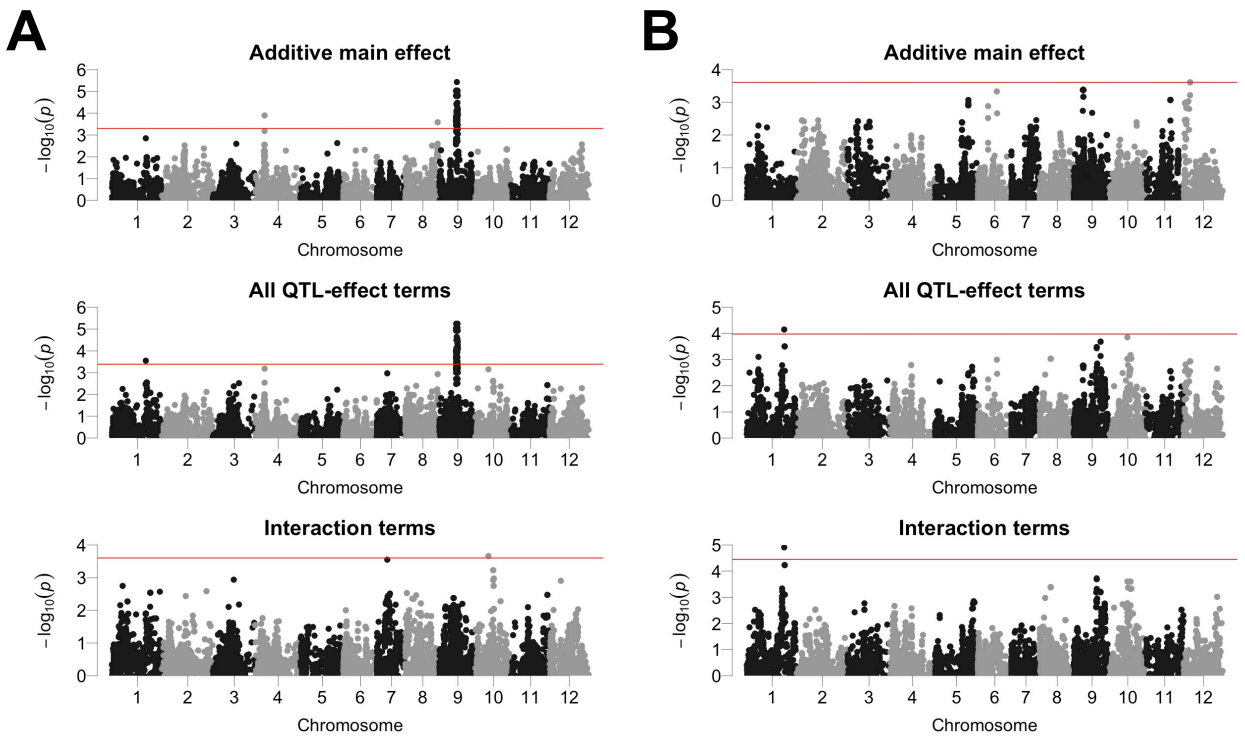
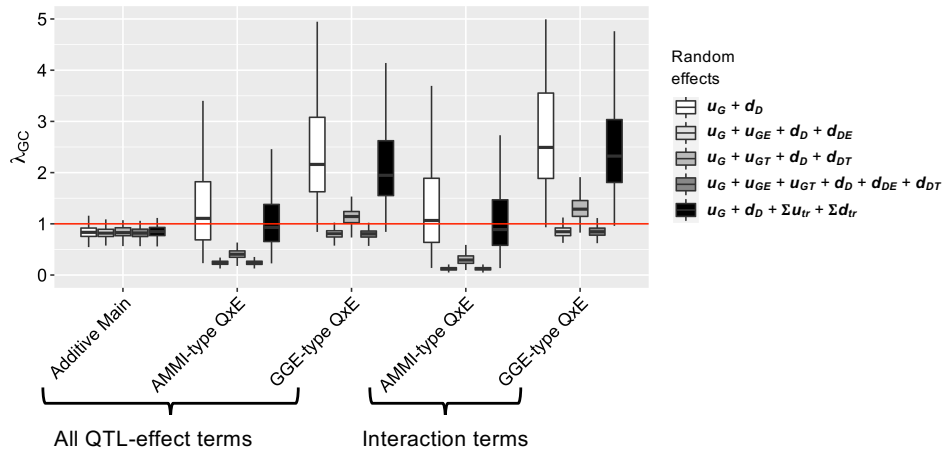

A
Additive main effect
All QTL-effect terms
Interaction terms
B
Additive main effect
All QTL-effect terms
Interaction terms
$-\log_{10}(p)$
Chromosome

Figure 4

A


λGC
5
4
3
2
1
0
Random effects
uG + dD
uG + uGE + dD + dDE
uG + uGT + dD + dDT
uG + uGE + uGT + dD + dDE + dDT
uG + dD + Σutr + Σdtr
Additive Main
AMMI-type QxE
GGE-type QxE
AMMI-type QxE
GGE-type QxE
All QTL-effect terms
Interaction terms

B

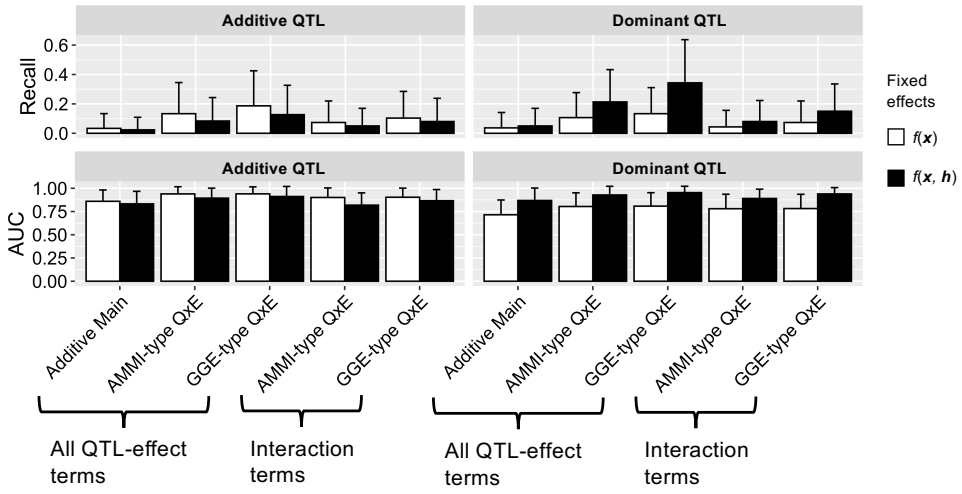

Additive QTL
Dominant QTL
Recall
0.6
0.4
0.2
0.0
AUC
1.00
0.75
0.50
0.25
0.00
Fixed effects
f(x)
f(x, h)
Additive Main
AMMI-type QxE
GGE-type QxE
AMMI-type QxE
GGE-type QxE
Additive Main
AMMI-type QxE
GGE-type QxE
AMMI-type QxE
GGE-type QxE
All QTL-effect terms
Interaction terms
All QTL-effect terms
Interaction terms

Figure S1

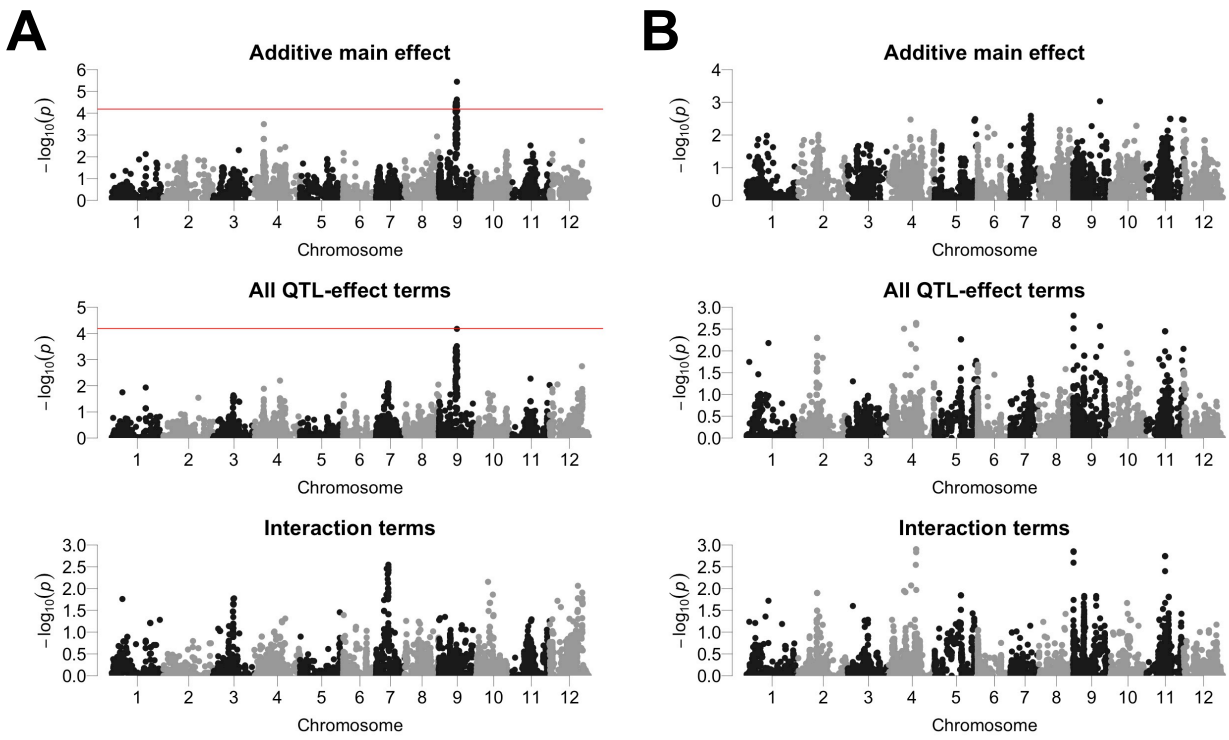

A
Additive main effect
All QTL-effect terms
Interaction terms
B
Additive main effect
All QTL-effect terms
Interaction terms
−log10(p)
Chromosome

Figure S2