An IBD-based mixed model approach for QTL mapping in multiparental populations

Wenhao Li (wenhao.li@wur.nl)
Wageningen University Plant Sciences https://orcid.org/0000-0001-5719-5775
Martin P. Boer
Wageningen University Plant Sciences
Chaozhi Zheng
Wageningen University Plant Sciences
Ronny V.L. Joosen
Rijk Zwaan Nederland BV
Fred A. van Eeuwijk
Wageningen University Plant Sciences

Research Article

Keywords: multiparental population designs, QTL mapping, IBD, Mixed effect model

Posted Date: February 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-233585/v1

An IBD-based mixed model approach for QTL mapping in multiparental populations

Wenhao Li [1], Martin P. Boer [1], Chaozhi Zheng [1], Ronny V.L. Joosen[2], and Fred A. Van Eeuwijk[1,*]

[1] Biometris, Wageningen University and Research Center, P.O Box 100, 6700 AC Wageningen, The Netherlands
[2] Rijk Zwaan Breeding B.V., PO Box 40, 2678 ZG De Lier, The Netherlands
* Correspondence: fred.vaneeuwijk@wur.nl

Abstract:

Key message The identity-by-descent (IBD)-based mixed model approach introduced in this study can detect quantitative trait loci (QTLs) referring to the parental origin and simultaneously account for multilevel relatedness of individuals within and across families. This unified approach is proved to be a powerful approach for all kinds of multiparental population (MPP) designs.

Multiparental populations (MPPs) have become popular for QTL detection. Tools for QTL mapping in MPPs are mostly developed for specific MPPs and do not generalize to other MPPs. We present an IBD-based mixed model approach for QTL mapping in all kinds of MPP designs, e.g., diallel, Nested Association Mapping (NAM), and Multiparental Advanced Generation Intercrosses (MAGIC) designs. The first step is to compute identity-by-descent (IBD) probabilities using a general Hidden Markov model framework, called reconstructing ancestry blocks bit by bit (RABBIT). Next, functions of IBD information are used as design matrices, or genetic predictors, in a mixed model approach to estimate variance components from multi-allelic genetic effects associated with parents. Family-specific residual genetic effects are added, and a polygenic effect is structured by kinship relations between individuals. Case studies of simulated diallel, NAM, and MAGIC designs proved that the IBD-based mixed model approach could increase true positive rates and mapping resolutions of QTLs in comparison to a widely used benchmark association mapping method. Successful analyses of various real data cases confirmed the wide applicability of our IBD-based mixed model methodology.

Keywords: Multiparental population designs; QTL mapping; IBD; Mixed effect model

1. Introduction

MPP designs have their unique advantages for QTL mapping over biparental populations and association panels. Crossing two parents in a biparental population can balance allele frequencies and increase the chance to detect rare QTLs, but the narrow genetic diversity from only two parents limits the number of detected QTLs (Y. Liu et al., 2000; Pascual et al., 2016). We can use association panels to broaden the genetic diversity, but the low-frequency variants may increase false positives (M. Malosetti et al., 2007; Xiao et al., 2016), and the potential population structure may mask the effects of causal variants (Flint-Garcia et al., 2003; M. Malosetti et al., 2007; Sul et al., 2018; Xiao et al., 2016). Experimental MPP designs, as a compromise between biparental populations and association panels, show broad genetic diversity with a controlled population structure. Such MPP designs like diallel (Giraud et al., 2017; Turner et al., 2018), NAM (Yu et al., 2008), and MAGIC populations (Gardner et al., 2016; B. E. Huang et al., 2015; X. Huang et al., 2011) have been proven to be promising populations for QTL mapping.

QTL mapping models can be classified into family-based (or linkage) and population-based (or linkage disequilibrium) approaches based on the specific design (Myles et al., 2009; Tobias Würschum, 2012; Xu et al., 2017). Most studies comparing and evaluating different statistical models were restricted to only one specific MPP design, such as a sugar beet random-cross design (T. Würschum et al., 2012; Tobias Würschum et al., 2011), a maize NAM population (Giraud et al., 2014), and a tomato MAGIC population (Gardner et al., 2016; B. E. Huang et al., 2015, 2011). A suitable QTL mapping approach for general MPP designs should account for polygenic effects while estimating QTL effects defined in terms of their parental origins. The multi-QTL effect model provides an example of such an approach with a cross-specific residual term for the EU-NAM maize data in Garin et al. (Garin et al., 2017). In this study, we focus on developing and evaluating unified QTL mapping models compatible with all kinds of MPPs, including diallel, NAM, and MAGIC designs.

Our proposal combines strategies from family-based and population-based mapping approaches. Family-based QTL mapping approaches were developed for biparental populations to detect bi-allelic QTLs. In the context of MPP designs, we can estimate multi-allelic effects referring to different parental origins. Despite the different MPP designs, parental origins of offspring alleles can always be estimated as functions from IBD probabilities between parents and offspring. To infer the precise genome composition inherited from parents to progenies, we need a sophisticated approach for IBD computation using the pedigree and whole-genome information. As IBD computations for specific MPP designs can be tedious and error-prone, a general pipeline is required (Broman et al., 2018). This study applied a general approach called RABBIT for IBD computations supporting QTL mapping for any MPP design (Zheng et al., 2015, 2018). The IBD information forms the basis for creating design matrices, or genetic predictors, for estimating QTL allele effects. Additional terms are added to model random genetic effects from family, like in (Garin et

al., 2017), and a polygenic effect that is structured by kinship similar to what is commonly done in genome-wide association studies (GWAS) (Marcos Malosetti et al., 2011; Stich et al., 2008).

We constructed several mixed models varying in whether they use identity-by-state (IBS) or IBD information as the basis of genetic predictors and how they account for the individual relatedness. We simulated diallel, NAM, and MAGIC designs from four inbred Arabidopsis parents to evaluate the performance of these models. The results show that our IBD-based mixed model approach accounting for multilevel relatedness of individuals can increase the true positive rate and the mapping resolution of detected QTL compared with simple IBD-based models and an IBS-based GWAS strategy. We also re-analyzed data from various empirical MPP designs, and the results is quite comparable to previous findings using alternative QTL mapping tools.

2. Methodology

2.1.Mixed models

An MPP design contains N individuals in F families that are derived from crosses between P parents. We developed the QTL mapping methodology in the linear mixed model framework. The contribution of a putative QTL to the phenotype is given by the product of an $N \times P$ design matrix M and a $P \times 1$ vector a of genetic effects. The element M_{ij} represents the genetic predictor for the change of the phenotype in the i^{th} *(i=1, 2, ..., N)* individual contributed by an offspring allele stemming from the j^{th} *(j=1, 2, ..., P)* parent. We constructed two types of models, depending on how the genetic predictor was calculated. Specifically, in the IBS-based model, the genetic predictors are given by observed numbers of IBS alleles, which are identical if the alleles have identical nucleotide base pairs. In the IBD-based models, the genetic predictors are given by the expected numbers of IBD alleles, which are identical if the alleles have been inherited from a common inbred parent, such that a_j denotes the haplotype effect of the j^{th} parent.

The genetic predictors in the IBD-based models were calculated using the RABBIT software (Zheng et al., 2015). RABBIT calculates the genetic predictors by haplotype reconstruction within the hidden Markov model (HMM) framework, where the prior transition probability matrix for modeling how the hidden states change along chromosomes can be calculated using a recursive algorithm on the breeding pedigree (Zheng et al., 2018). The flexibility of RABBIT follows from the applicability of this recursive algorithm to arbitrarily fixed pedigrees, and the genotypic data model is able to account for genotyping errors and missing values in parents and offspring. The principal outputs of RABBIT are the posterior probabilities of the hidden IBD states for each offspring at each locus, conditional on the genotypic data at all loci. For homozygous populations with inbred parents, the hidden states are the parental origins, and the genetic predictors are given by twice the parental origin probabilities. For heterozygous populations with inbred parents, the hidden states are given

by the pairwise combinations of the parental origins, and the exported posterior probabilities can be easily transformed into the parental origin probabilities.

We constructed five IBD-based models (Table 1) to estimate multi-allelic effects in a genome-wide scan. The models are defined for phenotypic data vectors coming in as genotypic means, best linear unbiased estimates, or BLUEs, obtained from a preliminary phenotypic analysis of trial data accounting for experimental design factors and spatial trends. We expected that for all models that incorporated cofactor and kinship corrections, the residual term ε would principally represent non-genetic within-trial variation.

The first two models are called *IBD.SQM_U* and *IBD.SQM_F*. They use IBD information as the basis for the genetic predictors in a simple single-locus QTL mapping model (SQM) with respective homogeneous or uniform (_U) and family-specific (_F) variance-covariance (VCOV) structures on residual terms. The first model *IBD.SQM_U* can be expressed as:

$$Y = X\beta + M_q a_q + \varepsilon$$

$$a_q \sim N(N, I_P \sigma_q^2)$$

$$\varepsilon \sim N(0, I_N \sigma_\varepsilon^2),$$

where Y is the $N \times 1$ column vector for the phenotypes of N individuals; X is the $N \times F$ design matrix with elements 1 or 0 indicating whether the i^{th} $(i = 1,2,\dots,N)$ individual belongs to the k^{th} $(k = 1,2,\dots,F)$ family or not, and β is the $F \times 1$ column vector of fixed family intercept effects; M_q is the $N \times P$ design matrix containing the expected number of parental alleles obtained by taking two times the IBD probability between parent and offspring at a putative QTL position, indexed by the subscript q. The $P \times 1$ column vector a_q contains the random parental effects at the putative QTL with VCOV structure equal to $I_P \sigma_q^2$, σ_q^2 being the genetic variance of the QTL effect; ε is the residual term with a homogeneous VCOV structure expressed as $I_N \sigma_\varepsilon^2$ with the residual variance σ_ε^2. The residual ε contains both genetic and non-genetic elements, from unidentified QTLs and within trial error variation, respectively.

The second model, *IBD.SQM_F* can be expressed as:

$$Y = X\beta + M_q a_q + \varepsilon$$

$$a_q \sim N(0, I_P \sigma_q^2)$$

$$\varepsilon \sim N(0, \oplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2),$$

where the residual term ε has family-specific VCOV structure written as $\oplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2$, in which $\sigma_{e_k}^2$ is the residual variance of the k^{th} family whose family size is n_k $(\sum_{k=1}^{F} n_k =$

N). For the standard MAGIC design with a single family ($F = 1$), *IBD.SQM_U* and *IBD.SQM_F* are equivalent models.

To account for QTLs elsewhere in the genome, say genome background, we can add a set of cofactors to the model *QTL.SQM_F*, as in composite interval mapping (Zeng et al.,1994; Jansen and Stam, 1994), to obtain a multi-QTL model (MQM) called *IBD.MQM_F*, which is expressed as:

$$Y = X\beta + \sum_{c \neq q} M_c a_c + M_q a_q + \varepsilon$$

$$a_q \sim N(0, I_P \sigma_q^2) \text{ and } a_c \sim N(0, I_P \sigma_c^2)$$

$$\varepsilon \sim N\left(0, \bigoplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2\right),$$

where the design matrix for a cofactor is M_c, and the column vector of random QTL effects at the cofactor is a_c, whose genetic variance is σ_c^2. M_c and a_c are structurally comparable to M_q and a_q, but they represent different positions in the genome.

As an alternative to the inclusion of explicit cofactors, we can include a polygenic term, g, into the *IBD.SQM_F* model, leading to the *IBD.Kin_F* model:

$$Y = X\beta + M_q a_q + g + \varepsilon$$

$$a_q \sim N(0, I_P \sigma_q^2)$$

$$g \sim N(0, K\sigma_g^2)$$

$$\varepsilon \sim N(0, \bigoplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2),$$

where the VCOV structure of the polygenic term g is $K\sigma_g^2$ in which σ_g^2 is the variance of the polygenic effect, and K is a $N \times N$ kinship matrix based on genotype information on the whole genome using IBS information (VanRaden, 2008). (We acknowledge that for consistency K should have been based on IBD information. However, in practice we found little difference between kinship corrections based on IBD and IBS, and therefore, for convenience, decided to implement the widely used Van Raden software for calculation of kinship matrices.) To reduce the computational burden and avoid proximal contamination, we applied the leave-one-chromosome-out (LOCO) method for kinship matrix calculation (Yang et al., 2014).

The last model, *IBD.MQMkin_F*, combines cofactors, a polygenic term, and a residual term with family-specific VCOV structure:

$$Y = X\beta + \sum_{c \neq q} M_c a_c + M_q a_q + g + \varepsilon$$

$a_q \sim N(0, I_P \sigma_q^2)$ and $a_c \sim N(0, I_P \sigma_c^2)$

$g \sim N(0, K\sigma_g^2)$

$\varepsilon \sim N\left(0, \bigoplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2\right)$.

The five IBD-based models are compared with a benchmark or reference model, a GWAS approach estimating fixed bi-allelic effects. This reference model is frequently used for population-based mapping. We will refer to this model as *IBS.Kin*:

$Y = X\beta + X_q\lambda_q + g + \varepsilon$

$g \sim N(0, K\sigma_g^2)$

$\varepsilon \sim N(0, I_N \sigma_\varepsilon^2)$,

where X_q is a $N \times 1$ vector column whose elements are 0, 1, and 2, indicating numbers of allele copies based on IBS information, and λ_q is the fixed bi-allelic effect at the QTL.

2.2.QTL detection procedure

To map QTLs with the above IBD-based mixed model approaches, we used log-likelihood ratio tests for variance components representing the putative QTLs as estimated by Residual Maximum Likelihood (Reml) (Morrell, 1998; Stram et al., 1994) in ASReml-R (Version 3.0) (Butler et al., 2018). A simple multiple testing correction was performed via a Bonferroni threshold placed at a genome-wide significance level of 0.01. Around cofactors, we set an exclusion window of 20cM within which no tests for further QTLs were performed. Genome scans for models with cofactors were repeated until the test profile (-log10(p)) stabilized. For the reference GWAS model, *IBS.Kin*, a Wald-test (Molenberghs et al., 2007) was used to determine the significance of QTLs.

3. Datasets

3.1.Simulated MPP data sets

3.1.1. The motivation for simulating MPPs

By simulating QTLs with different segregation configurations in various types of MPP designs, we aim to test the performance of IBD-based models versus GWAS models, where we looked at the true positive rate (TPR) and mapping resolution for major QTLs. Below in section 3.1.3, we will define the TPR and mapping resolution.

To create offspring populations with realistic marker profiles and segregation ratios, we based our simulations on four real Arabidopsis inbred lines with known genomes. These four

lines served as parents in crosses that simulated diallel, NAM, and MAGIC designs. With this number of parents, we still obtain enough detail for insightful simulations, while computation time per simulation remains low. Genomes for progenies were simulated by implementing a crossover process in which progenies inherited markers and QTLs from parents following one of the MPP designs. More details can be found in section 3.1.2.

Phenotypes for the offspring individuals contained a contribution from alleles at three major QTLs, positioned on different chromosomes and at markers with varying genotypic configurations across the parents, to investigate the impact of the number of segregating families and allele frequencies on QTL detection. Details are given below in section 3.1.2. Contributions of 24 minor QTLs were distributed across four chromosomes, including the three chromosomes with a major QTL to define a random polygenic effect. This polygenic effect was implicitly structured by family, depending on segregation or not in diallel and NAM families and allele frequency in MAGIC. The polygenic effect was further structured by the relations between the offspring individuals within families due to the transmission of QTL and marker alleles from parents to offspring. The polygenic effect was added to study the performance of family-dependent variance components. One chromosome did contain neither major nor minor QTLs to investigate error rates. An independent error was added to major and minor QTL effects to determine the heritability.

3.1.2. Details of the simulation

Four inbred Arabidopsis lines with known marker genotypes were chosen as parents for making our simulated crosses: Bla-1 (*parent1*), Br-0 (*parent2*), Got-7 (*parent3*), and Kas-2 (*parent4*). These parents were randomly selected from the Arabidopsis *HapMap* collection (Baxter et al., 2010). More information is available at http://bergelson.uchicago.edu/?page_id=790. SNPs of parental lines were called against the reference sequence. A consensus map with 462 markers was then created by merging the maps of parental variants based on the physical map. We selected positions at three markers for assigning major QTL allelic effects on the consensus map, named *simQTL1*, *simQTL2*, and *simQTL3*. For these three QTLs, a genotype coded as 11 was carried by respectively two, one, and three parents (Figure 1). Starting at 20cM above and below the major QTL position, minor QTLs were uniformly placed at distances of 10cM at chromosomes with a major QTL (from chromosomes 1 to 3). Chromosome 4 contained only minor QTLs. In total, 24 minor QTLs were placed at chromosomes 1 to 4. No QTLs were assigned to chromosome 5. This chromosome was used to assess the number of falsely discovered QTLs.

After deciding on simulated QTL positions, we simulated the genome of each progeny at the F6 generation from diallel, NAM, and MAGIC designs (Figure 2 Upper panel). Progeny were simulated from the parental genomes and crossing schemes using the tool RABBIT implemented in Mathematica (Zheng et al., 2015). Each chromosome in gametes in the generations from F1 to F6 was the random reshuffle of the unique parental genome due to crossovers. The number of crossovers on each chromosome followed a Poisson distribution,

with the mean being the chromosome length in Morgan. The positions of crossovers were uniformly distributed over the chromosomes. In the F6 generation, we obtained realized genotypes at marker positions by adding missing genotypes at a rate of 5% randomly on the 'true' simulated marker genotypes. In practice, we expect this missingness rate to be lower, but we wanted an assessment of the robustness of our procedure.

The total population size for each MPP design was fixed at 300. This number is, first of all, realistic and allows sufficiently fast calculations for different MPP configurations while retains sufficient power for QTL detection across the full population as well as within families. The six families in the diallel design named from *Diallel1* to *Diallel6* contained 50 progenies per family, and the three families named from *NAM1* to *NAM3* in the NAM design included 100 progenies per family. For the MAGIC population, only one family, named *MAGIC1,* had 300 progenies. To show the genetic relatedness between progenies, we performed principal component analysis (PCA) on the 'true' simulated genome from each MPP design (Figure 2 Bottom panel).

The phenotype of each progeny was the sum of the genetic effects of major and minor QTLs plus residual errors: $Y = az_1 + az_2 + az_3 + \sum_{i=1}^{24} bz_i + e$, where a and b are the major and minor additive effects, respectively; and z_i is the genotype indicator equal to 1 or -1 for the marker with homozygous genotype coded as 11 or 22, and equal to 0 for (residual) heterozygotes; the residual term e followed a normal distribution $N(0,1)$. For simplicity and convenience, we assigned only additive effects at the bi-allelic level to the simulated QTLs. To choose 'reasonably detectable' QTL effect sizes for major and minor QTLs, i.e., neither too low nor too high to compare the performance of the six models, we did a grid search on QTL effect sizes and calculated the corresponding realized heritability by $h^2 = \sigma_G^2/(\sigma_G^2 + 1)$ where σ_G^2 is the genetic variance contributed by both major and minor QTLs. As an example used for illustration in this paper, the effect size of major QTLs was taken to be 0.4, and the effect size of all minor QTLs was 0.1, which led to an h^2 in each design between 0.35 to 0.38. Appendix Figure S1 shows an example of the distribution of simulated phenotypes in each family of different MPP designs.

3.1.3. Assessment of performance of QTL mapping models

We simulated 500 realizations for each MPP design and assessed the performance of all six models on these 500 simulated datasets. As criteria for performance, we chose *TPR* and mapping resolution of the major QTL. TPR was defined as follows. The successful detection of a major QTL as a true positive had to meet two requirements. Firstly, the -log10(p) value for the likelihood ratio test for the variance component of the QTL effects at that position should exceed the threshold of 4.2, obtained by the Bonferroni correction on all 462 markers on the consensus map at a genome-wide significance level of 0.01. Secondly, the distance between the true position of the simulated major QTL and the peak marker with the highest -log10(p) value on the same chromosome should be within a certain distance. The TPR for each major QTL was the average percentage of runs where true QTLs were detected within

certain distances 5cM, 10cM, and 20cM. The mapping resolution of a major QTL in this study was indicated by the average genetic distance between the true simulated position and the detected position with the highest -log10(p) value on the same chromosome over 500 runs. A shorter distance indicates a higher mapping resolution.

3.2.Empirical MPP designs

Besides the three simulated MPP designs, we re-analyzed empirical diallel, NAM, and MAGIC designs collected from previous studies and the vegetable breeding company Rijk Zwaan. These six datasets are summarized in Table 2, and their crossing schemes are shown in Figure 3.

3.2.1. Maize MPP designs

In the first maize diallel design (Figure 3 A), recombinant inbred lines (RILs) were derived from crosses between four inbred parents (Coles et al., 2010). The four parents represent distinct germplasm groups in temperate (B73 and B97) and tropical (CML254 and Ki14) types of maize (K. Liu et al., 2003), and parents of different types were crossed with each other. In total, 569 progenies in four families were obtained with 1,339 genotyped markers. Traits of interest were days to anthesis (DTA), days to silking (DTS), plant height (PH), ear height (EH), and total leaf number (TLN) measured in long-day and short-day environments. The photoperiodic responses of those traits were the difference between long-day and short-day responses. The previous study used the package MCQTL to perform QTL mapping in joint and separate families (Jourjon et al., 2005).

The second maize MPP design is the Dent panel of the EU-NAM population (Figure 3 B). Ten families of 841 DH progenies were derived from 11 parents in which the central parent F353 was crossed with ten peripheral lines (Bauer et al., 2013). We used the consensus map of 22,122 Panzea markers that was also used in previous studies (Bustos-Korts et al., 2016; Garin et al., 2017; Giraud et al., 2014). We removed the marker if at least one of the 11 parents had a missing genotype at this marker, which led to 15,813 markers for the IBD calculation and QTL analysis. QTL mapping for DGY and PH was previously performed in a multi-QTL effect model that assessed mixture types of QTL effects at parental, ancestral, and bi-allelic levels (Garin et al., 2017).

The last maize MPP design is the eight-way MAGIC population (Figure 3 C) (Dell'Acqua et al., 2015). The eight maize inbred lines (A632, B73, B96, F7, H99, HP301, Mo17, W153R) were crossed in the format of 35 independent breeding funnels containing two-way, four-way, and eight-way crosses. According to the previous study (Dell'Acqua et al., 2015), the two-way cross $B96 \times HP301$ failed during the MAGIC population construction. A ninth parent (CLM91) was introduced in the two-way cross $B73 \times CML91$ to complement four-way crosses with failed two-way cross $B96 \times HP301$. Each funnel was advanced by single seed descent to the F6 generation with 529 progenies and 41,473 genotyped markers. 529

progenies were phenotyped in two different environments for days to pollen shed (PS), plant height (PH), ear height (EH), and grain yield (GY). In the previous study, two methods were applied for QTL mapping (Dell'Acqua et al., 2015): the linkage mapping approach using genotype probabilities as predictors and imposing a kinship VCOV structure on the polygenic term, and the association mapping approach estimating allelic additive effects.

For these empirical maize MPP designs, we did not use all the markers described above for analysis. Instead, we selected the markers at each 0.5cM for IBD computation and QTL mapping to remove co-located markers and speed up the analysis.

3.2.2. Tomato MPP designs

We had two tomato MPP designs provided by the breeding company Rijk Zwaan. The first one is a diallel F2 design constructed by crossing three inbred lines differing in fruit shape (Figure 3 D). The three inbred parents were crossed with each other to generate 248 F2 progenies with 459 genotyped markers. The second tomato MPP design combines two NAM F2 designs, where the two central parents are connected (Figure 3 E). We call the whole set-up a connected NAM design where 718 progenies were genotyped with 593 markers and phenotyped based on their resistance level.

The last tomato MPP design is an eight-way MAGIC population (Figure 3 F). In this population, eight inbred tomato lines with different molecular and physiological levels were selected as founders to capture the wide genetic diversity (Pascual et al., 2013). In the previous research, the trait of interest for QTL mapping was fruit weight (FW) measured in two locations for 397 F4 individuals with 1345 genotyped markers (Pascual et al., 2015). Interval mapping with adjusted P-values and GWAS approaches were applied for the QTL analysis. For the interval mapping, the previous study implemented R package mpMAP to estimate parental effects based on multipoint probabilities (B. E. Huang et al., 2011). The percentage of phenotypic variation was calculated by fitting all significant QTLs in a full mixed model (B. E. Huang et al., 2011; Pascual et al., 2015). For the GWAS approach, the kinship matrix was computed to describe the VCOV structure of the polygenic term in a mixed linear model using software TASSEL (Bradbury et al., 2007).

4. Results

4.1.Results of simulated MPP designs

4.1.1. MPP simulation

To illustrate relatedness in each simulated MPP design, we performed PCA on progenies using the simulated genome (IBS) (Figure 2 Bottom panel). The PCA plots of diallel and NAM designs show apparent family clusters. For the simulated MAGIC design, no visible family clusters were present.

Parental origins can be uncertain at multiple positions with non-segregating markers or missing genotypes. Using the pedigree information and genomes of parents and offspring, we reconstructed the parental origins represented by IBDs. We extracted the highest parental IBD probabilities at each marker position per progeny and averaged the maximum IBD probabilities of this position across all progenies. The average of the maximum IBD probabilities at all positions was above 0.9 in all simulated MPP designs (Appendix Figure S2).

4.1.2. Model performance on simulate MPP designs

By comparing the mapping results of the six models on simulated MPP designs, we aimed to prove the advantages of using IBD information for genetic predictors and the necessity of using the mixed model approach to account for the multilevel relatedness of offspring with and across families. Figure 4 demonstrates the performance of each model in terms of TPRs and mapping resolutions at the three major QTLs in the three simulated MPP designs. From the basic to the advanced IBD-based mixed models, we can observe the improved trend regarding TPR and mapping resolution.

On chromosome 5 where neither major nor minor QTLs were simulated, we counted how many markers were detected as QTLs (i.e., false positives) by each of the six models over 500 runs in each MPP design. Given 110 markers on chromosome 5, we expected the number of false positives over 500 runs using a threshold of 4.2 to be $110 \times 500 \times 10^{-4.2} = 3.47$. The reference model *IBS.Kin* identified 2, 3, and 5 positions as false QTLs in the respective simulated diallel, NAM, and MAGIC designs; among IBD-based models, only the advanced model *IBD.MQMkin_F* detected one false QTL on chromosome 5 in the simulated diallel design over 500 runs. It indicates the good control of the false positive rate using the Bonferroni threshold for all models.

In the simulated diallel design, the IBD-based models *IBD.Kin_F* and *IBD.MQMkin_F,* incorporating polygenic effects, generally performed better than other models considering both TPR and mapping resolution. For the *simQTL1* segregating in four out of six families, the TPR (39.4%) from *IBD.MQMkin_F* was higher than that at *simQTL2* (27.9%) and *simQTL3* (22.9%), both segregating in three out six families.

In the simulated NAM design, IBD-based models have advantages over the reference model *IBS.Kin* in improving TPRs and mapping resolutions for *simQTL1* and *simQTL2*. Notably, the advanced IBD-based model *IBD.MQMkin_F* detected *simQTL2*, which is segregating in all three families, with the highest TPR (59%) at a relatively high mapping resolution (4cM). For *simQTL1,* segregating in two out of three families, the TPR (44.8%) and mapping resolution (5.9cM) obtained from *IBD.MQMkin_F* were also better than other models. However, for *simQTL3* that segregated in only one family, all models identified this QTL with low TPRs and mapping resolutions, even though *IBD.MQMkin_F* performed slightly better.

The simulated MAGIC design seems to be the most promising MPP for detecting all three *simQTLs* — all of them were detected with relatively high TPRs and mapping resolutions. Especially for *simQTL1* with expected genotypes frequencies of 0.5, both the advanced IBD-based model *IBD.MQMkin_F* and the reference IBS-based model *IBS.Kin* successfully detected this QTL with a high TPR (around 77%) and high resolution (around 2.9cM). For *simQTL2* and *simQTL3*, advanced models *IBD.Kin_F* and *IBD.MQMkin_F* performed better than the rest of the models considering both TPRs and mapping resolutions.

4.2.Results of empirical MPP designs

4.2.1. Maize diallel design

The maize diallel design generated four families from four parents with multiple traits measured and analyzed in the previous study (Coles et al., 2010). In this research, we re-analyzed the photoperiodic responses of DTA, DTS, the difference in GDD between DTA and DTS (GDDASI), PH, EH, and TLN (Figure 5 A; Appendix Table S1 and Figure S3). The analysis of those traits using IBD-based models showed multiple shared QTLs, but the last model, *IBD.MQMkin_F* is superior to other models because it detected most QTLs with increased mapping signal and the relatively small Bayesian information criterion (BIC) value.

Coles et al. (2010) reported QTL mapping using separate biparental families and joint families and found that joint mapping detected more QTLs with higher resolution. These QTLs were found to coincide with key flowering time QTLs on chromosomes 1, 8, 9, and 10. Here we compare our results to the joint mapping of Coles et al. (2010): in the example of trait GDDTAP, we detected 7 QTLs on chromosomes 1, 2, 3, 4, 8, 9, and 10, which is comparable to the 6 QTLs detected on chromosomes 1, 2, 4, 8, 9, and10 in the joint model by Coles et al. (2010). For other traits (Appendix Figure S3), it also shows the advanced model *IBD.MQMkin_F* could detect most reported QTLs in the study by Coles et al. (2010). Due to the smaller BIC of the *IBD.MQMkin_F* model with more detected QTLs, we fitted these 7 detected QTLs in this advanced IBD-based mixed model to estimate the parental effects at those QTLs for trait GDDTAP (Figure 5 A). It shows that the parents of temperate type (B73 and B97) contribute negative effects at those QTLs while the other two parents of tropical type (CML254 and Ki14) contribute the positive effects at most of those detected QTLs.

4.2.2. Maize NAM design

In the maize NAM population, we identified two QTLs for DGY on chromosomes 6 and 8 using models without correction for genomic background, i.e., *IBD.SQM_U* and *IBD.SQM_F* (Figure 5 B; Appendix Table S1 and Figure S4). Including either cofactors, the polygenic term, or both of them increased the magnitude of the mapping signals for the two QTLs and allowed us to detect new QTLs on chromosomes 3 and 7 for trait DGY. As for the mapping results of PH (Appendix Table S1 and Figure S4), the advanced models *IBD.kin_F* and

IBD.MQMkin_F detected a new QTL on chromosome 5 and increased the magnitude of mapping signals for other QTLs that were also detected by basic models.

For both DGY and PH, our models detected QTLs that were detected by Garin et al. (2017) using a so-called multi-QTL effects model that combined genome-wide scans at bi-allelic, parental, and ancestral levels. Our study used a stringent threshold via Bonferroni correction, so some QTLs with relatively weaker signals were missed compared to the analysis by (Garin et al., 2017).

4.2.3. Maize MAGIC design

For the maize MAGIC population, we mapped the QTLs for trait PS, PH, EH, and GY (Figure 5 C; Appendix Table S1 and Figure S5). Using all five models, we detected one QTL on chromosome 8 for PS, and identified one QTL on chromosome 6 for both traits PH and GY, while no QTL was detected for EH, so the mapping profiles for EH were not shown. Dell'Acqua et al. (2015) used all 529 magic maize progenies for QTL mapping. However, in this study, we selected 303 progenies individuals in the population derived from the initial eight parents, while individuals derived from the later added ninth parent were not included because for those individuals, no pedigree data were available to calculate IBDs. We could confirm some major QTLs detected earlier by Dell'Acqua et al. (2015) in their analysis, but due to the smaller total population size in our analysis, we missed some QTLs, e.g., two QTLs for trait EH, one QTL on chromosome 8 for PH and PS.

4.2.4. Tomato diallel design

The diallel F2 design of two beef tomatoes and one round tomato generated a population of diverse fruit shapes, and we identified some QTLs for fruit shape (Figure 5 D; Appendix Table S1). In total, we detected three QTLs on chromosomes 1, 2, and 9. All three QTLs were detected using *IBD.SQM_F* and *IBD.MQMkin_F* models, while other models missed the QTL on chromosome 1 or 9 with relatively weak signals. The three detected QTLs using were fitted in the model with the smallest BIC value (*IBD.MQMkin_F)* to estimate the parental effects (Figure 5 D). It shows that the parental effect of one beef type tomato (parent B) contributed negatively to the fruit shape at all three QTLs, and the other beef type tomato (parent A) show negative effects at two QTLs on chromosomes 2 and 9 on which the round type tomato (parent C) show positive effects at these two QTLs.

4.2.5. Tomato NAM design

In the connected NAM F2 design, a resistant parent crossed with four susceptible lines, and one of the susceptible lines was crossed with another two resistant lines. All five IBD-based mixed models identified two QTLs with strong signals on chromosomes 1 and 6 (Figure 5 E; Appendix Table S1). Because all models detected those two QTLs with strong signals, the parental effects estimated by the five models show no big difference. As an example, we estimated the parental effect in the *IBD.MQMkin_F* model to show that the three resistant

parents (parent C, B, and D), at the strongest QTL on chromosome 6, contributed negatively to the disease score. Tomato breeders have successfully fine-mapped this QTL as a strong resistance gene.

4.2.6. Tomato MAGIC design

In the eight-way MAGIC F4 population, we mapped QTLs underlying fruit weight measured at two locations that we will refer to as A and B. The model *IBD.MQMkin_F* detected most QTLs for location A, and all models detected three consistent QTLs for location B (Figure 5 F; Appendix Table S1 and Figure S6). For location A, models that accounted for the genomic background by adding either cofactors or the polygenic term or both (*IBD.MQM_F*, *IBD.Kin_F*, or *IBD.MQMkin_F*), compared to *IBD.SQM_U* and *IBD.SQM_F*, allowed us to detect more QTLs on chromosomes 5 and 9. *IBD.Kin_F* and *IBD.MQMkin_F* identified two linked QTLs on chromosomes 2 and 11. The advanced model *IBD.MQMkin_F*, among the five models, has the smallest BIC value (Appendix Table S1).

The previous study by (Pascual et al., 2015) applied two approaches for QTL detection in this design. One was the interval mapping for founder effects based on the multipoint probability calculations, and another was a GWAS approach incorporating a polygenic term. Pascual et al. (2015) analyzed fruit weight measured at location A using interval mapping and detected nine QTLs on chromosomes 2, 3, 5, 7, 8, and 11. In our study, we detected eight QTLs with the *IBD.MQMkin_F* model. We fitted QTLs in the model *IBD.MQMkin* to estimate the percentage of phenotypic variation explained by the QTLs. The eight QTLs identified in our study slightly increased the explained percentage of phenotypic variation, from 51% in Pascual et al. (2015) to 56% now. For the fruit weight measured at location B, we detected the three QTLs on chromosomes 2, 3, and 11 in the same region that has been previously detected using interval mapping. The explained percentage of phenotypic variation of 33% was close to 34% in the previous study. The parental effects on those QTLs estimated from the *IBD.MQMkin_F* model can be conformed with parental performance. For instance, The parental effects at all detected QTLs of FW showed that parent *Cervil*, with the lightest fruit, contributed negative values to fruit weight in both A and B locations.

5. Discussion

5.1.MPPs show design-specific properties

Different MPPs are constructed for different goals. A diallel mating design of carrot was constructed to dissect the genetic architecture of shoot growth by estimating the general and specific combing abilities and non-additive effects (Turner et al., 2018); a maize NAM population was proven to be able to capture small effect QTLs when they were shared by families (Ogut et al., 2015); MAGIC populations allow a large set of QTLs segregating with higher resolution and thus can increase the chance of detecting QTLs (Mackay et al., 2014). Another study compared the different designs of biparental, multiparental, and association

panels in the context of the genome sequencing era to show their complementarity in genetic studies (Pascual et al., 2016).

This study simulated diallel, NAM, and MAGIC designs using four real Arabidopsis inbred lines. We focused on those MPP designs because they are often used in genetic research and breeding programs, and NAM and MAGIC designs are components of other more general designs that can also be analyzed using this framework.

QTL mapping results are impacted by the crossing schemes between parents in MPP designs and the genomic background. Probabilities for segregation differ between families within MPP designs owing to the specific crosses between parents. For our *simQTL3* segregating in only one family of the simulated NAM design, even the advanced IBD-based models could not improve the mapping results, while this QTL could be detected with higher TPR and mapping resolution in both diallel and MAGIC designs. The reason might be that the joint family QTL mapping in NAM designs favors the large-effect QTLs or QTLs shared by most families (Bajgain et al., 2016; Garin et al., 2017; Ogut et al., 2015). The other QTLs, *simQTL2* and *simQTL3* were expected to have the same allele frequency in the simulated MAGIC or diallel design, but it turned out that *simQTL2* could be detected with higher TPR and mapping resolution than *simQTL3*. The reason is that chromosome 2, where *simQTL2* is located, provided a more contrasting genomic background combining *parent1* with other parents than chromosome 3.

We expect multilevel relatedness between individuals as being full-sib, half-sib, or unrelated depending on the specific MPP design. In the simulated MAGIC design, each progeny's genome was the uniformly reshuffled genome of all parents (Dell'Acqua et al., 2015; Ongom et al., 2018; Pascual et al., 2015), and thus no apparent clustering or grouping was observed in the PCA plot. In simulated diallel and NAM design, we observed multilevel relatedness of offspring within and across families: the full-sib progenies gathering in their family clusters were more genetically correlated than the half-sib progenies sharing one common parent.

Different MPP designs require different QTL analysis models with the first question being the choice of using IBS or IBD information in the genetic predictors and a second question concerning how to deal with individual relatedness within and across families.

5.2.The IBD is informative to reflect the genome origins

We used the observed IBS with 5% of missing genotypes to estimate IBD probabilities and then compared IBS-based with IBD-based models. Parental origins can be ambiguous at non-segregating, missing, or mis-genotyped loci based on as IBS information. Therefore, we inferred IBD probabilities by using the pedigree information and the whole genome (Zheng et al., 2015, 2018). Generally, IBD-based models were more effective than the IBS-based model (*IBS.Kin*) concerning the TPRs and mapping resolutions for major QTLs.

A reliable approach for precise IBD computations is fundamental for inferring parental origins and performing IBD-based QTL mapping, but only a few methodologies are available, and most of them were limited to specific MPP designs (Broman et al., 2018; Verbyla et al., 2014). In this study, we used a general hidden Markov model framework to construct parental origins (Zheng et al., 2015, 2018), which was successfully extended to all kinds of MPP designs.

IBD information is not only useful as a basis for genetic predictors in the QTL mapping models but also valuable in consensus map construction. The traditional process of consensus map construction can be tedious in the MPP context, including marker cleaning, grouping, ordering (Taylor, 2018; Wu et al., 2008), and map integrating (Endelman et al., 2014). In future work, we can infer the recombination between markers for genetic map construction using IBD probabilities to simplify and optimize the MPP analysis framework from the beginning (Zheng et al., 2019).

5.3.Modeling family-specific VCOV structure on the residual term is recommendatory

Owing to the major and minor QTLs with varying segregation configurations in each family, the family-specific distribution of phenotypes motivated us to model a family-specific VCOV structure on the residual term to account for the family-genetic background. However, based on the simulated diallel and NAM designs, there is no substantial evidence to show the advantage of using *IBD.SQM_F* model over *IBD.SQM_U* model. Another study also reports that modeling a heterogeneous VCOV structure on the residual term may not always improve the mapping results (Garin et al., 2017). One of the reasons might be the limited family size, e.g., a family size of 50 or 100 in simulated diallel or NAM designs may not have been big enough to reveal the heterogeneity of variance components between families. Still, modeling a heterogeneous error can be advantageous in the fitting of single-locus QTL models in initial genome scans, whereas in the latter stages of the building of a multi-locus QTL model, the advantage of a heterogeneous error diminishes because most of the genetic effects have been incorporated in the QTL structure leaving the residual less heterogeneous.

Non-genetic factors can also cause family-specific variation. It is common in an MPP breeding program where each family is separately established and subjectively phenotyped by different breeders in specific locations. Therefore, we recommend modeling a family-specific VCOV structure on the residual term to account for the potential family background due to both genetic and non-genetic reasons.

5.4.Imposing a kinship structure on the polygenic effect accounts for individual relatedness

Including only significant positions as cofactors from the initial genome-wide scans can lead to ignoring part of the genetic variance and missing heritability (Myles et al., 2009). To deal with smaller QTLs that may go unnoticed, we incorporated a polygenic effect whose VCOV structure is described by a kinship matrix. A study on a three-way barley cross has shown

that the inclusion of the kinship VCOV structure containing co-ancestry information can avoid unrealistic marker-trait associations (Marcos Malosetti et al., 2011). In our study, the re-analysis of the empirical tomato MAGIC population with the polygenic term allowed us to detect more QTLs for fruit weight measured in location A with the relatively small BIC. In the simulated MAGIC design, adding a polygenic term (*IBD.Kin_F*) increased TPRs and mapping resolutions for all *simQTLs* compared with *IBD.SQM_F* and *IBD.MQM_F*.

Population-based mapping approaches incorporating individual relatedness are widely applied to association panels. In specific MPP designs (e.g., diallel and NAM), multilevel relatedness exists between individuals as being full-sib, half-sib, and unrelated within and across families. A priori no population structure is expected in standard MAGIC designs, but MAGIC lines may still show complicated realized genetic relationships. Multilevel relatedness can be corrected by using a general QK model where the Q matrix accounts for family structure, and the pairwise relationship matrix K deals with the individual relatedness (Yu et al., 2006). Likewise, this study modeled the family-specific residual term to correct for family structure and imposed a kinship VCOV structure on the polygenic term to incorporate multilevel relatedness.

5.5.The advanced IBD-based model works well for general MPP designs

To sum up, we can refer to conceptions from family-based and population-based mapping approaches to explain the efficiency of our approach. Family-based QTL mapping assumes QTL effects to be multi-allelic and referring to parental origins. For their estimation, we need design matrices that are functions of IBD probabilities. Popular population-based mapping strategies deal with multilevel relatedness by imposing family-specific and kinship-based VCOV structures on respectively the non-genetic residual and the polygenic terms in a mixed model approach. Family-based and population-based mapping approaches complement each other, and their synthesis in an advanced IBD-based mixed model approach (*IBD.MQMkin_F*) offers us a robust and comprehensive solution to map QTL in general MPP designs. In our simulation study, we observed no case where the *IBD.MQMkin_F* model performed significantly worse than other IBD-based models in terms of TPR and mapping resolution. Most results from empirical MPP designs also show that the unified *IBD.MQMkin_F* model detected most QTLs with relatively small BIC, and the major QTLs were comparable to those identified by previous studies using alternative mapping tools.

Declarations

Author contribution statement All authors contributed to developing the theoretical framework, editing the manuscript, and approving the final version. MPB configured the fundamental of fitting multi-QTL in the mixed model, and WHL extended this work to general MPPs with options to fit polygenic effects. CZ developed the software for the IBD computation and helped to interpret the output. RJ provided empirical data sets from Rijk Zwaan and gave constructive feedback to optimize the framework. FvE and MPB coordinated this research and provided critical feedback regarding the study design.

Funding This research was funded by Rijk Zwaan B.V.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Bajgain, P., Rouse, M. N., Tsilo, T. J., Macharia, G. K., Bhavani, S., Jin, Y., & Anderson, J. A. (2016). Nested Association Mapping of Stem Rust Resistance in Wheat Using Genotyping by Sequencing. *PLOS ONE*, *11*(5), e0155760. doi: 10.1371/journal.pone.0155760

Bauer, E., Falque, M., Walter, H., Bauland, C., Camisan, C., Campo, L., Meyer, N., Ranc, N., Rincent, R., Schipprack, W., Altmann, T., Flament, P., Melchinger, A. E., Menz, M., Moreno-González, J., Ouzunova, M., Revilla, P., Charcosset, A., Martin, O. C., & Schön, C.-C. (2013). Intraspecific variation of recombination rate in maize. *Genome Biology*, *14*(9), R103. doi: 10.1186/gb-2013-14-9-r103

Baxter, I., Brazelton, J. N., Yu, D., Huang, Y. S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J., Borevitz, J. O., Nordborg, M., Vitek, O., & Salt, D. E. (2010). A Coastal Cline in Sodium Accumulation in Arabidopsis thaliana Is Driven by Natural Variation of the Sodium Transporter AtHKT1;1. *PLoS Genetics*, *6*(11), e1001193. doi: 10.1371/journal.pgen.1001193

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. doi: 10.1093/bioinformatics/btm308

Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, Ś., Yandell, B. S., & Churchill, G. A. (2018). R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multi-parent Populations. *Genetics*, genetics.301595.2018. doi: 10.1534/genetics.118.301595

Bustos-Korts, D., Malosetti, M., Chapman, S., Biddulph, B., & van Eeuwijk, F. (2016). Improvement of Predictive Ability by Uniform Coverage of the Target Genetic Space. *G3 (Bethesda, Md.)*, *6*(11), 3733–3747. doi: 10.1534/g3.116.035410

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2018). *ASReml-R Reference Manual Version 4 ASReml estimates variance components under a general linear mixed model by residual maximum likelihood (REML)*. Retrieved from http://www.homepages.ed.ac.uk/iwhite/asreml/uop.

Coles, N. D., McMullen, M. D., Balint-Kurti, P. J., Pratt, R. C., & Holland, J. B. (2010). Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. *Genetics*, *184*(3), 799–812. doi: 10.1534/genetics.109.110304

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A. L., Aung, H. H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G. A., Inzé, D., Morgante, M., & Pè, M. E. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome Biology*, *16*(1), 167. doi: 10.1186/s13059-015-0716-z

Endelman, J. B., & Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics*, *30*(11), 1623–1624. doi:

10.1093/bioinformatics/btu091

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler, E. S. (2003). Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology*, *54*(1), 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

Gardner, K. A., Wittern, L. M., & Mackay, I. J. (2016). A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnology Journal*, *14*(6), 1406–1417. doi: 10.1111/pbi.12504

Garin, V., Wimmer, V., Mezmouk, S., Malosetti, M., & van Eeuwijk, F. (2017). How do the type of QTL effect and the form of the residual term influence QTL detection in multi-parent populations? A case study in the maize EU-NAM population. *Theoretical and Applied Genetics*, *130*(8), 1753–1764. doi: 10.1007/s00122-017-2923-3

Giraud, H., Bauland, C., Falque, M., Madur, D., Combes, V., Jamin, P., Monteil, C., Laborde, J., Palaffre, C., Gaillard, A., Blanchard, P., Charcosset, A., & Moreau, L. (2017). Reciprocal genetics: Identifying QTL for general and specific combining abilities in hybrids between multiparental populations from two maize (zea mays L.) heterotic groups. *Genetics*, *207*(3), 1167–1180. doi: 10.1534/genetics.117.300305

Giraud, H., Lehermeier, C., Bauer, E., Falque, M., Segura, V., Bauland, C., Camisan, C., Campo, L., Meyer, N., Ranc, N., Schipprack, W., Flament, P., Melchinger, A. E., Menz, M., Moreno-González, J., Ouzunova, M., Charcosset, A., Schön, C. C., & Moreau, L. (2014). Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic QTL for hybrid performance in the flint and dent heterotic groups of maize. *Genetics*, *198*(4), 1717–1734. doi: 10.1534/genetics.114.169367

Huang, B. E., & George, A. W. (2011). R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics*, *27*(5), 727–729. doi: 10.1093/bioinformatics/btq719

Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., Leung, H., Varshney, R. K., & Cavanagh, C. R. (2015). MAGIC populations in crops: current status and future prospects. *Theoretical and Applied Genetics*, *128*(6), 999–1017. doi: 10.1007/s00122-015-2506-0

Huang, X., Paulo, M. J., Boer, M., Effgen, S., Keizer, P., Koornneef, M., & Van Eeuwijk, F. A. (2011). Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(11), 4488–4493. doi: 10.1073/pnas.1100465108

Jourjon, M.-F. o., Jasson, S., Marcel, J., Ngom, B., & Mangin, B. (2005). MCQTL: multi-allelic QTL mapping in multi-cross design. *Bioinformatics*, *21*(1), 128–130. doi: 10.1093/bioinformatics/bth481

Liu, K., Goodman, M., Muse, S., Smith, J. S., Buckler, E., & Doebley, J. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites.

693 *Genetics*, *165*(4), 2117–2128. Retrieved from
694 http://www.ncbi.nlm.nih.gov/pubmed/14704191

695 Liu, Y., & Zeng, Z. B. (2000). A general mixture model approach for mapping quantitative
696 trait loci from diverse cross designs involving multiple inbred lines. *Genetical Research*,
697 *75*(3), 345–355. doi: 10.1017/S0016672300004493

698 Mackay, I. J., Bansept-Basler, P., Barber, T., Bentley, A. R., Cockram, J., Gosman, N.,
699 Greenland, A. J., Horsnell, R., Howells, R., O'Sullivan, D. M., Rose, G. A., & Howell,
700 P. J. (2014). An eight-parent multiparent advanced generation inter-cross population for
701 winter-sown wheat: creation, properties, and validation. *G3 (Bethesda, Md.)*, *4*(9),
702 1603–1610. doi: 10.1534/g3.114.012963

703 Malosetti, M., Van Der Linden, C. G., Vosman, B., & Van Eeuwijk, F. A. (2007). A mixed-
704 model approach to association mapping using pedigree information with an illustration
705 of resistance to Phytophthora infestans in potato. *Genetics*, *175*(2), 879–889. doi:
706 10.1534/genetics.105.054932

707 Malosetti, Marcos, van Eeuwijk, F. A., Boer, M. P., Casas, A. M., Elía, M., Moralejo, M.,
708 Bhat, P. R., Ramsay, L., & Molina-Cano, J. L. (2011). Gene and QTL detection in a
709 three-way barley cross under selection by a mixed model with kinship information using
710 SNPs. *Theoretical and Applied Genetics*, *122*(8), 1605–1616. doi: 10.1007/s00122-011-
711 1558-z

712 Molenberghs, G., & Verbeke, G. (2007). Likelihood ratio, score, and wald tests in a
713 constrained parameter space. *American Statistician*, *61*(1), 22–27. doi:
714 10.1198/000313007X171322

715 Morrell, C. H. (1998). Likelihood Ratio Testing of Variance Components in the Linear
716 Mixed-Effects Model Using Restricted Maximum Likelihood. *Biometrics*, *54*(4), 1560.
717 doi: 10.2307/2533680

718 Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., & Buckler, E. S.
719 (2009). Association mapping: critical considerations shift from genotyping to
720 experimental design. *The Plant Cell*, *21*(8), 2194–2202. doi: 10.1105/tpc.109.068437

721 Ogut, F., Bian, Y., Bradbury, P. J., & Holland, J. B. (2015). Joint-multiple family linkage
722 analysis predicts within-family variation better than single-family analysis of the maize
723 nested association mapping population. *Heredity*, *114*(6), 552–563. doi:
724 10.1038/hdy.2014.123

725 Ongom, P. O., & Ejeta, G. (2018). Mating Design and Genetic Structure of a Multi-Parent
726 Advanced Generation Intercross (MAGIC) Population of Sorghum (Sorghum bicolor
727 (L.) Moench). *G3: Genes, Genomes, Genetics*, *8*(1), 331–341. doi:
728 10.1534/G3.117.300248

729 Pascual, L., Albert, E., Sauvage, C., Duangjit, J., Bouchet, J.-P., Bitton, F., Desplat, N.,
730 Brunel, D., Le Paslier, M.-C., Ranc, N., Bruguier, L., Chauchard, B., Verschave, P., &
731 Causse, M. (2016). Dissecting quantitative trait variation in the resequencing era:

complementarity of bi-parental, multi-parental and association panels. *Plant Science*, *242*, 120–130. doi: 10.1016/J.PLANTSCI.2015.06.017

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J.-P., Le, Q. H., Chauchard, B., Verschave, P., & Causse, M. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnology Journal*, *13*(4), 565–577. doi: 10.1111/pbi.12282

Pascual, L., Xu, J., Biais, B., Maucourt, M., Ballias, P., Bernillon, S., Deborde, C., Jacob, D., Desgroux, A., Faurobert, M., Bouchet, J.-P., Gibon, Y., Moing, A., & Causse, M. (2013). Deciphering genetic diversity and inheritance of tomato fruit weight and composition through a systems biology approach. *Journal of Experimental Botany*, *64*(18), 5737–5752. doi: 10.1093/jxb/ert349

Stich, B., Möhring, J., Piepho, H. P., Heckenberger, M., Buckler, E. S., & Melchinger, A. E. (2008). Comparison of mixed-model approaches for association mapping. *Genetics*, *178*(3), 1745–1754. doi: 10.1534/genetics.107.079707

Stram, D. O., & Lee, J. W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, *50*(4), 1171. doi: 10.2307/2533455

Sul, J. H., Martin, L. S., & Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. In PLoS Genetics (Vol. 14, Issue 12). Public Library of Science. doi: 10.1371/journal.pgen.1007309

Taylor, J. (2018). *Statistics for the Australian Grains Industry Technical Report Series Efficient linkage map construction using R/ASMap*. Retrieved from http://cran.uvigo.es/web/packages/ASMap/vignettes/asmapvignette.pdf

Turner, S. D., Maurizio, P. L., Valdar, W., Yandell, B. S., & Simon, P. W. (2018). Dissecting the Genetic Architecture of Shoot Growth in Carrot (Daucus carotaL.) Using a Diallel Mating Design. *G3 (Bethesda, Md.)*, *8*(2), 411–426. doi: 10.1534/g3.117.300235

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. doi: 10.3168/JDS.2007-0980

Verbyla, A. P., George, A. W., Cavanagh, C. R., & Verbyla, K. L. (2014). Whole-genome QTL analysis for MAGIC. *Theoretical and Applied Genetics*, *127*(8), 1753–1770. doi: 10.1007/s00122-014-2337-4

Wu, Y., Bhat, P. R., Close, T. J., & Lonardi, S. (2008). Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph. *PLoS Genetics*, *4*(10), e1000212. doi: 10.1371/journal.pgen.1000212

Würschum, T., Liu, W., Gowda, M., Maurer, H. P., Fischer, S., Schechert, A., & Reif, J. C. (2012). Comparison of biometrical models for joint linkage association mapping. *Heredity*, *108*(3), 332–340. doi: 10.1038/hdy.2011.78

Würschum, Tobias. (2012). Mapping QTL for agronomic traits in breeding populations. In Theoretical and Applied Genetics (Vol. 125, Issue 2, pp. 201–210). doi:

10.1007/s00122-012-1887-6

Würschum, Tobias, Maurer, H. P., Schulz, B., Möhring, J., & Reif, J. C. (2011). Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. *Theoretical and Applied Genetics*, *123*(1), 109–118. doi: 10.1007/s00122-011-1570-3

Xiao, Y., Tong, H., Yang, X., Xu, S., Pan, Q., Qiao, F., Raihan, M. S., Luo, Y., Liu, H., Zhang, X., Yang, N., Wang, X., Deng, M., Jin, M., Zhao, L., Luo, X., Zhou, Y., Li, X., Liu, J., … Yan, J. (2016). Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytologist*, *210*(3), 1095–1106. doi: 10.1111/nph.13814

Xu, Y., Li, P., Yang, Z., & Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *Crop Journal*, *5*(2), 175–184. doi: 10.1016/j.cj.2016.06.003

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*(2), 100–106. doi: 10.1038/ng.2876

Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*, *178*(1), 539–551. doi: 10.1534/genetics.107.074245

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. doi: 10.1038/ng1702

Zheng, C., Boer, M. P., & van Eeuwijk, F. A. (2015). Reconstruction of Genome Ancestry Blocks in Multiparental Populations. *Genetics*, *200*(4), 1073–1087. doi: 10.1534/genetics.115.177873

Zheng, C., Boer, M. P., & van Eeuwijk, F. A. (2018). Recursive Algorithms for Modeling Genomic Ancestral Origins in a Fixed Pedigree. *G3 (Bethesda, Md.)*, g3.200340.2018. doi: 10.1534/g3.118.200340

Zheng, C., Boer, M. P., & van Eeuwijk, F. A. (2019). Construction of Genetic Linkage Maps in Multiparental Populations. *Genetics*, genetics.302229.2019. doi: 10.1534/genetics.119.302229

Figures and Tables

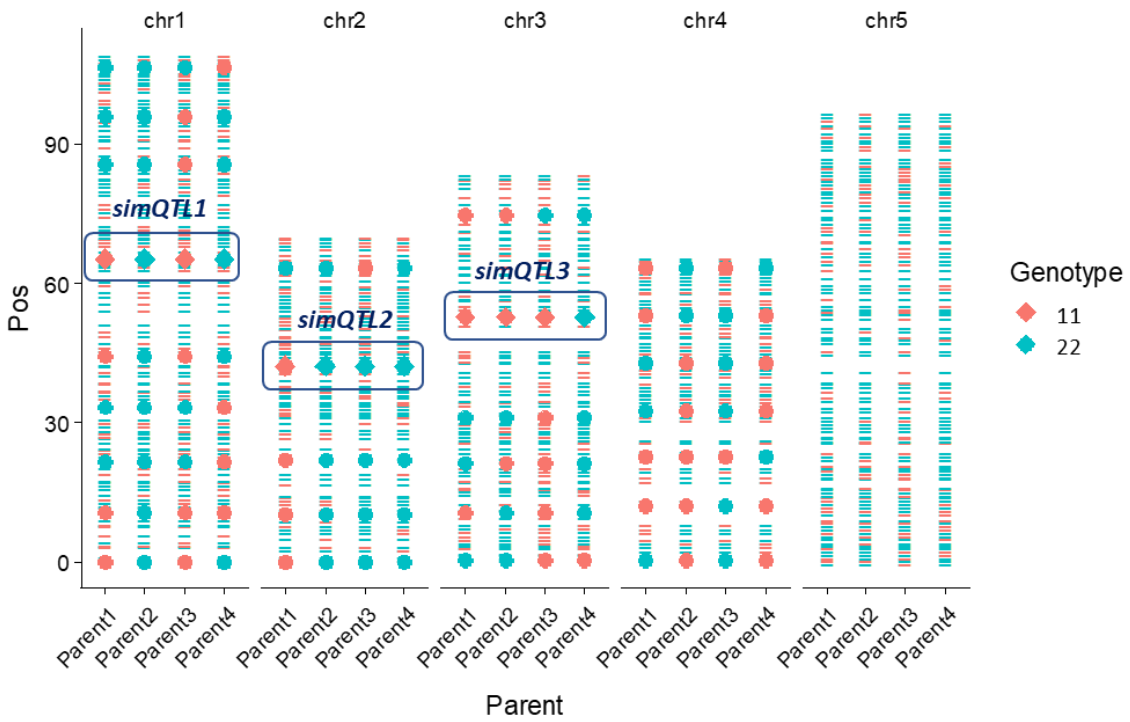


Figure 1 Simulated positions and genotypes for three major QTLs (diamonds) with an additive allelic substitution effect of 0.4 and the allele labelled as 1 increasing the trait. For 24 minor QTLs (dots) the additive allelic substitution effect was 0.1 with the allele labelled as 1 again increasing the trait.

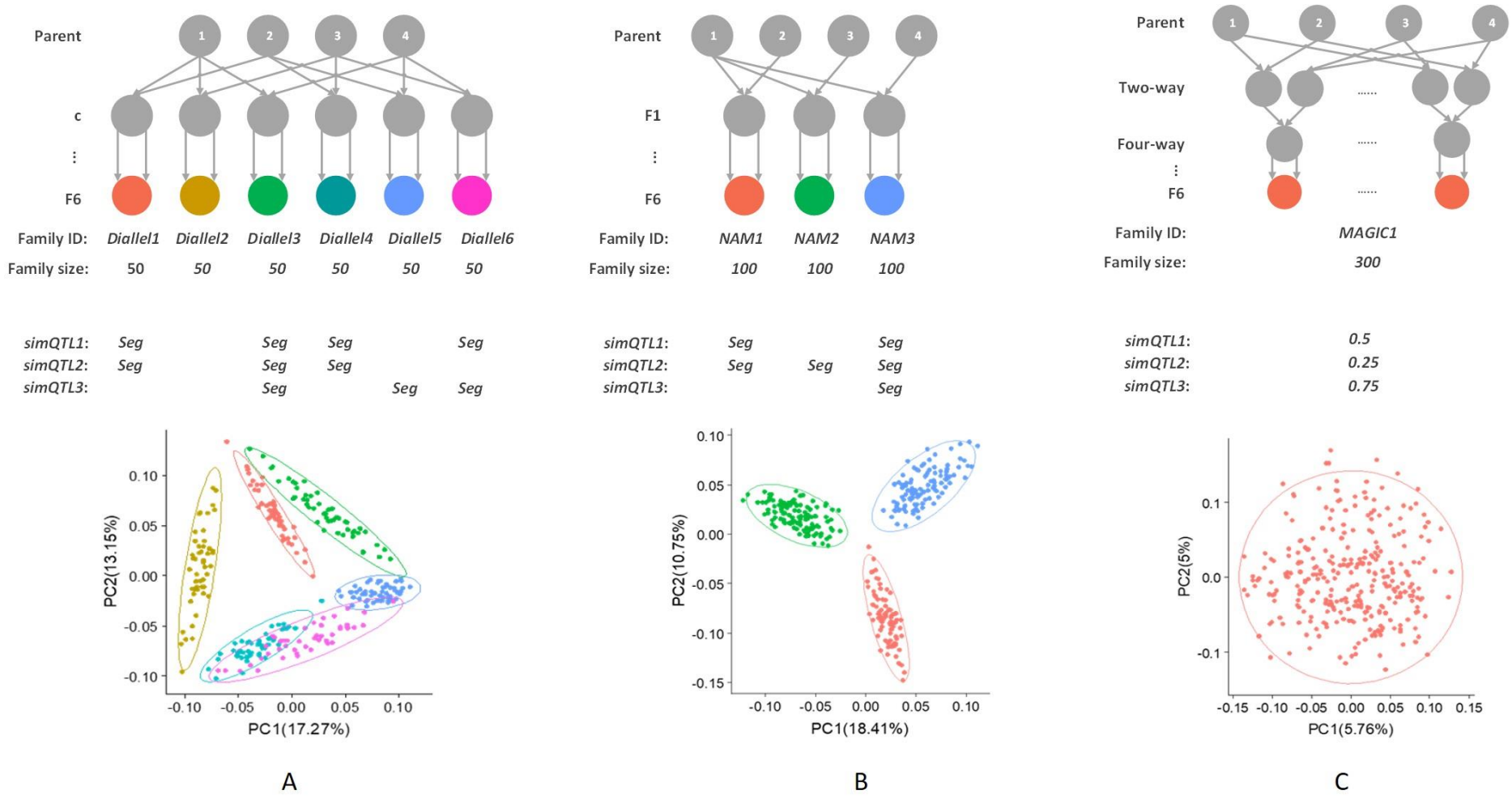


Figure 2 ***Upper panel*** *the crossing schemes of simulated MPP designs.* ***Middle panel*** *Families with segregating QTLs or allele frequencies at those QTLs.* ***Bottom panel*** *PCA plots for progenies based on simulated genome data for* ***A****.diallel,* ***B****. NAM, and* ***C****. MAGIC designs.*

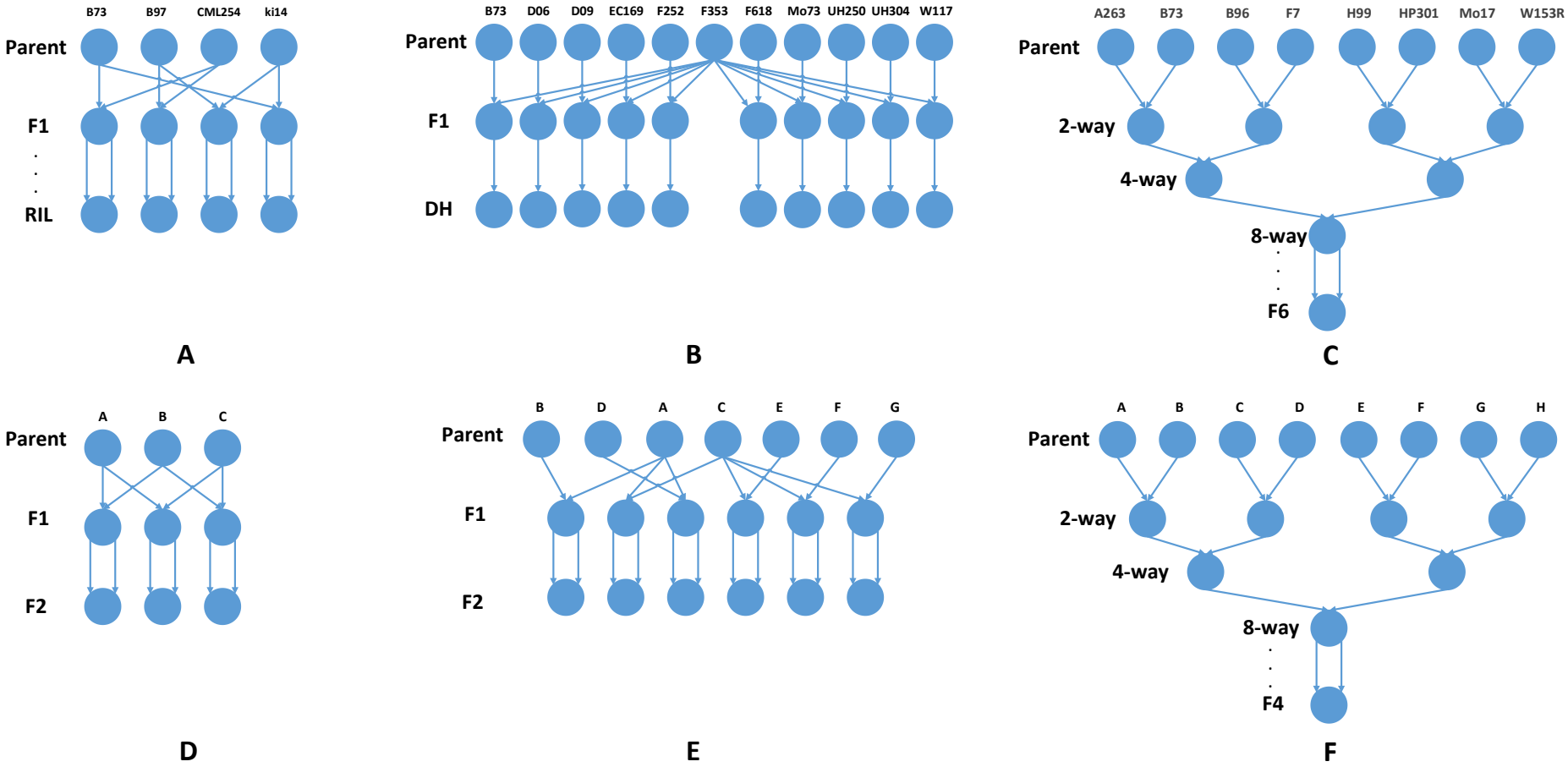


Figure 3 The crossing schemes of empirical MPP designs: A. maize diallel RIL design B. maize NAM DH design C. maize eight-way MAGIC F6 design D. tomato diallel F2 design E. tomato connected NAM F2 design, and F. tomato eight-way MAGIC F4 design.

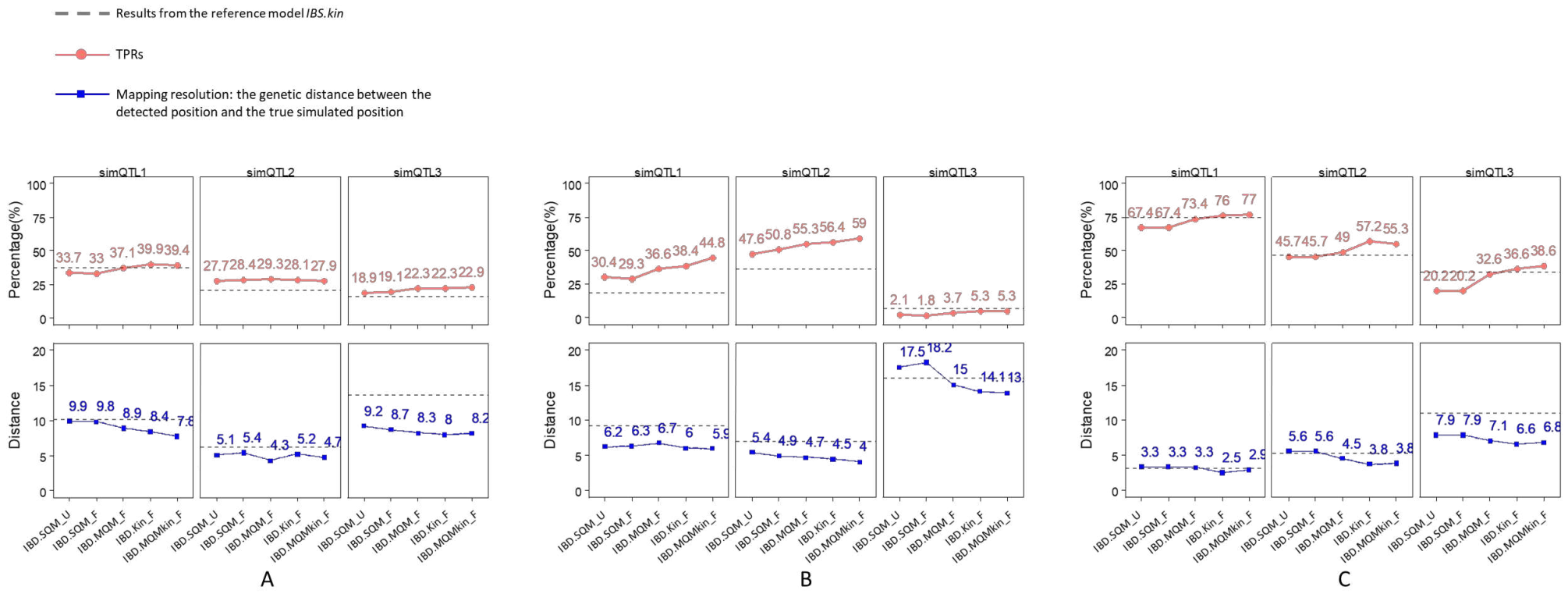
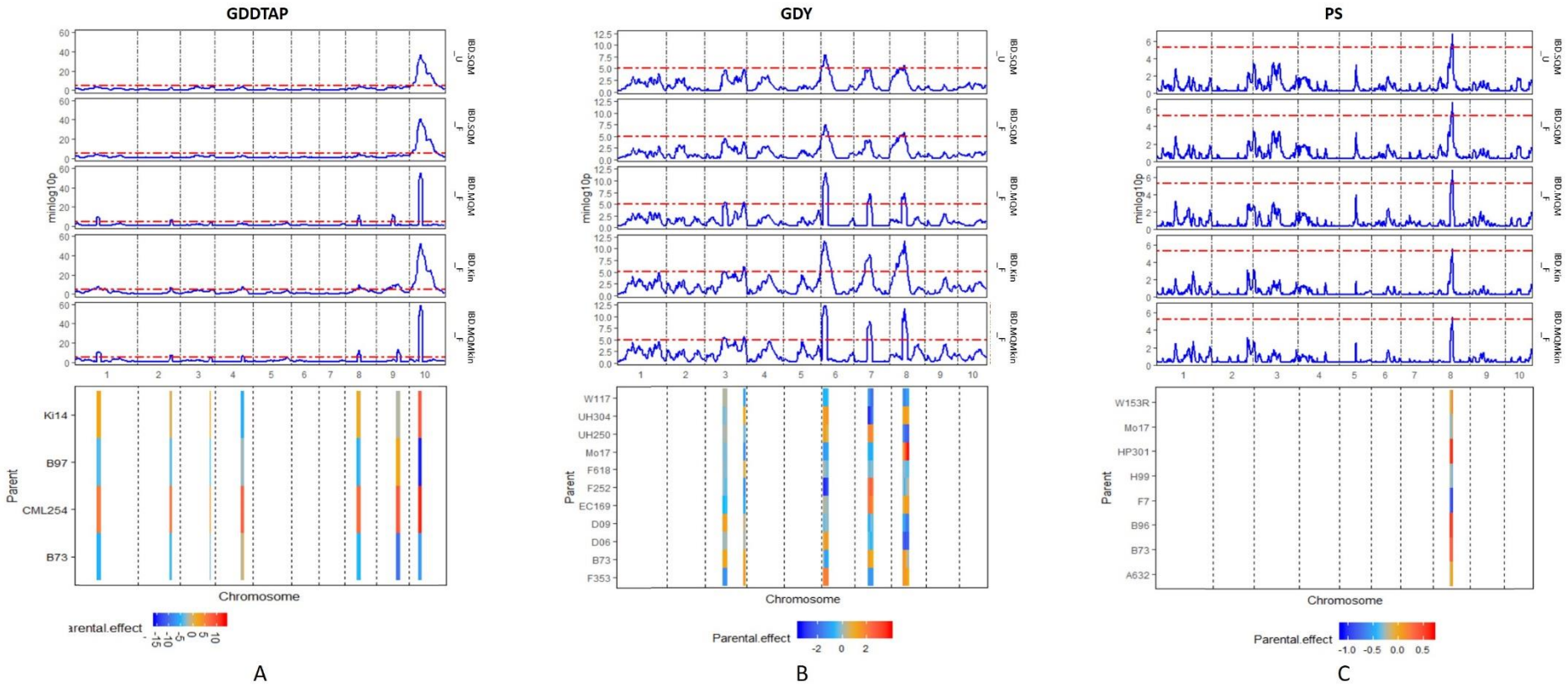


Figure 4 Model comparison in terms of True Positive Rates (dots with red lines) in percentages and mapping resolutions (squares with blue lines) in cM for simulated QTLs in A. diallel B.NAM, and C. MAGIC designs.


GDDTAP
GDY
PS
minlog10p
IBD.SQM_U
IBD.SQM_F
IBD.MQM_F
IBD.Kin_F
IBD.MQMKin_F
Parent
Ki14
B97
CML254
B73
W117
UH304
UH250
Mo17
F618
F252
EC169
D09
D06
B73
F353
W153R
Mo17
HP301
H99
F7
B96
B73
A632
Chromosome
Parental.effect
A
B
C

817

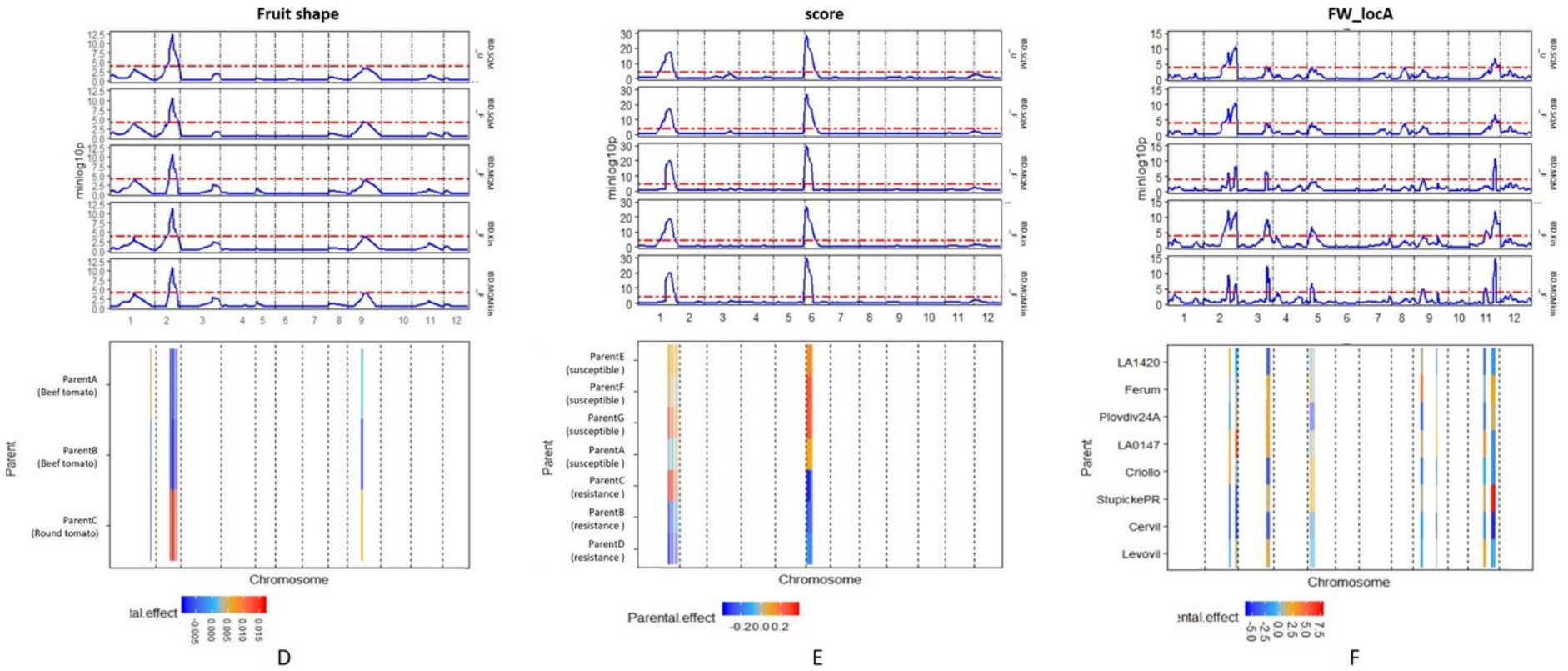


Figure 5 Mapping results of some traits in the empirical MPP designs: ***A.*** *maize diallel,* ***B.*** *maize NAM,* ***C.*** *maize MAGIC,* ***D.*** *tomato diallel,* ***E.*** *tomato NAM and* ***F.*** *tomato MAGI using the five IBD-based mixed models.* ***Upper panel*** *QTL profiles from the five IBD-based mixed model approaches.* ***Bottom panel*** *Estimation of parental effects at QTLs detected by model selected based on BIC. Mapping results of other traits are provided in the Appendix.*

822 *Table 1 Overview of QTL mapping models*

Model name	Genetic predictor	Genome background	Residual structure	Formula	VCOV structure of random terms
IBD.SQM_U	*IBD*	-	*Homogeneous (Uniform)*	$Y = X\beta + M_q a_q + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $\varepsilon \sim N(0, I_N \sigma_\varepsilon^2)$
IBD.SQM_F	*IBD*	-	*Family-specific*	$Y = X\beta + M_q a_q + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2)$
IBD.MQM_F	*IBD*	*Cofactors*	*Family-specific*	$Y = X\beta + \sum_{c \neq q} M_c a_c + M_q a_q + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $a_c \sim N(0, I_P \sigma_c^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2)$
IBD.Kin_F	*IBD*	*Polygenic term*	*Family-specific*	$Y = X\beta + M_q a_q + g + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $g \sim N(0, K\sigma_g^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2)$
IBD.MQMkin_F	*IBD*	*Cofactors polygenic term*	*Family-specific*	$Y = X\beta + \sum_{c \neq q} M_c a_c + M_q a_q + g + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $a_c \sim N(0, I_P \sigma_c^2)$ $g \sim N(0, K\sigma_g^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^{F} I_{n_k} \sigma_{e_k}^2)$
IBS.kin (GWAS model)	*IBS*	*Polygenic term*	*Homogeneous (Uniform)*	$Y = X\beta + X_q \lambda_q + g + \varepsilon$	$g \sim N(0, K\sigma_g^2)$ $\varepsilon \sim N(0, I_N \sigma_\varepsilon^2)$

823

824 *Table 2 Summary of empirical MPP datasets.*

Crop	MPP designs	Family types	# of families	# of progenies	# of markers	Traits	Reference
Maize	Diallel	RIL	4	569	1339	Days to anthesis (DTA), days to silking (DTS), plant height (PH), ear height (EH), and total leaf number (TLN)	(Coles et al., 2010)
	NAM	DH	11	841	22122	Dry grain yield (DGY), PH	(Garin et al., 2017)
	MAGIC	F6	8	303	41473	Pollen shed (PS), plant height (PH), ear height (EH), and grain yield (GY)	(Dell'Acqua et al., 2015)
Tomato	Diallel	F2	3	248	456	Fruit shape	-
	NAM	F2	7	718	593	Resistance	-
	MAGIC	F4	8	397	1345	Fruit weight (FW)	(Pascual et al., 2015)

825

Appendix

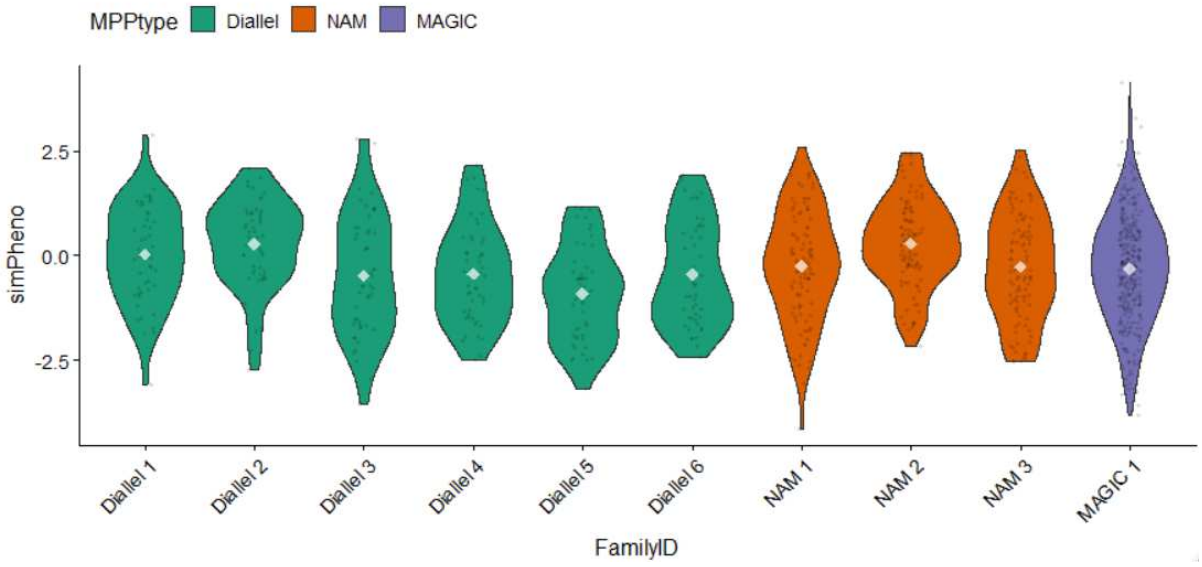


Figure S1 Example of simulated phenotypic distributions within families of different MPP designs.

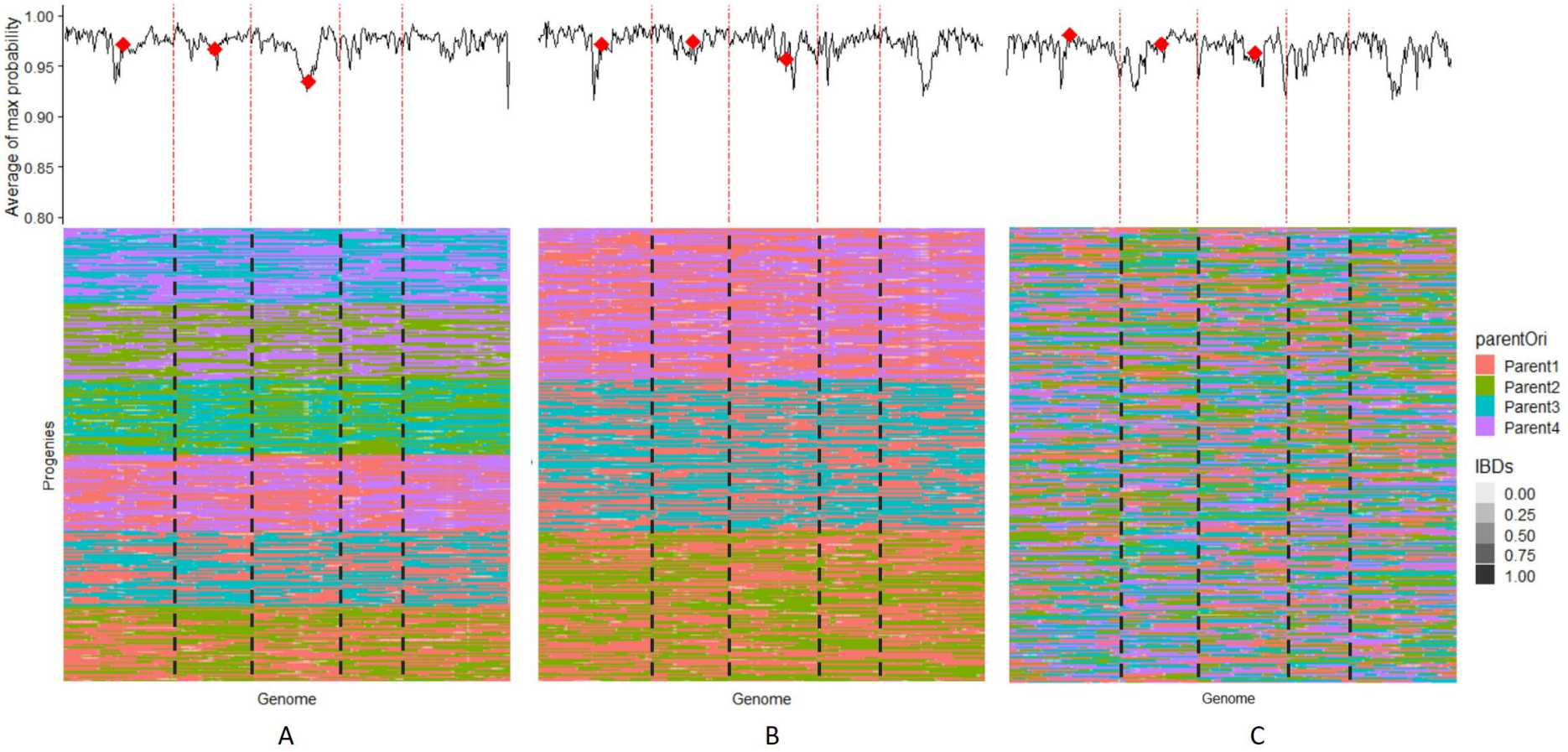
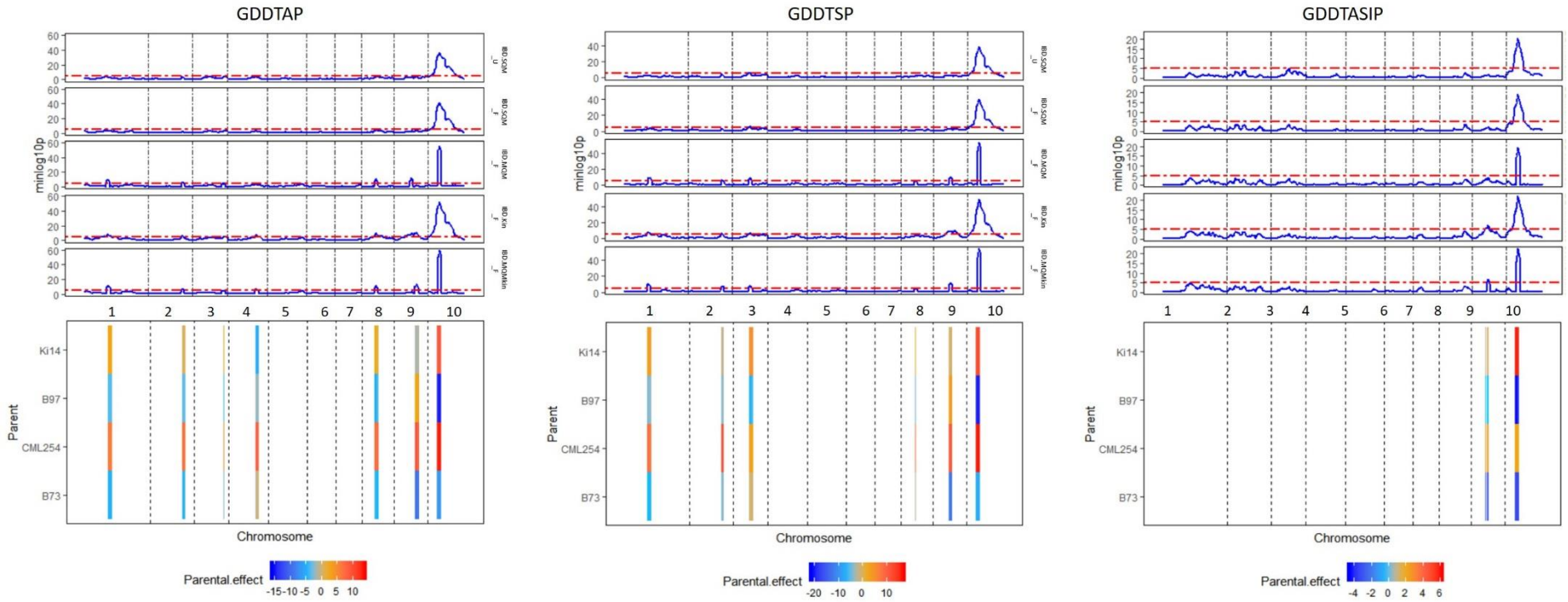


Figure S2 ***Upper panel*** *Average maximum IBD probabilities across the genome, where diamonds in red indicate positions of simulated major QTLs.* ***Bottom panel*** *Visualization of progeny genome construction referring to parent origins in IBDs for simulated* ***A****. diallel* ***B****. NAM, and* ***C****. MAGIC designs.*


GDDTAP
GDDTSP
GDDTASIP
minlog10p
Parent
Ki14
B97
CML254
B73
Chromosome
Parental.effect

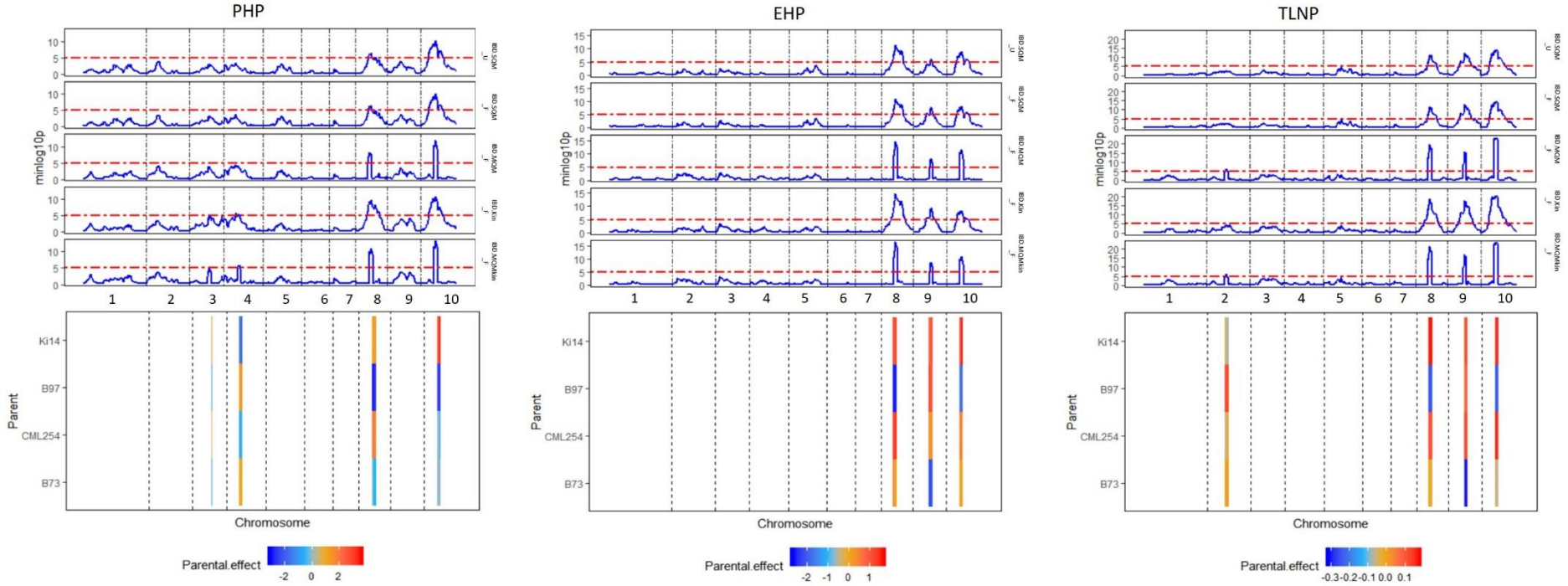


Figure S3 Mapping results for the empirical maize diallel design using the five IBD-based mixed models. ***Upper panel*** *QTL profiles from the five IBD-based mixed model approaches.* ***Bottom panel*** *Estimation of parental effects at QTLs detected by a model selected on the basis of BIC.*

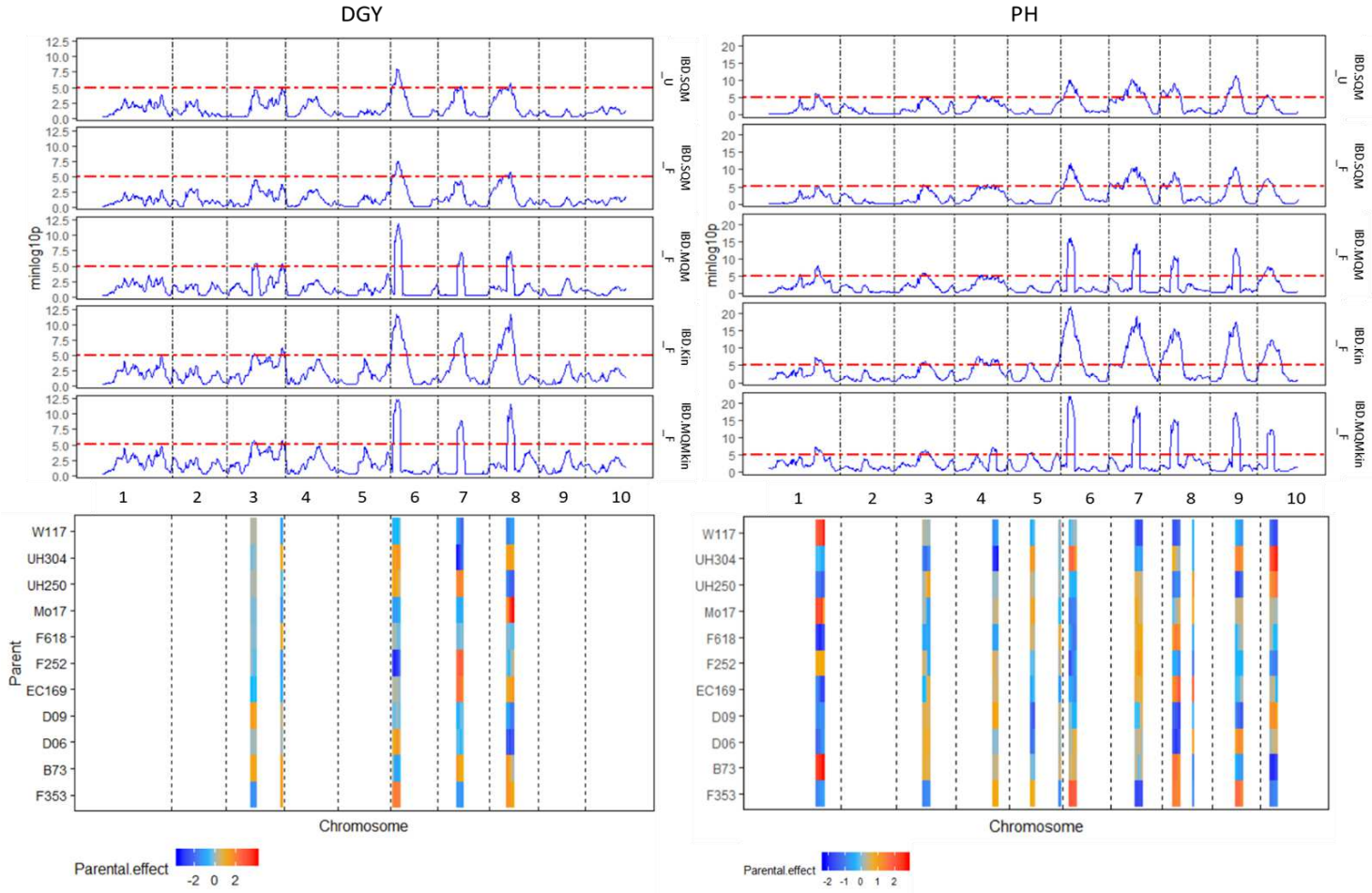


Figure S4 Mapping results for the empirical maize NAM using the five IBD-based mixed models. ***Upper panel*** *QTL profiles from the five IBD-based mixed model approaches.* ***Bottom panel*** *Estimation of parental effects at QTLs detected by the model with the smallest BIC among the five models.*

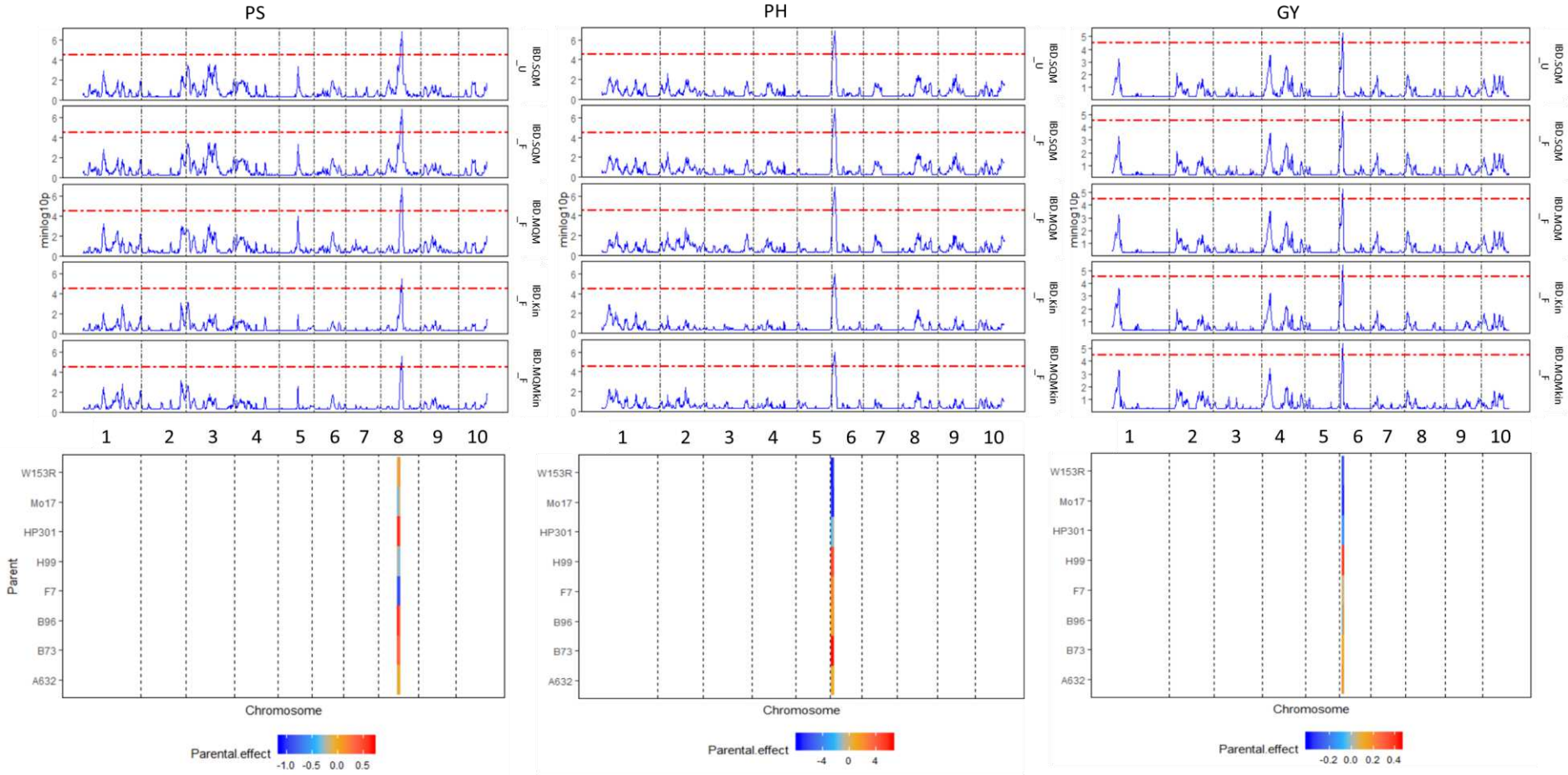


Figure S5 Mapping results for the empirical maize MAGIC using the five IBD-based mixed models. ***Upper panel*** *QTL profiles from the five IBD-based mixed model approaches.* ***Bottom panel*** *Estimation of parental effects at QTLs detected by the model with the smallest BIC among the five models.*

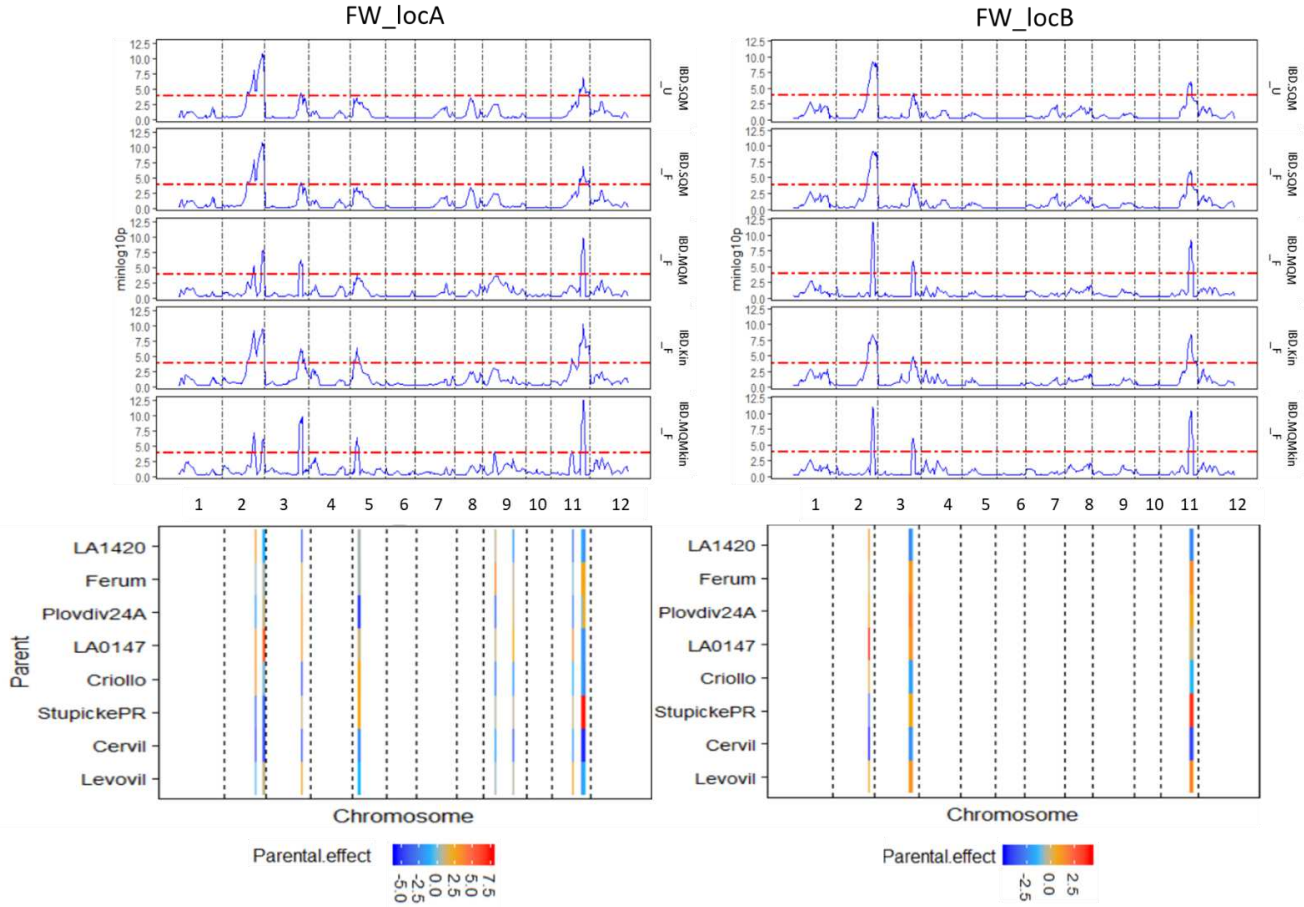


Figure S6 Mapping results for the empirical tomato MAGIC design using the five IBD-based mixed models. ***Upper panel*** *QTL profiles from the five IBD-based mixed model approaches.* ***Bottom panel*** *Estimation of parental effects at QTLs detected by the model with the smallest BIC among the five models.*

Table S1 Summary of QTL mapping results. The number of QTLs and BIC apply to the final multi-QTL model fitted on the basis of genome scans with mentioned IBD-based mixed models.

Trait	IBD-based mixed models	Number of identified QTLs	BIC
	Maize diallel		
GDDTAP	*IBD.SQM_U*	3	4522
	IBD.SQM_F	3	4479
	IBD.MQM_F	6	4416
	IBD.Kin_F	7	4404
	IBD.MQMkin_F	7	4401
GDDTSP	*IBD.SQM_U*	2	4846
	IBD.SQM_F	3	4803
	IBD.MQM_F	6	4707
	IBD.Kin_F	6	4689
	IBD.MQMkin_F	6	4695
GDDTASIP	*IBD.SQM_U*	1	3937
	IBD.SQM_F	1	3928
	IBD.MQM_F	1	3928
	IBD.Kin_F	2	3881
	IBD.MQMkin_F	2	3881
PHP	*IBD.SQM_U*	2	3388
	IBD.SQM_F	2	3398
	IBD.MQM_F	2	3397
	IBD.Kin_F	5	3356
	IBD.MQMkin_F	4	3354
EHP	*IBD.SQM_U*	3	2835
	IBD.SQM_F	3	2852
	IBD.MQM_F	3	2852
	IBD.Kin_F	3	2842
	IBD.MQMkin_F	3	2842
TLNP	*IBD.SQM_U*	3	224
	IBD.SQM_F	3	243
	IBD.MQM_F	4	225
	IBD.Kin_F	3	197
	IBD.MQMkin_F	4	201
	Maize NAM		
DGY	*IBD.SQM_U*	3	4791
	IBD.SQM_F	2	4842
	IBD.MQM_F	5	4819

	IBD.Kin_F	5	4698
	IBD.MQMkin_F	5	4698
	IBD.SQM_U	8	4289
	IBD.SQM_F	8	4314
PH	*IBD.MQM_F*	9	4312
	IBD.Kin_F	10	4219
	IBD.MQMkin_F	11	4225
	Maize MAGIC		
	IBD.SQM_U	1	669
	IBD.SQM_F	1	669
PS	*IBD.MQM_F*	1	669
	IBD.Kin_F	1	672
	IBD.MQMkin_F	1	672
	IBD.SQM_U	1	663
	IBD.SQM_F	1	663
PH	*IBD.MQM_F*	1	663
	IBD.Kin_F	1	667
	IBD.MQMkin_F	1	667
	IBD.SQM_U	1	648
	IBD.SQM_F	1	648
GY	*IBD.MQM_F*	1	648
	IBD.Kin_F	1	647
	IBD.MQMkin_F	1	647
	Tomato diallel		
	IBD.SQM_U	1	1429
	IBD.SQM_F	3	1429
Fruit shape	*IBD.MQM_F*	2	1433
	IBD.Kin_F	3	1433
	IBD.MQMkin_F	3	1431
	Tomato NAM		
	IBD.SQM_U	2	145
	IBD.SQM_F	2	165
Disease score	*IBD.MQM_F*	2	165
	IBD.Kin_F	2	165
	IBD.MQMkin_F	2	165
	Tomato MAGIC		
	IBD.SQM_U	3	2719
FT_locA	*IBD.SQM_F*	3	2719
	IBD.MQM_F	4	2703

	IBD.Kin_F	5	2667
	IBD.MQMkin_F	8	2649
	IBD.SQM_U	3	2346
	IBD.SQM_F	3	2346
FT_locB	*IBD.MQM_F*	3	2346
	IBD.Kin_F	3	2327
	IBD.MQMkin_F	3	2327

851

Figures

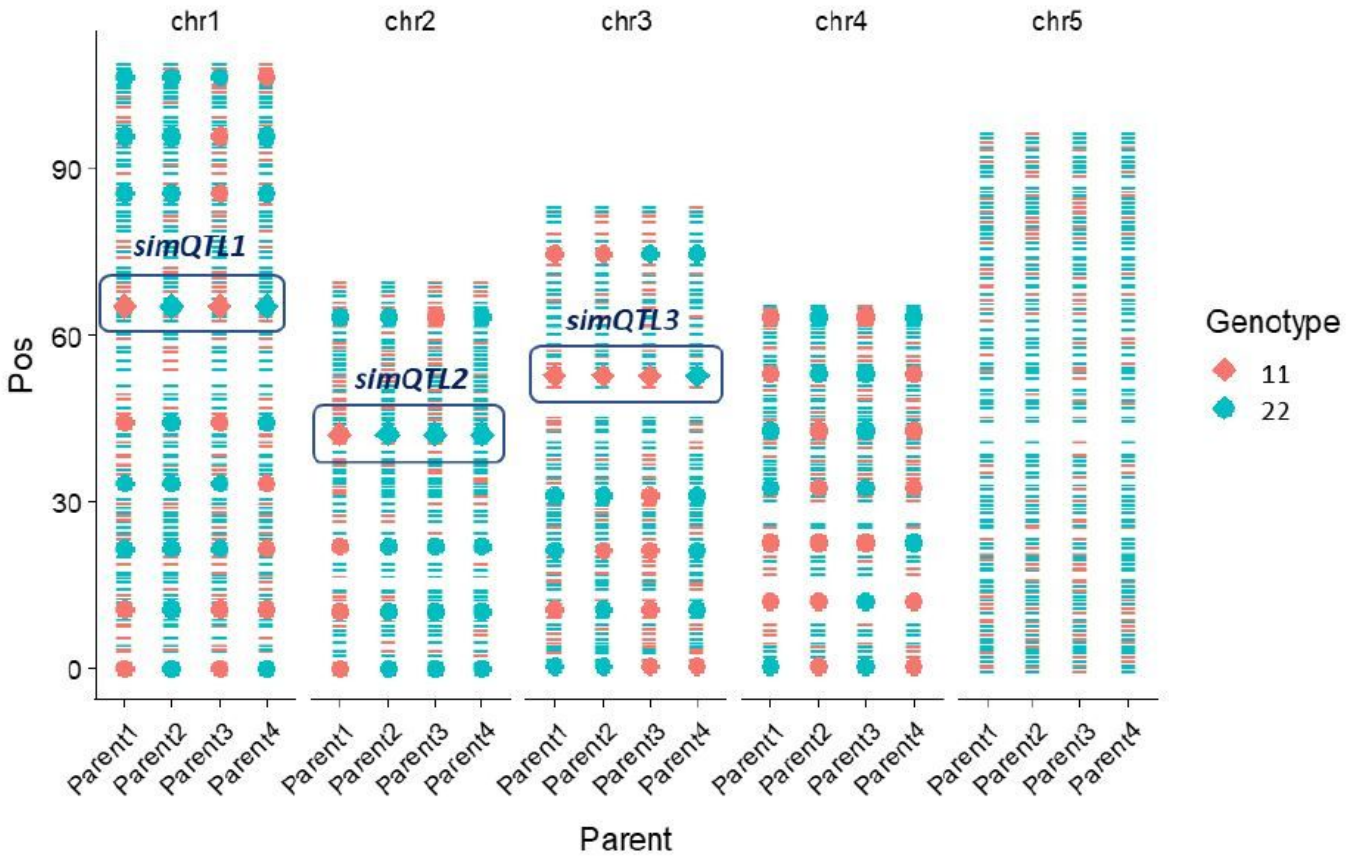


Figure 1

Simulated positions and genotypes for three major QTLs (diamonds) with an additive allelic substitution effect of 0.4 and the allele labelled as 1 increasing the trait. For 24 minor QTLs (dots) the additive allelic substitution effect was 0.1 with the allele labelled as 1 again increasing the trait.

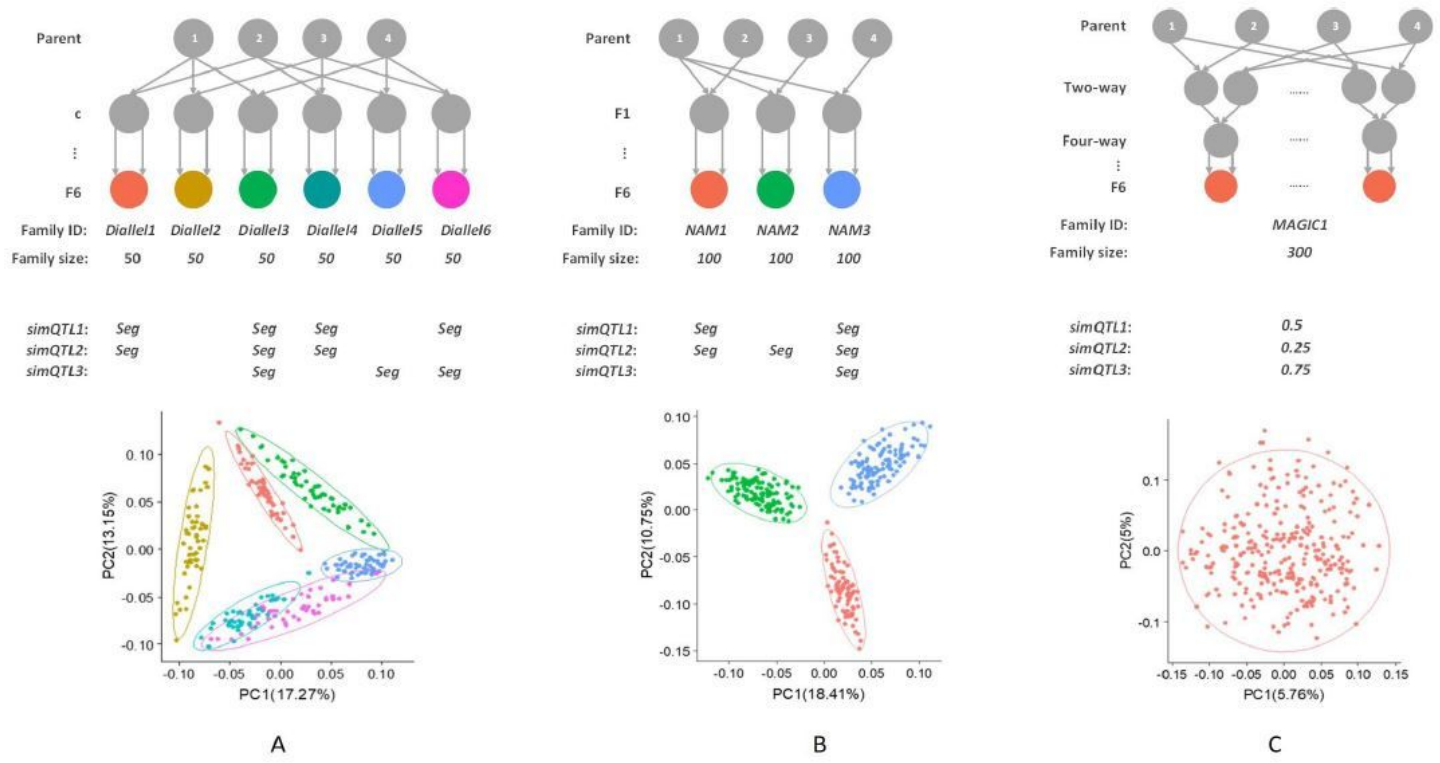


Figure 2

Upper panel the crossing schemes of simulated MPP designs. Middle panel Families with segregating QTLs or allele frequencies at those QTLs. Bottom panel PCA plots for progenies based on simulated genome data for A.diallel, B. NAM, and C. MAGIC designs.

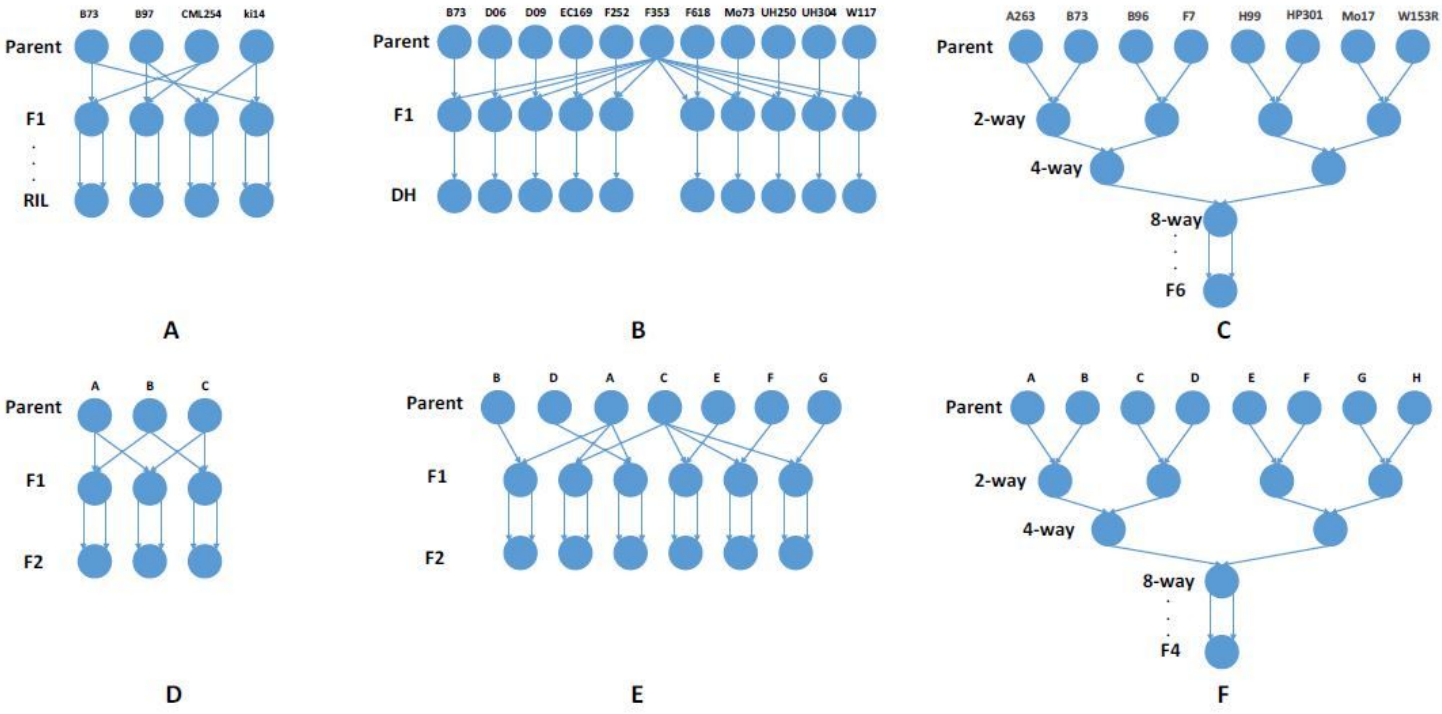


Figure 3

The crossing schemes of empirical MPP designs: A. maize diallel RIL design B. maize NAM DH design C. maize eight-way MAGIC F6 design D. tomato diallel F2 design E. tomato connected NAM F2 design, and F. tomato eight-way MAGIC F4 design.

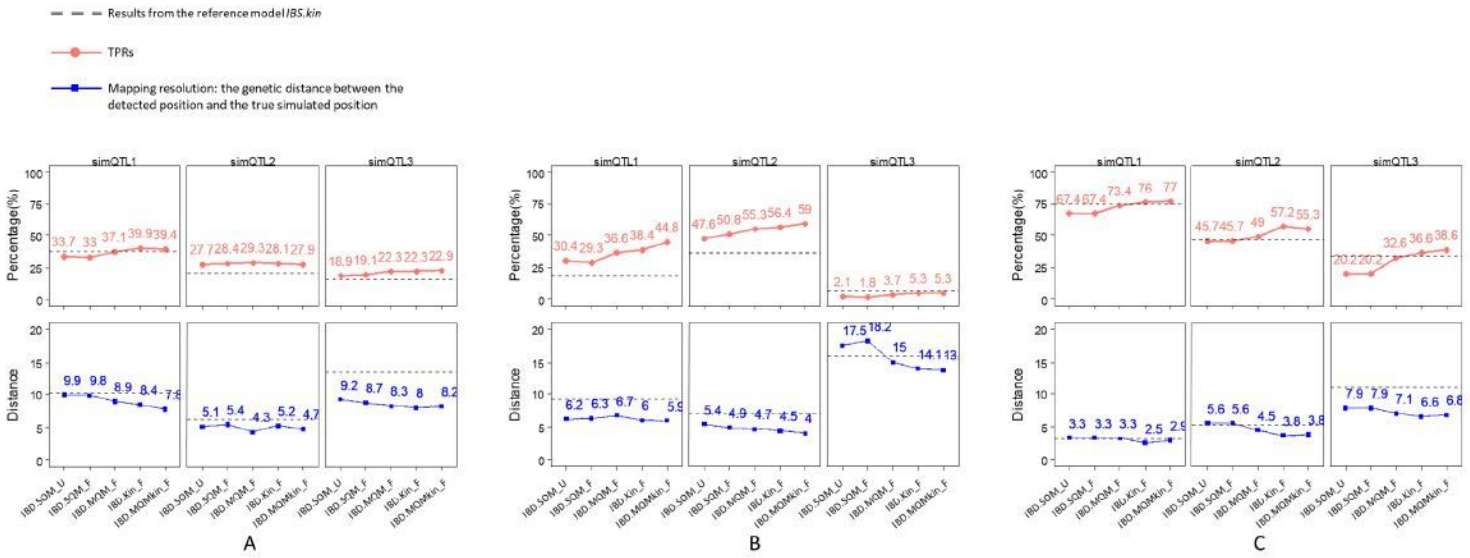


Figure 4

Model comparison in terms of True Positive Rates (dots with red lines) in percentages and mapping resolutions (squares with blue lines) in cM for simulated QTLs in A. diallel B.NAM, and C. MAGIC designs.

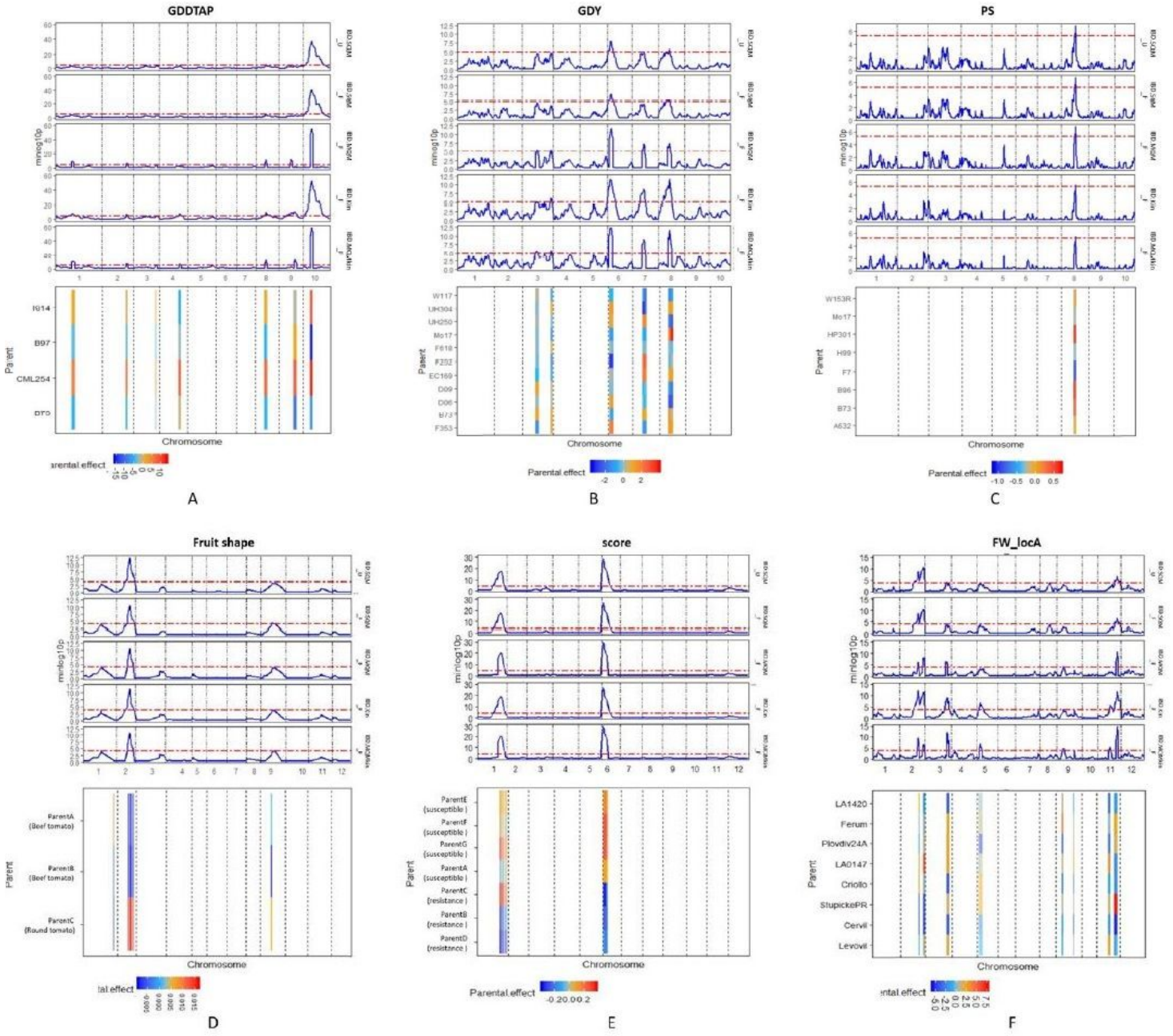


Figure 5

Mapping results of some traits in the empirical MPP designs: A. maize diallel, B. maize NAM, C. maize MAGIC, D. tomato diallel, E. tomato NAM and F. tomato MAGI using the five IBD-based mixed models. Upper panel QTL profiles from the five IBD-based mixed model approaches. Bottom panel Estimation of parental effects at QTLs detected by model selected based on BIC. Mapping results of other traits are provided in the Appendix.