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

KEYWORDS: Genome-wide association study (GWAS), Quantitative trait locus (QTL), Genotype-by-environment interaction (GEI), QTL-by-environment interaction (QEI), 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 (GEIs) are important for not only a precise understanding of genotype–phenotype relationships but also improving the environmental adaptability of crops. Although many formulae have been proposed to model GEI effects, a comprehensive comparison of their efficacy in genome-wide association studies (GWASs) has not been performed. Therefore, the advantages and disadvantages of the formulae are not well recognized. In this study, linear mixed models (LMMs) consisting of various combinations of foreground fixed genetic and background random genetic effect terms were constructed. Next, the power to detect quantitative trait loci (QTLs) with GEI effects was compared across the LMMs by using simulation. The fixed genetic effect terms of the genotype main effects and GEI (GGE) model were preferred over those based on the additive main effects and multiplicative interaction (AMMI) model because the latter showed *p*-value inflation, whereas the former yielded theoretically expected *p*-value distribution. With regard to the background random genetic effects, inclusion of genotype-by-trial interaction effects has been recommended to achieve high power and robustness when phenotype data are obtained from multiple environments and multiple trials. The recommended form of LMM was applied to GWASs performed using real agronomic trait data of tomato F_1

varieties (*Solanum lycopersicum* L.) that were obtained from two different cropping seasons. The GWASs detected QTLs with not only persistent effects across the cropping seasons, but also cropping season-specific effects. Thus, the application of GWAS strategy to phenotypic data from multiple environments and trials allowed the detection of more QTLs with GEI effects.

INTRODUCTION

Phenotypes are determined by not only genetic potential but also environmental growth conditions. More specifically, the rank of phenotypic values has been shown to often change 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, and genotype-by-environment interactions (GEIs) occur when the slope of the reaction norm differs for different genotypes in an environment (Malosetti *et al.* 2013). Analysis of GEIs is important for not only a precise understanding of genotype–phenotype relationships but also the designing of crop varieties that fit a certain environment.

To date, various statistical approaches have been proposed for GEI analysis (Malosetti *et al.* 2013). The simplest approach is analysis of variance, which compares the mean and variance of phenotypic values of genotypes in multiple environments. However, practically, more flexible methods are being used for GEI analysis. In the additive main effects and multiplicative interaction (AMMI) model, the genetic contribution for a phenotypic value is divided into fixed genetic main effects (i.e., genetic effect irrelevant to environments) and GEI effects (i.e., reaction norms to

environments; Gauch 1988). The reaction norms in the AMMI model are expressed as multiplicative interaction terms consisting of products of genotypic score and hypothetical environmental score. Next, principal component (PC) analysis of these scores is conducted. In the PC analysis, genotypic scores are explained by environmental scores and vice versa. Therefore, GEIs in the AMMI model can be explained by a two-axis biplot (i.e., PC1 and 2). This enables a graphical representation of the interactivity between genotypes and environments. In particular, a genotype will show the highest performance in the closest environment in the biplot. The genotype main effects and GEI (GGE) model resembles the AMMI model, but focuses more on the total genetic variation in each environment (Yan *et al.* 2000). In the GGE model, both genetic main effects and reaction norms are scored together; therefore, genotypes in the GGE biplot represent the overall performance, whereas the AMMI biplot represents only the GEI part of the phenotype.

Availability of molecular markers has enabled the analysis of interactions between environmental conditions and each gene and/or quantitative trait locus (QTL). The QTL-by-environment interaction (QEI) analysis requires the separation of overall GEI effects into the effect of a focal QTL and those of the other loci (i.e., background

genetic effects). This separation is achieved using linear mixed models (LMMs) that account for background genetic effects as random effects (Boer *et al.* 2007; Mathews *et al.* 2008; Sabadin *et al.* 2012). In the LMMs, background random genetic effects follow a multivariate normal distribution whose variance–covariance is a product of the genetic relationship matrix (GRM) and the variance–covariance matrix of the environmental values or design matrix of categorically classified environmental conditions. Random effects are being increasingly considered to account for GEI effect in genomic selection (GS) studies. In GS, a training population, which has been phenotyped and genotyped, is used to construct a model that predicts the genetic potential of unphenotyped individuals by using genome-wide genotype data (Meuwissen *et al.* 2001). GS models that account for GEI effects need to be constructed for increasing predictability and designing genotypes that are robust to multiple environments and/or suitable for specific environments (Crossa *et al.* 2017). In the GS models that account for GEI effects, the overall random genetic effects are separated into main (environment-independent) and environment-specific random effects. The main random effects follow a multivariate normal distribution with GRM, as the variance–covariance and environment-specific random effects follow a multivariate normal distribution with a product of GRM and variance–covariance matrix for environmental differences (Heslot *et al.* 2014; Jarquín *et*

al. 2014; Lopez-Cruz *et al.* 2015; Cuevas *et al.* 2017). In the abovementioned method, environment-specific random genetic effects have common variance across the environments. e Sousa *et al.* (2017) extended the model to allow different variances across environments, which resulted in an increase in the prediction accuracy in some experiments.

Genome-wide association studies (GWASs) that analyze associations between molecular marker genotypes and phenotypic variance by using a diverse population set are now widely used for the identification of genes and/or QTLs associated with traits that have complicated genetic architecture (Lipka *et al.* 2015). The most standard GWAS method is LMM (Yu *et al.* 2006). Therefore, methods developed for GEI analysis in GS studies are directly applicable to GWASs. For example, in StructLMM (Moore *et al.* 2019), a product of GRM and variance–covariance matrix of environmental values is used to model the background random genetic effects. LMMs that allow different genotypic variances between environments, which was proposed by e Sousa *et al.* (2017), have been implemented in GxEMM (Dahl *et al.* 2020). These methods were designed for human and animal genetic studies in which a numerous samples (genotypes) are generally available. Currently increasing genotype, phenotype

and environmental data make the methods promising approach to reveal QEI effects of each gene.

Because plants cannot move from their habitat, considering GEI effects is essential when the phenotypes are measured in multiple environments (e.g., location and cropping season). One of the disadvantages of plant genetic studies against human and animal genetic studies is the small number of samples (genotypes) in an experiment. In general, less than 1,000 genotypes are phenotyped in each experiment owing to the limited field area size and work force (Atwell *et al.* 2010; Yano *et al.* 2016). Nonetheless, in plant genetic studies, phenotype measurement of a genotype can be repeated, leading to more accurate phenotypic evaluation. However, due to limitation in field size and work force, the replication trials are often performed multiple years. The unstable and extreme climate changes in recent years suggest genotype-by-trial interaction (GTI) effects on phenotypic values (Kornhuber *et al.* 2019). Thus, plant genetic studies often involve not only ordinary GEI effects that have been long discussed (Malosetti *et al.* 2013), but also unexpected GTI effects that are merely disturbance factors of genetic studies. That is, a QTL has an effect in an environment, but the effect size differs across trials because of unexpected GTI effects, and the QTL

might fail to represent the effect in some trials. This could be an unusual assumption because it violates the ordinary hypothesis that a gene must have a persistent effect in the same environment. However, we believe that such effects can occur in plant genetic studies under recent unstable climate conditions (Kornhuber *et al.* 2019). For example, a linear relationship has been shown to exist between temperature and various plant phenotypes such as height and flowering time (e.g., Millet *et al.* 2019), but phenotyping under excessively high and/or low temperatures might not follow a linear relationship because such conditions are stressful for plant growth (e.g., Yuan *et al.* 2019).

In GS studies, various statistical models for GEI analysis have been proposed, and their efficacy under unstable environmental conditions has been analyzed extensively (Heslot *et al.* 2014; Jarquín *et al.* 2014; Lopez-Cruz *et al.* 2015; Cuevas *et al.* 2017; e Sousa *et al.* 2017). However, the efficacy of statistical models for QTL detection (i.e., GWASs) has been less investigated. In the present study, we explored efficient GWAS methods to detect QTLs with QEI effects. To model the reaction norms of a focal QTL, we compared two methods that are analogous to the formula in the AMMI (Gauch 1988) and GGE models (Yan *et al.* 2000). Regarding the background random genetic effects, we compared the methods that have been used in GS studies

(Heslot *et al.* 2014; Jarquín *et al.* 2014; Lopez-Cruz *et al.* 2015; Cuevas *et al.* 2017; e Sousa *et al.* 2017). The combinations of modeling methods for focal QTLs and the background effects were compared to identify LMMs that represent robustness and high power to detect QTLs with QEI effects by conducting simulation experiments. Next, we performed GWASs by using the efficient LMMs to detect QEI effects by using real phenotype data of tomato (*Solanum lycopersicum* L.). Our results indicated that the implementation of both GEI and GTI effects in GWASs was necessary to achieve high QTL detection power for phenotypic data obtained from multiple environments and trials.

MATERIALS AND METHODS

Assumptions

Numerous studies have analyzed the GEI effects for various situations (e.g., Moore *et al.* 2019; Dahl *et al.* 2020). To clarify the aim of the present study and the differences from previous studies, we describe the assumptions in this study as follows:

1. The target traits are controlled by several major QTLs and numerous minor QTLs, and the major QTLs are the target of detection.
2. The number of samples (genotypes) is relatively smaller than the standards used in human and animal genetic studies.
3. Phenotypic values of a genotype were obtained for several trials that were performed across different years. Therefore, the phenotypic values were affected by the GTI effects.
4. The effect of QTLs may not be persistent between trials in the same environment.
5. Differences in environments or trials are represented as categorical variables.

Assumptions 1 to 3 are common situations in plant genetic studies. Assumptions 2 and 3 are different from those in ordinary human and animal genetic studies. Assumption 4 is the problem that we specifically addressed in this study. The β of column “Foreground”

in Table 1 represents the numerical expression of this assumption. Assumption 5 will be removed if numerical environmental values are available (e.g., growth temperature, humidity, etc.).

LMMs

We modeled QEI and GEI effects by using LMMs. In the LMMs, the total genetic effects were divided into foreground and background genetic effects:

$$y = \mu + f_{\text{fixed}}(x) + f_{\text{random}}(g) + \varepsilon \quad (1)$$

where $f_{\text{fixed}}(x)$ and $f_{\text{random}}(g)$ are the foreground and background genetic effects, respectively; y, μ, and ε indicate a vector for phenotypic values, population-wide mean, and residuals, respectively. In the standard GWAS, eq. 1 is described as

$$y = \mu + x\beta + f_{\text{random}}(g) + \varepsilon \quad (2).$$

where x is a vector of SNP genotype values coded as {0, 1, 2} = {aa, Aa, AA}. β models the SNP effect. Without consideration of differences in environments or trials, $f_{\text{random}}(g)$ is g that models the genetic background effects as random effects with

$$g \sim MVN(0, [Z_g G Z_g'] \sigma_g^2) \quad (3),$$

where G and Z_g represent the GRM and incidence matrix for the phenotype and genotype, respectively. σ_g^2 is the variance for g. In the GRM, the element G_{jk} is

defined as

$$G_{jk} = \sum_{i=1}^{m} (x_{ij} - 2p_i)(x_{ik} - 2p_i)/2p_i(1 - p_i) \quad (4),$$

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 m is the total number of markers. The GRM was calculated using the *A.mat* function in the R package "rrBLUP" (Endelman 2011). When the dominance effect terms were included, eq. 2 was extended as

$$y = \mu + x\beta + d\gamma + f_{\text{random}}(g) + \varepsilon \quad (5),$$

where d is an indicator of SNP genotype heterozygosity coded as {0, 1, 0} = {aa, Aa, AA}. γ is the dominance effect for the SNP. The following sections discuss the extension of the basic LMMs to model the QEI and GEI effects. The extensions were performed for both foreground and background genetic effects.

Foreground genetic effects: In this study, the foreground GEI effects represent the reaction norms of a focal QTL. Hereafter, we refer to the foreground GEI effects as QEI effects. To model QEIs as fixed effects in LMMs, we used two types of formulae. The first is

$$y = \mu + x\alpha + \sum_{l}^{L} x\beta_l + f_{\text{random}}(g) + \varepsilon \quad (6),$$

where α is the SNP effect common to all environments, and β_l is the SNP effect

specific to the *l*-th environment or experimental trial. In the equation, $x\alpha$ is analogous to the additive main genetic effect in the AMMI model (Gauch 1988). Therefore, we named the fixed effect formula in eq. 6 (i.e., $x\alpha + \sum_l^L x\beta_l$) as fAMMI. Equation 6 can be extended to include dominance effects.

$$y = \mu + x\alpha + d\tau + \sum_l^L x\beta_l + \sum_l^L d\gamma_l + f_{\text{random}}(g) + \varepsilon \quad (7).$$

In eq. 7, τ is a dominance effect irrelevant to environmental differences. γ_l is the dominance effect specific to the *l*-th environment or trial. We named the fixed effect formula in eq. 7 as fAMMI_d. The second formula used to model QEI is

$$y = \mu + \sum_l^L x\beta_l + f_{\text{random}}(g) + \varepsilon \quad (8).$$

Unlike eq. 6, eq. 8 does not include the additive main genetic effect term (i.e., $x\alpha$). Therefore, this QEI effect formula is analogous to the GGE model (Yan *et al.* 2000). We named the fixed effects in eq. 7 (i.e., $\sum_l^L x\beta_l$) as fGGE. The extension of eq. 8 to include dominance effects is

$$y = \mu + \sum_l^L x\beta_l + \sum_l^L d\gamma_l + f_{\text{random}}(g) + \varepsilon \quad (9).$$

We named the fixed effects formula in eq. 9 as fGGE_d.

Background genetic effects: Appropriate statistical modeling for background genetic effects has been the key for efficient GWASs. In the standard GWASs performed using

LMMs, genetic background effects are modeled as random effects as in eq. 3 (Yu *et al.* 2006). The first background genetic effect model in the present study is the same as the standard GWAS, which is represented in eq. 3. Hereafter, we refer to this standard background genetic effect model as rgeG, in which random genetic effects are shown as g. Notably, rgeG does not include reaction norms for the GEIs. Therefore, we extended eqs. 2 and 5 as

$$y = \mu + f_{\text{fixed}}(x) + g + ge + \varepsilon \quad (10)$$

where ge models the GEI effects with

$$ge \sim MVN\left(0, \left[Z_g G Z_g'\right] \circ [Z_e Z_e']\sigma_{ge}^2\right) \quad (11).$$

Z_e represents the incidence matrix for the phenotype and focal environmental differences (e.g., location and cropping season). Thus, eq. 10 allows different random genetic effects between environments. The same formula for random genetic effects was used in GWASs and implemented in StructLMM (Moore *et al.* 2019). In this study, we named this formula (i.e., $g + ge$) rgeGE. The third random genetic effect model is

$$y = \mu + f_{\text{fixed}}(x) + g + gt + \varepsilon \quad (12)$$

where gt models the reaction norm for GTI effects with

$$gt \sim MVN(0, \left[Z_g G Z_g'\right] \circ [Z_t Z_t']\sigma_{gt}^2) \quad (13).$$

Z_t represents the incidence matrix for the phenotype and trials. Thus, eq. 12 allows

independent genetic random effects between trials and has the potential to capture unexpected GTI effects mentioned in assumption 4. We hereafter refer to this random genetic effect model as rgeGT. Equations 10 and 12 can be combined as follows:

$$y = \mu + f_{\text{fixed}}(x) + g + gt + ge + \varepsilon \quad (14).$$

Eq. 14 can model both environment-specific and trial-specific random genetic effects. We named this random effect modeling rgeGTE. The last random genetic effect modeling is

$$y = \mu + f_{\text{fixed}}(x) + g + \sum_{tr}^{TR} gt_{tr} + \varepsilon \quad (15)$$

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

$$\sum_{tr}^{TR} gt_{tr} \sim N(0, \sigma_{gt1}^2 \begin{bmatrix} Z_g G Z_g' & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} + \sigma_{gt2}^2 \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ 0 & Z_g G Z_g' & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$
$$+ \cdots + \sigma_{gtTR}^2 \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & Z_g G Z_g' \end{bmatrix}) \quad (16)$$

This random genetic effect design is the same as that of e Sousa *et al.* (2017) and Dahl *et al.* (2020), which allows different heritability between trials. We named this random effect modeling rgeGTd.

Fitting the LMMs: All calculations and parameter estimations for the LMMs were performed using R package "gaston" (Perdry and Dandine-Roulland 2018). The LMMs in the present study included multiple random effects. We tentatively generalized the background genetic random effects as

$$bg_r \sim MVN\left(0, \left[Z_g G Z_g'\right] \circ [Z_r Z_r']\sigma_r^2\right) \quad (17),$$

where bg_r is the *r*-th random genetic effect, and Z_r is the incident matrix for phenotype and *r*-th random effects. σ_r^2 is the variance for bg_r. To solve the LMMs with R random effects, we estimated the genetic variances (i.e., $\sigma_1^2 \sim \sigma_r^2 \sim \sigma_R^2$) without any fixed genetic effects, that is,

$$y = \mu + \sum_r^R bg_r + \varepsilon \quad (18).$$

Equation 18 was solved using the AIREML algorithm implemented in the *lmm.aireml* function. The estimated genetic variances were used to calculate the weight for each random effect as

$$w_s = \hat{\sigma}_s^2 \,/\, \sum_r^R \hat{\sigma}_r^2 \quad (19),$$

where w_s is the weight of the *s*-th random effect. The variance–covariance matrix for the integrated random effects K was calculated as

$$K = \sum_r^R \left[Z_g G Z_g'\right] \circ [Z_r Z_r']w_r \quad (20).$$

Next, the integrated background random genetic effect u was derived as

$$u \sim MVN(0, K\sigma_K^2) \quad (21).$$

where σ_K^2 is the variance for u. Thus, the LMMs used for the calculation of the test statics in the present study can be expressed as follows:

$$y = \mu + f_{\text{fixed}}(x) + u + \varepsilon \quad (22).$$

Therefore, variance component analysis for each LMM was performed against u and ε during the GWASs.

Statistical tests: The statistical significance of the fixed effects in each LMM was evaluated using the log-likelihood (LL) ratio test (LRT). The tests and formulae for the calculation of deviance and degree of freedom are described in Table 2. Hypotheses QEI and QTI in Table 2 are the test for significance of the full (i.e., main and QEI effects) model against the main genetic effects. Therefore, these effects will be significant when the QTLs have different reaction norms between environments or trials (Malosetti *et al.* 2013). The LL of maximum likelihood (ML) estimates was calculated using the *lmm.diago.profile.likelihood* function. The *p*-value of each test was calculated using the chi-square test with deviance and degrees of freedom.

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 model. As a complement to the LRT, we performed the Wald test for each coefficient included in the LMMs. The scores for the Wald test were calculated as

$$W_i = \hat{\beta}_i^2 / var(\hat{\beta}_i) \quad (23),$$

where $\hat{\beta}_i$ and W_i indicate the estimated effect size of SNP i and the Wald score, respectively. Next, the p-value of W_i was calculated on the basis of the chi-square test with degree of freedom = 1.

In GWASs, determining a stringently applicable genome-wide significant threshold is difficult (Bian and Holland 2017). In this study, we used two methods to determine the genome-wide significant thresholds. The first method is the arbitrary significant threshold (e.g., $-\log_{10}(p) \geq 2$). The second is the false discovery rate (FDR), which is a popular method in GWASs (Benjamini and Hochberg 1995; Storey and Tibshirani 2003). In this study, the FDR rate was calculated using the method proposed by Storey and Tibshirani (2003), which was implemented in R package "qvalue" (Storey *et al.* 2020).

Genomic inflation factor

In the LRT, $-\log_{10}(p)$ values from the chi-square tests represented inflation or deflation from the expected distribution. The degree of inflation or deflation was evaluated using

the genomic inflation factor (λ_{GC}; Devlin and Roeder 1999), which was calculated as

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

where $p_{obs}[q]$ indicates the q-quantile of the observed p-values.

Simulation experiments

To evaluate the QTL detection power, we designed simulation experiments. The detailed conditions are described in Table 1. The simulated phenotypes were designed using real genotype data of F_1 tomato varieties (Yamamoto *et al.* 2016). The simulated phenotypic values were generated using the following equation:

$$y = \sum_i^3 x\beta_i + \{(g \times w_g) + (ge \times w_{ge}) + (gt \times w_{gt})\}/\omega + \varepsilon/\varphi \quad (25)$$

where $\sum_i^3 x\beta_i$ is the foreground major genetic effects that are the target of detection. We randomly selected three major QTLs from 16,782 SNP markers. The proportion of variance (PVE) explained by each QTL in the present study refers to the PVE of the QTL in phenotypic values in all environments and trials. The PVE was set to 0.1 for all the simulated major QTLs. In this study, we assumed four reaction norms, namely, “persistence,” “divergence,” “convergence,” and “crossover” (Malosetti *et al.* 2013). Persistence means the QTL has a persistent effect across environments or trials, and thus has no QEI or QTI effects. Divergence means that the QTL has different degrees of

effect size between trials (i.e., the direction of β is the same across trials, but the values are different). Convergence means that the QTL shows effect in an environment, but not in the other environments or trials (i.e., β is zero in some environments or trials). Crossover means that the direction of the QTL effect is different between environments. Persistence, divergence, and convergence were used for simulated QTI effects, whereas convergence and crossover were used for simulated QEI effects (Table 1). In eq. 25, $(g \times w_g)$, $(ge \times w_{ge})$, and $(gt \times w_{gt})$ are the background polygenic effects that follow eqs. 3, 11, and 13, respectively. The polygenic effect values were generated using the *mvrnorm* function in R by using the corresponding variance–covariance matrix (R Core Team 2019). The contribution of the three polygenic effects was adjusted by multiplying the generated values from *mvrnorm* and specified weight parameters (w_g, w_{ge}, and w_{gt} in Table 1). In all simulated phenotypes, heritability was set to 0.5. The residual values ε were generated using the *rnorm* function in R (R Core Team 2019). In eq. 25, ω and φ were optimized to adjust the phenotypic values to satisfy the given PVE and heritability. The optimal ω and φ were determined using the *optimize* function in R.

Because QTLs may have a dominance effect (e.g., Krieger *et al.* 2010), we simulated the dominance effect of QTLs. To model the dominance effects, we coded a

vector of SNP genotype values for the simulated QTLs as {0, 3, 2} = {aa, Aa, AA}, and then, the coded SNP genotype values were used in eq. 25 to generate the simulated phenotypic values. In generation of simulated phenotypic values including dominance-effect QTLs, SNPs with minor genotype count < 3 were excluded.

To enable the direct comparison of GWAS results from different experimental conditions, we used a receiver operating characteristic (ROC) curve; the GWAS power was compared on the basis of the area under the ROC curve (AUC). 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 the top signals predicted by the GWAS agree with the true signals. The *roc* function in "pROC" package (Robin *et al.* 2011) was used to calculate the AUC from $-\log_{10}(p)$ values.

In addition, we used F-measure, a harmonic mean of recall and precision, calculated as

$$\text{F measure} = 2/(\text{Recall}^{-1} + \text{Precision}^{-1}) \quad (26).$$

A high F-measure is achieved only by balancing high precision and high recall. In the simulation experiments, F-measures were calculated using a genome-wide significant threshold from FDR 0.2 or $-\log_{10}(p) \geq 2$.

Tomato genotype and phenotype data

We applied the LMMs to real genotype and phenotype data in tomato. The tomato data used in this study were obtained from 96 big-fruited F_1 varieties that were also used in previous studies (Yamamoto *et al.* 2016 and 2017). Genotype data consisting of 16,782 SNP markers were obtained using Axiom myDesign genotyping arrays (Affymetrix Co., Ltd., Santa Clara, CA, USA). All the SNP markers had minor allele frequency of >0.05 and missing value rate of 0. Phenotyping was performed in winter and summer cropping seasons with four and two trials, respectively. All plants were grown hydroponically by using a high-wire system in a greenhouse at the National Agriculture and Food Research Organization, Institute of Vegetable and Tea Science in Tsu, Japan. One plant was grown for each variety in each trial. Among the phenotypes obtained (Yamamoto *et al.* 2016), we focused on the average fruit weight and fruit set ratio. The fruit set ratio indicates the ratio of flowers that reached fruit set. The values were transformed using the empirical logit transformation.

RESULTS

Genomic inflation factor

In this study, the genomic inflation factor (λ_{GC}) was calculated to assess the deviation of test statistics from the expected null distribution. In the tests for Main (Table 2), $-\log_{10}(p)$ values were close to the theoretically expected distribution (i.e., $\lambda_{\mathrm{GC}} = 1$) except when rgeGTd was the random genetic effect (Figure 1 and Table S1). The tests that included QEI and QTI fixed effects (i.e., fAMMI and fGGE) yielded different distributions depending on the random genetic effects. The λ_{GC} values inflated and fluctuated when the random effects were rgeG or rgeGTd (Figure 1 and Table S1). When the fixed effects included QTI, the inflation and fluctuation of λ_{GC} values were also noted in rgeGE (Full.T and QTI in Figure 1 and Table S1). The use of rgeGT or rgeGTE successfully restrained the λ_{GC} fluctuations (Figure 1 and Table S1). When the fixed effects were fGGE, the use of rgeGT or rgeGTE resulted in theoretically expected λ_{GC}, whereas it was deflated in fAMMI (Figure 1 and Table S1). $\lambda_{\mathrm{GC}} = 1$ is necessary to calculate the FDR (Storey and Tibshirani 2003). Therefore, these results indicate that GGE is preferable to model QEI or QTI effects in LMMs.

QTI effects in a single environment

This study aimed to compare the power between the tests (Table 2) when the target QTL effects and heritability are unstable owing to GTI effects (column N.env = 1 in Table 1). The power between the tests was compared using AUC and F-measure with genome-wide significant threshold FDR of 0.2 and $-\log_{10}(p) \geq 2$. In this analysis, we did not calculate F-measure for the tests using fAMMI because the deflation in λ_{GC} did not meet the prior assumptions for FDR calculation (Table S1; Storey and Tibshirani 2003). Although we generated conditions that had different contribution rates in background random genetic effects (column Background in Table 1), the differences had little effect on the power to detect QTLs (Table S2 and S3). Conversely, the method to model random genetic effects had a significant impact on power (Figure 2). In particular, rgeGT represented higher AUC and F-measure than the other random genetic effect models (Figure 2). Condition which simulated QTLs with a persistent effect across trials fitted the assumption in test Main (Table 2). As expected, Main showed the highest AUC and F-measure under Persistence conditions (Figure 2). The tests for QTI yielded lower AUC and F-measures than those for Main and Full.T except for condition crossover (Figure 2). The decrease in AUC and F-measures in QTI was the most evident when the QTL had a persistent effect across trials (i.e., Persistence in Figure 2), whereas

the decrease in divergent and convergent QTLs was small in AUC and moderate in F-measures (Figure 2). These results suggested that the comparison of results between Main and QTI enables the detection of unexpected QTI effects in a single environment.

QEI effects between multiple environments

With regard to the QEI effects between two or three environments, we generated two reaction norm scenarios—convergence and crossover—(Table 1). Again, power between the tests was compared using AUC and F-measure with genome-wide significant threshold FDR of 0.2 and $-\log_{10}(p) \geq 2$; we did not calculate F-measure for tests using fAMMI because of the deflation in λ_{GC} (Figure 1 and Table S1). When the QEI effects were convergent, the tests for Main yielded equivalent AUC as those with tests including QEI or QTI terms ("Full," "QEI," and "QTI" in Figure 3A). The advantage of the tests including QEI or QTI terms became clearer when the power was compared using F-measure (Figure 3B and C). In most cases, power was higher in QEI (i.e., Full.E and QEI in Figure 3B and C) than in QTI (i.e., Full.T and QTI in Figure 3B and C). When the QEI effects were crossover, the tests for Main almost lost the power (i.e., AUC ≈ 0.6 in Figure 3A and F-measure ≈ 0 in Figure 3B and C). Conversely, tests including QEI or QTI terms ("Full," "QEI," and "QTI" in Figure 3) represented the

power in both AUC (>0.7 in Figure 3A) and F-measure (Figure 3B and C). Tests including QEI terms (i.e., Full.E and QEI in Figure 3, Table S2 and Table S3) showed slightly higher power than those including QTI terms (i.e., Full.T and QTI in Figure 3, Table S2 and Table S3). Notably, Full.E and QEI had higher power when the background random genetic effect was rgeGE or rgeGTE, whereas Full.T and QTI had higher power with rgeGT or rgeGTE (Figure 3). These results indicated that, for higher power, LMMs including QEI in the fixed effects require rgeGE, whereas those with QTI require rgeGT. In addition, the inclusion of both GEI and GTI in the random genetic effects (i.e., rgeGTE) had no drawback in terms of the detection power. Similar to the results from a single environment (Figure 2), power in rgeG and rgeGTd was smaller than that in the other random genetic effect models (Figure 3).

Power to detect QTLs with dominance effects

The above simulation experiments assumed the detection of additive-effect QTLs. In this study, we also performed simulation experiments for detecting QTLs with dominance effects. The dominance effect was designed on the basis of real dominance-effect QTLs in tomato (Krieger *et al.* 2010). The LMMs including dominance effect terms showed slightly deflated λ_{GC} values (Figure 4A). We compared

the power to detect the dominance effect of QTLs by using AUC (Figure 4B and Table S4). Under our simulation conditions, LMMs including dominance effect terms had advantage over those with additive effect terms for detecting dominance-effect QTLs (Figure 4B and Table S4). On the other hand, LMMs including dominance effect terms showed inferior power for the detection of additive-effect QTLs (Figure 4B and Table S4).

GWASs of tomato agronomic traits

The results from λ_{GC} indicated that fGGE is better than fAMMI for fixed effect design (Figure 1). Regarding the background random genetic effects, rgeGTE showed stable λ_{GC} and higher QTL detection power without any drawbacks (Figures 1–3). Therefore, we used the LMM consisting of fGGE and rgeGTE for real agronomic trait data in tomato. We focused on QTLs with FDR > 0.05, which are often used as genome-wide significant thresholds in GWASs (e.g., Endelman 2011). There was no significant difference between LMMs with or without dominance effect terms (Figure S1). We elaborate this result in the discussion section. Because the data for winter cropping season was obtained from four trials, we assessed the potential unexpected GTI effects under environmental condition (Figure 5A). This situation resembles the simulation

experiment that assumes a single environment with multiple trials (N.env. = 1 in Table 1 and Figure 2). For average fruit weight, test for Main detected a significant signal on chromosome 9, whereas the test for Full.T detected significant signals on chromosomes 4 and 9 (Figure 5B). In the QTI test, chromosome 4 retained high $-\log_{10}(p)$ (>4), whereas the peak for chromosome 9 disappeared. The estimated SNP effects and Wald scores suggested that the QTL on chromosome 4 was effective only in 2011 (Table 3). Conversely, the QTL on chromosome 9 showed a persistent effect across trials (Table 3). Thus, the test for QTI effects enabled the detection of trial-specific QTLs. Next, we determined QEI effects between winter and summer cropping seasons. The significant signal for average fruit weight was again detected on chromosome 9 in the tests for Main and Full.E (Figure 5C). This signal disappeared in the test for QEI (Figure 5C). Both the estimated effect size and Wald scores for this signal indicated that the QTL on chromosome 9 had an effect in both winter and summer cropping seasons (Table 4). Fruit set ratio is recognized as a trait that is highly sensitive to environmental differences (Adams *et al.* 2001). For the fruit set ratio, a significant signal was noted on chromosomes 1 in the test for QEI, whereas no significant signal was detected in the tests for Main and Full.E (Figure 5D). The estimated SNP effect size and Wald scores indicated that the QTL is effective only in the summer cropping season (Table 4). In fact,

these QTLs were not detected in the tests performed using the data from the winter cropping season (Table 3 and Figure S2).

DISCUSSION

In this study, we explored efficient LMMs to detect QTLs with QEI effects. For this, we used various combinations of foreground (i.e., fAMMI and fGGE) and background genetic effect modeling methods (i.e., rgeG, rgeGE, rgeGT, rgeGTE, and rgeGTd). The efficiency of background genetic effect modeling has been extensively compared in GS studies (e.g., e Sousa *et al.* 2017). However, to the best of our knowledge, GWASs performed using LMMs have not been compared, although the modeling methods are directly applicable to LMMs in GWASs. Although LMMs have been used for the detection of QEI effects in some previous studies (e.g., Boer *et al.* 2007; Mathews *et al.* 2008; Sabadin *et al.* 2012), these studies did not focus on using modeling methods for conducting genome-wide analyses. Therefore, the advantages and disadvantages of these methods are not yet known.

Because the inclusion of QEI terms from numerous environments results in curse of dimensionality problem, we simulated up to three environments (Table 1). If the similarity between the environments can be expressed by a numeric relationship matrix, the use of a random effect term from the matrix to design the reaction norm of a focal QTL might be a powerful approach to address the curse of dimensionality (Moore

et al. 2019). Although the method in Moore *et al.* (2019) was mainly designed for human and animal genetics (data where numerous phenotypes, genotypes, and environmental data are available), it could be applicable for plant genetic studies because the number of data depositions consisting of multiple phenotypes, environments, and condition data have been increasing (Pommier *et al.* 2019). Nevertheless, the methods and approaches used in this study (i.e., GWASs for QEI effects in a limited number of environments) are still valuable when the number of environments of interest is limited and when the conditions that affect QTL function cannot be represented as continuous numeric values.

Inflation or deflation of λ_{GC} causes various problems in GWASs (Devlin and Roeder 1999). For example, although FDR is a popular method to determine genome-wide significant thresholds in GWASs (Benjamini and Hochberg 1995; Storey and Tibshirani 2003), its calculation requires theoretically expected distribution in *p*-values (i.e., λ_{GC} = 1). Genomic control (GC) corrects inflated or deflated *p*-values by using the λ_{GC} value (Devlin and Roeder 1999). However, GC does not always work well with GWAS results. Therefore, λ_{GC} = 1 is desirable even if no difference in the *p*-value rank exists. In this study, we compared the QTL detection power between

fAMMI and fGGE (Figures 1–3). fAMMI showed deflated λ_{GC} although the AUCs were equivalent to those of fGGE (Figures 1–3). This is mainly attributed to the degree of freedom in the LRT (Table 2). Because of the additive main effect term ($x\alpha$ in eq. 6), the LRT that uses fAMMI has one more degree of freedom than that of fGGE (eq. 7), although the power of explanation for phenotypic variance is basically the same between fAMMI (eq. 6) and fGGE (eq. 7). Thus, fGGE is preferable for genome-wide analysis of QEI effects. One of the desirable features of fAMMI is the separation of genetic effects into main and QEI effects (Gauch *et al.* 1998). Therefore, fixed effect terms analogous to fAMMI have been used in previous studies (Boer *et al.* 2007; Mathews *et al.* 2008), although they were used for the analysis of some focal QTLs rather than for genome-wide analysis. To perform exhaustive QEI effect analysis with minimal oversight, our findings suggest that (1) a genome-wide analysis with LMMs should be performed using fGGE and (2) QEI effect should be assessed using fAMMI.

For the construction of LMMs, we compared various modeling methods for the background random effects (i.e., rgeG, rgeGE, rgeGT, rgeGTE, and rgeGTd in Figures 1–3). When the fixed effect terms included QTI effects (i.e., Full.T and QTI in Figures 1–3), the random effects rgeG and rgeGE resulted in inflated λ_{GC} values (Figure 1 and

Table S1) and inferior power to detect QTLs (Figures 2 and 3). These λ_{GC} inflations in the QTI fixed effects disappeared when the random effect terms included GTI effects (rgeGT and rgeGTE in Figures 1–3). Interestingly, the inclusion of more random effect terms (i.e., rgeGTE) did not affect the λ_{GC} values and QTL detection power (Figures 1–3). Therefore, when the phenotype data are obtained from multiple environments and multiple trials, we recommend the use of rgeGTE for random genetic effects in the LMMs. rgeGTd was less effective in controlling λ_{GC} values and power to detect QTLs (Figures 1–3), although its efficacy has been confirmed in previous studies (e Sousa *et al.* 2017; Dahl *et al.* 2020). One of the possible reasons could be the differences in methods for variance component estimation during the LMM fitting. In this study, we performed direct estimation by using AIREML without any fixed genetic effect terms (Gilmour *et al.* 1995; Perdry and Dandine-Roulland 2018). Conversely, for example, the variance component estimation and fitting method in e Sousa *et al.* (2017) were based on a Bayesian approach that updates the parameters by using a Gibbs sampler (Pérez and de Los Campos 2014). In general, use of Bayesian approach can yield more reliable convergence in the parameters. Because model fitting in e Sousa *et al.* (2017) was used to construct prediction models for GS, one model fitting process per trait was sufficient. Conversely, because this was a GWAS, we needed to perform model fitting for every

marker in the data, which might become computationally unfeasible as the number of markers increases. Therefore, we did not use the Bayesian approach for model fitting in this study, although it has the potential to increase the power of QTL detection.

In this study, the effect size of simulated QTLs with dominance effects was determined on the basis of real dominance-effect QTLs reported in a previous study (Krieger *et al.* 2010). When we assessed the power of the tests including the dominance effect terms (Table 2), the λ_{GC} values were slightly deflated in all the tests (Figure 4A). This is likely because of the many fixed terms, which resulted in greater degree of freedom in the subsequent chi-square test. The inclusion of dominance effect terms in the LMMs showed advantage in detecting dominance-effect QTLs simulated in this study (Figure 4B). This is reasonable because models including dominance effect terms generally represent a better fit to phenotypic values with a dominance effect (Letter *et al.* 2007). Nevertheless, LMMs including dominance effect terms did not detect dominance-effect QTLs for real agronomic trait data in the present study (Figure S1). Some previous studies suggested that the contribution of dominance effects in quantitative traits is less evident, and the use of an additive effect model is sufficient for genome-wide analysis (e.g., Varona *et al.* 2018). In addition, problems in the inclusion

of dominance effect terms has been discussed by Balding (2006). That is, the existence of few homozygotes results in the neglect of homozygote genotypes and linear regression fitting between the other homozygote and heterozygote genotypes (Balding 2006). Moreover, the inclusion of QEI terms with dominance effects leads to inflation of *p*-values owing to the cost of degree of freedom, which results in decrease of power for detection of additive QTLs. Although this problem can be addressed by increasing the number of samples (genotypes), phenotyping numerous samples is often difficult, especially in plant genetic studies (Atwell *et al.* 2010; Yano *et al.* 2016). Thus, further studies are required to determine the efficacy of the inclusion of dominance effect terms, especially in plant genetic studies.

The findings from the simulation experiments can be summarized as follows. First, LMMs using fGGE represent *p*-values closer to theoretically expected distributions than those using fAMMI. Second, rgeGTE is preferable to model the background random genetic effects if the phenotypic data of an environmental condition were obtained from multiple trials. Therefore, we recommend the following steps for GWASs to detect QEI effects:

1. Tests for QTI effects for each environment to detect (or remove) QTLs affected by

unexpected QTI effects.

2. Tests for QEI effects to detect QTLs with QEI effects.

Application of the above steps for real phenotypic data in tomato revealed loci with QEI effects, although further experiments are neccesary to confirm the QEI effects of the detected loci (Figure 5, Tables 3 and 4). Thus, the application of the strategy developed in this study to phenotypic data from multiple environments and trials has the potential to contribute to the identification of more QTLs with QEI effects.

ACKNOWLEDGMENTS

This work was supported by the Japan Science and Technology Agency (JST) PRESTO (Grant number, JPMJPR16Q9) and JSPS KAKENHI (Grant number, 19H02951) to E.Y.

COMPETING INTERESTS

The authors declare no competing interests.

LITERATURE CITED

Adams, S. R., K. E. Cockshull, and C. R. J. Cave, 2001 Effect of temperature on the growth and development of tomato fruits. Ann. Bot. 88: 869-877.

Atwell, S., Y. S. Huang, B. J. Vilhjálmsson, G. Willems, M. Horton *et al.*, 2010 Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465: 627-631.

Balding, D. J., 2006 A tutorial on statistical methods for population association studies. Nat. Rev. Genet. 7: 781-791.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B57: 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.

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.

e 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.

Endelman, J. B., 2011 Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4: 250−255.

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.

Heslot, N., D. Akdemir, M. E. Sorrells, and J. L. Jannink, 2014 Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor. Appl. Genet. 127: 463−480.

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.

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.

Kornhuber, K., S. Osprey, D. Coumou, S. Petri, V. Petoukhov *et al.*, 2019 Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. Environ. Res. Lett. 14: 054002.

Lettre, 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.

Lipka, A. E., C. B. Kandianis, M. E. Hudson, J. Yu, J. Drnevich *et al.*, 2015 From association to prediction: statistical methods for the dissection and selection of complex traits in plants. Curr. Opin. Plant Biol. 24: 110-118.

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.

Millet, E. J., W. Kruijer, A. Coupel-Ledru, S. A. Prado, L. Cabrera-Bosquet *et al.*, 2019 Genomic prediction of maize yield across European environmental conditions. Nat. Genet. 51: 952−956.

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.

Pommier, C., C. Michotey, G. Cornut, P. Roumet, E. Duchêne *et al.*, 2019 Applying FAIR principles to plant phenotypic data management in GnpIS. Plant Phenomics 2019: 1671403.

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.

Sabadin, P. K., M. Malosetti, M. P. Boer, F. D. Tardin, F. G. Santos *et al.*, 2012 Studying the genetic basis of drought tolerance in sorghum by managed stress trials and

adjustments for phenological and plant height differences. Theor. Appl. Genet. 124: 1389−1402.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100: 9440−9445.

Storey, J. D., A. J. Bass, A. Dabney, and D. Robinson, 2020 *qvalue:* Q-value estimation for false discovery rate control. R package version 2.20.0. Available at: http://github.com/jdstorey/qvalue.

Varona, L., A. Legarra, M. A. Toro, and Z. G. Vitezica, 2018 Non-additive effects in genomic selection. Front. Genet. 9: 78.

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.

Yano, K., E. Yamamoto, K. Aya, H. Takeuchi, P. C. Lo *et al.*, 2016 Genome-wide

association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat. Genet. 48: 927−934.

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.

Yuan, Y., J. E. Cairns, R. Babu, M. Gowda, D. Makumbi *et al.*, 2019 Genome-wide association mapping and genomic prediction analyses reveal the genetic architecture of grain yield and flowering time under drought and heat stress conditions in maize. Frontiers Plant Sci. 9: 1919.

FIGURE LEGENDS

Figure 1

Box and whisker plots for genomic inflation factors in the simulated phenotypes. The title in each facet indicates the simulation conditions described in Table 1. The labels on the *x*-axis indicate the name of the test in Table 2. The values were obtained from 100 simulations. The red horizontal lines indicate the theoretically expected value (i.e., $\lambda_{GC} = 1$).

Figure 2

Power to detect quantitative trait loci (QTLs) in the simulated phenotypes that assume single environment with multiple trials (N.env = 1 in Table 1). The title in each facet indicates the simulation condition on the QTL effects. The labels in *x*-axis indicate the name of the test in Table 2. The values were obtained from 100 simulations. (A) Box and whisker plots for area under the receiver operating characteristic curve (AUC). (B) Box and whisker plots for F-measures calculated using false discovery rate (FDR) of 0.2 as the genome-wide significant threshold. (C) Box and whisker plots for F-measures calculated using $-\log_{10}(p) \geq 2$ as the genome-wide significant threshold.

Figure 3

Power to detect quantitative trait loci (QTLs) in the simulated phenotypes that assume multiple environments with multiple trials. The title in each facet indicates the simulation condition described in Table 1. The labels in *x*-axis indicate the name of the test in Table 2. The values were obtained from 100 simulations. (A) Box and whisker plots for area under the receiver operating characteristic curve (AUC). (B) Box and whisker plots for F-measures calculated using false discovery rate (FDR) of 0.2 as the genome-wide significant threshold. (C) Box and whisker plots for F-measures calculated using $-\log_{10}(p) \geq 2$ as the genome-wide significant threshold.

Figure 4

Comparison of the linear mixed models (LMMs) with or without dominance effect terms. Additive and Dominance in the title of each facet indicate additive-effect and dominance-effect QTLs, respectively. The labels in *x*-axis indicate the name of the test in Table 2. The values were obtained from 100 simulations. (A) Box and whisker plots for genomic inflation factor. The red horizontal lines indicate the theoretically expected

value (i.e., λ_{GC} = 1). (B) Box and whisker plots for area under the receiver operating characteristic curve (AUC).

Figure 5

Genome-wide association studies (GWASs) for tomato agronomic traits. (A) Schematic representation for the periods of phenotyping. (B) Manhattan plots for average fruit weight in the winter cropping season. The title in each plot indicates the name of the test in Table 2. The red horizontal lines indicate the genome-wide significant threshold (false discovery rate (FDR) < 0.05). (C) Manhattan plots for the average fruit weight in the winter and summer cropping seasons. The title in each plot indicates the name of the test in Table 2. The red horizontal lines indicate the genome-wide significant threshold (FDR < 0.05). (D) Manhattan plots for the fruit set ratio in the winter and summer cropping seasons. The red horizontal lines indicate the genome-wide significant threshold (FDR < 0.05).

Figure S1

Genome-wide association studies (GWASs) using LMMs including dominance effect terms for tomato agronomic traits. (A) Manhattan plots for the average fruit weight in

the winter and summer cropping seasons. The title in each plot indicates the name of the test in Table 2. The red horizontal lines indicate the genome-wide significant threshold (FDR < 0.05). (B) Manhattan plots for the fruit set ratio in the winter and summer cropping seasons.

Figure S2

Manhattan plots for the fruit set ratio in the winter cropping season. The title in each plot indicates the name of the test in Table 2.

Table 1. Parameters used for simulated phenotypes.

N.env	N.trials	Background[c]	RN[a]	Foreground[b]						
		w_g:w_{ge}:w_{gt}		Trial	1	2	3	4	5	6
1	3	3 : 0 : 1	Persistence	Env.	A	A	A	-	-	-
				β	1	1	1	-	-	-
1	3	1 : 0 : 3	Persistence	Env.	A	A	A	-	-	-
				β	1	1	1	-	-	-
1	3	1 : 0 : 1	Persistence	Env.	A	A	A	-	-	-
				β	1	1	1	-	-	-
1	3	3 : 0 : 1	Divergence	Env.	A	A	A	-	-	-
				β	0.2	0.5	1	-	-	-
1	3	1 : 0 : 3	Divergence	Env.	A	A	A	-	-	-
				β	0.2	0.5	1	-	-	-
1	3	1 : 0 : 1	Divergence	Env.	A	A	A	-	-	-
				β	0.2	0.5	1	-	-	-
1	3	3 : 0 : 1	Convergence	Env.	A	A	A	-	-	-
				β	0	1	1	-	-	-
1	3	1 : 0 : 3	Convergence	Env.	A	A	A	-	-	-
				β	0	1	1	-	-	-
1	3	1 : 0 : 1	Convergence	Env.	A	A	A	-	-	-
				β	0	1	1	-	-	-
1	3	3 : 0 : 1	Crossover	Env.	A	A	A	-	-	-
				β	-0.2	0.5	1	-	-	-
1	3	1 : 0 : 3	Crossover	Env.	A	A	A	-	-	-
				β	-0.2	0.5	1	-	-	-
1	3	1 : 0 : 1	Crossover	Env.	A	A	A	-	-	-
				β	-0.2	0.5	1	-	-	-
2	4	1 : 1 : 1	Convergence	Env.	A	A	B	B	-	-
				β	0	0	0.5	1	-	-
2	4	1 : 1 : 1	Crossover	Env.	A	A	B	B	-	-
				β	-0.5	-1	0.5	1	-	-
3	6	1 : 1 : 1	Convergence	Env.	A	A	B	B	C	C
				β	0.5	1	0.5	1	0	0
3	6	1 : 1 : 1	Crossover	Env.	A	A	B	B	C	C
				β	-0.5	-1	0.5	1	0	0

[a] RN, Reaction norm of the major genes.

[b] β in the column indicates the effect size of the major QTLs in the trial. Env. indicates the identifier of environment in the trial.

[c] Values in the column indicate the relative weight of random genetic effects (see Methods).

Table 2. Tests and arguments used for the chi-square test in this study.

Tests	Formula for the deviance	Degree of freedom	Details
Main	$-2 \times (LL_{Eq.2} - LL_{Eq.18})$	1	Additive effect only
$Main_d$	$-2 \times (LL_{Eq.5} - LL_{Eq.18})$	2	Additive and dominance effects
Full.E (fAMMI)	$-2 \times (LL_{Eq.6} - LL_{Eq.18})$	N.env + 1	
$Full_d$.E ($fAMMI_d$)	$-2 \times (LL_{Eq.7} - LL_{Eq.18})$	(2 × N.env) + 2	β models SNP effect for each environment
Full.E (fGGE)	$-2 \times (LL_{Eq.8} - LL_{Eq.18})$	N.env	
$Full_d$.E ($fGGE_d$)	$-2 \times (LL_{Eq.9} - LL_{Eq.18})$	2 × N.env	
Full.T (fAMMI)	$-2 \times (LL_{Eq.6} - LL_{Eq.18})$	N.trials + 1	
$Full_d$.T ($fAMMI_d$)	$-2 \times (LL_{Eq.7} - LL_{Eq.18})$	(2 × N.trials) + 2	β models SNP effect for each trial
Full.T (fGGE)	$-2 \times (LL_{Eq.8} - LL_{Eq.18})$	N.trials	
$Full_d$.T ($fGGE_d$)	$-2 \times (LL_{Eq.9} - LL_{Eq.18})$	2 × N.trials	
QEI (fAMMI)	$-2 \times (LL_{Eq.6} - LL_{Eq.2})$	N.env	
QEI_d ($fAMMI_d$)	$-2 \times (LL_{Eq.7} - LL_{Eq.5})$	2 × N.env	
QTI (fAMMI)	$-2 \times (LL_{Eq.6} - LL_{Eq.2})$	N.trials	
QTI_d ($fAMMI_d$)	$-2 \times (LL_{Eq.7} - LL_{Eq.5})$	2 × N.trials	Tests for significance of QEI or QTI effects
QEI (fGGE)	$-2 \times (LL_{Eq.8} - LL_{Eq.2})$	N.env − 1	
QEI_d ($fGGE_d$)	$-2 \times (LL_{Eq.9} - LL_{Eq.5})$	(2 × N.env) − 2	
QTI (fGGE)	$-2 \times (LL_{Eq.8} - LL_{Eq.2})$	N.trials − 1	
QTI_d ($fGGE_d$)	$-2 \times (LL_{Eq.9} - LL_{Eq.5})$	(2 × N.trials) − 2	

N.env and N.trials indicate the number of environments and trials in the test, respectively. $LL_{Eq.n}$ indicates the loglikelihood of Eq. *n* in the main text.

804

Table 3. Statistics obtained from the GWAS for phenotypic data from the winter cropping season.

Trait	SNP	Chr	pos (cM)	$-\log_{10}(p)$				Trial			
				Main	Full.T	QTI		2011	2012	2013	2014
Fruit weight	AX-107553846	9	52.7	5.6	4.85	0.87	β[a]	-0.59	-1.05	-1.23	-1.08
							Wald[b]	1.59	4.01	5.47	4.32
	AX-95783660	4	47.3	0.46	3.93	4.24	β	1.27	-0.12	0.39	-0.37
							Wald	3.05	0.12	0.52	0.49
Fruit set	AX-107525913	1	117.8	1.12	0.32	0.03	β	-0.42	-0.35	-0.34	-0.62
							Wald	0.54	0.41	0.43	0.98

[a] Estimated SNP effect size

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

Table 4. Statistics obtained from the GWAS for phenotypic data from the winter and summer cropping seasons.

Trait	SNP	Chr	pos (cM)	$-\log_{10}(p)$				Environment	
				Main	Full.E	QEI		Sum	Win
Fruit weight	AX-107553846	9	52.7	5.10	5.34	1.50	β[a]	-0.58	-1.03
							Wald[b]	1.90	6.56
	AX-95783660	4	47.3	0.49	0.21	0.00	β	0.27	0.28
							Wald	0.37	0.46
Fruit set	AX-107525913	1	117.8	0.05	4.20	4.96	β	0.87	-0.40
							Wald	2.39	0.93

[a] Estimated SNP effect size

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

806

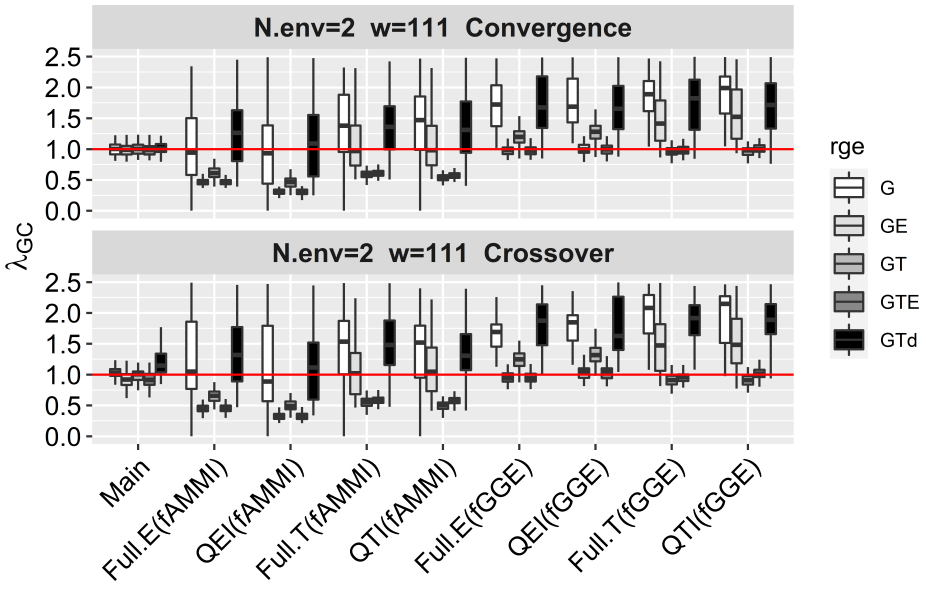
N.env=2 w=111 Convergence
N.env=2 w=111 Crossover
λGC
2.5
2.0
1.5
1.0
0.5
0.0
Main
Full.E(fAMMI)
QEI(fAMMI)
Full.T(fAMMI)
QTI(fAMMI)
Full.E(fGGE)
QEI(fGGE)
Full.T(fGGE)
QTI(fGGE)
rge
G
GE
GT
GTE
GTd

Figure 1

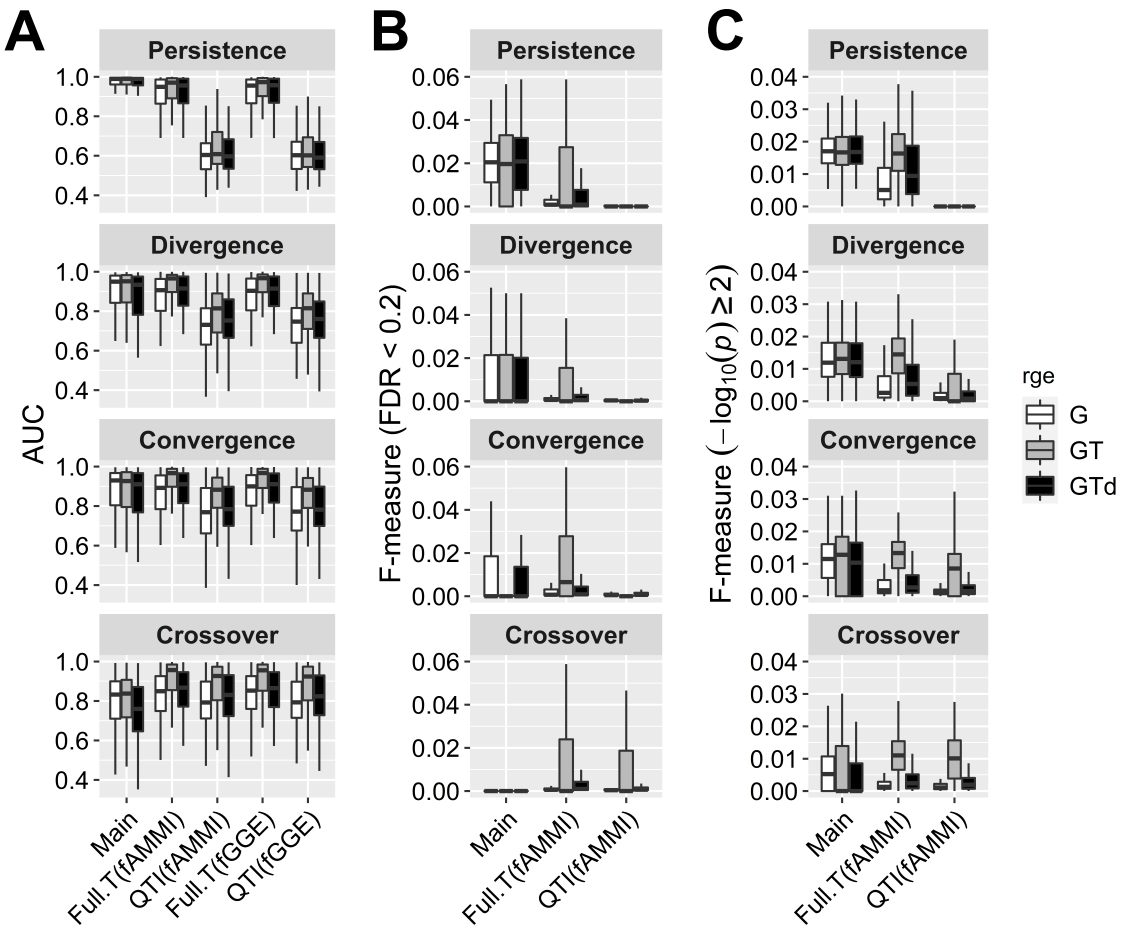
A
B
C
Persistence
Divergence
Convergence
Crossover
AUC
F-measure (FDR < 0.2)
F-measure (−log10(p) ≥ 2)
Main
Full.T(fAMMI)
QTI(fAMMI)
Full.T(fGGE)
QTI(fGGE)
rge
G
GT
GTd

Figure 2

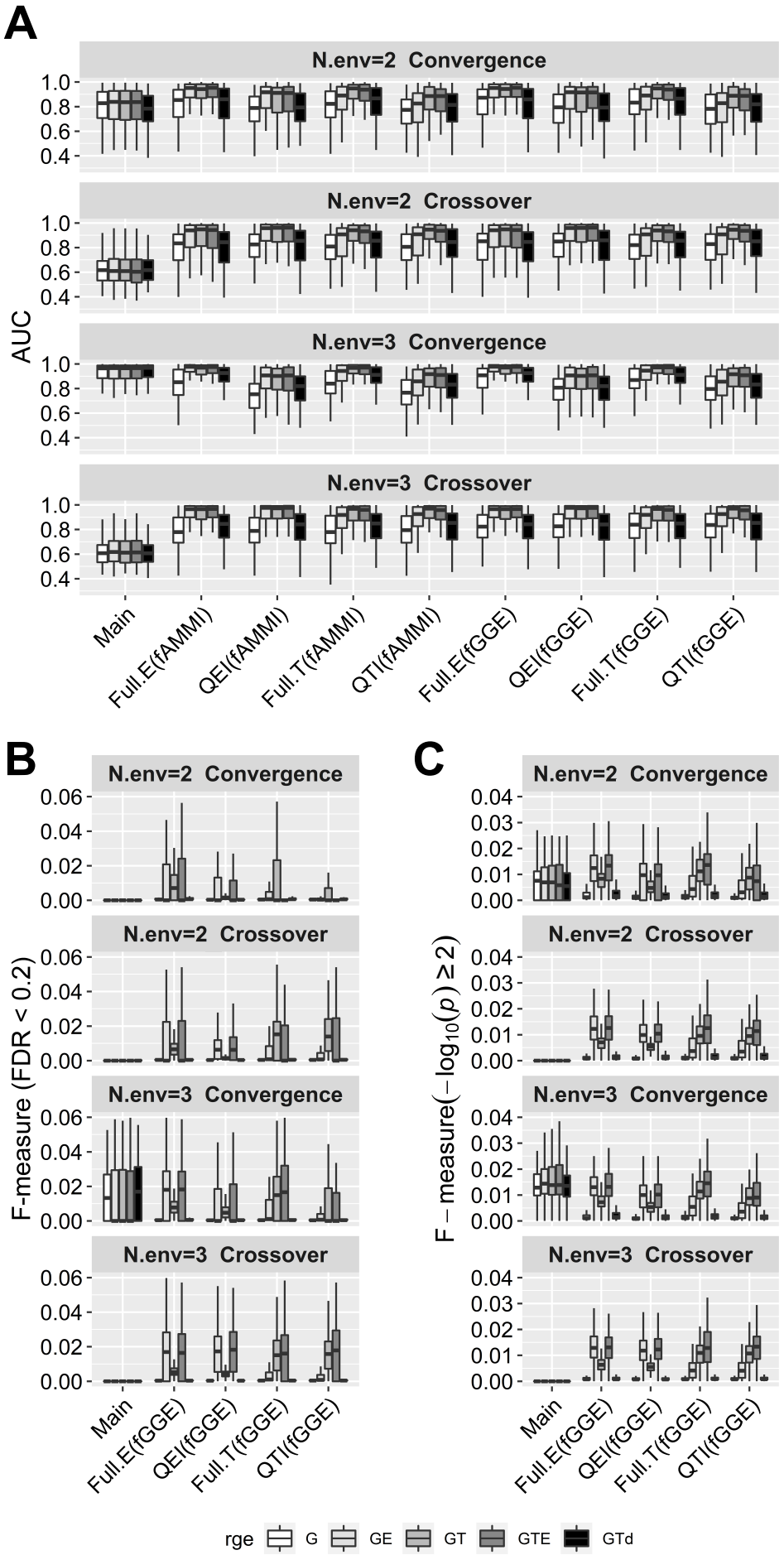
A
N.env=2 Convergence
N.env=2 Crossover
N.env=3 Convergence
N.env=3 Crossover
AUC
1.0
0.8
0.6
0.4
Main
Full.E(fAMMI)
QEI(fAMMI)
Full.T(fAMMI)
QTI(fAMMI)
Full.E(fGGE)
QEI(fGGE)
Full.T(fGGE)
QTI(fGGE)
B
F-measure (FDR < 0.2)
0.06
0.04
0.02
0.00
C
F − measure(−log10(p) ≥ 2)
0.04
0.03
0.02
0.01
0.00
rge
G
GE
GT
GTE
GTd

Figure 3

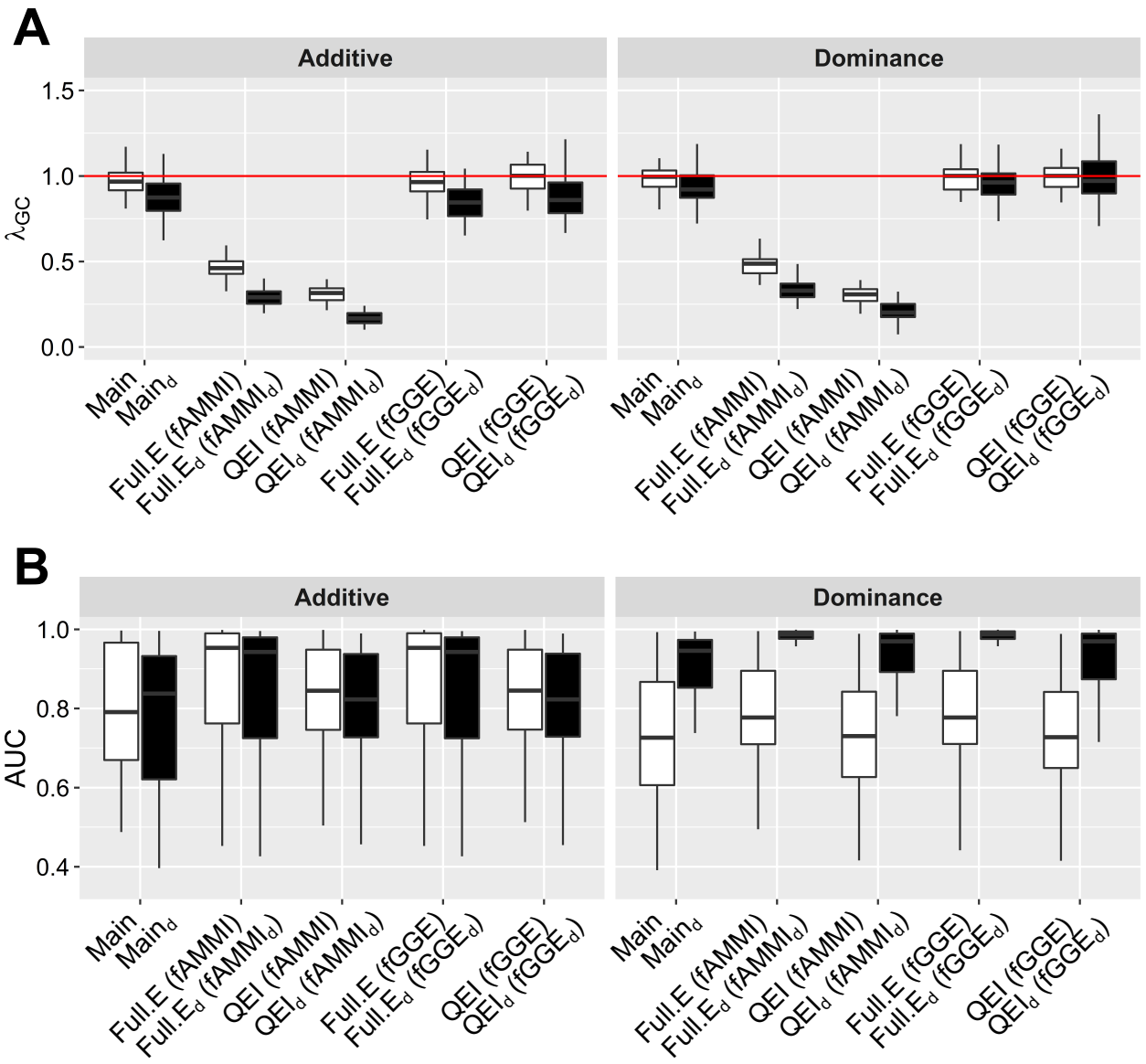
A
Additive
Dominance
1.5
1.0
0.5
0.0
λGC
Main
Main_d
Full.E (fAMMI)
Full.E_d (fAMMI_d)
QEI (fAMMI)
QEI_d (fAMMI_d)
Full.E (fGGE)
Full.E_d (fGGE_d)
QEI (fGGE)
QEI_d (fGGE_d)
B
Additive
Dominance
1.0
0.8
0.6
0.4
AUC

Figure 4

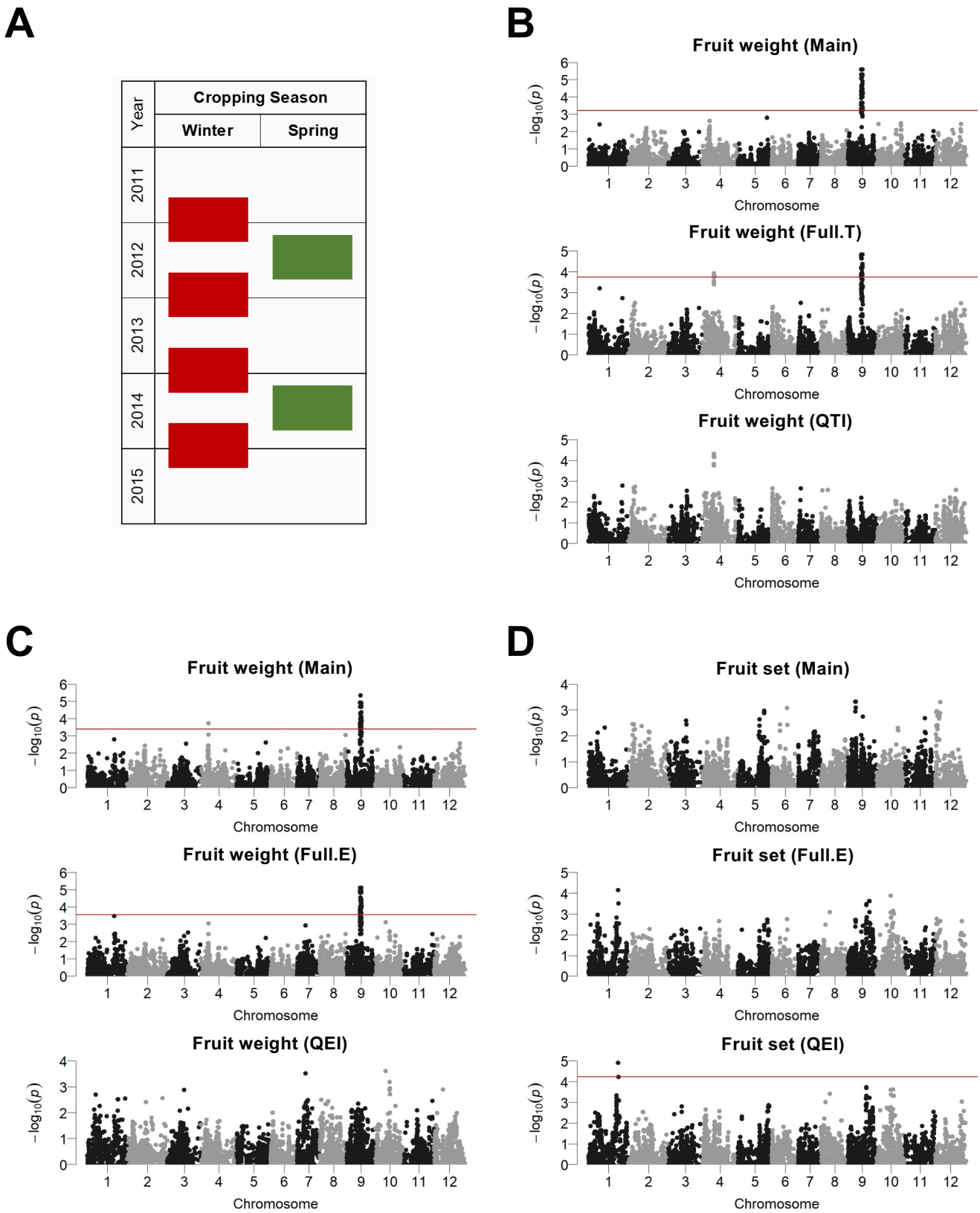
A
Year
Cropping Season
Winter
Spring
2011
2012
2013
2014
2015
B
Fruit weight (Main)
Fruit weight (Full.T)
Fruit weight (QTI)
C
Fruit weight (Main)
Fruit weight (Full.E)
Fruit weight (QEI)
D
Fruit set (Main)
Fruit set (Full.E)
Fruit set (QEI)
$-\log_{10}(p)$
Chromosome

Figure 5

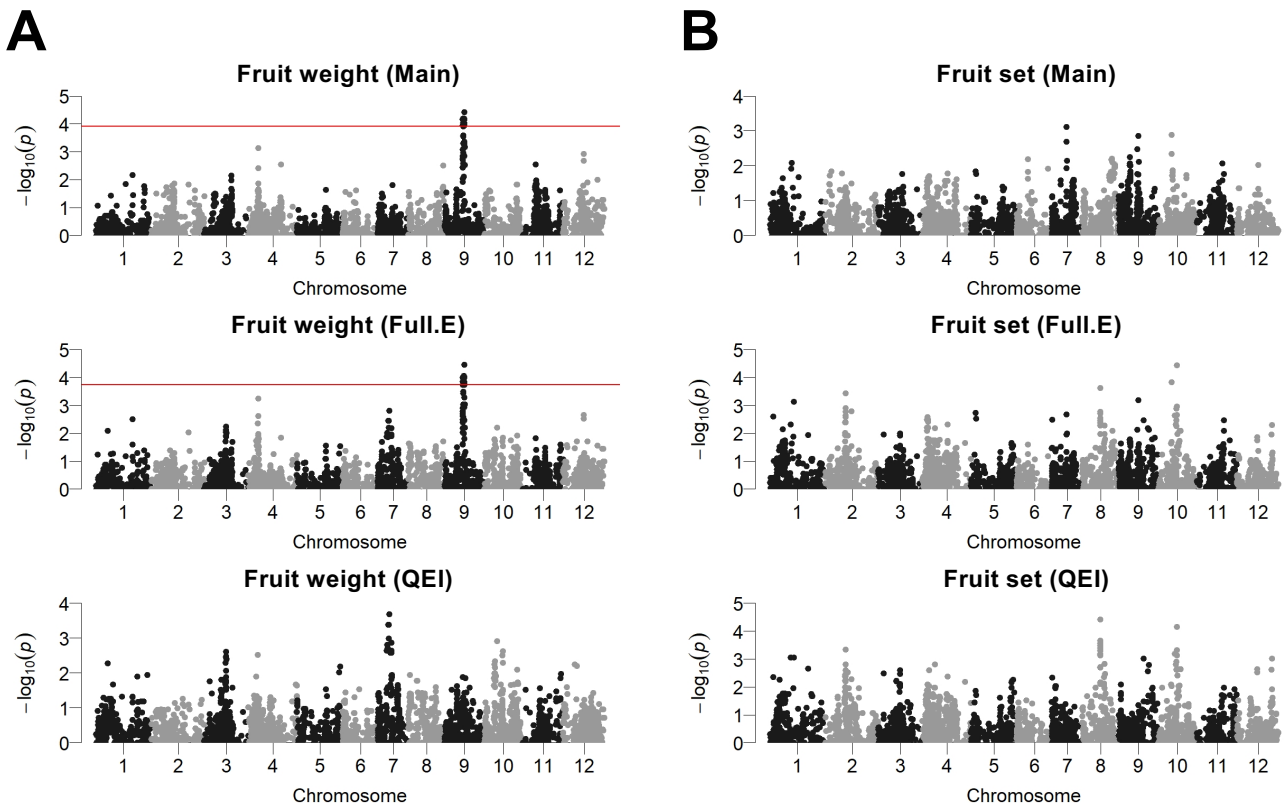
A
Fruit weight (Main)
Fruit weight (Full.E)
Fruit weight (QEI)
B
Fruit set (Main)
Fruit set (Full.E)
Fruit set (QEI)
−log10(p)
Chromosome

Figure S1

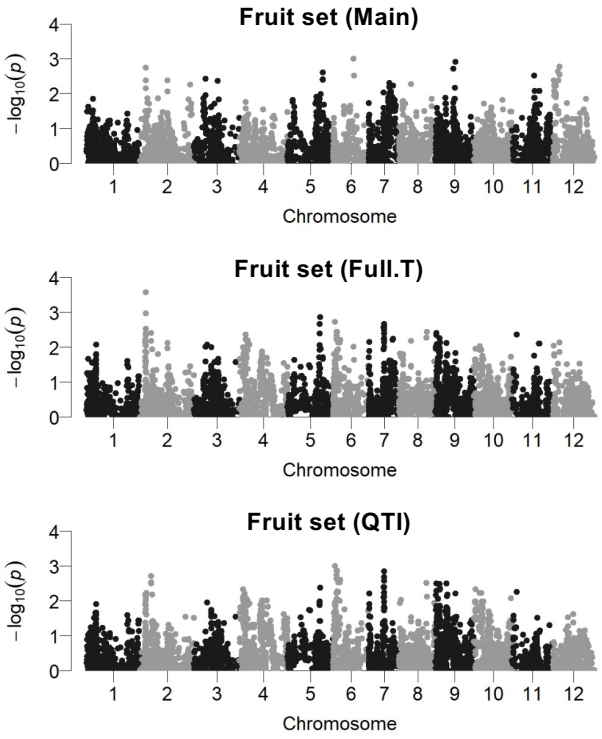
Fruit set (Main)
Fruit set (Full.T)
Fruit set (QTI)
−log10(p)
0
1
2
3
4
1 2 3 4 5 6 7 8 9 10 11 12
Chromosome

Figure S2