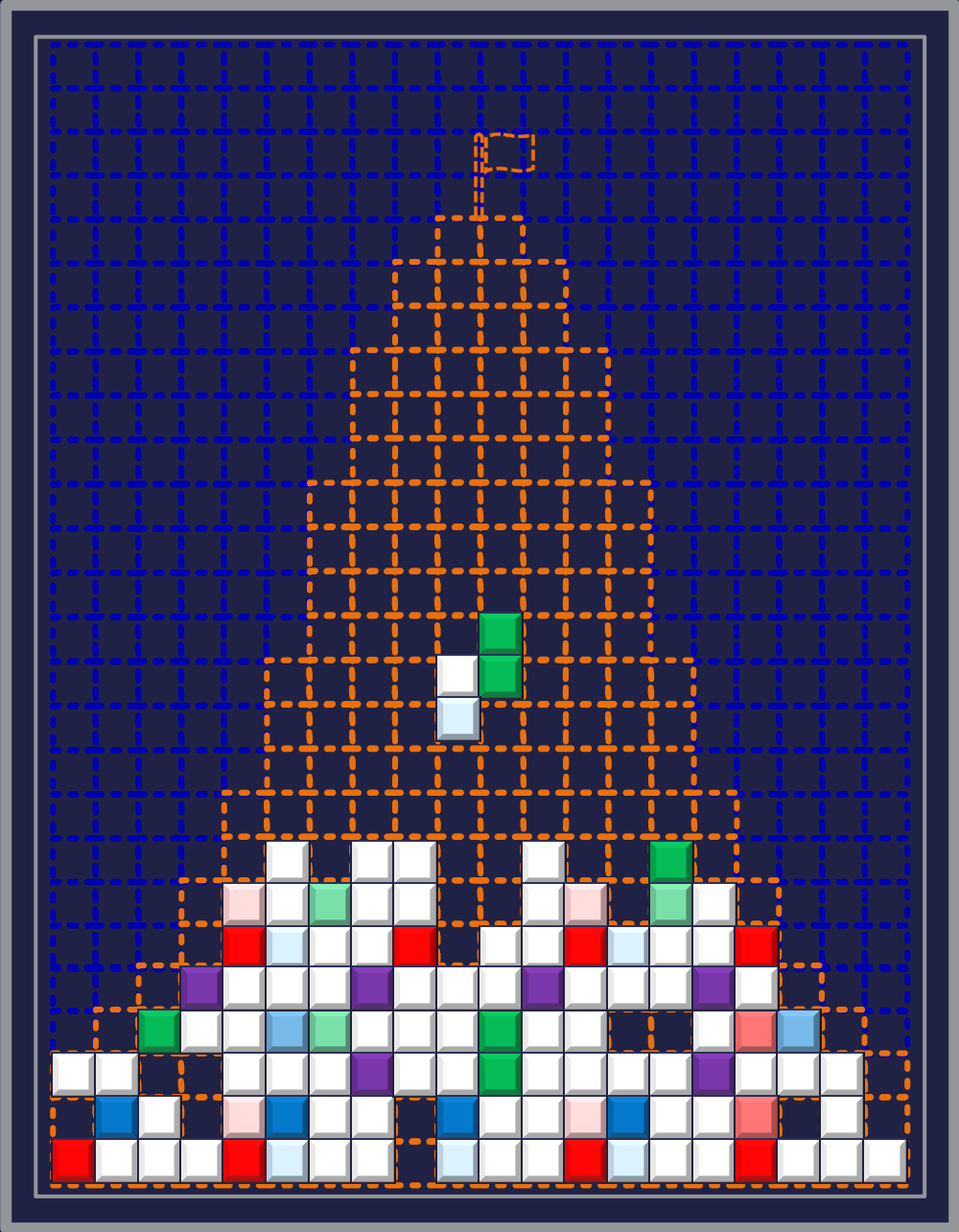


STATISTICAL MODELS AND TOOLS FOR QTL MAPPING IN MULTI-PARENT POPULATIONS



Wenhao Li



Propositions

1. A multi-parent population is a novel tool to answer traditional questions.
(this thesis)
2. QTL mapping and genomic prediction complement rather than compete with each other.
(this thesis)
3. It is dangerous to easily accept conclusions from favourably looking statistical analyses.
4. Good science combines theoretical innovation with practical utility.
5. For testing your own understanding, teaching others works better than examining yourself.
6. Taking a power nap in the middle of the day should be promoted in the Dutch working environment.

Propositions belonging to the thesis, entitled

Statistical Models and Tools for QTL Mapping in Multi-parent Populations

Wenhao Li

Wageningen, 28th February 2023

Statistical Models and Tools for QTL Mapping in Multi-parent Populations

Wenhao Li

Thesis Committee

Promotor

Prof. Dr F.A. van Eeuwijk
Professor of Applied Statistics
Wageningen University & Research

Co-promotors

Dr M.P. Boer
Researcher, Applied Statistics
Wageningen University & Research

Dr R.V.L. Joosen
Researcher, Quantitative Genetics
Rijk Zwaan Breeding B.V.

Other members

Dr D. Borchardt, KWS group, Germany
Dr M. Dell'Acqua, Scuola Superiore Sant'Anna, Italy
Dr C.A. Maliepaard, Wageningen University & Research
Prof. Dr B.J. Zwaan, Wageningen University & Research

This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PE&RC)

Statistical Models and Tools for QTL Mapping in Multi-parent Populations

Wenhao Li

Thesis

Submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday 28th February 2023

at 4 p.m. in the Omnia Auditorium

Wenhao Li

Statistical Models and Tools for QTL Mapping in Multi-parent
Populations 165 pages

PhD thesis, Wageningen University, Wageningen, The Netherlands (2023)
With references, with summary in English

ISBN: 978-94-6447-583-8

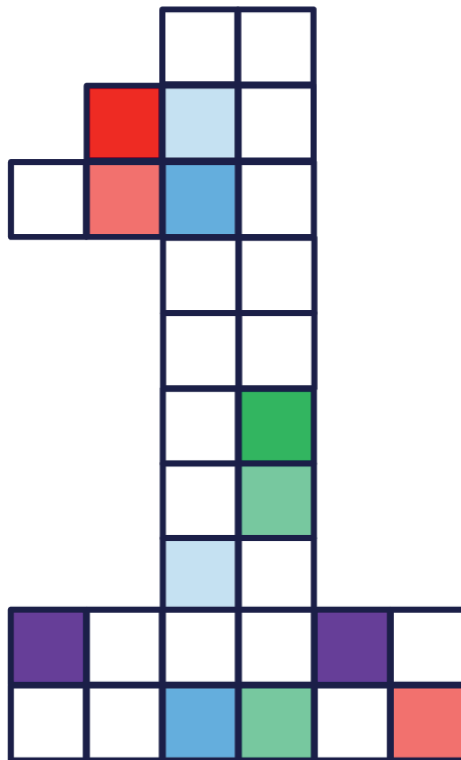
DOI: <https://doi.org/10.18174/586884>

Table of contents

Chapter 1	
General introduction.....	1
Chapter 2	
An IBD-based mixed model approach for QTL mapping in multi-parent populations.....	15
Chapter 3	
statgenMPP: an R package implementing a generic IBD-based mixed model approach for QTL mapping in a wide range of multi-parent populations.....	51
Chapter 4	
A one-dimensional mixed model genome scan approach for detecting QTL-by-genetic-background interactions in diallel and nested association mapping designs.....	61
Chapter 5	
An IBD-based mixed model approach to detection of QTL-by-Environment interactions with multi-parent populations.....	95
Chapter 6	
General discussion.....	125
Bibliography.....	139
Summary.....	157
Acknowledgement.....	159
About the author.....	163
PE&RC training and education statement.....	165

Chapter 1

General introduction



1. Background of QTL mapping

1.1. Definition and applications of QTL mapping

A quantitative trait locus (QTL) is a genome fragment or region that contains genes underlying quantitative traits (Geldermann 1975; Abiola et al. 2003; Collard et al. 2005). The definition of QTLs remained a theoretical concept until the occurrence of DNA or molecular markers. Two seminal papers about QTL mapping are Beckmann and Soller (1983) and Lander and Botstein (1989). In QTL mapping, genome fragments are characterized by molecular markers that tag genes along the genome. QTL mapping or analysis is the attempt to discover the co-segregation marker and trait variation by testing for association between marker and trait value (Kearsey 1998).

QTL analysis helps to discover the genetic basis underlying phenotypic variation and provides insights into the physiological and biological mechanisms producing such variation (Nadeau and Frankel 2000; Korstanje and Paigen 2002; Abiola et al. 2003; Collard et al. 2005). From the perspective of plant breeding, the selection of superior lines relying on visible phenotypes shifts to the investigation of QTL-linked molecular markers associated with phenotypic variation (Collard and Mackill 2008), which activity is referred to as molecular plant breeding (Moose and Mumm 2008). Thereafter, the creation, selection, and fixation of desirable genotypes in the breeding program are greatly facilitated by using molecular markers, which is known under the name of marker-assisted breeding/selection (MAB/MAS) (Ribaut and Hoisington 1998; Dekkers and Hospital 2002).

1.2. Rationale and basis of QTL mapping

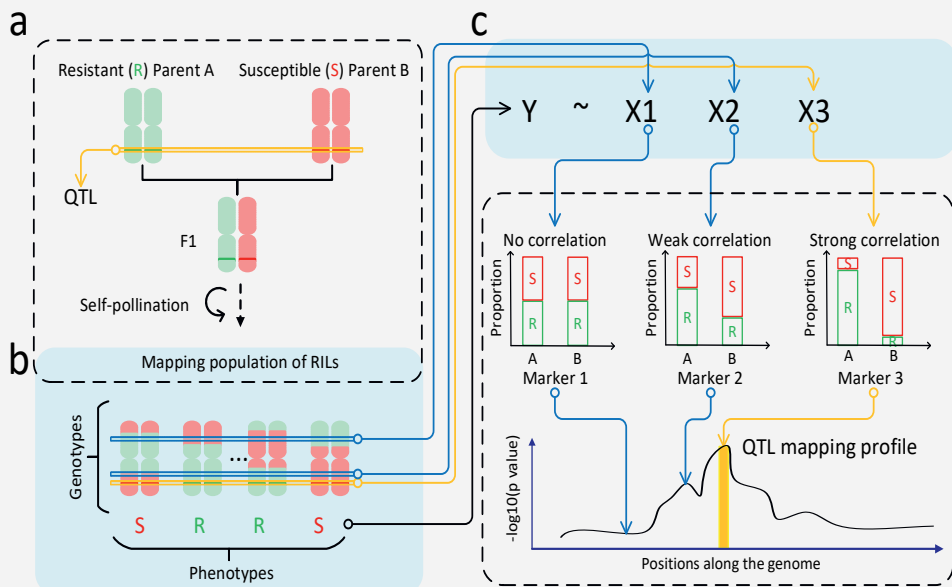
The principle of QTL mapping is conceptually simple: estimate the correlation or association between genetic variants and phenotypic variation (Leal 2001). To reach such a point, a mapping or segregating population is required as a platform to obtain phenotypic and genotypic data, and a suitable statistical method is required to estimate the correlation or association. We use a simple case (see Box 1) as a starting point to demonstrate the main steps of QTL mapping before touching on complex scenarios.

For the case of quantitative traits, the general linear model to describe the relationship between phenotype and a single marker genotype, is $Y = \mu + Xa + \varepsilon$, for response Y , with mean μ , a design vector, or genetic predictor, X , containing the expected number of allele copies from the reference parent, the allelic substitution effect a , and a normally distributed error (Kearsey 1998). This model is fitted along the genome to identify genomic locations, QTLs, for which the association between the observed response and the genetic predictor is significant. The term QTL-heritability is often used to describe the percentage of phenotypic variation explained by the genetic effect. From the perspective of statistics, the goodness of fit or coefficient of determination describes how well the observations (phenotypes) are

described by the genetic predictor. The single QTL model above can be extended to become a multi-QTL model.

Box 1 QTL mapping for a simple case using a bi-parental population (BPP) to map a single resistance gene.

- Create a mapping population of recombinant inbred lines (RILs) by crossing two inbred parent lines (A and B) with contrasting phenotypes, resistance (R) vs. susceptibility (S), followed by multiple rounds of self-pollinations starting at the F1.
- Obtain phenotypic and genotypic information. Genome origins from parent A and B are indicated by the colours red and green; R and S denote resistant and susceptible lines.
- Assess the correlation between trait and marker variation to identify QTLs responsible for resistance. The phenotype determines the response variable (Y) which is regressed on the marker genotype as explanatory variables (X). The $-\log_{10}(p)$ -values for the marker-trait association are plotted along the genome and the regions that exceed a predefined threshold identify potential QTLs. For example, marker 3 has a large signal that shows evidence for a QTL.



1.3. Challenges and opportunities for QTL mapping

Phenotypes like resistances with an essentially binary response can have a simple monogenic basis. Most traits, like yield and plant height, are regulated by many genes and corresponding QTLs can have small effects. Furthermore, gene actions may be non-additive and depend on the environment. The model introduced above will then not be able to identify all QTLs. Second, traditional BPPs contain very limited genetic diversity and this may again impede the study of the genetic basis of complex traits. Lastly, QTL mapping usually estimates QTL effects at the bi-allelic level in a specific BPP background, but breeders may have more interest in a wider range of alleles that are tested against a diverse breeding background.

The use of multi-parent populations (MPPs) provides opportunities to address the challenges occurring for BPPs, which will be addressed in the section “Mapping populations”. Methodologies for QTL mapping in MPPs should be able to define QTL effects for the alleles corresponding to the included parents. As a consequence, the genetic predictors, X , should reflect the multiple alleles at each position along the genome (section “Genetic predictors”). Further, addressing the genetic basis of complex traits, QTL methodology for MPPs should include modelling opportunities for non-additive effects (section “Phenotypes and genetic architecture”). Lastly, the methodology should be applicable across a wide range of MPPs (section “Statistical models and tools”).

2. Mapping populations

2.1. Bi-parental populations (BPPs)

Experimental mapping populations are initially and routinely derived from crossing two inbred parents. The desirable and complementary phenotypes in the two parents are assumed to depend on genetic variants whose locations we aim to identify as QTLs. Analysing the offspring of a cross between two parents in QTL analysis of a BPP has the advantages of balancing allele frequencies for rare genetic variants and increasing the mapping power of QTLs (Jansen 2004). The use of BPPs is common in molecular breeding for introgressing and pyramiding desirable genes (Moose and Mumm 2008). However, analysis of a single BPP by QTL mapping may provide insufficient insight into the genetic basis underlying a target trait. The narrow genetic diversity captured by only two parents may lead to a limited number of identified QTLs and the estimated effects of QTLs in a specific BPP will be poorly transferrable to other genetic backgrounds (Melchinger et al. 1998; Liu and Zeng 2000; Crepieux et al. 2005; Pascual et al. 2016).

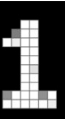
2.2. Diversity panels

In comparison to BPPs, diversity panels represent a wider genetic variability. For BPPs, linkage analysis is the standard for QTL mapping, whereas linkage disequilibrium, i.e., the association between genotype and phenotype forms the basis for QTL mapping in diversity panels, like in genome-wide association studies (GWAS). The inclusion of large numbers of unrelated and lowly related genotypes increases genetic diversity together with a large number of accumulated historical recombinations in diversity panels is beneficial to the resolution of QTL mapping in diversity panels (Korte and Farlow 2013; Visscher et al. 2017). However, diversity panels can exhibit low minor allele frequencies, and the panel contains population structure, with subpopulations differing in allele frequencies. Population structure can act as a confounding factor in association studies when allele frequency differences between subpopulations coincide with phenotypic differences. Therefore, in association mapping corrections for population structure are required (Yu et al. 2006; Malosetti et al. 2007; Price et al. 2010; Visscher et al. 2012; Xiao et al. 2016).

2.3. Multi-parent populations

The use of multi-parent populations, or MPPs, for QTL studies is the central topic of this thesis. MPPs, as an intermediate between BPPs and diversity panels, increase genetic diversity by using multiple parents and balancing population structure through experimental crossing schemes (Scott et al. 2020; Arrones et al. 2020). Genetic studies using MPPs have recently been applied to a wide range of crops and vegetables, such as rice (Kitony et al. 2021; Zheng et al. 2022; Liang et al. 2022), maize (Swarts et al. 2021; Odell et al. 2022; Michel et al. 2022), potato (Amadeu et al. 2021), wheat (Rollar et al. 2021a; Hu et al. 2022), barley (Dang et al. 2020; Hautsalo et al. 2021; Grieco et al. 2022), cowpea (Ravelombola et al. 2021, 2022), chickpea (Thudi et al. 2014), tomato (Diouf et al. 2018; Campanelli et al. 2019), eggplant (Gramazio et al. 2019).

In this thesis, we classify the MPP designs into two categories (Figure 1). The first category contains all designs with connected families such as the Nested Association Mapping population (NAM) (Yu et al. 2008; Gage et al. 2020), where a central parent is crossed with multiple peripheral parents, and diallel designs (Coles et al. 2010; Turner et al. 2018; Seye et al. 2019) where parents are crossed with each other in a factorial structure. Other possible multi-cross populations can be the combination of diallel and NAM structures. The second category refers to single offspring populations with no apparent population structure obtained from crossing multiple parents. The best example of such crosses is the Multi-parent Advanced Generation Intercrosses (MAGIC) populations (Beyer et al. 2008; Huang et al. 2015; Arrones et al. 2020). Previous studies have shown the benefits and necessities of MPPs in plant breeding and genetic studies. For example, when the family size for each BPP is small, using connected families can increase population size and power for QTL mapping. In a NAM population, QTLs can be mapped in multiple connected families and their effects can be studied in their dependence on genetic background by evaluating QTL by family



effects. In MAGIC populations, QTLs can be mapped with high resolution and progeny individuals containing desirable alleles at multiple QTLs can be directly selected.

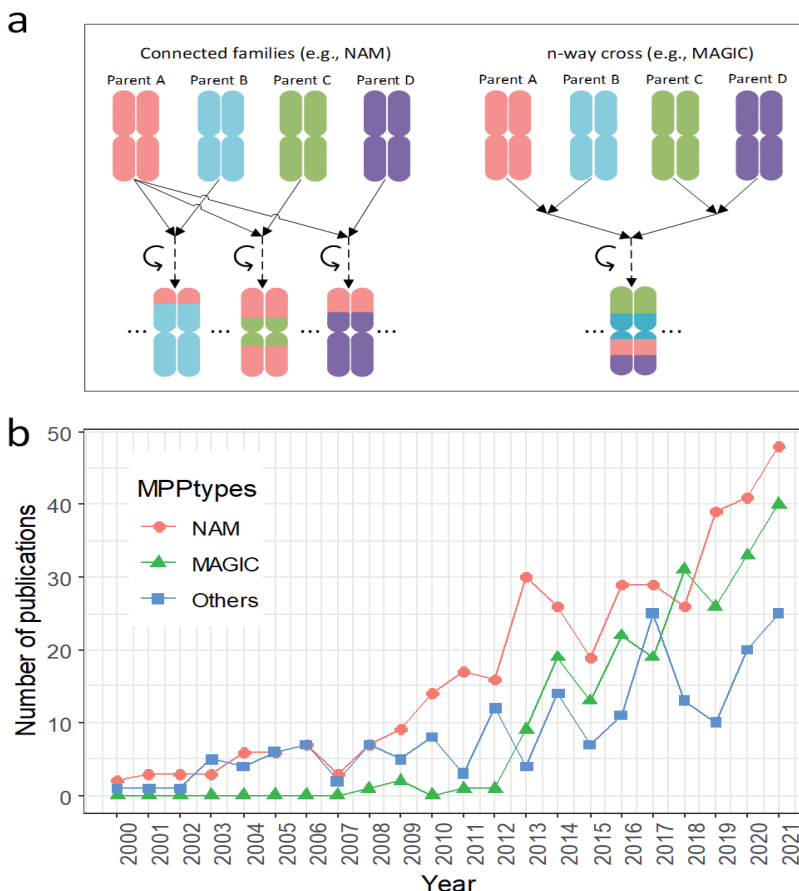


Figure 1. **a.** The crossing scheme of common MPP designs. **b.** Numbers of MPP publications regarding MPP types.

3. Genetic predictors

The genetic similarity between two individuals can be measured in terms of identity-by-state (IBS) and identity-by-descent (IBD). Identity in IBS means that the same sequence of DNA occurs in both individuals. Identity in IBD means that segments of DNA in the genome of both individuals are inherited from a common ancestor origin (Browning 2008; Thompson 2013; Sticca et al. 2021). Figure 2 illustrates how to calculate IBD probabilities between parents and offspring on a genome segment for a four-way cross. In multi-cross MPPs, like NAM and diallel populations, the genome of an offspring contains contributions from two parents, while in multi- or n-way cross MPP designs, the genome composition of



an offspring individual is a mosaic of multiple parental genomes. In BPPs, QTL mapping is commonly done by linear regression models with genetic predictors that contain the expected number of alleles of the reference parent, see section “Rationale and basis of QTL mapping”. For MPPs, this same principle can be used if we develop or have access to a method for calculating the IBD probabilities between parents and offspring (Broman et al. 2019). The calculation of IBDs in this thesis is based on so-called inheritance vectors from parents to offspring. For the calculation of these inheritance vectors, we need the crossing scheme of the parents, a joint linkage map, and genome (marker) information on parents and offspring. For IBD calculations for multi-parent populations, we used the framework developed by Zheng et al. (2014, 2015) and Boer and van Rossum (2021).

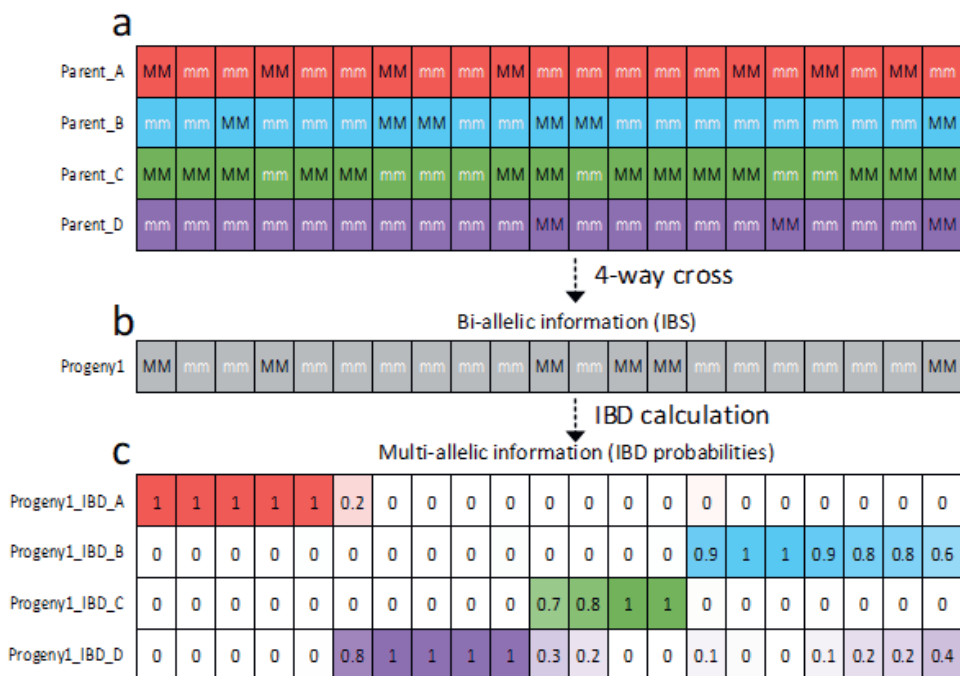
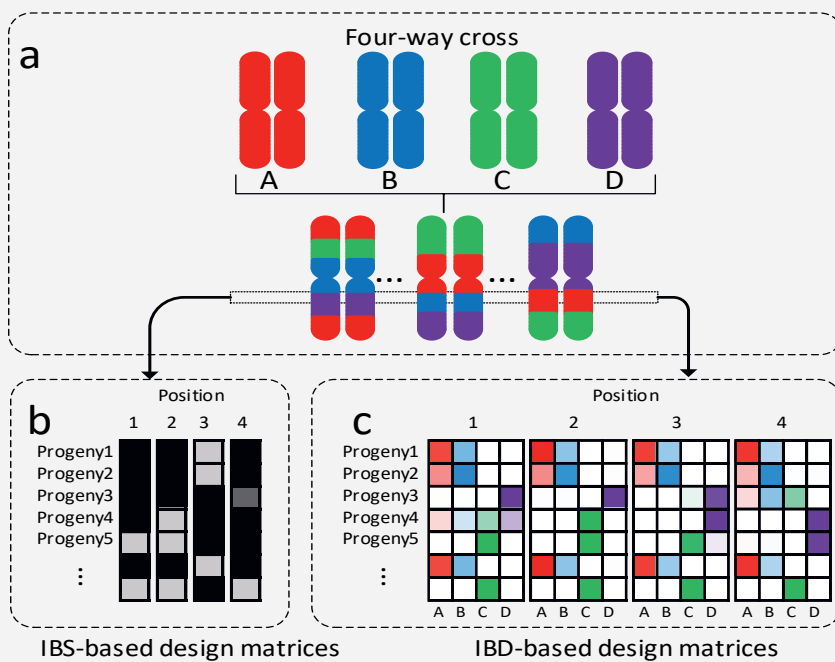


Figure 2. Comparison of IBS and IBD on a genomic segment for a RIL in a simulated four-way cross. **a.** Biallelic marker genotypes (mm and MM) of four parents expressed in IBS fashion. **b.** Biallelic marker genotypes in offspring in IBS fashion. **c.** IBD calculations: the probability of inheritance of offspring marker alleles from individual parents. The expected number of alleles received from a parent follows from the IBD calculations by multiplying the IBD information by 2.

For diversity panels, genetic predictors or design vectors and matrices are based on IBS. For BPPs, genetic predictors are mostly based on IBD information. As MPPs are an intermediate between BPPs and diversity panels, the genetic predictors in the design matrices for QTL mapping can use either IBS or IBD information. A key difference between statistical models for QTL mapping in MPPs developed in this thesis and previous methodologies is that we advocate using multi-allelic IBD information as genetic predictors. Box 2 demonstrates structures of design matrices based on IBS and IBD information in the case of a four-way cross. The IBS information, which for SNPs has usually just two states, will define a design vector or matrix in a linear mixed model without referring to the parental origins of the alleles. IBD information, which will automatically be multi-allelic for MPPs, leads to multi-column design matrices whose columns are labelled by the parents and whose entries are calculated on the basis of the probabilities that the alleles originate from a particular parent. In Chapter 2, we employ a general framework of the Hidden Markov Model for IBD calculation and develop a generic IBD-based approach for QTL mapping and demonstrate their advantages concerning QTL mapping power and resolution in comparison to an IBS-based GWAS approach in various types of simulated and empirical MPP designs.

Box 2 Design matrices based on IBS and IBD information in a four-way cross of RILs.

Using a piece of the chromosome containing 4 positions in a real four-way cross of RILs as an example (a), for IBS-based models (b), the design matrices along the genome simply reflect the marker phenotypes. When there are two observable allelic states, the entries of the design vectors or matrices contain only 0s and 2s. For IBD-based models (c), the design matrices contain the expected number of allele copies from a particular parent. The colours indicate the parental origins and the colour intensities indicate the number of alleles, i.e., twice the probability of IBD.



4. Phenotypes and genetic architecture

QTL mapping can be challenging when traits are the outcome of complex genetic architectures. A complex trait is regulated by a large number of QTLs, whose effects show intra- and inter-locus interactions as well as with the environment. (Holland 2007; Ehrenreich 2017). Interactions between QTLs can occur in various forms. Epistasis is the modification of a QTL effect in relation to other QTLs or genetic backgrounds. QTL-by-environment interactions (QEIs) occur when QTL effects differ across environments. For example, due to the epistatic effect between QTL1 and QTL2 (Figure 3a), the allelic substitution effect (the difference of effects between two alleles) of QTL1 depends on the alleles at QTL2. Due to QEI (Figure 3b), the allelic substitution effect of QTL1 is inconsistent

over multiple environments. Epistasis and QEI are two forms of non-additive behavior of a QTL, and ignoring them may lead to missing heritability (Manolio et al. 2009) and constrain marker-assisted selection applications (Gimelfarb and Lande 1994; Long et al. 2011).

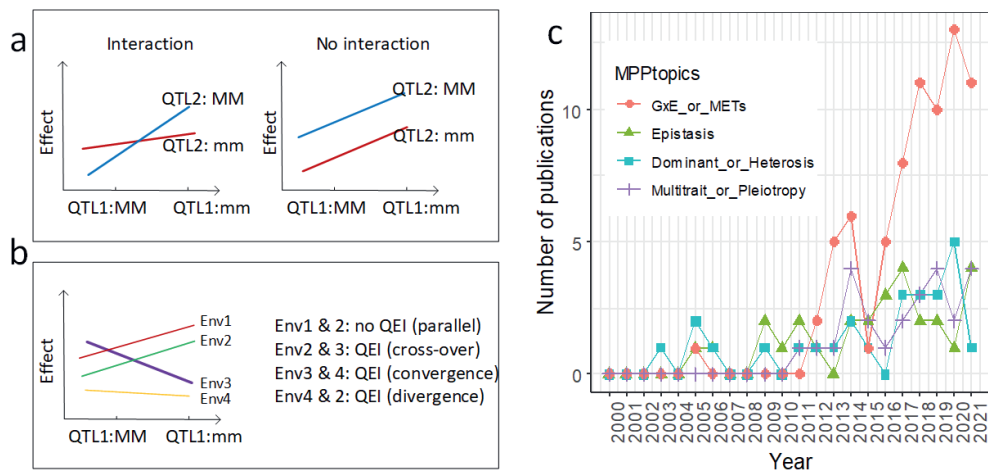


Figure 3. **a.** QTL1-by-QTL2 interaction versus absence of such an interaction. **b.** Illustration of QTL by environment interaction. **c.** Numbers of MPP publications referring to genetic architectures.

The deciphering of complex genetic architectures requires suitable mapping populations, collection of informative genotypes and phenotypes, and statistical methodology to model non-additive effects. Complex architectures have been studied in both experimental families and diversity panels (Korte and Farlow 2013; Xu et al. 2017). Additional attempts were made in the context of MPP designs (Figure 3c). “Diversity-based QTL mapping” is advocated to unravel complex genetic architectures (Holland 2007). The wide germplasm in MPPs increases the chance to combine genetic variations with non-additive effects (Churchill et al. 2004; McMullen et al. 2009) and multiple connected families offer the chance to evaluate genetic effects in different genetic backgrounds (Jannink and Jansen 2001; Blanc et al. 2006; Han et al. 2016). In this thesis, we develop appropriate statistical tools to explore the potential of using diverse MPPs to study epistasis (Chapter 4) and QEI (Chapter 5).

5. Statistical models and tools

For QTL mapping, experimental mapping populations and diversity panels are often treated differently (Mackay and Powell 2007; Mitchell-Olds 2010; Xu et al. 2017). For mapping populations, family-based or linkage mapping approaches are the default. For diversity panels, we can use population-based, linkage disequilibrium (LD), or association mapping approaches. In the context of MPP designs, both family-based and population-based approaches using respectively IBD and IBS information can be adopted (Table 1). We postulate that the approach integrating techniques of both family-based and population-based models and using IBDs as genetic predictors to estimate random-QTL effects is robust

in various scenarios (Chapter 2). To the set of tools shown in Table 1, we add in Chapter 3 the R package *statgenMPP* that integrates IBD calculation for genetic predictors by a Hidden Markov Model with mixed model approaches for QTL mapping for a wide range of MPP designs.

Table 1 Examples of statistical models for QTL mapping in MPP designs

Tool/software	Reference	Genetic predictor	Application in MPP designs
R/qtl2	(Broman et al. 2019)	IBS IBD	MAGIC (Abdulcina et al. 2019; Riaz et al. 2020)
mppR	(Garin et al. 2017)	IBS IBD	NAM (Garin et al. 2017, 2020)
mpQTL	(Thérèse Navarro et al. 2022)	IBS IBD	NAM (polyploid species) (Thérèse Navarro et al. 2022)
TASSEL	(Bradbury et al. 2007)	IBS	NAM (Tao et al. 2020; Olatoye et al. 2020; Swarts et al. 2021); MAGIC (Rida et al. 2021; López-Malvar et al. 2022)
PlabMQTL	(Utz et al. 2014)	IBS	Diallel (Galiano-Carneiro et al. 2021; Zhu et al. 2021b, a)
(MV)MPWGAI M	(Verbyla et al. 2014a, b)	IBD	MAGIC (Verbyla et al. 2014a, b)
WGNAM	(Paccapelo et al. 2022)	IBD	NAM (Paccapelo et al. 2022)
GAPIT	(Lipka et al. 2012)	IBS	MAGIC (Islam et al. 2016; Abdelraheem et al. 2021); NAM (Altendorf et al. 2021; Sandhu et al. 2021)
HAPPY	(Mott et al. 2000)	IBD	MAGIC (Kover et al. 2009)
mpMap	(Huang and George 2011)	IBD	MAGIC (Rollar et al. 2021b; Burgos et al. 2021)
GAPL	(Zhang et al. 2019)	IBD	MAGIC (Diaz et al. 2021)
IciMapping	(Meng et al. 2015)	IBS	NAM (Zhao et al. 2022; Hu et al. 2022)
FarmCPU	(Liu et al. 2016)	IBD	MAGIC (Satturu et al. 2020; Hautsalo et al. 2021)
BLINK	(Huang et al. 2019)	IBS	MAGIC (Hautsalo et al. 2021; Ravelombola et al. 2022)
MagicQTL	(Wei and Xu 2016)	IBD	MAGIC (Liang et al. 2022)
QTLRel	(Cheng et al. 2011)	IBS	MAGIC (Dell'Acqua et al. 2015)

6. Objectives and outline of the thesis

The overall objective of this thesis is to develop methodologies for QTL mapping in MPP designs (Figure 4). The specific objective, methodology, and used data sets for each chapter are detailed below:

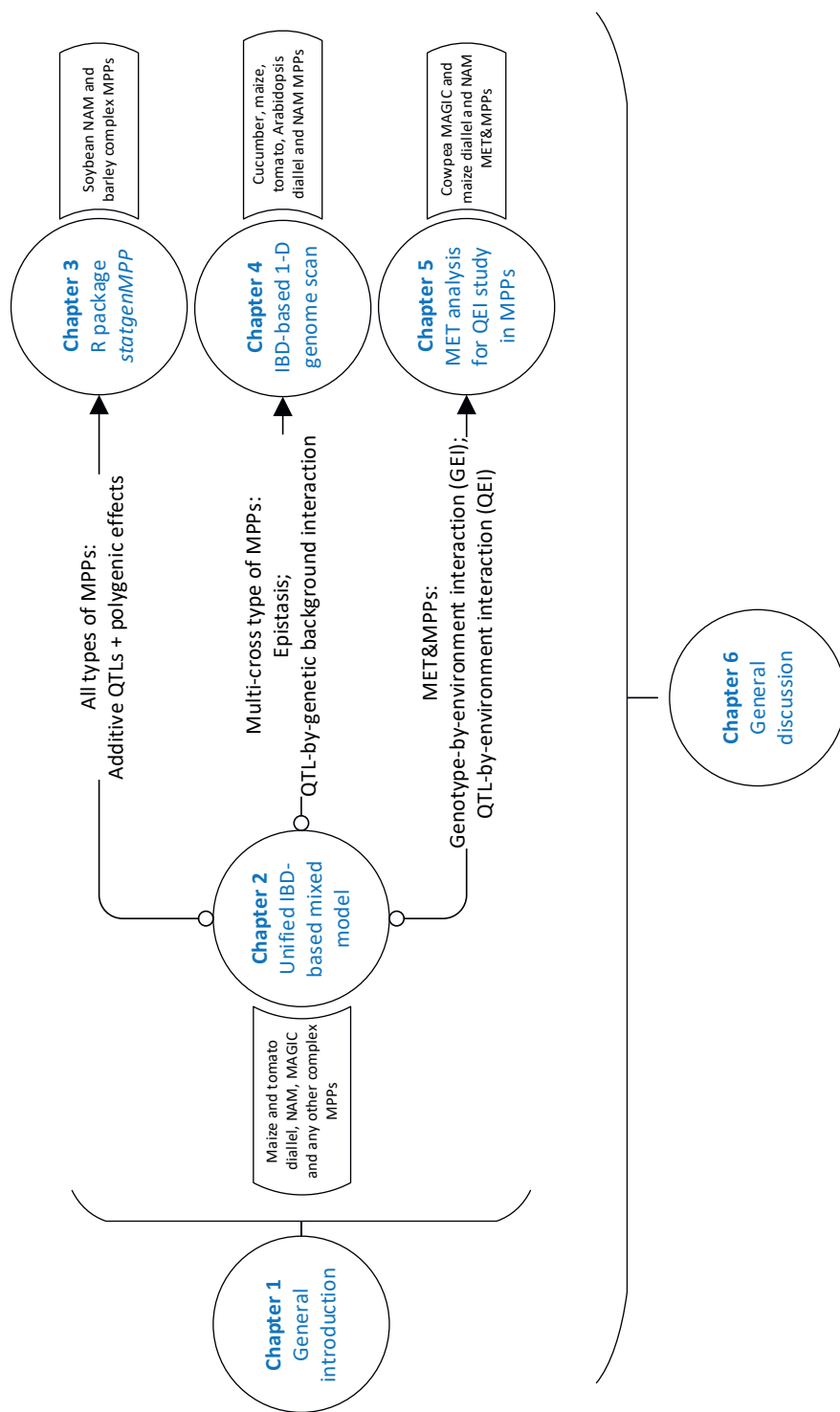


Figure 4. The outline of addressed topics, developed methodologies, and used MPP datasets in each chapter of the thesis.

Chapter 2: We aim to develop IBD-based mixed model approaches applicable to a wide range of MPPs including diallel, NAM, MAGIC, and other complex MPPs. IBD probabilities are calculated using a general framework of the Hidden Markov Model (Zheng et al. 2014, 2015). Functions of IBD probabilities to indicate the expected number of allele copies at parent origins are used as genetic predictors at individual loci to estimate variance components for multi-allelic genetic effects possibly associated with parental contributions in mixed model approaches while correcting for individual relatedness and family structures by marker-based kinship relations and family-specific residual terms. This approach can increase mapping power, mapping resolution, and QTL numbers in comparison to the benchmark IBS-based GWAS approach or other existing software tools simulated and empirical diallel, NAM, and MAGIC populations of *Arabidopsis*, tomato, and maize.

Chapter 3: A generic IBD-based mixed model approach is implemented in an easy-to-use R package called *statgenMPP*. The package *statgenMPP* incorporates the functionalities of IBD calculation from an R package called *statgenIBD* or software called *RABBIT* and functionalities of solving mixed models for QTL mapping from the R package *LMMsolver*. The flexibility and convenience of *statgenMPP* are demonstrated by analyzing public data sets of soybean NAM and barley complex multi-cross populations.

Chapter 4: For multi-cross populations (diallel and NAM populations), a one-dimensional (1D) genome scan approach is developed to detect any form of inter-locus interactions or epistasis in the form of QTL-by-genetic-background interaction. The nature of a QTL being family-specific or parent-specific is modeled and determined. In contrast to the generic method modelling only additive (parent-specific) QTL effects, the approach using a combination of family-specific and parent-specific genetic effects improves the mapping power of epistatic QTLs in simulated data sets and increases the number of identified QTLs and the proportion of explained genetic variance in empirical NAM and diallel MPPs of cucumber, tomato, maize, and *Arabidopsis*.

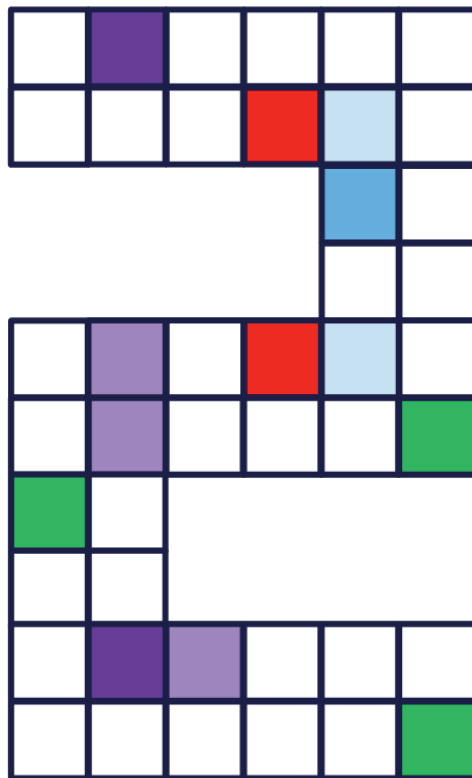
Chapter 5: For data from multi-environment trials using multi-parent populations (MET&MPPs), an approach for MET&MPP analysis is developed to identify QTLs corresponding to the differential phenotypic responses over multiple environments. Random QTL effects with different variance structures are defined following the stability of QTL effects across environments and families: environment-consistent or environment-specific within the specific family or across families. The approach is demonstrated on empirical MET&MPP data in a cowpea MAGIC, a maize diallel, and some NAM populations.

Chapter 6: We discuss the usefulness of MPPs over BPPs and diversity panels in the application of plant breeding programs. We further discuss the advantages of using IBD probabilities as genetic predictors and the efficiency of using mixed model approaches to estimate random-QTL effects. The possibilities of multi-trait analysis and the analysis of further complex genetic architectures such as dominance are explored. Lastly, the

possibilities of extending the current IBD-based mixed model approaches to regularized regression approaches and genomic prediction are discussed.

Chapter 2

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



This chapter has been published as:

Li, Wenhao, Martin P. Boer, Chaozhi Zheng, Ronny VL Joosen, and Fred A. Van Eeuwijk. "An IBD-based mixed model approach for QTL mapping in multiparental populations." Theoretical and Applied Genetics 134, no. 11 (2021): 3643-3660.

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 multi-parent population (MPP) designs.

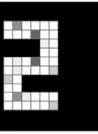
Multi-parent populations (MPPs) have become popular for quantitative trait loci (QTL) detection. Tools for QTL mapping in MPPs are mostly developed for specific MPPs and do not generalize well to other MPPs. We present an IBD-based mixed model approach for QTL mapping in all kinds of MPP designs, e.g., diallel, Nested Association Mapping (NAM), and Multi-parent Advanced Generation Intercross (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 for multiallelic 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 advanced IBD-based multi-QTL mixed model approach incorporating both kinship relations and family-specific residual variances (IBD.MQMkin_F) is robust across a variety of MPP designs and allele segregation patterns in comparison to a widely used benchmark association mapping method, and in most cases, outperformed or behaved at least as well as other tools developed for specific MPP designs in terms of mapping power and resolution. Successful analyses of real data cases confirmed the wide applicability of our IBD-based mixed model methodology.

1. Introduction

MPP designs have their unique advantages for QTL mapping over bi-parental populations and association panels. Crossing two parents in a bi-parental 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 (Liu and Zeng 2000; Pascual et al. 2016). We can use association panels to broaden the genetic diversity, but low-frequency variants may increase false positives (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; Malosetti et al. 2007; Xiao et al. 2016; Sul et al. 2018). Experimental MPP designs, as a compromise between bi-parental 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 (Huang et al. 2011, 2015; Gardner et al. 2016) populations 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; 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 (Würschum et al. 2011, 2012), a maize NAM population (Li et al. 2011; Giraud et al. 2014), and a tomato MAGIC population (Pascual et al. 2015). Several requirements can be formulated for the QTL mapping methodology involving general MPP designs. Firstly, it will be convenient to define QTL effects in terms of their origins while allowing residual polygenic and non-genetic effects to have heterogeneous variances. The multi-QTL effect (MQE) model with a mixture of bi-allelic, ancestral, and parental QTL effects and cross-specific residual proposed by Garin et al. (2017) provides an example of such an approach for the EU-NAM maize data collection. Secondly, in addition to a kinship or marker-structured polygenic term to control background genetic variation, it is attractive to have further control using some form of cofactors as in classical composite interval mapping. A good example of the latter is the inclusive composite interval mapping (ICIM) approach as described by Li et al. (2007), Zhang et al. (2019), and Shi et al. (2019), with an application to an eight-way MAGIC design in cow pea.

In the current paper, we aim at developing a unified QTL mapping framework 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 bi-parental populations to detect bi-allelic QTLs. In the context of MPP designs, we can estimate multi-allelic effects referring to different parental origins. Despite differences in 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). Our 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, to which QTL allele effects can be estimated. Additional terms are added to model random genetic effects from families, like family-specific residuals (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) (Stich et al. 2008; Malosetti et al. 2011).

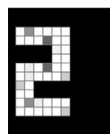
We constructed a series of models varying in whether they adopt identity-by-state (IBS) or IBD information as the basis for genetic predictors and how they account for individual relatedness. We simulated diallel, NAM, and MAGIC designs from four inbred *Arabidopsis* parents to evaluate the performance of our models and compared that performance where possible with that of alternative approaches, such as IciMapping (Meng et al. 2015) for our simulated NAM design, and the integrated genetic analysis software for multi-parent pure-line population (GAPL) (Zhang et al. 2019) for the simulated MAGIC design. We compared further QTL mapping results obtained by our IBD-based mixed model approach for our simulated diallel and NAM designs with those of mppR (Garin et al. 2017). Overall, for the simulated data, our IBD-based mixed model approach performed well for mapping power and resolution and was competitive in comparison to alternative approaches, where our method is applicable to a wider set of MPPs. Further demonstrations of our approach are provided for various empirical MPP data sets.

2. Methodology

We developed the QTL mapping methodology in the linear mixed model framework. First, a linkage map for the MPP design is required as input for IBD calculations and QTL identification procedure. Best is to construct a consensus map following a protocol of marker cleaning, grouping, ordering (Wu et al. 2008; Taylor 2018) and map integration (Endelman and Plomion 2014) for MPP designs. For an MPP design contains N individuals in F families that are derived from crosses between P parents, 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} ($i=1, 2, \dots, N$) individual contributed by an offspring allele stemming from the j^{th} ($j=1, 2, \dots, P$) 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. In the IBD-based models, the genetic predictors are given by the expected numbers of IBD alleles, where the QTL allelic effect, a_j , will denote the haplotype effect of the j^{th} parent.

2.1. IBD probability calculation

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 can 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 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. Figure 1 gives a schematic impression of the IBD calculations performed by RABBIT.



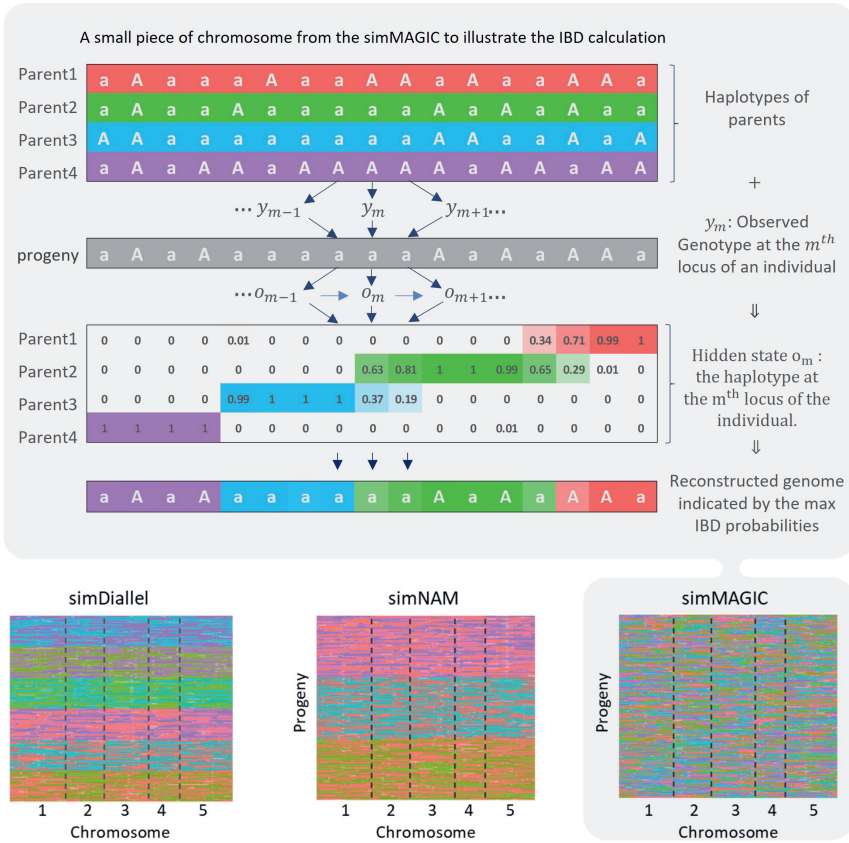


Figure 1. **Upper panel** An example of a MAGIC design to illustrate the framework of IBD calculations underlying the construction of design matrices for mixed model QTL mapping. The assessment of chromosome segments in the offspring of having been transmitted from one of the parents follows on the estimation of transmission probabilities of alleles from parents to offspring. For this example, MAGIC parents and offspring are assumed to be inbred. Therefore, haplotypes and genotypes coincide with respect to allelic composition. The labels 1 and 2 refer to alternative alleles in the parental haplotypes and to alternative genotypes in the offspring. For the hidden parental states in the offspring, the transmission probabilities for respective parents are shown. The parental contribution with the highest transmission probability determines the IBD status in the offspring individuals. **Lower panel** Graphical genotype heat maps showing parent of origin information for offspring in simulated diallel, NAM, and MAGIC populations obtained from thresholding IBD probabilities

2.2. Mixed model construction

We constructed five IBD-based models (Table 1) to estimate multi-allelic effects in a genome-wide scan. We compared the IBD-based modelling with a benchmark GWAS model based on IBS information. 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.

Table 1. Overview of mixed models used for IBD-based QTL mapping.

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	$Y = X\beta + M_q a_q + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^F I_{n_k} \sigma_{\varepsilon_k}^2)$
IBD.MQM_F	IBD	Cofactors	Family-specific	$Y = X\beta + \sum_{c \neq q} M_c a_c + M_q a_q + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $a_c \sim N(0, I_P \sigma_c^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^F I_{n_k} \sigma_{\varepsilon_k}^2)$
IBD.Kin_F	IBD	Polygenic term	Family-specific	$Y = X\beta + M_q a_q + g + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $g \sim N(0, K \sigma_g^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^F I_{n_k} \sigma_{\varepsilon_k}^2)$
IBD.MQMkin_F	IBD	Cofactors polygenic term	Family-specific	$Y = X\beta + \sum_{c \neq q} M_c a_c + M_q a_q + g + \varepsilon$	$a_q \sim N(0, I_P \sigma_q^2)$ $a_c \sim N(0, I_P \sigma_c^2)$ $g \sim N(0, K \sigma_g^2)$ $\varepsilon \sim N(0, \oplus_{k=1}^F I_{n_k} \sigma_{\varepsilon_k}^2)$
IBS.Kin (GWAS model)	IBS	Polygenic term	Homogeneous (Uniform)	$Y = X\beta + X_q \lambda_q + g + \varepsilon$	$g \sim N(0, K \sigma_g^2)$ $\varepsilon \sim N(0, I_N \sigma_\varepsilon^2)$

The first two models are called *IBD.SQM_U* and *IBD.SQM_F*. We use IBD information as the basis for the genetic predictors in a simple single-locus QTL mapping model (SQM) with respectively 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:

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

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

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

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 the 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:

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

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

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

where the residual term ε has a family-specific VCOV structure written as $\bigoplus_{k=1}^F I_{n_k} \sigma_{\varepsilon_k}^2$, in which $\sigma_{\varepsilon_k}^2$ 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, i.e., 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:

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

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

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

where the design matrix for a cofactor is M_c , and the column vector of the random QTL effect 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:

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

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

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

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

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 the family-specific VCOV structure:

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

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

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

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

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 employed in population-based mapping. We will refer to this model as *IBS.Kin*:

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

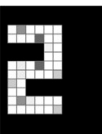
$$g \sim N(0, K \sigma_g^2)$$

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

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.

2.3. QTL detection procedure

To identify QTLs with the above IBD-based mixed model approaches, we used a likelihood ratio test (LRT), $LRT = -2 \left(\ln \left(\frac{\max L_0}{\max L_A} \right) \right)$, comparing the likelihood under the alternative ($\sigma_a^2 \neq 0$) and null ($\sigma_a^2 = 0$) hypotheses to evaluate variance components representing potential QTLs, with $\max L_0$ and $\max L_A$ being the Residual Maximum Likelihood (REML) (Gleeson and Cullis 1987) under null and alternative hypotheses estimated by ASReml-R (Version 3.0) (Butler et al. 2009). The LRT for the genetic variance (σ_a^2) of a set of random QTL effects for a single QTL approximates a $0.5\chi_0^2 + 0.5\chi_1^2$ mixture distribution (Self and Liang 1987). The p-value corresponding to the LRT statistic was expressed as a $-\log_{10}(p)$. 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 $-\log_{10}(p)$ profile along the genome stabilized. For the reference GWAS model *IBS.Kin*, we utilized ASReml-R to incorporate the kinship matrix and performed a Wald-test (Molenberghs and Verbeke 2007) to determine significant QTLs.



3. Data sets

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 mapping power and resolutions for major QTLs. In the section “Model performance assessment on simulated data”, we will define mapping power and 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 four-way MAGIC designs. With this number of parents, we still obtain enough details 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.

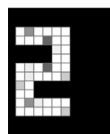
Phenotypes for the offspring individuals contained a contribution from alleles at three major QTLs, positioned on different chromosomes and at markers with varying genotypic configurations across the parents, to investigate the impact of the number of segregating families and allele frequencies on QTL detection. Details are given in the section “Details of the simulation”. 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. An independent error was added to major and minor QTL effects to determine 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 linkage map with 462 markers was created from bi-parental linkage maps for offspring populations involving the above parents. We selected positions at three markers, namely *simQTL1*, *simQTL2*, and *simQTL3*, on the consensus map to assign major QTL allelic effects. Marker genotypes in the parents were coded as 11 and 22 for homozygous reference and alternative alleles. For the three *simQTLs*, a homozygous

genotype coded as 11 was carried by respectively two, one, and three parents (Figure 2A). Starting at 20cM above and below the major QTL position, minor QTLs were uniformly placed at distances of 10cM 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 2B). Progeny was 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.



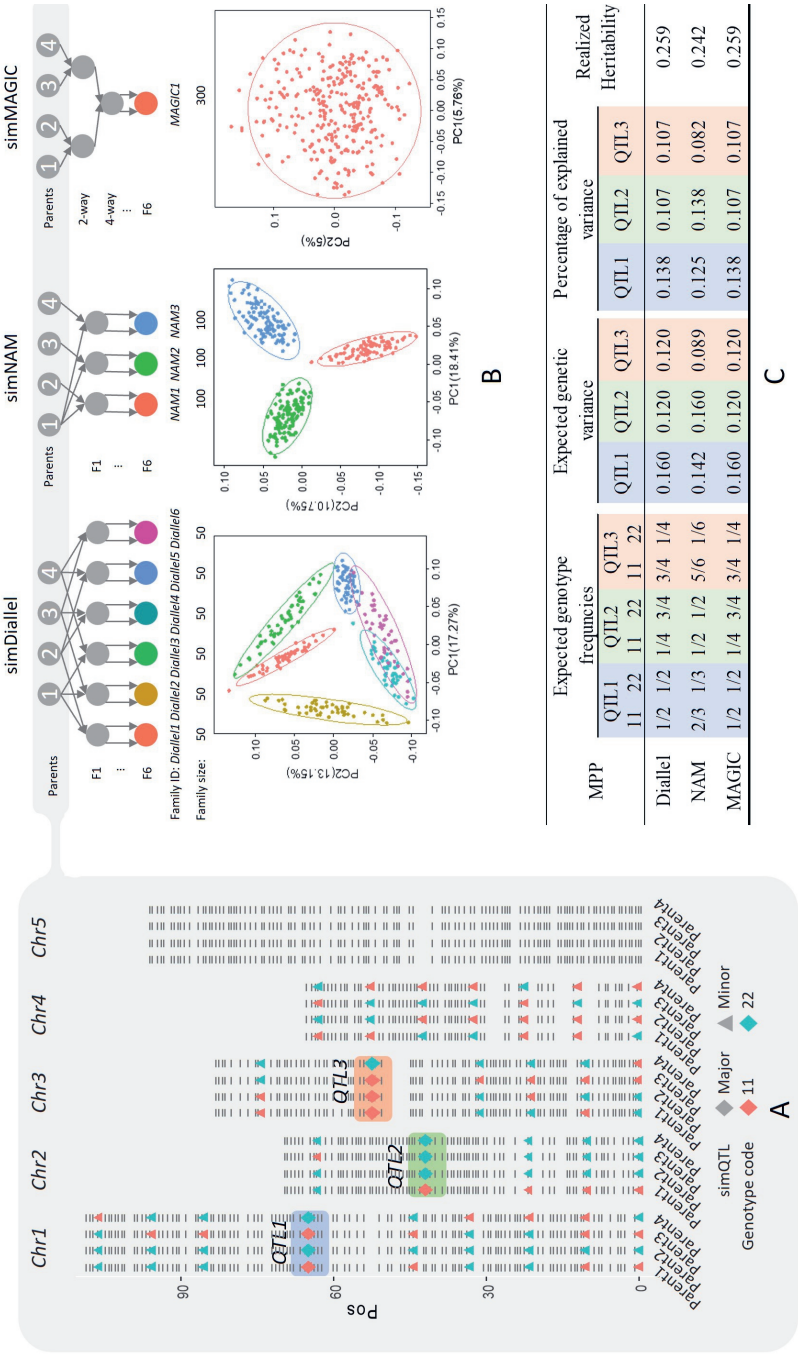


Figure 2. **A** Marker positions and genotypes for the four real inbred *Arabidopsis* genotypes used for simulating different MPP designs. Three major QTLs (diamonds) were simulated with an additive allelic substitution effect of 0.4 and the allele labeled as 1 increasing the trait; 24 minor QTLs (triangles) were simulated with the additive allelic substitution effect 0.1 with the allele labeled as 1 again increasing the trait. **B** Crossing schemes of simulated diallel, NAM, and MAGIC designs using the four parents with PCA plots for progenies based on simulated genome data. **C** Summary of expected genotype frequencies and genetic variance of each simulated major QTL and realized heritability of all major QTLs.

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 retaining sufficient power for QTL detection across the full population as well as within families. The six families in the diallel design, named *Diallel1* to *Diallel6*, contained 50 progenies per family, and the three NAM families, named *NAM1* to *NAM3*, 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 2B).

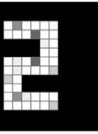
The phenotype of each progeny was the sum of the genetic effects of major and minor QTLs plus residual errors: $Y = az_1 + az_2 + az_3 + \sum_{i=1}^{24} bz_i + \varepsilon$, where a and b are the major and minor additive effects, respectively; and z_i is the genotype indicator equal to 1 or -1 for the marker with homozygous genotype coded as 11 or 22, and equal to 0 for residual heterozygotes coded as 12 with very low frequencies in F6 generations; the residual term ε 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 ‘realistic’ QTL effect sizes for major and minor QTLs, i.e., neither too low nor too high to compare the performance of the models, we did a grid search on QTL effect sizes. As an example, and 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 chosen to be 0.1. Supplementary Figure S1 shows an example of the distribution of simulated phenotypes in families of different MPP designs. We can calculate the expected genetic variance at the q^{th} major QTL based on the formula

$$var(simQTL_q) = a^2 \{E(Z_q^2) - E(Z_q)\}^2 \text{ in simulated MPP designs with equal family sizes.}$$

The frequency of Z_q relies on the specific MPP design. The percentage of explained variance is $var(simQTL_q)/var(Y)$. The heritability of the simulated trait is a function of both major and multiple minor QTLs. We calculated the realized heritability from the genetic effect contributed by major QTLs’ additive effects over 500 replications (Figure 2C).

3.1.3. Model performance assessment on simulated data

We simulated 500 replications for each MPP design and assessed the performance of all six models on these 500 simulated data sets. As criteria for performance, we chose mapping power and resolutions of major QTLs. The successful detection of a major QTL had to meet two requirements. Firstly, the $-\log_{10}(p)$ 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 $-\log_{10}(p)$ value on the same chromosome should be within the QTL window size (20cM). The mapping resolution of a major QTL in our study was indicated by the average genetic distance between the true simulated position and the detected position with the highest $-\log_{10}(p)$ value on the same chromosome over 500 runs. A shorter distance indicates a higher mapping resolution.






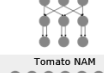

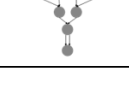
We compared our IBD-based mixed models to other approaches for genetic study in certain types of MPP designs. One of them is called ICIM (Li et al. 2007; WANG 2009) that is implemented in IciMapping (Meng et al. 2015) and GAPL (Zhang et al. 2019). Specifically, for the NAM design, we employed the joint inclusive composite interval mapping (JICIM) approach (Li et al. 2011) in the software IciMapping. For the MAGIC design, we used the *PLQ* function in the software GAPL for background-controlled QTL mapping (Zhang et al. 2017). However, the ICIM-based approaches cannot be applied to diallel designs. Another package for QTL mapping in MPP designs is mppR (<https://cran.r-project.org/web/packages/mppR/index.html>). The MQE model of mppR defines QTL effects for genomic positions at di-allelic, parental, and ancestral levels and chooses the type of QTL effect that produces the highest significance (Garin et al. 2017). The MQE model can be applied to diallel and NAM designs but cannot handle MAGIC designs. QTL mapping for simulated NAM and MAGIC designs was performed using IciMapping and GAPL by setting the mapping method ICIM-ADD to estimate only additive effects with the LOD threshold 4.2 that was the same as the Bonferroni-corrected threshold used in our mixed models. For QTL mapping in simulated diallel and NAM designs by mppR, we also applied the same threshold 4.2 to map QTLs with different effect types using the MQE mapping model. QTL mapping power and resolutions for all alternative software packages were summarized within the same QTL window size of 20 cM as we used in our mixed models.

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. Six data sets are summarized in Table 2. Given the available genotypic and phenotypic information and the consensus linkage map from previous studies, we calculated IBDs using RABBIT and mapped QTLs by the IBD-based mixed models using ASReml-R following the framework described in "Methodology".

Following Verbyla (2019), for each combination of data and model, we calculated the Bayesian information criterion (BIC) to identify empirically an appropriate mixed model: $BIC = (DF_{fixed} + DF_{var}) \times \ln(n - r + DF_{fixed}) - 2\ln(L_{max})$, where n is the number of observations (population size) and r is the rank of the fixed effects design matrix (related to the number of families); DF_{fixed} and DF_{var} are the degree of freedoms for fixed (families) and random parameters (QTL effect variances), respectively; L_{max} is the estimated residual maximum likelihood. Parental effects at QTL candidates were estimated from the model with the smallest BIC value.

Table 2. Summary of empirical maize and tomato data sets of diallel, NAM, and MAGIC designs collected from previous studies.

MPP designs	# of progenies	# of markers	Traits	Reference
 <p>Parents F1 RIL</p> <p>Maize Diallel</p>	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)
 <p>Parents F1 DH</p> <p>Maize NAM</p>	841	22122	Dry grain yield (DGY), PH	(Bauer et al. 2013; Garin et al. 2017)
 <p>Parents 2-way 4-way 8-way F2 F4 F8 F16</p> <p>Maize MAGIC</p>	303	41473	Pollen shed (PS), plant height (PH), ear height (EH), and grain yield (GY)	(Dell'Acqua et al., 2015)
 <p>Parents F1 F2</p> <p>Tomato Diallel</p>	248	456	Fruit shape	-
 <p>Parents F1 F2</p> <p>Tomato NAM</p>	718	593	Resistance	-
 <p>Parents 2-way 4-way 8-way F2 F4</p> <p>Tomato MAGIC</p>	397	1345	Fruit weight (FW)	(Pascual et al., 2015)

3.2.1. Maize MPP designs

In the first maize diallel design, 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 (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 applied the package MCQTL to perform QTL mapping in joint and separate families (Jourjon et al. 2005).

The second maize MPP design was the Dent panel of the EU-NAM population. 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 from previous studies (Giraud et al. 2014; Bustos-Korts et al. 2016; Garin et al. 2017). Markers were removed when one or more parents had missing genotype information, which led to 15,813 markers for the IBD calculation and QTL analysis. QTL mapping for DGY and PH was previously performed in MQE model that allowed 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 (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 $B73 \times CLM91$ 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. However, our study utilized only 303 progenies derived from the initial eight parents after excluding progenies derived from the ninth parent (CLM91), because the pedigree information related to the ninth parent was not available for us to calculate IBDs.

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. 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. 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. 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 (Huang and George 2011). The percentage of phenotypic variation was calculated by fitting all significant QTLs in a full mixed model (Huang and George 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 2B). 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 averaged maximum IBD probabilities at all positions were above 0.9 in all simulated MPP designs demonstrating the high quality of the parental genomic reconstructions in the offspring populations.

4.1.2. Model performance assessment

By comparing the QTL mapping results of the six models on simulated MPP designs, we evaluated the efficiency of using IBD information for genetic predictors and the validity of using the mixed model approach to account for the multilevel relatedness of offspring within and across families. Figure 3 demonstrates the performance of each model in terms of mapping power and resolutions at the three major QTLs in each MPP design. Moving from basic to advanced IBD-based mixed models, we can observe the increasing trend regarding mapping power and resolution for major QTLs. 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 QTLs 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. The mapping tools ICIMapping and GAPL detected 3 and 4 false QTLs on chromosome 5 in simulated NAM and MAGIC designs, respectively, while with mppR, 5 and 2 false QTLs were detected on chromosome 5 in simulated NAM and diallel. Therefore, no large deviations from the expected number of false positives were observed for the alternative mapping methods either.



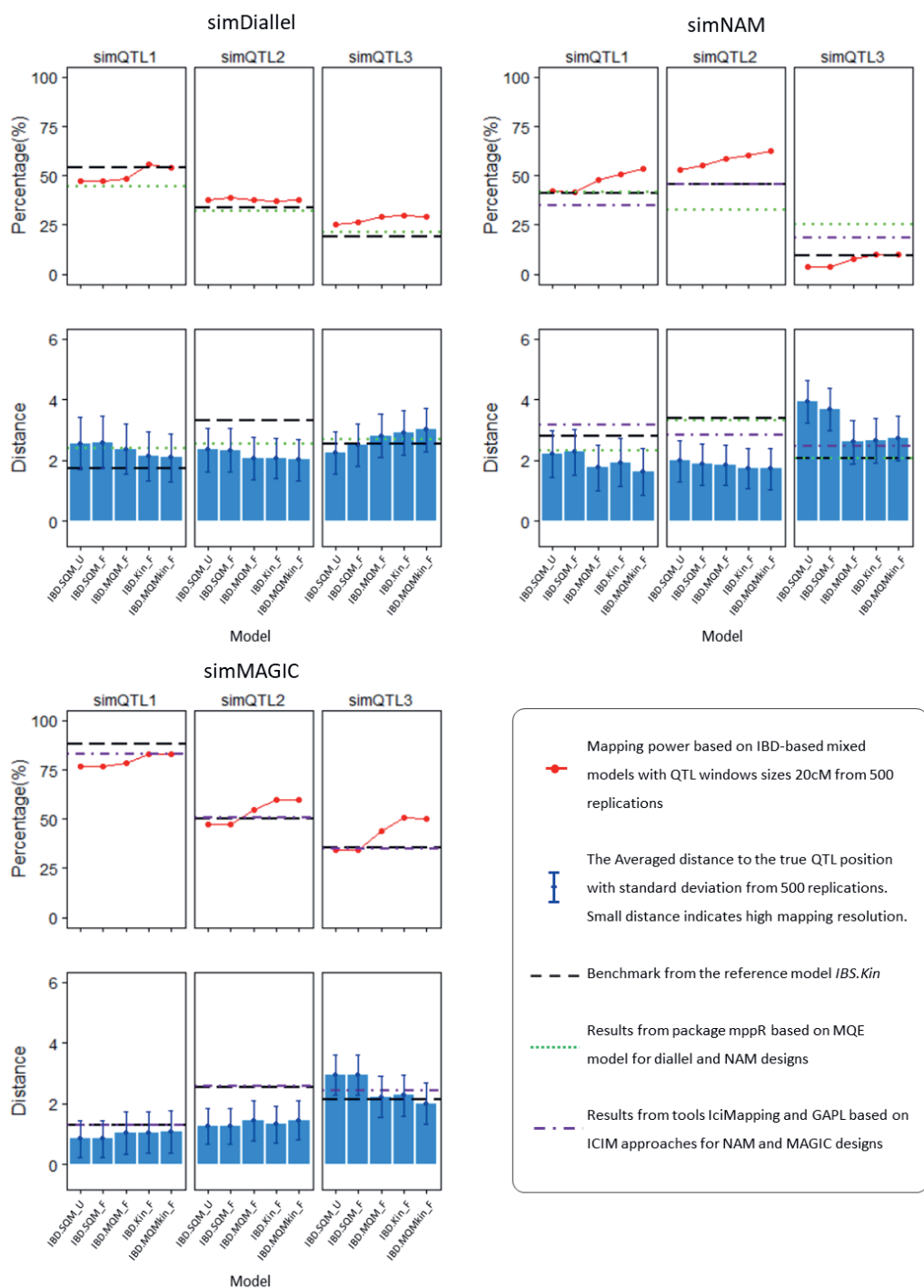


Figure 3. The model performance assessment is based on simulated MPP designs in terms of mapping power (**upper panel**) and mapping resolution (**lower panel**). IBD-based mixed models are compared with the multiple QTL (MQE) model in the mppR package for simDiallel and simNAM designs, and ICIM-based models implementing IciMapping and GAPL tools for respective simNAM and simMAGIC designs.

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 mapping power and resolution. For the *simQTL1* segregating in four out of six families, the mapping power by *IBD.MQMkin_F* was higher than that at *simQTL2* and *simQTL3*, both segregating in three out of six families.

In the simulated NAM design, IBD-based models have advantages over the reference model *IBS.Kin* in improving mapping power and resolution 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 mapping power and resolution. For *simQTL1*, segregating in two out of three families, mapping power and resolution obtained from *IBD.MQMkin_F* were also higher than other models. However, for *simQTL3* that segregated in only one family, *IBD.MQMkin_F* detected this QTL with slightly higher mapping power than other IBD-based models but the mapping resolution was lower than the *IBS.Kin* model.

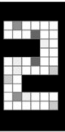
The simulated MAGIC design seems to be the most promising MPP for detecting all three *simQTLs* — all of them were detected with relatively high mapping power and resolutions over diallel and NAM designs. Especially for *simQTL1* with an expected genotype frequency of 0.5, both the advanced IBD-based model *IBD.MQMkin_F* and the reference IBS-based model *IBS.Kin* successfully detected this QTL with high mapping power and resolution. For *simQTL2* and *simQTL3*, advanced models *IBD.Kin_F* and *IBD.MQMkin_F* performed better than the rest of the models considering both mapping power and resolutions.

When comparing our IBD-based mixed model approach to alternative approaches from the literature, we observed (see Figure 3) that results of *IBD.MQMkin_F* were comparable to ICIM-based approaches for simulated NAM and MAGIC designs, while the same *IBD.MQMkin_F* was comparable to the MQE model for simulated diallel and NAM designs.

In the simulated diallel design, the mapping power of *simQTL1* and *simQTL3* by *IBD.MQMkin_F* were higher than the MQE model using the mppR package, while no apparent difference occurred between *IBD.MQMkin_F* and MQE for *simQTL2* in terms of mapping power. For the mapping resolution, *IBD.MQMkin_F* and other IBD-based models performed better than the MQE model on all *simQTLs*.

In the simulated NAM design, we can compare *IBD.MQM_F* with JICIM approach in the software IciMappig and MQE model in the mppR. For *simQTL1* and *simQTL2*, *IBD.MQM_F* performed better than both JICIM and MQE models with higher mapping power and resolution. However, for *simQTL3* that segregated in only one of three families, *IBD.MQMkin_F* detected this QTL with lower mapping power and resolution than JICIM and MQE, because of the test for a variance component being underpowered for a single biparental population.

For the simulated MAGIC design, we see that *IBD.MQMkin_F* detected *simQTL2* and *simQTL3* with higher mapping power and resolution than the ICIM-based approach in the



GAPL software. Both approaches detected *simQTL1* with equally high mapping power, and the mapping resolution is slightly higher by using *IBD.MQMkin_F* than GAPL.

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 4A; Supplementary Figure S2). 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 BIC value (Supplementary Table S1).

Coles et al. (2010) reported QTL mapping using separate bi-parental 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, and 10 in the joint model by Coles et al. (2010). For other traits (Supplementary S2), it also shows the advanced model *IBD.MQMkin_F* could detect most of the 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 (Supplementary Table S1), we fitted these 7 detected QTLs in this advanced IBD-based mixed model to estimate the parental effects at those QTLs for trait GDDTAP. 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 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 4B; Supplementary Figure S3). Including either cofactors, the polygenic term, or both 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, 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 the MQE model that combined genome-wide scans at bi-allelic, parental, and ancestral levels. Our study adopted 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 4C; Supplementary Figure S4) 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, but we selected 303 progenies individuals in the population derived from the initial eight parents. 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

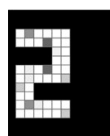
The diallel F2 design of two beef tomatoes and one round tomato generated a population of diverse fruit shapes, and we identified some QTLs related to fruit shape (Figure 4D). In total, we detected three QTLs on chromosomes 1, 2, and 9. All three QTLs were detected using *IBD.SQM_F* and *IBD.MQMkin_F* 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 (Supplementary Table S1). 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 4E). 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 *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.

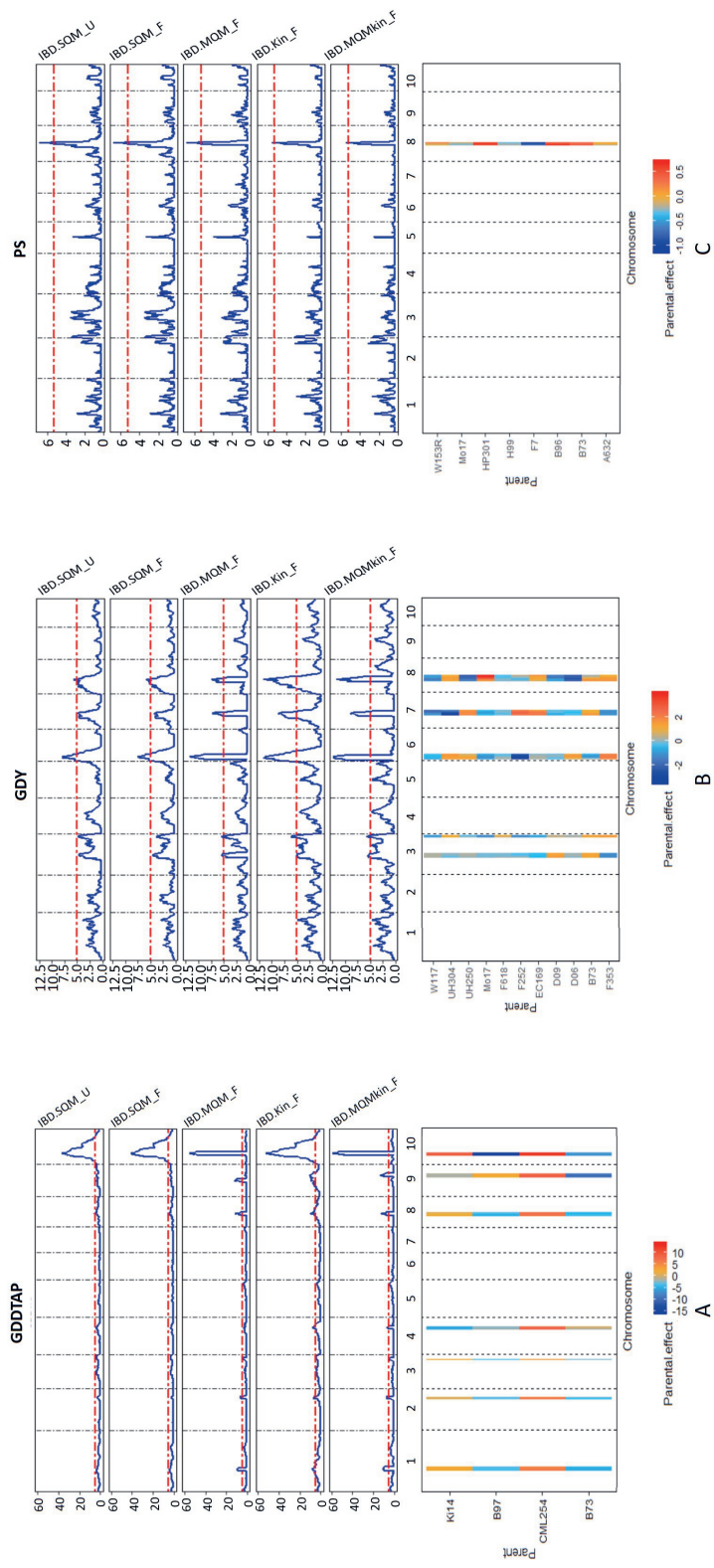
4.2.6. Tomato MAGIC design

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 4F; Supplementary Figure S5). 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 (Supplementary 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 *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 (Pascual et al. 2015). The explained percentage of phenotypic variation of 33% was close to 34% in the previous study. The parental effects on those QTLs are 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.



(Figure caption on next page)

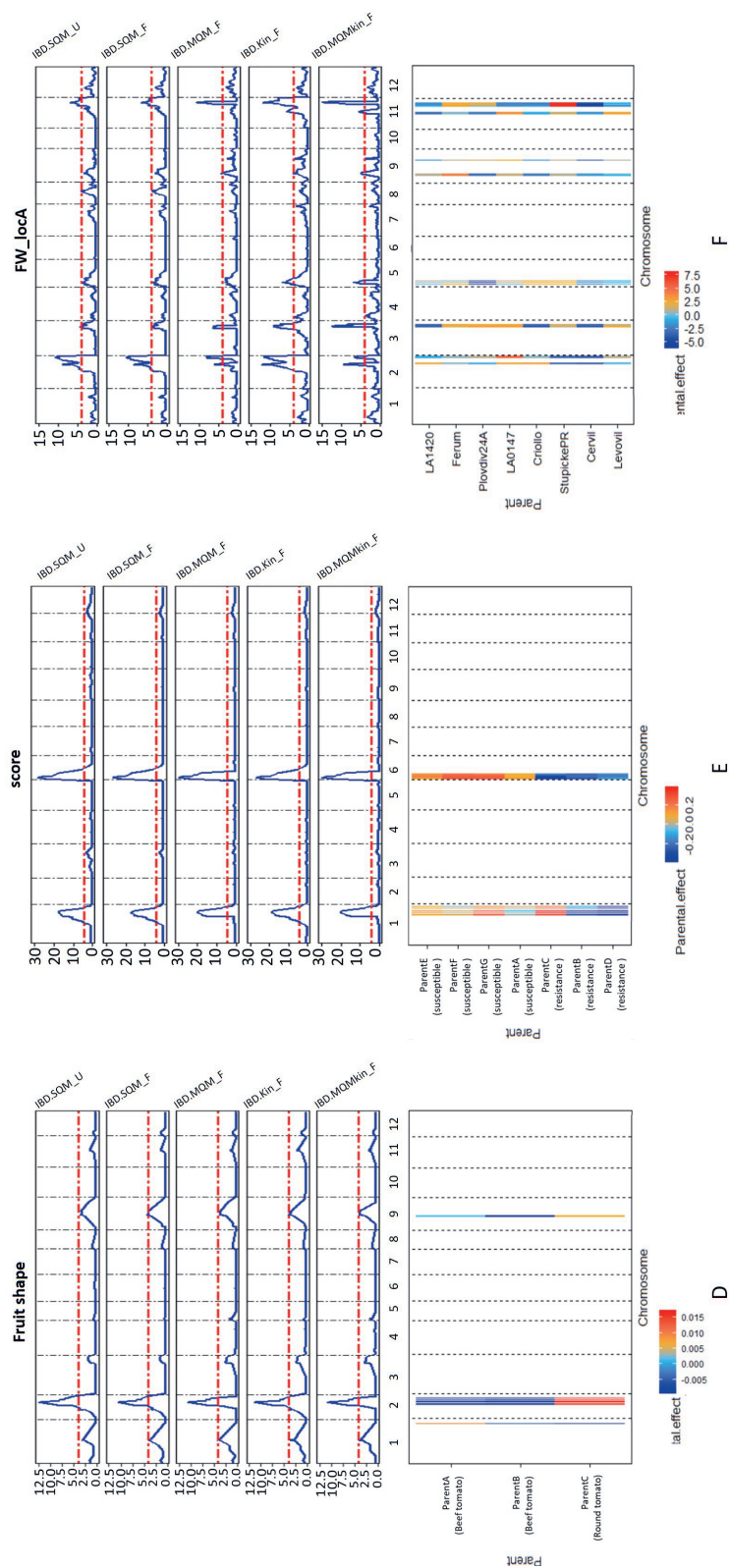


Figure 4. Mapping results of some selected traits as examples in the empirical MPP designs: **A.** maize diallel, **B.** maize NAM, **C.** maize MAGIC, **D.** tomato diallel, **E.** tomato NAM, and **F.** tomato MAGIC using the five IBD-based mixed models. **Upper panel** QTL profiles from the five IBD-based mixed model approaches. **Lower panel** Estimation of parental effects at QTLs detected by a model selected with the smallest BIC among the five models. The BIC values of other models and mapping results of other traits are provided in the supplementary material.

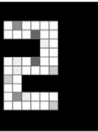
5. Discussion

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 combining 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 bi-parental, multi-parent, and association panels in the context of the genome sequencing era to show their complementarity in genetic studies (Pascual et al. 2016). Our 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 our 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 owing to the specific crosses between parents. For *simQTL3* segregating in only one family of the simulated NAM design, even the advanced IBD-based models could not remarkably improve the mapping results, whereas this QTL could be detected with higher mapping power and 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 (Ogut et al. 2015; Bajgain et al. 2016; Garin et al. 2017). *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 mapping power and 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. The problem for the IBD-based mixed models with *simQTL3* in the NAM MPP was that this QTL segregated in one bi-parental cross only and the test for a variance component related to a QTL will be underpowered when the QTL represents too few allelic effects (Crainiceanu and Ruppert 2004). In such cases, a conventional Wald test for a fixed QTL substitution effect would have been more adequate.

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 (Pascual et al. 2015; Dell'Acqua et al. 2015; Ongom and Ejeta 2018), 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 IBS information, whereas IBD probabilities inferred by the pedigree information and the whole genome can reduce uncertainty for genome origins (Zheng et al. 2015, 2018). Generally, IBD-based models were more effective than the IBS-based model (*IBS.Kin*) concerning mapping power and 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 (Verbyla et al. 2014b; Broman et al. 2018). 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 (Wu et al. 2008; Taylor 2018), and map integrating (Endelman and Plomion 2014). In future work, we can utilize IBD probabilities to infer the recombination between markers for genetic map construction (Zheng et al. 2019), as thus, we simplify and optimize the MPP analysis framework from the beginning.

5.3. Modeling the 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

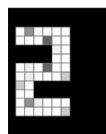
limited family size, e.g., a family size of 50 or 100 in simulated diallel or NAM designs may not have been big enough to reveal the heterogeneity of variance components between families. Still, modeling a heterogeneous error can be advantageous in the fitting of single-locus QTL models in initial genome 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 separate 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. For choosing an appropriate VCOV model for a trait in a particular MPP we proposed a model selection procedure based on BIC.

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 (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 mapping power and 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 mapping 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 individual relatedness (Yu et al. 2006). Likewise, our 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 the estimation of parental effects at the QTL, we need to design matrices that are functions of IBD probabilities. Popular population-based mapping strategies employ the mixed model approach to deal with multilevel relatedness by imposing family-specific and kinship-based VCOV structures on the non-genetic residual and the polygenic term, respectively. 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 mapping power and resolution, and this model is also competitive with other tools developed for specific MPP designs. 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.

Supplementary material

Table S1 Summary of QTL mapping results of empirical MPP designs. The number of QTLs and BIC apply to the final multi-QTL model fitted based on 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
PH	IBD.SQM_U	8	4289
	IBD.SQM_F	8	4314
	IBD.MQM_F	9	4312
	IBD.Kin_F	10	4219
	IBD.MQMkin_F	11	4225

<i>Maize MAGIC</i>			
PS	IBD.SQM_U	1	669
	IBD.SQM_F	1	669
	IBD.MQM_F	1	669
	IBD.Kin_F	1	672
	IBD.MQMkin_F	1	672
PH	IBD.SQM_U	1	663
	IBD.SQM_F	1	663
	IBD.MQM_F	1	663
	IBD.Kin_F	1	667
	IBD.MQMkin_F	1	667
GY	IBD.SQM_U	1	648
	IBD.SQM_F	1	648
	IBD.MQM_F	1	648
	IBD.Kin_F	1	647
	IBD.MQMkin_F	1	647
<i>Tomato diallel</i>			
Fruit shape	IBD.SQM_U	1	1429
	IBD.SQM_F	3	1429
	IBD.MQM_F	2	1433
	IBD.Kin_F	3	1433
	IBD.MQMkin_F	3	1431
<i>Tomato NAM</i>			
Disease score	IBD.SQM_U	2	145
	IBD.SQM_F	2	165
	IBD.MQM_F	2	165
	IBD.Kin_F	2	165
	IBD.MQMkin_F	2	165
<i>Tomato MAGIC</i>			
FT_locA	IBD.SQM_U	3	2719
	IBD.SQM_F	3	2719
	IBD.MQM_F	4	2703
	IBD.Kin_F	5	2667
	IBD.MQMkin_F	8	2649
FT_locB	IBD.SQM_U	3	2346
	IBD.SQM_F	3	2346
	IBD.MQM_F	3	2346
	IBD.Kin_F	3	2327
	IBD.MQMkin_F	3	2327

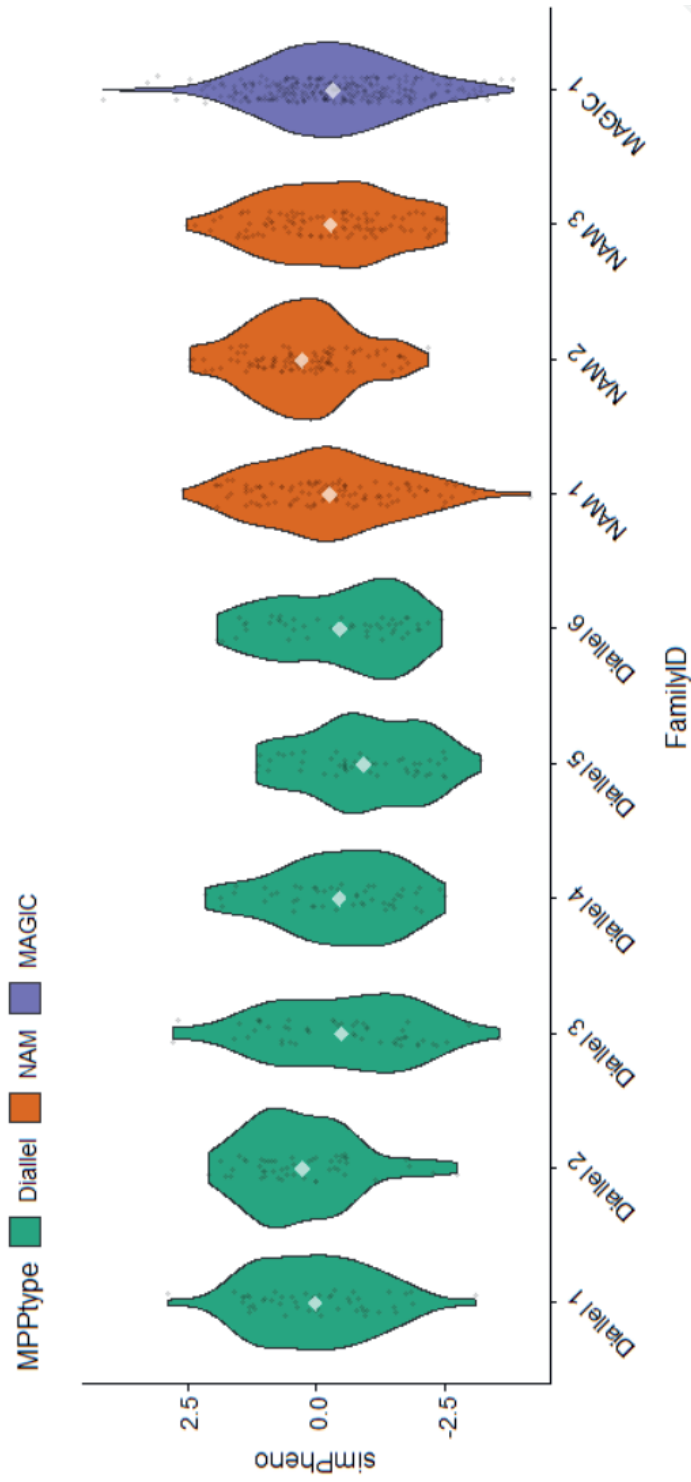
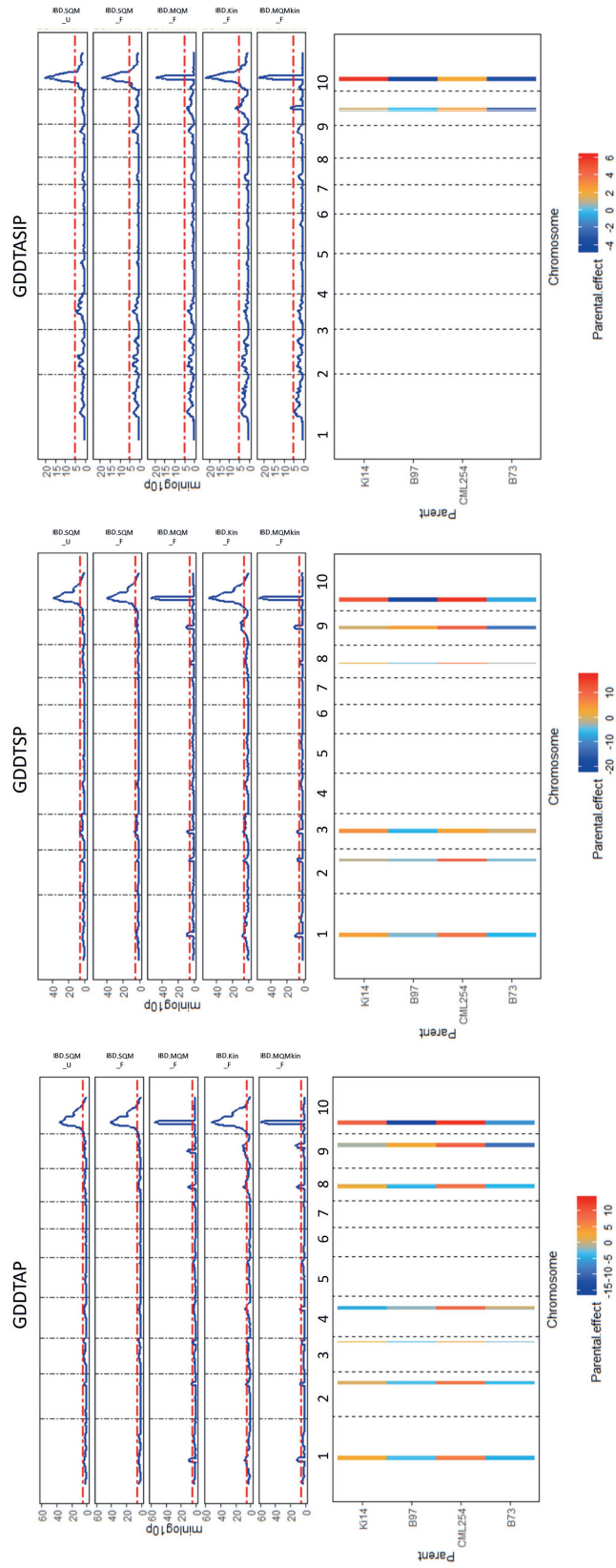


Figure S1 An example taken out of a set of 500 replications demonstrates the family-specific phenotype distributions in simulated diallel, NAM, and MAGIC designs.



(Figure caption on next page)

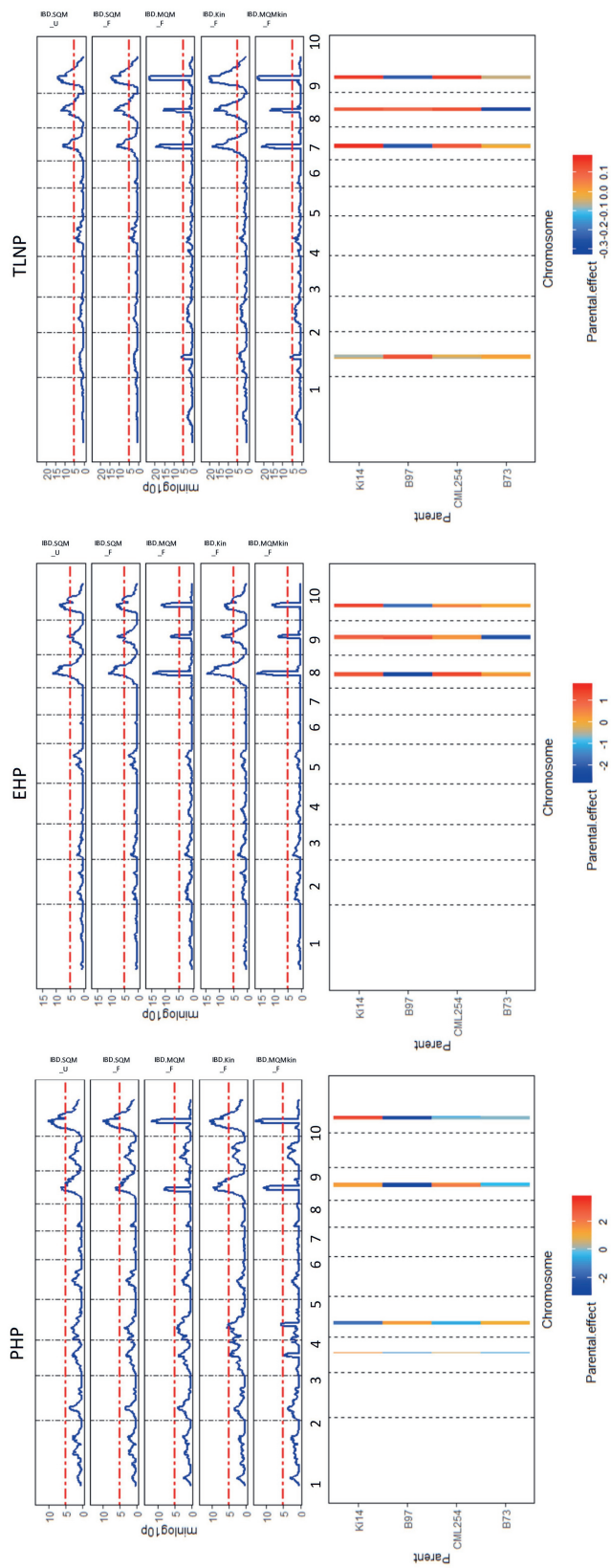


Figure S2. Mapping results for the empirical maize diallel design using the five IBD-based mixed models. **Upper panel** QTL profiles from the five IBD-based mixed model approaches. **Lower panel** Estimation of parental effects at QTLs detected by a model selected with the smallest BIC among the five models.



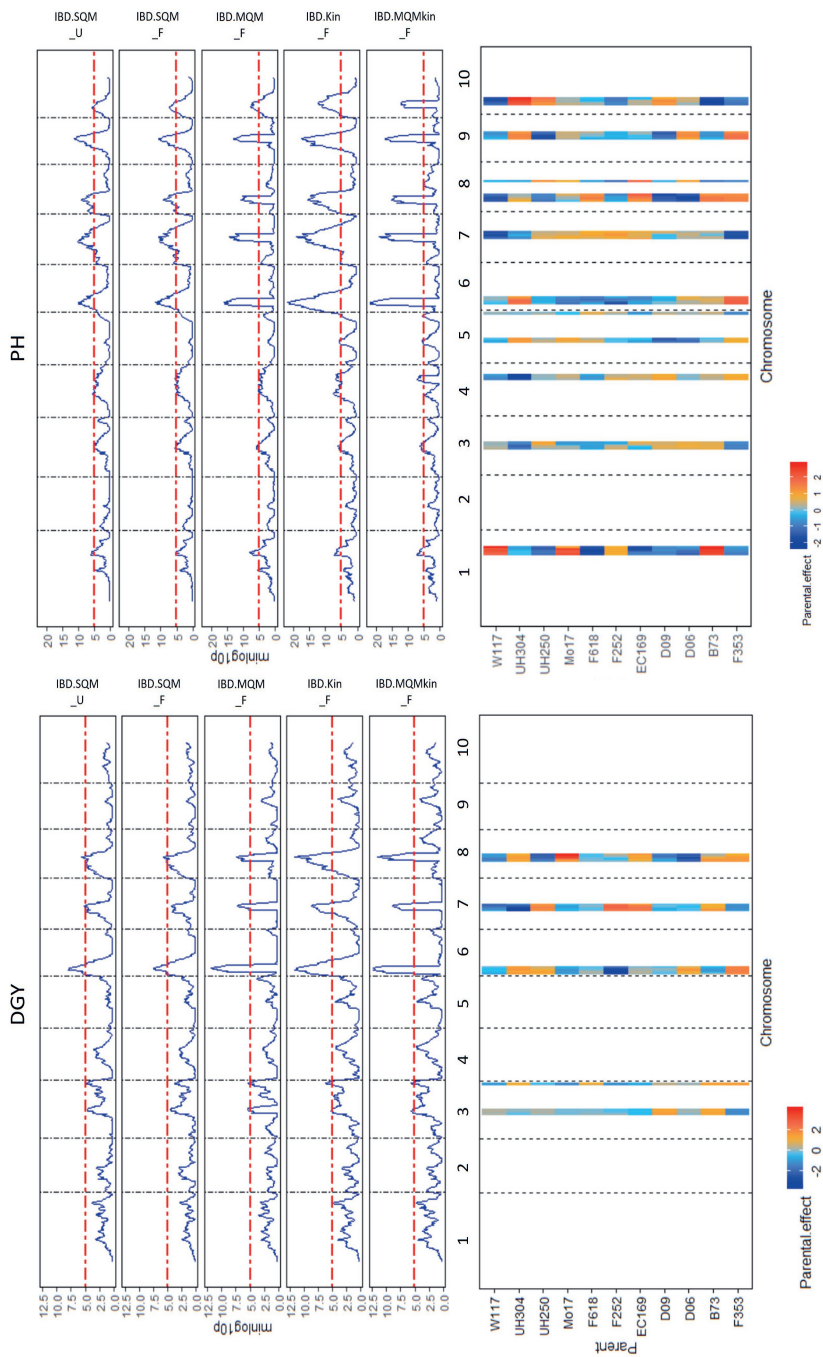


Figure S3. Mapping results for the empirical maize NAM using the five IBD-based mixed models. **Upper panel** QTL profiles from the five IBD-based mixed model approaches. **Lower panel** Estimation of parental effects at QTLs detected by the model with the smallest BIC among the five models.

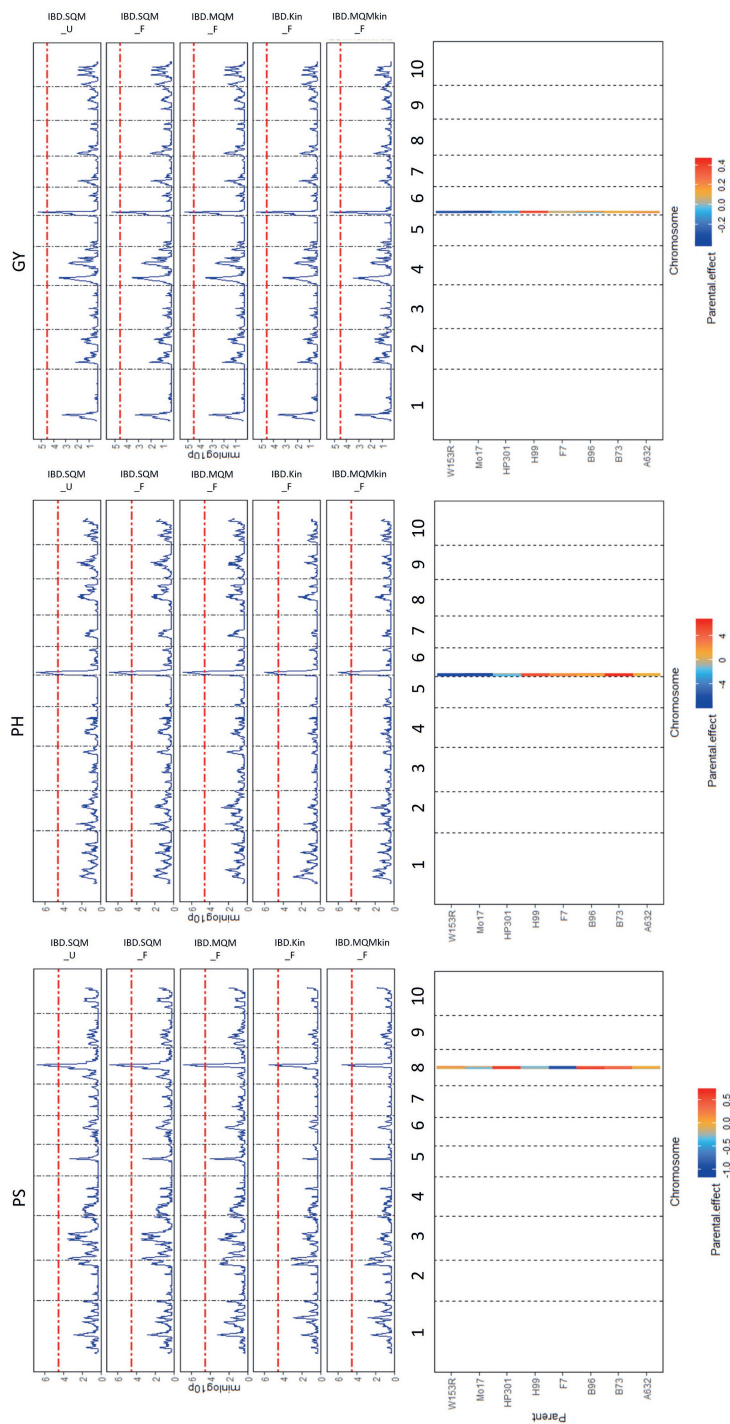


Figure S4. Mapping results for the empirical maize MAGIC using the five IBD-based mixed models. **Upper panel** QTL profiles from the five IBD-based mixed model approaches. **Lower panel** Estimation of parental effects at QTLs detected by the model with the smallest BIC among the five models.



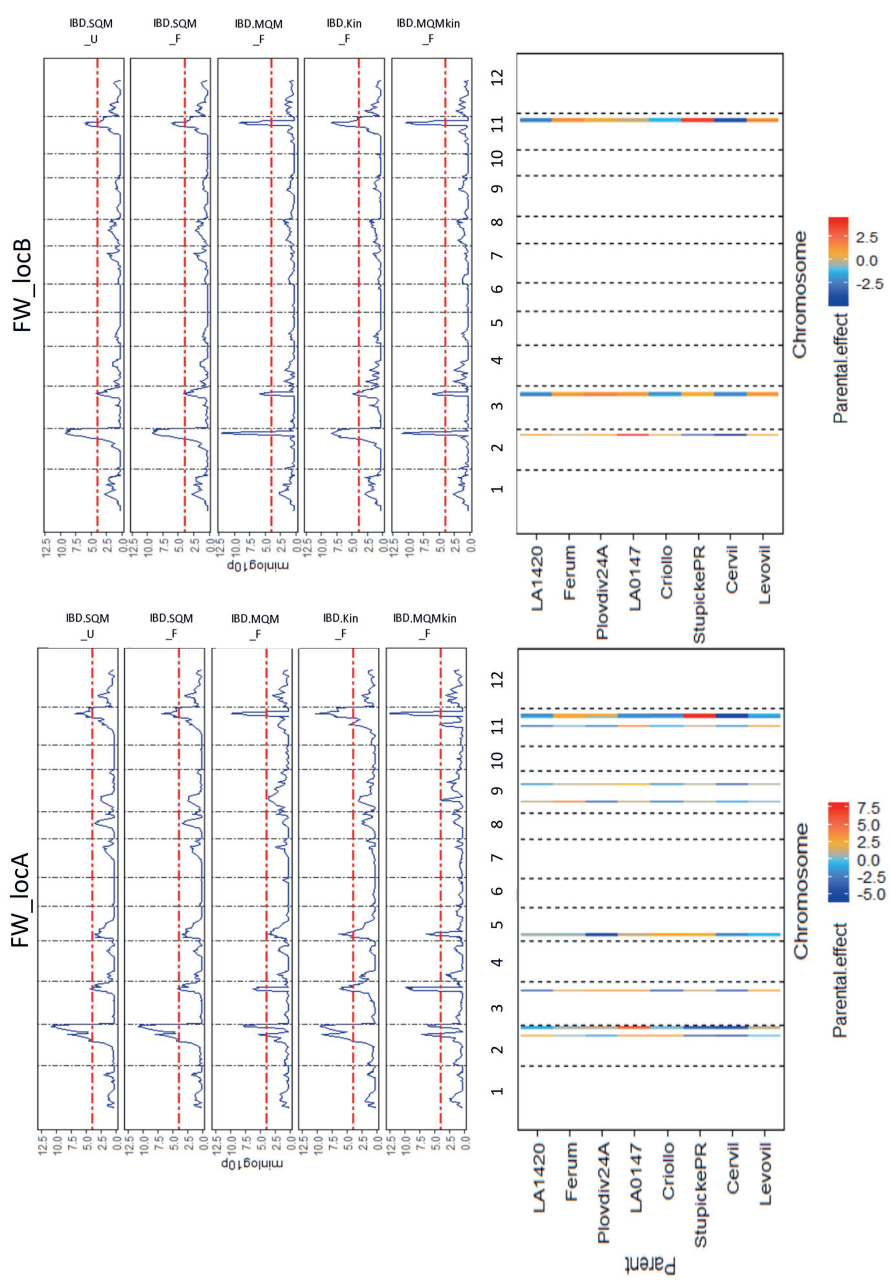
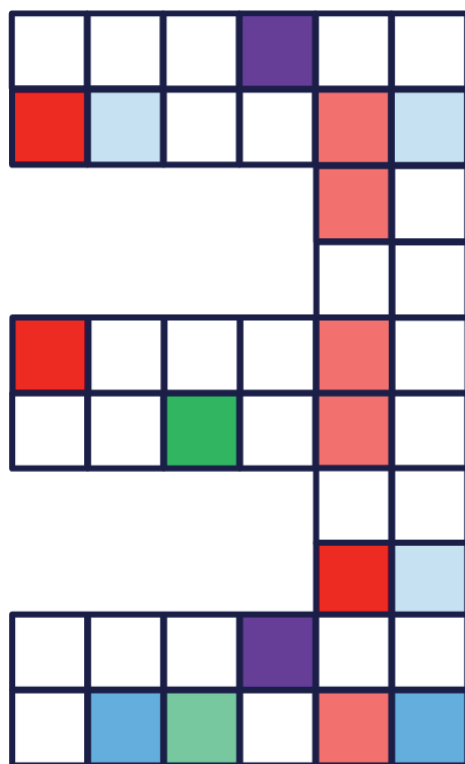


Figure S5. Mapping results for the empirical tomato MAGIC design using the five IBD-based mixed models. **Upper panel** QTL profiles from the five IBD-based mixed model approaches. **Lower panel** Estimation of parental effects at QTLs detected by the model with the smallest BIC among the five models.

Chapter 3

statgenMPP: an R package implementing a generic IBD-based mixed model approach for QTL mapping in a wide range of multi-parent populations



This chapter is based on the following published paper:

Li, Wenhao, Martin P. Boer, Bart-Jan van Rossum, Chaozhi Zheng, Ronny VL Joosen, and Fred A. Van Eeuwijk. "statgenMPP: an R package implementing an IBD-based mixed model approach for QTL mapping in a wide range of multi-parent populations." *Bioinformatics* 38, no. 22 (2022): 5134-5136.

Abstract

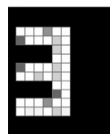
Multi-parent populations (MPPs) are popular for QTL mapping because they combine wide genetic diversity in parents with easy control of population structure, but a limited number of software tools for QTL mapping are specifically developed for general MPPs. We developed an R package called *statgenMPP*, adopting a unified identity-by-descent (IBD)-based mixed model approach for QTL analysis in MPPs. The package offers easy-to-use functionalities of IBD calculations, mixed model solutions, and visualizations for QTL mapping in a wide range of MPP designs, including diallel, nested-association mapping populations (NAM), multi-parent advanced genetic inter-cross (MAGIC) populations and other complicated MPPs with known crossing schemes.

Availability and implementation: The R package *statgenMPP* is open-source and freely available on CRAN at <https://CRAN.R-project.org/package=statgenMPP>.

1. Introduction

Multi-Parent Population (MPP) designs capture wide genetic diversity and overcome the drawbacks of low minor-allele frequencies and population structures (Scott et al. 2020; Arrones et al. 2020). General MPPs, e.g., diallel (Xu 2007; Turner et al. 2018), NAM (Yu et al. 2008), and MAGIC (Huang and George 2011; Huang et al. 2015; Gardner et al. 2016) are nowadays widely used in genetic studies and plant breeding programs. Statistical models that can be used for MPP designs are either family-based (linkage mapping) or population-based (linkage disequilibrium) methods (Würschum et al. 2012; Giraud et al. 2014; Xu et al. 2017), but a limited number of software tools is specifically developed for MPP designs.

In this application note, we present an easy-to-use R package, *statgenMPP*, adopting an Identity-By-Descent (IBD)-based mixed model approach for QTL mapping. Compared to other tools, *statgenMPP* integrates a framework of IBD calculation using R package *statgenIBD* or Mathematica software *RABBIT* (Zheng et al. 2014, 2015; Boer and van Rossum 2021a) with linear mixed models using R package *LMMsolver* (Boer and van Rossum 2021b). The IBD-based mixed model approach estimates random QTL effects in relation to IBD probabilities of parental origins across the offspring genome (Wei and Xu 2016; Li et al. 2021) while accounting for polygenic and family background genetic variation, which has been proven to increase the mapping power and resolution of QTLs for simulated and empirical MPPs (Li et al. 2021).



2. Methods

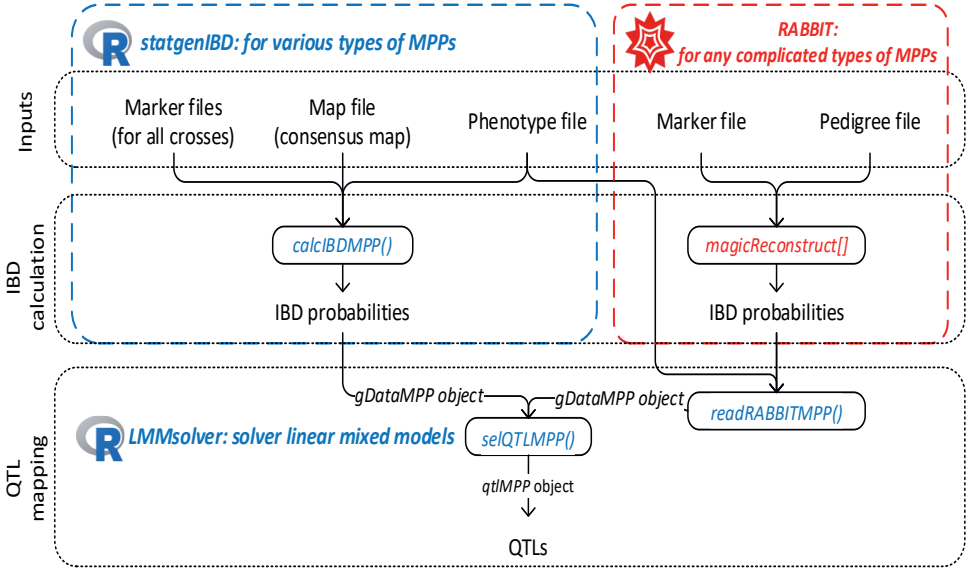


Figure 1. The workflow of IBD-based QTL mapping implemented in statgenMPP.

2.1. IBD calculation

The framework of Hidden Markov Models (HMM) and inheritance vectors (Zheng et al. 2014, 2015; Boer and van Rossum 2021a) is employed for the IBD calculation. For a wide range of MPP designs, IBD probabilities are calculated by the function `calcIBDMPP()` at a customizable grid (cM) for a wide range of population types. For complex MAGIC MPPs and any other MPPs with complicated pedigree structures, the IBD probabilities can be first calculated using `RABBIT` (Zheng 2019) and then imported by the function `readRABBITMPP()`.

2.2. QTL mapping

IBD probabilities between parents and offspring, as design vectors, or genetic predictors, are fitted in a mixed model for QTL mapping using the function `selQTLMPP()`. The IBD-based mixed model approach is described by Li et al. (2021). To test each position on a one-dimensional grid along the genome, a single locus QTL model is fitted whose effects are modeled as random in a mixed effects model:

$$\underline{Y} = X\beta + M_q \underline{a}_q + \underline{g} + \underline{\varepsilon}$$

$$\underline{a}_q \sim MVN(0, I_p \sigma_a^2), \underline{g} \sim MVN(0, K \sigma_g^2), \underline{\varepsilon} \sim MVN(0, \bigoplus_{k=1}^F I_{n_k} \sigma_{\varepsilon_k}^2)$$

Y is a vector with phenotypes; X is the design matrix indicating to which family each individual belongs; β is a vector of fixed family intercepts; M_q is the design matrix containing the expected number of parental alleles as a function of IBD probabilities; a_q is the vector of random parental effects with the variance-covariance (VCOV) structure $I_P \sigma_q^2$, in which I_P is the identity matrix for P parents and σ_q^2 is the genetic variance of the QTL effects; it is optional to include a polygenic term g whose VCOV is described by the kinship matrix K ; the residual term ε has a family-specific VCOV structure $\bigoplus_{k=1}^F I_{n_k} \sigma_{\varepsilon_k}^2$ in which $\sigma_{\varepsilon_k}^2$ is the residual variance of individuals in the k -th family with family size n_k ($k = 1, \dots, F$ where F is the total family number).

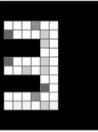
The linear mixed model is fitted and variance components are estimated based on restricted maximum likelihood (REML). Variance components corresponding to putative QTLs ($\sigma_q^2 = 0$ vs. $\sigma_q^2 \neq 0$) are evaluated by likelihood ratio tests (LRT) that approximate a mixture of χ^2 distributions (Self and Liang 1987). Multiple rounds of genome QTL scans can be performed until either 1) no new QTL outside of a certain window size is found with a $-\log_{10}(p)$ value below a predefined threshold or 2) a predefined maximum number of QTLs is reached.

3. Applications

We demonstrate the main functionalities of *statgenMPP* using two publicly available MPP designs — a soybean NAM population and a barley complex population. Other examples with full details are described in the vignette of *statgenMPP*. All computations were performed in (64-bit) R 4.2.1 (R Core Team 2022) and a 3.10GHz Intel Core i5 processor computer with 16GB of RAM and Windows 10 operating system. We used the parallel option of *statgenMPP* using 4 cores, and using the default values, not including a kinship matrix.

Example 1: Soybean NAM population

We selected six families (Figure 2a) with a total population size of 732 genotypes from the soybean NAM project (<https://www.soybase.org/SoyNAM/index.php>) (Xavier et al. 2018) for analysis to demonstrate the functionalities of *statgenMPP*. The central parent IA3023 was crossed with six peripheral parents, including two high-yielding elite lines (TN05.3027 and X4J105.3.4), two diverse-ancestry lines (LG03.2979 and LG03.3191), and two high-yielding lines in drought (PI427136 and PI398881), which in total generated six bi-parental F5 families. The consensus map contained 4,289 markers (Song et al. 2017) and was used to calculate IBD probabilities on a regular 5 cM grid of evaluation points. Multiple traits were measured at nine locations across three years. We analyzed the mean seed weight ("mean_seedWT ") across all trials. The total computation time was 1.03 min.



Soybean NAM population

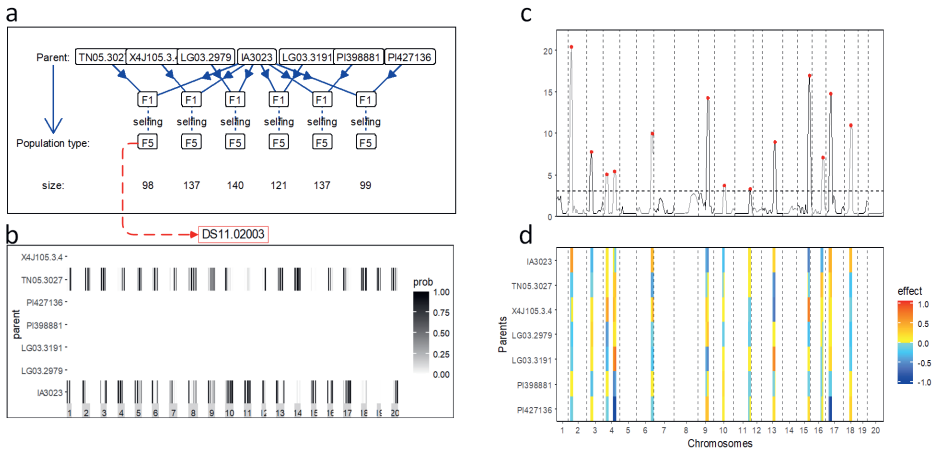


Figure 2. QTL mapping for seed weight in a soybean NAM population using *statgenMPP*. **a.** the pedigree or crossing design of the MPP. **b.** an example to show the parent genome blocks of the first progeny from the first family. **c.** QTL mapping profile for the trait of interest. **d.** estimated QTL effects at parent origins.

Example 2: Barley complex design

The second example is a complex barley MPP design (Liller et al. 2017), with a total population size of 916 genotypes. This MPP design consists of five parents. Four wild parents were crossed with the cultivar Morex and then backcrossed with Morex once. Individuals from the four families from the backcrosses were then crossed with each other as in a full diallel design, which generated six F6 families through five generations of selfing (Figure 3a). IBD calculation for the complex design was performed by RABBIT, and then the output was imported by the *readRABBITMPP()* function in *statgenMPP*. The trait of interest for this population is awn length (“Awn_length”). The total computation time was 1.43 min.

The QTL mapping results for awn length in this barley complex population are shown in Figure 3. According to the computed IBD probabilities at parent origins (Figure 3b), a large proportion of genome blocks for a progeny are inherited from the parent Morex, and the rest of the parent genome blocks are related to the other two parents used for backcross (Figure 3a). The previous study (Liller et al. 2017) can confirm all 12 QTLs for awn length (Figure 3c) identified by *statgenMPP*, among which the QTL on chromosome 7 with strong signal was successfully fine-mapped. Estimated parent effects at the major QTL on chromosome 7 (Figure 3d) show large genetic variation in awn length and parents Morex and HID369 positively contribute to awn length. It explains the substantial variation in awn length from filed observations that HID4, HID64, and HID382 have a significantly shorter awn length than that of Morex and HID369. Further details on how to use the package for visualization can be found in the *statgenMPP* vignette.

Supplementary material

Table S1: Summary of QTLs for seed weight in soybean NAM population.

evalPos	chr	pos	mrkNear	-log10p	varExpl	Parent effect						
						IA3023	TN05.3027	X4J105.3.4	LG03.2979	LG03.3191	PI398881	PI427136
EXT_2_19.13	2	19.13	Gm02_6011261_T_C	20.37	0.03	0.57	-0.15	0.01	-0.23	-0.16	0.05	-0.08
EXT_3_24.21	3	24.21	Gm03_4682537_T_C	7.72	0.03	-0.34	-0.06	0.19	0.33	0.13	-0.32	0.08
EXT_4_19.88	4	19.88	Gm04_6064289_C_A	5.05	0.05	0.14	0.15	0.51	-0.43	-0.32	0.09	-0.13
EXT_4_59.63	4	59.63	Gm04_42365585_A_G	5.35	0.16	-0.03	0.34	0.35	0.17	0.76	-0.52	-1.05
EXT_6_81.81	6	81.81	Gm06_46292681_G_T	9.96	0.05	0.45	-0.48	0.27	-0.14	-0.2	0.13	-0.04
EXT_9_49.12	9	49.12	Gm09_36517311_T_C	14.23	0.05	-0.44	0	0.25	0.25	-0.47	0.06	0.36
EXT_10_53.65	10	53.65	Gm10_36744954_C_T	3.73	0.02	-0.31	0.01	0.09	-0.13	0.05	0.2	0.11
EXT_11_81.49	11	81.49	Gm11_36823232_T_C	3.24	0.01	0.18	0.08	-0.05	-0.15	0.07	-0.08	-0.05
EXT_13_63.98	13	63.98	Gm13_30719300_A_G	8.95	0.09	0.26	-0.53	-0.65	0.02	0.69	-0.01	0.21
EXT_15_63.36	15	63.36	Gm15_40823560_G_A	16.91	0.08	-0.62	-0.03	0.76	-0.32	0.13	-0.04	0.11
EXT_16_57.84	16	57.84	Gm16_30876174_T_C	7.07	0.03	-0.38	0.34	0.03	0.08	-0.05	-0.06	0.05
EXT_17_24.36	17	24.36	Gm17_5062060_C_T	14.7	0.11	0.49	0.22	0.18	0.1	0.26	-0.19	-1.06
EXT_18_33.32	18	33.32	Gm18_10013065_G_A	10.91	0.03	0.38	-0.3	-0.05	-0.29	-0.12	0.13	0.25

Table S2: Summary of QTLs for awn length in barley complex population.

evalPos	chr	pos	mrkNear	-log10p	varExpl	Parent effect				
						Morex	HID4	HID64	HID369	HID382
2_1053	1	47.06	2_1053	10.05	0.02	-4.48	1.72	3.56	-1.43	0.63
2_1187	2	27.21	2_1187	4.86	0.01	3.71	-1.01	-0.2	-0.81	-1.68
1_1522	2	46.06	1_1522	4.73	0.05	-0.06	7.25	-2.55	-5.23	0.58
2_0340	2	68.99	2_0340	6.14	0.02	-3.83	1.96	-0.81	-0.44	3.12
1_1501	3	42.17	1_1501	13.42	0.05	4.66	-2.88	-1.65	-5.95	5.83
1_1070	4	51.72	1_1070	5.83	0.06	-4.79	-1.89	8.33	0.84	-2.5
1_0611	4	102.75	1_0611	4.05	0.03	-0.91	-2.47	3.53	-4.2	4.05
2_0645	5	56.84	2_0645	6.86	0.03	5.22	-2.56	-1.82	1.23	-2.06
1_0094	5	94.26	1_0094	14.48	0.09	-8.45	1.7	-3.08	9.93	-0.1
1_0120	6	1.67	1_0120	4.41	0.04	2.63	-1.39	-5.9	1.99	2.66
2_0687	6	88.03	2_0687	5.7	0.03	-4.33	2.9	3.67	1.14	-3.39
1_1012	7	121.04	1_1012	65.39	0.26	11.38	5.77	-9.66	7.65	-15.14

S3: R codes for the analysis of the soybean NAM population using *statgenMPP*.

```
## Inputs

phenoDat <- read.csv("soybeanNAM_phenotype.csv",stringsAsFactors = F)

crosses <- c("IA3023xTN05.3027","IA3023xX4J105.3.4",

            "IA3023xLG03.2979","IA3023xLG03.3191",

            "IA3023xPI398881", "IA3023xPI427136")

## step 1: IBD calculations

soybeanNAMObj<- calcIBDMPP(crossNames = crosses,

                           markerFiles = paste0(crosses, ".txt"),

                           mapFile = "map.txt",

                           pheno = phenoDat,

                           popType = "F5",

                           evalDist = 5)

### Plot pedigree (Figure 2a)

plot(soybeanObj, plotType = "pedigree")

### Plot parent genome blocks for 1st progeny as example (Figure 2b)

plot(soybeanObj, plotType = "singleGeno", genotype = "DS11.02003")

## step 2: QTL mapping

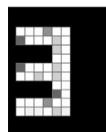
soybeanNAMObjRes<- selQTLMPP(MPPobj = soybeanNAMObj,

                             trait = "mean_seedWT",

                             threshold=3)

### Plot QTL profile and effects (Figure 2c and Figure 2 c and d)

plot(soybeanObjRes, plotType = "QTLProfileExt")
```



S4: R codes for the analysis of the barley complex population using *statgenMPP*.

```
## Step1: Read IBDs from RABBIT output

barleyPheno <- read.csv("barleyComplex_Pheno.csv")

barleyObj <- readRABBIT(infile = "barleyComplex_magicReconstruct.csv",

                        pedFile = "barleyComplex_pedInfo.csv", #optional

                        pheno = barleyPheno)

### Plot pedigree (Figure 3a)

plot(barleyObj, plotType = "pedigree")

### Plot parent genome blocks for 1st progeny as example (Figure 3b)

plot(barleyObj, plotType = "singleGeno", genotype = "AB-1-1-1-1")

## step 2: QTL mapping

barleyObjRes <- selQTLMPP(MPPobj = barleyObj,

                          trait = "Awn_length",

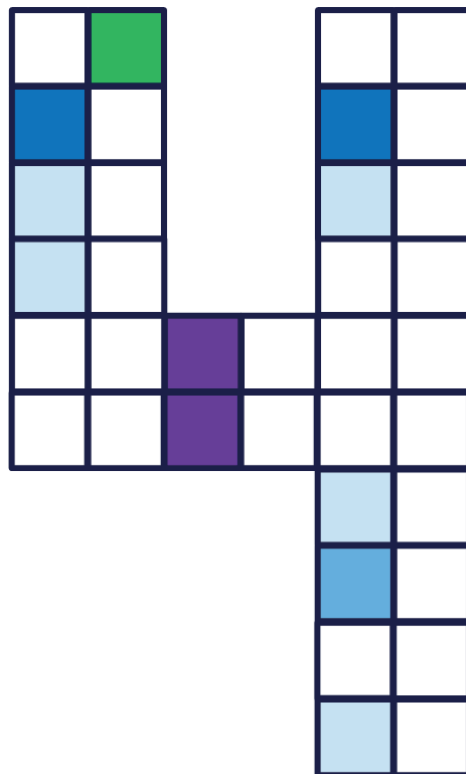
                          threshold = 4)

### Plot QTL profile and effects (Figure 3c and Figure 3 c and d)

plot(barleyObjRes, plotType = "QTLProfileExt")
```

Chapter 4

A one-dimensional mixed model genome scan approach for detecting QTL-by-genetic-background interactions in diallel and nested association mapping designs



Abstract:

Key message To study epistasis in the form of QTL-by-background interactions in multi-parent populations, we present a one-dimensional genome scan by mixed models in which the design matrices are derived from identity-by-descent (IBD) probabilities between parents and offspring.

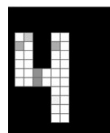
QTLs in multi-parent population (MPP) designs can be modelled as having parent-specific or family-specific allelic effects. QTLs with family-specific allelic effects point to epistatic interactions. We present a one-dimensional genome scan approach by mixed models using design matrices based on IBD probabilities to identify parent-specific and family-specific QTLs. We fit both parent- and family-specific QTLs along a genomic grid of evaluation points and use an information criterion to decide on the nature of significant QTLs. This combined QTL detection procedure for parent- and family-specific jointly is more powerful than an existing approach testing for parent-specific QTLs only. We demonstrate the properties and application of our approach on simulated and real data.

1. Introduction

Epistasis is the phenomenon that allelic effects at two or more genes or quantitative trait loci (QTLs) interact. The evidence in favor of the importance of epistasis is mixed. Some studies (Crow 2010; Bloom et al. 2013) show that additive genetic effects are sufficient to explain phenotypic variation in traits of interest, and other studies (Brem et al. 2005; Bloom et al. 2015) report that pairs of interacting QTLs contribute to the variation in the target trait. Epistasis can involve more than two QTLs in complex genetic architectures (Taylor and Ehrenreich 2014, 2015). The high-order interaction between multiple QTLs occurs in regulatory systems where a central QTL interacts with multiple loci (Pettersson et al. 2011; Forsberg et al. 2017) or multiple QTLs interact in a network (Rowe et al. 2008). Models for epistasis will help to dissect the genetic architecture underlying complex traits, enhance the understanding of genotype-to-phenotype relations, and may be beneficial to genomic prediction (Forsberg et al. 2017).

Detecting epistasis is statistically challenging. A first complication in testing for epistasis involves the choice of a multiple testing correction because of the large number of tests that is implied in screens for epistatic interactions (Cordell 2009; Sham and Purcell 2014,). A large number of tests is required for two-dimensional screens on epistasis, or digenic epistasis (Carlborg et al. 2003, 2004; Li et al. 2008) and even more tests are required for higher-order epistasis. Computational difficulties will arise and the installment of adjusted testing thresholds (Aflakparast et al. 2014; Sham and Purcell 2014) will hinder the detection of epistatic QTLs. A second complication arises in the requirement of large sample sizes for studying epistasis directly (Bloom et al. 2013, 2015), since low frequencies for particular multi-locus genotypes will decrease detection power for epistasis (Carlborg and Haley 2004). As an alternative, one can use a stepwise approach and first try to detect QTLs with additive effects and later look at interactions between the earlier detected additive QTLs (He and Zhang 2011; Stratz et al. 2013). Although this two-step process is computationally efficient and successful in detecting QTLs with large additive effects, epistatic QTLs with weak additive effects will pass undetected in the first step, and then there is no chance to evaluate their epistasis with other QTLs (Fijneman et al. 1996; Holland et al. 1997).

Multi-parent populations (MPPs) have become popular for QTL studies, owing to the inclusion of broad genetic diversity while controlling population structure. Examples of such MPP designs are diallel populations (Giraud et al. 2017), Nested Association Mapping (NAM) populations (Yu et al. 2008), and Multi-parent Advanced Generation InterCross (MAGIC) populations (Huang et al. 2015). MPP designs offer new opportunities to study epistasis. Recombinations between multiple parents increase the chances of observing epistatic QTLs, and a controlled population structure balances frequencies of multi-locus genotypes. Various studies have considered the exploitation of epistasis in the context of MPP designs. For instance, a mouse eight-way cross capturing wide genetic diversity was believed to present abundant recombination events and sufficient population size for high-resolution mapping of epistasis (Churchill et al. 2004). A maize NAM design containing 25 families was characterized by tremendous diversity and was expected to harbor epistasis (McMullen et al. 2009). In a simulation study of a diallel design, the general combining ability (additive effects) and specific combining ability (non-additive effects, including



epistasis) were evaluated (Verhoeven et al., 2006). Still, few studies have investigated the development of epistatic mapping methods in MPPs (Wang et al. 2020).

An important property of MPP designs, like NAM and diallel populations, is that locus-by-locus epistatic interactions can express themselves as QTL-by-genetic-background interactions, where the genetic background refers to different families or offspring populations with a common parent. In the MPP context, QTL-by-background interaction is QTL-by-population interaction or QTL-by-family interaction. What we mean here is that the presence of QTL-by-QTL interactions can be inferred from the inconsistency of QTL allele substitution effects for an individual parent across the families in which the parent is involved. Therefore, epistasis can be detected by 1-dimensional (1D) genome scans on QTL-by-family or QTL-by-background interactions. In these 1D scans, QTLs are identified as interacting with the genetic background because of family-specific QTL expression. As an early example of this approach, Jannink and Jansen (2001) simulated a diallel design to study epistasis via QTL-by-genetic-background interactions. Likewise, other authors (Blanc et al., 2006; Coles et al., 2010; Han et al., 2016) investigated the consistency of parental allelic effects across families versus the alleles being family-specific in maize diallel designs. In contrast to our work here, these studies tackled only specific diallel designs with a limited number of parents, and no attention was given to NAMs and more general MPP designs for mapping multi-allelic QTLs.

In this paper, we develop an approach based on a 1D QTL-by-background scan to detect epistatic interactions in MPPs. Following previous work (Li et al. 2021), we will extend Identity By Descent (IBD-) based mixed models to testing for family-specific QTLs, i.e., QTLs whose effects change in relation to the genetic background, in both simulated and empirical diallel and NAM designs. The mixed model approach estimates genetic variance components for multi-allelic QTLs whose magnitude may depend on the family, i.e., genetic background. The QTL design matrices in the mixed model follow from the application of a general Hidden Markov Model for parental genome reconstruction in offspring individuals in MPPs (Zheng et al. 2015, 2018). We first simulated simple digenic epistatic interactions in standard three-parent diallels and four-parent NAM designs to demonstrate the feasibility of our approach and evaluate its mapping power for epistatic and additive effects. Next, we applied our approach to real data with NAM and diallel MPPs and assessed its performance by looking at the numbers of detected QTLs and the proportions of genetic variance explained by random QTL effects.

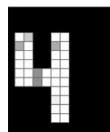
2. Methodology

2.1. Single locus QTL models for the one-dimensional genome-wide scan on QTL-by-background interactions

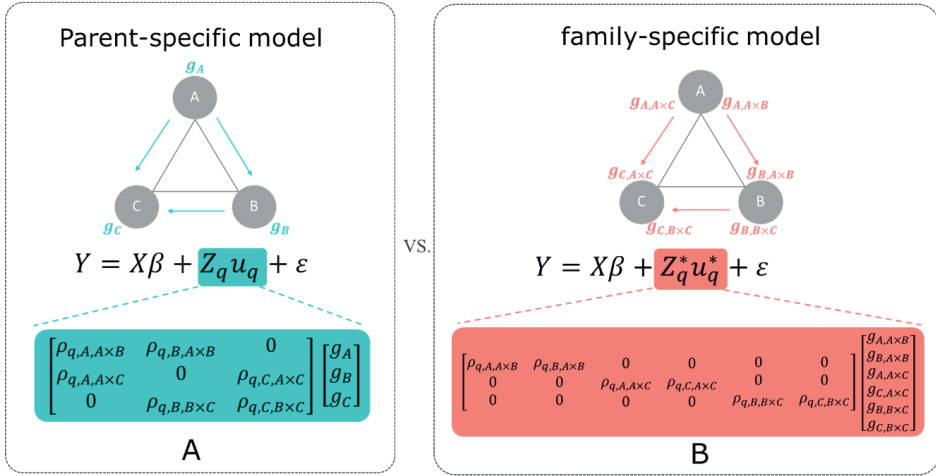
In our 1D approach to epistasis, we fit a mixed model at a regular evaluation grid along the genome, where the marker information in parents and offspring is translated into genetic design matrices whose entries are functions of IBD probabilities between parents and offspring following a Hidden Markov Model framework, called reconstructing ancestry

blocks bit by bit (RABBIT) (Zheng et al. 2015, 2018), using Mathematica software (<https://github.com/chaozhi/RABBIT>). As an alternative to RABBIT, for less complex MPPs, the R package statgenIBD (Boer and van Rossum 2021a) can be used. The position of the parents in the crossing scheme that has led to the MPP determines the IBD probabilities and the entries of the design matrices. Further, the number of columns in a genetic design matrix follows from the number of parents and whether parent specific, i.e., consistent QTL effects are fitted, or cross/family-specific QTL effects, i.e., QTLs whose effects interact with the background.

We will describe mixed models for parent- and family-specific QTLs in diallel and NAM designs. We illustrate for a simple three-parent diallel and a four-parent NAM MPPs the composition of the IBD design matrices. Figure 1 shows the topology of the crossing scheme, QTL effects, and corresponding mixed model for a three-parent full diallel with parent-specific (Figure 1a) and family-specific (Figure 1b) QTL effects. Figure 1c and 1d present the same information for a four-parent NAM design. As can be observed in Figure 1, the main difference between the parent- and family-specific designs is that in the parent-specific design the number of QTL effects is equal to the number of parents, whereas in the family-specific design the number of QTL effects is equal to the number of families multiplying two, because each parent has a unique effect in each family.



simDiallel



simNAM

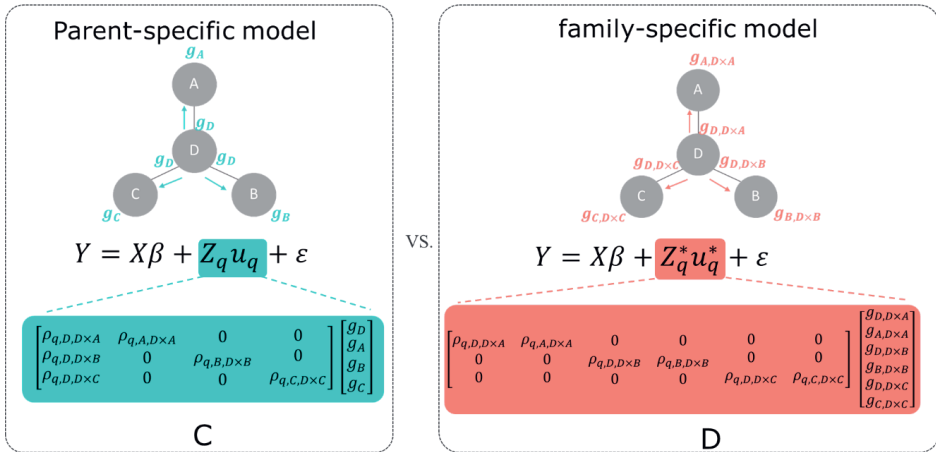


Figure 1. Topology of the crossing scheme and corresponding mixed model for parent-specific and family-specific QTL effects in the examples of a three-parent full diallel and a four-parent NAM MPPs.

2.1.1. Parent-specific QTL model

For an MPP design with K families derived from P parents, and a total of N individuals, the statistical model for a single locus parent-specific QTL can be expressed as:

$$Y = X\beta + Z_q u_q + \varepsilon, \quad \text{Eq.1 (parent-specific model)}$$

$$G_q = \text{var}(u_q) = I_P \sigma_q^2,$$

$$R = \text{var}(\varepsilon) = \bigoplus_f^K I_{n_f} \sigma_{e_f}^2.$$

Y is the $N \times 1$ column vector of phenotypes for all N individuals in K families. X is an incidence matrix of individuals by families, $N \times K$, matrix with elements 1 and 0 indicating whether the i^{th} ($i = 1, 2, \dots, N$) individual belongs to the f^{th} ($f = 1, 2, \dots, K$) family or not, and the $K \times 1$ column vector β represents fixed family means.

Z_q is an offspring by parents, $N \times P$, design matrix at position q in the genome with rows corresponding to families, and columns corresponding to parents (Figure 1a and Figure 1c). The elements of Z_q , $\rho_{q,m,f}$, are vectors that contain the expected number of alleles coming from parent m , for all of the individuals in family f , with values equal to twice the IBD probabilities between parents and offspring, i.e., the expected number of copies for the allele coming from the parent in that column. The $P \times 1$ vector u_q contains the allelic effects for the P parents at this specific putative QTL position, see also (Jannink and Jansen 2001). Note that for obtaining allele substitution effects, differences between parental allelic effects need to be calculated. The variance-covariance matrix for u_q is $G_q = \text{var}(u_q) = I_P \sigma_q^2$, where σ_q^2 is the genetic variance at the putative QTL, which is homogeneous across the families.

The residual term ε follows a family-specific variance-covariance (VCOV) structure as $\text{var}(\varepsilon) = \bigoplus_f^K I_{n_f} \sigma_{e_f}^2$ in which I_{n_f} is the identity matrix for the f^{th} family whose family size is n_f ($\sum_{f=1}^K n_f = N$), and $\sigma_{e_f}^2$ is the residual variance for that family.

2.1.2. Family-specific QTL model

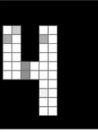
For the family-specific QTL model, the QTL allelic substitution effects are assumed to depend on the family and contain QTL by genetic-background interactions, a form of epistasis. For a single QTL at position q with family dependent QTL effects the model is:

$$Y = X\beta + Z_q^* u_q^* + \varepsilon \quad \text{Eq. 2 (family-specific model)}$$

$$G_q^* = \text{var}(u_q^*) = \bigoplus_{f=1}^K I_{2} \sigma_{q,f}^2$$

$$R = \text{var}(\varepsilon) = \bigoplus_f^K I_{n_f} \sigma_{e_f}^2.$$

$Z_q^* = \bigoplus_{f=1}^K [\rho_{q,m,f} \quad \rho_{q,n,f}]$ is a block diagonal design matrix based on IBD probabilities between parents and offspring. The column vectors $\rho_{q,m,f}$ and $\rho_{q,n,f}$ contain the expected number of alleles coming from parent m and n , respectively, for all of the individuals in family f . The $2K \times 1$ column vector u_q^* contains the QTL allelic substitution effects that are



family dependent and have a family dependent variance $\sigma_{q,f}^2$, as can be seen from the definition of $\text{var}(\mathbf{u}_q^*)$.

2.2. QTL detection procedure

As a first step in QTL detection with the above described models, we calculate functions of IBD probabilities that are used as genetic predictors in the design matrices of the QTL models. Inputs to these calculations are marker and pedigree information, together with the crossing scheme of the MPP. These calculations can be performed for ‘regular’ families, like diallel and NAM, by the R package *stagenIBD* (Boer and van Rossum 2021a). We chose to estimate the IBD probabilities at a grid of 1cM along the genome, which proved to deliver a good compromise between computational effort and resolution for later QTL mapping.

The genome-wide QTL mapping process comprises three steps shown in the flowchart of Figure 2. In the first step, we test for variance component estimates for parent- and family-specific QTL effects using Residual Maximum Likelihood (REML) as implemented in the package *ASReml-R* (Butler et al. 2018) with R version 4.1. Variance components are tested by log-likelihood ratio tests assuming a $0.5\chi_0^2 + 0.5\chi_1^2$ mixture distribution for the test statistic (Patterson and Thompson 1971; Self and Liang 1987). As a correction for multiple testing, we use a significance threshold that is determined by a Bonferroni correction at a genome-wide level of 0.05, where the correction factor is based on the number of evaluation positions. Effectively, the correction factor is then equal to the map length in cM. This turned out to give a decent protection against type I errors in prior simulations that we performed.

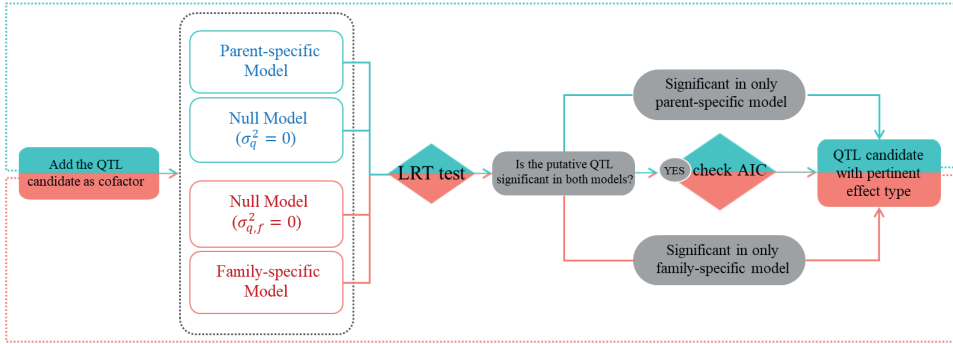


Figure 2. QTL detection procedure using cP&F approach implemented in a mixed model framework. QTLs are examined by a log-likelihood ratio test for the significance of a variance component representing allelic variation in a model with parental QTL effects and in another model for family-specific QTL effects. When one of these two tests is significant, the QTL is included in a cofactor set with the pertinent type of QTL effect. When both types of QTL effects are significant, the QTL is added to the cofactor set for the effect with the lowest AIC.

When a QTL is identified by one of our two model classes only, parent-specific or family-specific, the nature of the QTL is clear. When a QTL is found significant by both models, we

choose the model with the smallest Akaike Information Criterion (AIC) value. Next, the identified QTL with its pertinent effect type is added to a set of cofactors for a later genome scan using composite interval mapping. Around cofactors, we set an exclusion window of 20cM within which no test for further QTLs will be performed.

After identification of the first set of candidate QTLs, the procedure above is repeated as in composite interval mapping and iterated until no new QTLs are found anymore, upon which the final set of QTLs is determined together with their nature as parent-specific and family-specific. Detection of one or more family-specific QTLs will point to epistatic interactions being present. If wanted, locus-by-locus epistatic interactions can then be further investigated with models containing at least one of the family-specific QTLs. As an additional characterization of the detected QTLs, we calculated the contributions of the QTLs to the goodness-of-fit or coefficient-of-determination of the QTL model following the approach described by (Piepho 2019).

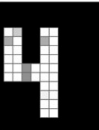
3. Simulated diallel and NAM design

We simulated diallel and NAM designs to examine the effectiveness of our 1D approach to detect family-specific QTLs and with that indications for locus-by-locus epistatic interactions. We first simulated the genome of four inbred *Arabidopsis* parents. This number of parents allows us to simulate simple NAM and diallel designs (Figure 1) with just enough details for insightful simulations while easing the computational burden. We simulated three major QTLs with additive effects and digenic epistasis between two of them, for the sake of easy demonstration and verification. The QTLs were chosen such that interesting frequency distributions for the multi-locus QTL genotypes would arise and that we could assess the performance of our 1D approach to detecting epistatic QTLs. Details of the simulation process are described below.

3.1. Genome simulation

The genome of each parent comprised five chromosomes mimicking the genome of *Arabidopsis*, and the genetic length of each chromosome was 100cM. In total, 500 markers were evenly distributed at 1cM distances along the genome. The (Identity-By-State, IBS) genotypes of the fully inbred and therefore homozygous parents on each marker were randomly determined (binomially, probability 0.5), and coded as being either 1 or 2. Heterozygous genotypes in the progeny were coded as 1|2. Those codes are made to meet the input format required by *statgenIBD* for subsequent IBD calculations. From here onwards, we express genotype (IBS) codes by 1, 2, and 1|2, for MM, mm, and Mm, respectively.

We named the four inbred parents with simulated genomes as parents A, B, C, and D. The *simDiallel* design (Figure 1a and Figure 1b) was generated by crossing parents A, B, and C with each other. For *simNAM* design (Figure 1c and Figure 1d), the parents A, B, and C were crossed as peripheral parents with the central parent D. Three F8 families with 100 offspring individuals were generated for the *simDiallel* and *simNAM* designs. We tested in preliminary



runs for our simulations different numbers of family sizes, and found a family size of 100, giving 300 offspring individuals together for the three families in *simDiallel* and *simNAM*, to provide a realistic scenario with appropriate power for revealing epistatic QTLs and discriminating between simulation scenarios. For family sizes below 50, it was hard to detect any family-specific QTLs. For each progeny in the F8 generation, its genome was generated as the result of crossovers from the parental genome. The crossover process was simulated by the function *GeneDropping* in the tool *RABBIT* (Zheng et al. 2015).

3.2. QTL simulation

Three markers ('M57', 'M266', and 'M440'), on chromosome 1 (56cM), 3 (65cM), and 5 (39cM), were selected as QTLs (*simQTLs*) to form various segregation configurations. Table 1 lists parental genotypes at the three *simQTLs* and whether these *simQTLs* segregate inside the bi-parental families. Each bi-parental family for both *simDiallel* and *simNAM* designs had one non-segregating *simQTL* to form a genetic background for the other two segregating *simQTLs*.

The genetic effect was defined as a bi-allelic effect at the three *simQTLs*, among which *simQTL1* and *simQTL3* interacted. The phenotype of the i^{th} progeny was simulated based on the genetic model:

$$y_i = \sum_{q=1}^3 z_{iq} a_q + z_{i1} z_{i3} \delta_{1,3} + e_i,$$

where a_q is the additive effect of the q^{th} QTL and $\delta_{1,3}$ is the additive-by-additive interaction effect between *simQTL1* and *simQTL3*. z_{iq} is the genotype indicator of the q^{th} QTL in the i^{th} individual. $z_{iq} = 1$ and -1 for respectively the *MM* and *mm* genotypes, and $z_{iq} = 0$ for the heterozygote genotype, *Mm*.

For the magnitude of the additive effects a_q , we simulated two sets, one without additive values, $a_1 = a_2 = a_3 = 0$, and the other with the additive values $a_1 = a_2 = a_3 = 0.3$. The two sets were simulated to explore the impacts of epistatic effects on the mapping results with or without main additive QTL effects and assess the control of false positives without the main and interaction effects. For both the diallel (*simDiallel*) and the NAM population (*simNAM*), we ran simulations with (0.3) and without (0.0) QTL additive effects combined with epistatic effects of size 0.0, 0.1, 0.2, and 0.3 between *simQTL1* and *simQTL3*. The residual error e_i followed a normal distribution with a mean of 0 and a variance of 1.

For each of the 16 simulation scenarios (Table 1), we simulated 500 replicates to test our 1D approach for parent-specific and family-specific QTL effects. We calculated the expected genetic variances explained by *simQTLs* and their heritabilities (Table 1 and Supplementary Material A). Expected genetic heritabilities for the *simQTLs* ranged between 0.00 and 0.25 per family.

3.3. Assessment of model and QTL test performance

We compared our 1D approach allowing for both parent- and family-specific QTLs with a 1D scan in which we allow exclusively parent-specific QTLs. This latter scan leads to the

multi-locus QTL model that was described as the *IBD.MQM_F* model in Li et al. 2021. *IBD.MQM_F* refers to an IBD-based multi-QTL model with a family-specific residual structure. From here onwards, for ease of reference, we will call the procedure based on the *IBD.MQM_F* model the rP approach, with r for reference and P for parent-specific effects. The new 1D approach that combines parent-specific and family-specific effects, we will call cP&F.

For the simulated data, we found in preliminary analyses that nothing was gained by performing full genome-wide scans in comparison to performing tests for QTLs at the positions of the three *simQTLs*. Therefore, for the simulations, we report the power of tests for QTLs at those positions, comparing the rP and cP&F approaches, where we chose a simple test level of 0.01. Mapping power was defined as the proportion of times a *simQTL* was found significant over 500 replications. For the simulations without additive and epistatic effects at the position of *simQTL2*, the mapping power should give an estimate of the false positive rate.

As a complement to the detection of family-specific QTLs, we can perform genome scans for detecting QTL-by-QTL interactions after obtaining cP&F results. We investigated a few simulated progeny data sets for which both *simQTL1* and *simQTL3* were identified as family-specific QTLs (family AxB in *simDiallel* and family DxC in *simNAM*). The design matrices in those scans for QTL-by-QTL interactions were constructed by multiplying the genetic predictor for an identified family-specific *simQTL* with each of a full set of genetic predictors along the rest of the genome. Tests were done for epistasis after correction for additive QTL effects. We applied F-tests for fixed *simQTL*-by-marker interactions implementing a genome-wide Bonferroni-corrected test level of 0.05.

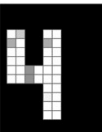


Table 1. Segregation of simulated QTLs and expected heritability for each simQTL per family in simDiallel and simNAM designs. Among the simulated genome of parents A, B, C, and D, three simQTLs were selected to generate varying segregation scenarios. The heritability for each simQTL per family was calculated for a series of interaction effects (0, 0.1, 0.2, and 0.3) in combination with no additive effects (set 1) versus additive effects of 0.3 (set 2).

MPP	Genotypes at simQTLs		segregation or fixed genotype	Expected heritability by simQTL1 2 3 per family given main additive and epistasis effects											
				set1: $a_1 = a_2 = a_3 = 0$						set2: $a_1 = a_2 = a_3 = 0.3$					
	Parent1: simQTL1 2 3	Parent2: simQTL1 2 3		$\delta_{1,3} = 0$	$\delta_{1,3} = 0.1$	$\delta_{1,3} = 0.2$	$\delta_{1,3} = 0.3$	$\delta_{1,3} = 0$	$\delta_{1,3} = 0.1$	$\delta_{1,3} = 0.2$	$\delta_{1,3} = 0.3$				
SimDiallel	A: $M_1M_1 M_2M_2 M_3M_3$	B: $m_1m_1 M_2M_2 m_3m_3$	seg M_2M_2 seg	0 0 0	0.01 0 0.01	0.04 0 0.04	0.08 0 0.08	0.08 0 0.08	0.08 0 0.08	0.11 0 0.11	0.14 0 0.14				
	A: $M_1M_1 M_2M_2 M_3M_3$	C: $m_1m_1 m_2m_2 M_3M_3$	seg seg M_3M_3	0 0 0	0.01 0 0	0.04 0 0	0.08 0 0	0.08 0.08 0	0.13 0.07 0	0.19 0.07 0	0.25 0.06 0				
	B: $m_1m_1 M_2M_2 m_3m_3$	C: $m_1m_1 m_2m_2 m_3m_3$	m_1m_1 seg seg	0 0 0	0 0 0.01	0 0 0.04	0 0 0.08	0 0.08 0.08	0 0.08 0.04	0 0.08 0.01	0 0.08 0				
	D: $M_1M_1 m_2m_2 m_3m_3$	A: $M_1M_1 M_2M_2 M_3M_3$	M_1M_1 seg seg	0 0 0	0 0 0.01	0 0 0.04	0 0 0.08	0 0.08 0.08	0 0.07 0.13	0 0.07 0.19	0 0.06 0.25				
SimNAM	D: $M_1M_1 m_2m_2 m_3m_3$	B: $m_1m_1 M_2M_2 m_3m_3$	seg seg m_3m_3	0 0 0	0.01 0 0	0.04 0 0	0.08 0 0	0.08 0.08 0	0.04 0.08 0	0.01 0.08 0	0 0.08 0				
	D: $M_1M_1 m_2m_2 m_3m_3$	C: $m_1m_1 m_2m_2 m_3m_3$	seg m_2m_2 seg	0 0 0	0.01 0 0.01	0.04 0 0.04	0.08 0 0.08	0.08 0 0.08	0.08 0 0.08	0.11 0 0.11	0.14 0 0.14				

4. Empirical MPP designs

We also applied our 1D genome scan approach, cP&F, to a series of empirical data sets collected on MPP designs. Again, the results were compared to those of a scan for only parent-specific QTLs, rP. Mixed models for QTL detection were fitted at a genome-wide evaluation grid of genetic predictors at 1 cM distances. The two approaches were specifically compared for the number and nature of the detected QTLs and the explained genetic variance by the model. Table 2 summarizes empirical MPP data sets, which included three NAM and three diallel MPPs. The NAM designs differ in family size, with a first NAM having large families and a second having relatively small families. A third NAM is a concatenation of two NAMs with a connecting cross between the two central parents. For the diallels, we looked at a full diallel with three parents, a partial diallel for four parents with four crosses in round robin style and a partial diallel as a concatenation of two full diallels with three parents, and an extra cross connecting the two full diallels.

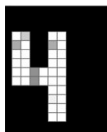
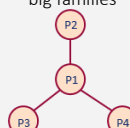

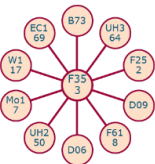



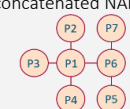

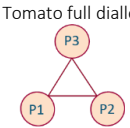


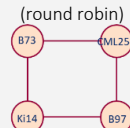

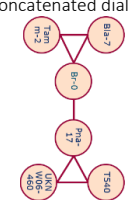








Table 2. Summary of various real MPP designs and QTL mapping results by using reference parent-specific (rP) model and the combined parent-specific and family-specific (cP&F) model.

MPP designs	Family type	Number of progenies	Traits	Number of identified QTLs*	Explained variance	
					rP	cP&F
<p>Cucumber NAM with big families</p> 	F2	963	Overall_4		0.38	0.38
<p>Maize EU-NAM with small families</p> 	DH	841	Dry matter yield (DMY)		0.37	0.41
			Plant height (PH)		0.56	0.61
			Day to silking (DtSILK)		0.70	0.72
<p>Tomato two concatenated NAMs</p> 	F2	718	Infection severity (sore)		0.33	0.33
<p>Tomato full diallel</p> 	F2	248	Area		0.60	0.62
			Proximal angel		0.80	0.80
<p>Mazi partial diallel (round robin)</p> 	RIL	569	Plant height (PHP)		0.51	0.51
<p>Arabidopsis two concatenated diallels</p> 	DH	507	Ear height (EHP)		0.43	0.43
			Total leaf in number (TLNP)		0.41	0.45
			Cauline leaf number in the cold condition (CL_c)		0.36	0.38
			Cauline leaf number in warm condition (CL_w)		0.36	0.32
			Flowering time in the cold condition (FT_c)		0.22	0.51
			Flowering time in the warm condition (FT_w)		0.56	0.60

*Number of identified QTLs is presented in the Venn diagram. The numbers of QTLs detected by the rP model are in the grey circles, and those detected by the cP&F model are in white circles where red highlighted numbers are numbers of family-specific QTLs and blue highlighted numbers are numbers of parent-specific QTLs.

4.1. NAM designs

4.1.1. Cucumber NAM with big families

A central inbred parent was crossed with three other inbred parents by the vegetable seed company Rijk Zwaan. Each of the three F2 families generated more than 300 progenies, summing up to a total of 963 progenies in the whole design, with a consensus map containing 268 markers. The trait of interest was related to the leaf trait that was measured by four breeders, and the average of their measurements called “Overall_4” was the phenotype we used for QTL mapping. The cucumber NAM design has appreciable family sizes (more than 300 per family), which distinguished this population from other empirical MPP designs with smaller family sizes. The larger families allowed us to perform both a 1D scan for parent- and family-specific QTLs across the MPP and two-dimensional (2D) scans for locus-by-locus epistatic interactions within the bi-parental families. For the 2D scan, we used the tool *r/qtl* (Broman et al. 2003), in which the thresholds for additive and interacting QTLs were determined by 1000 permutations at 5% significance levels.

4.1.2. Maize EU-NAM with small families

Our maize NAM design data were taken from the EU-NAM project. There were 10 families with a total of 841 doubled hybrid (DH) progenies from 11 parents among 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 used in previous studies (Giraud et al. 2014; Bustos-Korts et al. 2016; Garin et al. 2017). The traits of interest for QTL mapping were dry matter yield (DMY), plant height (PH) (Garin et al. 2017), and day-to-silking (DtSILK).

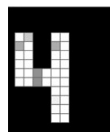
4.1.3. Tomato with two concatenated NAMs

The tomato NAM design was created by the vegetable seed company Rijk Zwaan and is a concatenation of two NAM F2 designs, where the two central parents were crossed with each other as well. The first central parent was crossed with three peripheral parents and the second central parent was crossed with two other peripheral parents. 718 progenies were obtained from the 7 families. The consensus map contained 593 markers for IBD calculation and QTL mapping was done for a trait related to disease resistance expressed as a score with low scores indicating resistance and high scores indicating severe infection.

4.2. Diallel designs

4.2.1. Tomato full diallel

A tomato diallel design was constructed by vegetable seed company Rijk Zwaan from three inbred tomato lines differing in fruit shapes. The three inbred parents were crossed with each other to generate a total of 248 F2 progenies with 459 genotyped markers on the



consensus map. The traits we analyzed were related to fruit shapes like fruit area and proximal angle.

4.2.2. Maize round robin (partial diallel)

In our maize diallel design, the four parents represented two distinct germplasm groups, a temperate group with B73 and B97, and a tropical group with CML254 and Ki14 (Liu et al. 2003). Four families were created (Coles et al. 2010) in a round robin style (B73 × CML254, CML254 × B97, B97 × Ki14, Ki14 × B73). In total, 569 recombinant inbred lines (RILs) were obtained with 1,339 genotyped markers. We chose for analysis the photoperiod response defined as the difference between long- and short-day conditions for plant height (PH), ear height (EH), and total leaf number (TLN).

4.2.3. *Arabidopsis* two concatenated full diallels (partial diallel)

The *Arabidopsis* concatenated partial diallel design was constructed from six *Arabidopsis* accessions, namely Bla-1, br-0, Pna-17, T540, Tamm-2, and UKNW06-460. Founders were selected from the HapMap data set of *Arabidopsis thaliana* (http://bergelson.uchicago.edu/?page_id=790). Three parents Bla-1, br-0, and Tamm-2 were crossed in a full diallel design, and a second full diallel was created from the three parents T540, Pna-17, and UKNW06-460. These two diallel designs were connected by the family Br-0 × Pna-17. F1 hybrids were obtained in each bi-parental family by crossing the two parents, and haploids were propagated to obtain DH lines. In total, 504 DH lines were generated from 7 families. A genetic map with 1,805 SNP markers was constructed by merging several maps using a physical map. A temperature treatment and a vernalization treatment were applied to the DH lines. We analyzed the date of flowering time (FT) and the absolute cauline leaf (CL) number in cold and warm conditions.

5. Results of simulations diallel and NAM designs

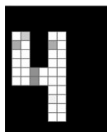
5.1. How digenic epistatic interactions translate into QTLs with family-specific effects

The settings from which we simulated phenotypic data containing digenic interactions to demonstrate how digenic interactions in families of MPPs translate into family-specific QTLs are presented in Table 1, with parental genotypes at the three QTLs, segregation in the families for those QTLs, and expected heritabilities (proportions of explained phenotypic variance) for the three *simQTLs* in each of the families for different scenarios with respect to additive (no additive QTL effects versus QTL effects of 0.3 units) and digenic epistatic effects (0.0, 0.1, 0.2, and 0.3 units). Derivations for expected genetic variances for phenotypes and QTLs are given in Supplementary Material A.

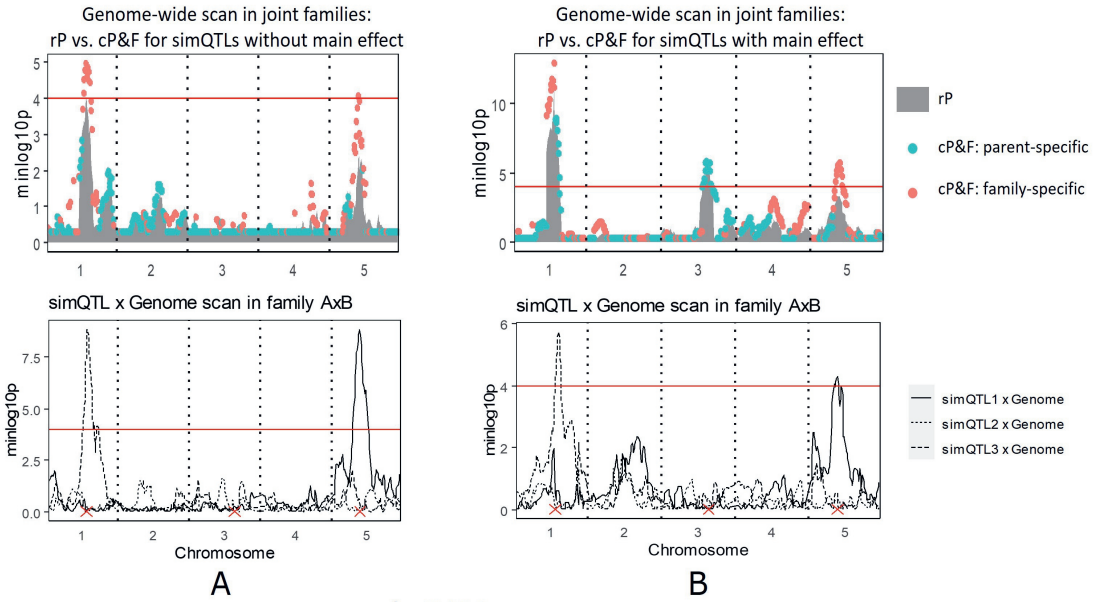
Figure 3 shows the results of genome scans for two scenarios for both the simulated diallel and NAM. The scenarios contain both a digenic epistatic interaction on chromosomes 1 and 3 of size 0.3, and in one case there is no additive QTL effect at any of the three QTLs (Figure 3a and Figure 3c), while in the other case all three QTLs have an effect of 0.3 (Figure 3b and Figure 3d). We use these specific simulations to illustrate our point that digenic locus by locus interactions translate into QTLs with family-specific variances, or, equivalently, to QTL by background interactions.

We now compare mapping profiles for the cP&F approach versus the reference rP approach. For the scenario with epistasis but no additive effect ($a_1 = a_2 = a_3 = 0$) (Figure 3a and Figure 3c), the mapping signals were not strong enough to be picked up by the rP approach that searched only for parent-specific (additive) QTLs. In contrast, the interacting *simQTL1* and *simQTL3* were detected as family-specific QTLs in the cP&F approach. A scan for locus-by-locus interaction confirmed the expected epistatic interactions between *simQTL1* and *simQTL3* in the families *simNAM*: $D \times C$ and *simDiallel*: $A \times B$, where both epistatic *simQTLs* were known to segregate (see Table 1).

In the presence of additive QTL effects of size 0.3 at the three simulated QTLs (Figure 3b and Figure 3d), the reference model rP detected the QTL without epistatic interactions, *simQTL2*, as well as the QTLs whose total variance was increased by the epistatic interaction (*simDiallel*:*simQTL1* and *simNAM*:*simQTL3*). The QTLs whose variance was decreased by the epistatic interactions (*simDiallel*:*simQTL3* and *simNAM*:*simQTL1*) were likely to be missed by the reference model rP, although scans for epistatic interactions pointed to the existence of those QTLs. These QTLs with decreased genetic variance in some families because of epistasis were detected as family-specific QTLs by the cP&F approach, while the purely additive *simQTL2* was correctly identified as parent-specific QTL. The locus-by-locus scan showed no significant signal of digenic interactions with other positions on the genome.



simDiallel



simNAM

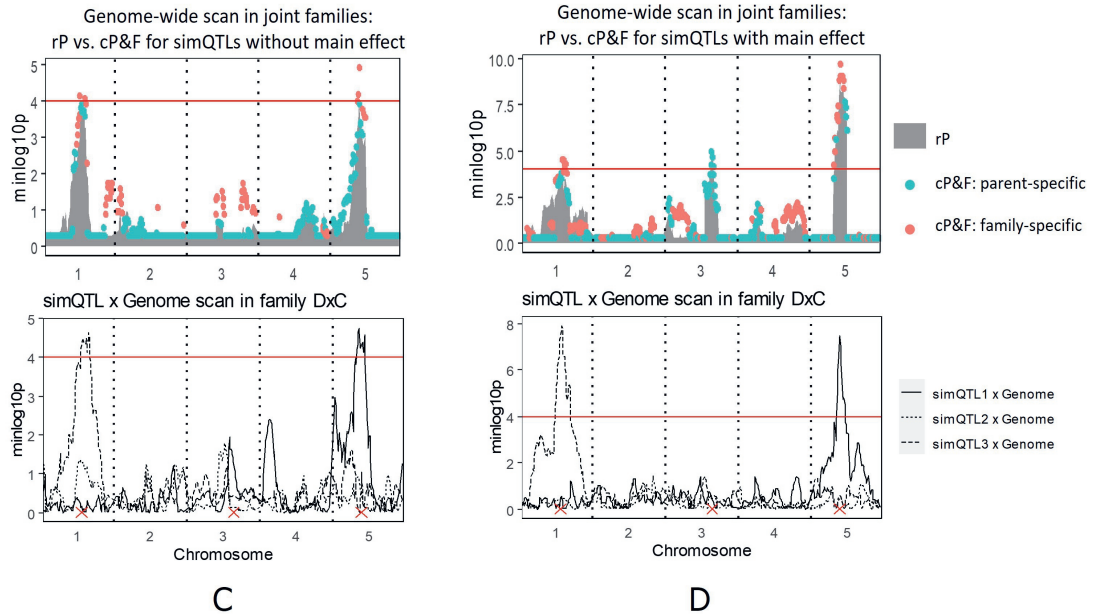


Figure 3. The comparison of the reference parent-specific (rP) model (grey area) and the combined parent-specific and family-specific (cP&F) model (blue and red points) for the genome-wide scan (upper panel) in *simDiallel* (A and B) and *simNAM* (C and D) designs. *simQTL1*-by-*simQTL3* interactions are validated in specific families by *simQTLs*-by-genome scan (lower panel) under the scenarios without (A and C) and with (B and D) main additive effect at all *simQTLs*.

5.2. Model performance assessment

Figure 4 gives the mapping power for cP&F with parent- and family-specific QTLs versus rP with only parent-specific QTLs for a series of epistatic interaction effects between QTLs on chromosomes 1 and 3 ($\delta_{1,3} = 0, 0.1, 0.2, \text{ and } 0.3$) in combination with the absence or presence of additive effects.

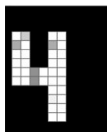
5.2.1. No additive QTL effects

In absence of any additive effects (Figure 4a and Figure 4c), for both *simDiallel* and *simNAM* designs, the non-epistatic *simQTL2* was hardly detected, the (false positive) detection rate being around 0.01, in correspondence with the test level of 0.01. The epistatically interacting QTLs, *simQTL1* and *simQTL3*, were hardly detected by the reference method, rP, without family-specific QTLs, whereas the proposed cP&F approach improved the mapping power for *simQTL1* and *simQTL3* to around 0.50 for the epistatic effects of size 0.3 in both *simDiallel* and *simNAM* designs.

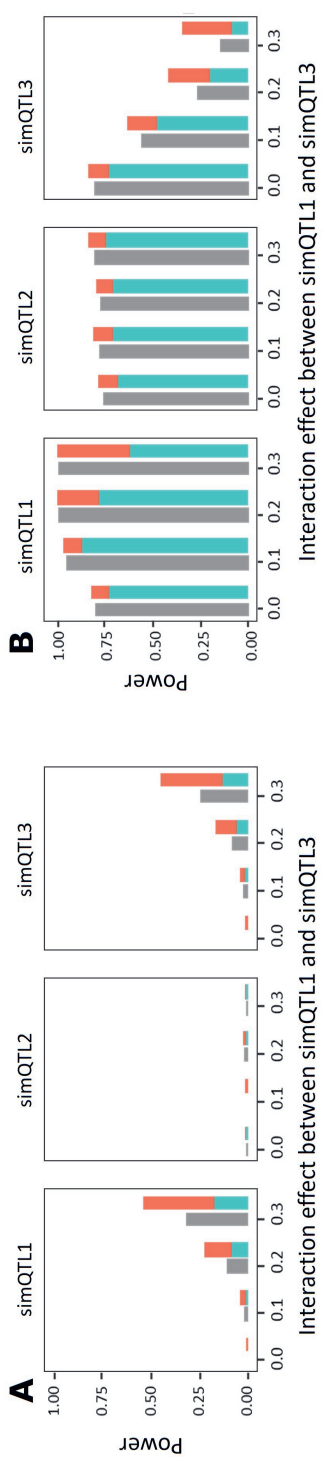
5.2.2. With additive QTL effects

When additive effects of 0.3 were assigned to the three *simQTLs* (Figure 4b and Figure 4d), combined with increasing epistatic effects between *simQTL1* and *simQTL3*, the mapping power for *simDiallel:simQTL1* and *simNAM:simQTL3* climbed to as high as 1 (rounded off) for both the rP and cP&F approach. With the epistatic effect increasing from 0.0 to 0.3, the cP&F approach improved the detection of *simDiallel:simQTL1* and *simNAM:simQTL3*, both being family-specific QTLs.

For *simDiallel:simQTL3* and *simNAM:simQTL1*, their mapping power dramatically declined using the rP approach. The cP&F approach gained power in comparison to rP for *simDiallel:simQTL3* and *simNAM:simQTL1*, both increasingly being identified as family-specific QTLs with increasing epistatic effect. Still, at these QTLs, power decreased with increasing epistasis, although less fast for cP&F. For the purely additive *simQTL2*, mapping power did not change with the size of the epistatic effect at the other two QTLs.



simDiallel



simNAM

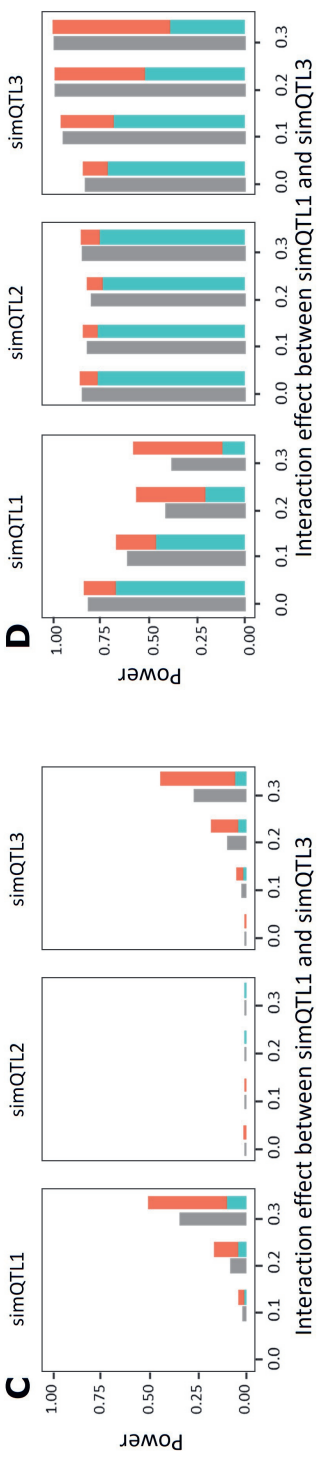


Figure 4. Performance assessment of combined parent-specific and family-specific (cP&F) model vs. reference parent-specific (rP) model regarding the mapping power (percentage) for three simulated QTLs (simQTLs) under varying strength of interaction between simQTL1 and simQTL3 in scenarios without (A and C) and with (B and D) main effect QTLs in simDiallel (A and B) and simNAM (C and D) designs.

6. Results of empirical MPP designs

6.1. NAM designs

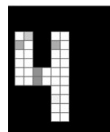
6.1.1. Cucumber NAM with big families

The same four QTLs were detected by both rP and cP&F approaches (Table 2; Supplementary Material B, Figure S1). These four QTLs were defined as parent-specific QTLs by the cP&F approach, so the two models lead to the same fraction of explained variance ($R^2 = 0.38$). Because the cucumber design had large family sizes, we also employed composite interval mapping and two-dimensional testing for each bi-parental family using the tool r/qtl (Broman et al. 2003). The composite interval mapping with the default setting to fit three cofactors detected respectively two, three, and four QTLs in the first, second, and third family, which jointly verifies all the four QTLs detected using the MPP design. No LOD signal for epistasis in two-dimensional testing exceeded the significance threshold derived by the permutation of 1000 replicates, which indicated no evidence for interacting QTLs. The r/qtl results confirmed the findings of the cP&F approach in which we identified four parent-specific QTLs without evidence of epistasis.

6.1.2. Maize EU-NAM with small families

The maize EU-NAM design analysis was a direct extension of the simulated four-parent NAM design to the eleven-parent NAM design (Table 1; Supplementary Material B, Figure S2). Five QTLs for trait DMY were detected by the rP approach with a fraction of explained variance 0.37, whereas 0.41 of the variance is explained by seven QTLs detected by the cP&F approach. All five QTLs detected by rP were detected by cP&F and one of them was identified as a family-specific QTL. In addition, two extra QTLs were detected by cP&F, all as family-specific QTLs. For trait PH, the reference model, rP, detected nine QTLs accounting for the fraction of explained variance of 0.56. The cP&F approach detected 13 QTLs, among which five QTLs were identified as family-specific QTLs, together explaining 0.61 of the variance. For trait DtSIKL, the rP model detected 10 QTLs, explaining a fraction of variance of 0.70. The cP&F approach detected 14 QTLs and seven of them were identified as family-specific QTLs, leading to a total explained fraction of variance of 0.72.

Most QTLs we found for DMY, PH, and DtSILK by the cP&F approach were comparable to earlier results using combined linkage disequilibrium and linkage analysis approaches (Giraud et al. 2014). Our results also coincided with those of the multi-QTL effect (MQE) model, which assumes a combination of bi-allelic, parental, and ancestral levels (Garin et al. 2017). However, these previous analyses rarely reported epistasis for this NAM maize population.



6.1.3. Tomato two concatenated NAMs

We detected QTLs for the phenotype “score”, a measure of disease infection severity (Table 1; Supplementary Material B, Figure S3). Both the cP&F approach and the reference model, rP, detected the same two QTLs on chromosomes 1 and 6 with strong signals. The cP&F model defined these two QTLs as parent-specific, leading to the same fraction of explained variance as in the reference model, 0.33. The parent-specific QTL on chromosome 6 was confirmed by fine mapping by the breeding company and no epistasis was reported.

6.2. Diallel designs

6.2.1. Tomato full diallel

The traits fruit area and proximal angle were correlated traits that produced a common QTL identified on chromosome 2 by both approaches (Table 1; Supplementary Material B, Figure S4). For fruit area, another QTL was detected on chromosome 1 by the rP model ($R^2 = 0.60$) and this QTL was identified as a family-specific QTL by the cP&F approach, with a slightly increased fraction of explained variance of 0.62. For the trait proximal angle, both the rP and the cP&F approach detected a QTL on chromosome 2 as parent-specific and explained a fraction of the genetic variance of 0.80. Small family sizes in this design made that the family-specific QTL on chromosome 1 for fruit area could not be shown to have interactions with other loci.

6.2.2. Maize round robin

In the maize diallel design, we mapped QTLs for the photoperiodic responses (the difference between long-day and short-day conditions) for plant height, ear height, and total leaf number (PHP, EHP, and TLNP) (Table 1; Supplementary Material B, Figure S5). For all PHP and EHP, we detected no family-specific QTLs, both rP and cP&F approaches detected the same parent-specific QTLs, which explained fractions of variance of 0.51 and 0.43 for PHP and EHP, respectively. All of these QTLs could be confirmed from a previous study (Coles et al. 2010) except for an extra QTL that we detected on chromosome 4 for PHP, with a relatively weak signal. No interaction was reported from the work of Coles et al. (2010) in the form of neither digenic nor QTL-by-family interactions, which corresponds with our results that no family-specific QTLs were detected. For trait TLNP, three QTLs were detected by reference model rP (with fractions of explained variance of 0.41) and the cP&F model detected the same three QTLs plus an extra QTL (with fractions of explained variance of 0.45). Two of the four QTLs detected by cP&F were identified as family-specific QTLs.

6.2.3. *Arabidopsis* two concatenated diallels

In the *Arabidopsis* connected diallel design, QTLs were detected for CL and FW traits under cold and warm treatments (Table 1; Supplementary Material B, Figure S6). For trait CL in the cold condition (CL_c), two QTLs were identified by the rP model explaining 0.36 of the variance. The cP&F approach identified one of the rP QTLs as family-specific, together with another family-specific QTL, which explained 0.38 of variance. Under the warm treatment (CL_w), three QTLs were detected by the rP model and two of these were detected by the cP&F approach, among which the first QTL on chromosome 5 was identified as a family-

specific QTL. In this example, cP&F ($R^2 = 0.32$) detected one QTL less than the rP model ($R^2 = 0.36$). For the trait FT under cold (FT_c), the rP model ($R^2 = 0.22$) detected three QTLs on chromosome 5, and two of them might be linked. The distance between the QTLs should be bigger than 10 cM as we set a QTL exclusion window of 10 cM at either side of a cofactor in the CIM procedure. Besides the three family-specific QTLs on chromosome 5, the cP&F approach identified some extra family-specific QTLs on chromosome 2 ($R^2 = 0.51$). Under the warm treatment (FT_w), two linked QTLs were detected by both the reference model ($R^2 = 0.56$) and cP&F approach and there are family-specific QTLs on chromosome 3 with relatively weak signals by cP&F explaining 0.60 of variance.

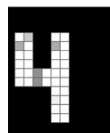
7. Discussion

We simulated digenic epistatic interactions in *simDiallel* and *simNAM* designs to show that such interactions cause family dependent variation in QTL variances and can be detected by 1D genome scans in which we compare models with family-specific QTL variances with those that contain a homogeneous variance across families. Our simulation study further demonstrated that when one of two interacting QTLs is fixed for a particular allele within a family, the additive allelic effects at the other interacting QTL are increased or decreased within the same family. In this scenario, we cannot detect the epistatic signal using a 2D scan in bi-parental families, because such 2D scans require segregation at both epistatic loci, but we can still detect a suggestion for possible epistasis via our test for family-specific QTL effects, which is a reflection of QTL-by-genetic-background interaction.

For digenic epistatic interactions, it is relatively straightforward to calculate expected genetic variances and heritabilities (Supplementary Material A). In practice, many types of epistasis can influence the genetic variance within and across families. For m QTLs, the number of parameters to test including main effects, two-way interactions, to m -way interactions is $2^m - 1$. Consequently, rather than investigating individual epistatic interactions, it is easier to evaluate the combined effects of epistatic QTLs in MPPs by testing for heterogeneity of QTL variances between families.

Two-dimensional genome scans for identifying two-locus epistatic QTLs are typically restricted to bi-parental families (Carlborg et al. 2003; Wei et al. 2010) and require large families. A 1D scan on the heterogeneity of family variance can be successfully followed up by the locus-by-locus interaction test, as shown for some illustrative simulated examples presented in Figure 3. For the empirical data in our MPP examples, we didn't find any locus-by-locus interactions in follow-up scans. The epistatic patterns that caused the heterogeneity of the family variances may have been of higher-order and too complex for our locus-by-locus scans (Papa et al. 2013). Furthermore, the bi-parental families were rather small for investigating epistatic patterns.

To improve the power of our scan for family-specific QTL variances, we used a composite interval mapping approach with cofactors. Although our work builds on earlier studies that estimated QTL effects within and across families (Jannink and Jansen 2001; Blanc et al. 2006; Bardol et al. 2013; Han et al. 2016), a difference is that in these earlier studies cofactors were either homogeneous across families, i.e., parent-specific (Blanc et al. 2006) or nested within families, i.e., family-specific (Jannink and Jansen 2001) when testing for QTL-by-genetic-



background interactions. By allowing both types of effects to occur for the QTLs in our cofactor set, we hoped to reach a more powerful compromise between model parsimony and adequacy. A similar approach combining parent-specific and family-specific cofactors in a fixed linear model context was proposed by Garin et al. (2017) and proved to be beneficial to QTL detection power.

When a large number of parents and families are included in MPPs, the estimation of fixed QTL effects can be unstable (Gatti et al. 2014; Wei and Xu 2016), and family-specific models may require many more parameters than parent-specific models. If the QTL effects are modeled as fixed effects, the F-test method to test the QTL-by-genetic-background QTLs may favor the parent-specific model (what is called the connected model by Blanc et al., 2006) with fewer degrees of freedom. When the genetic terms are modeled as random terms, we can use AIC values to examine the trade-off between the model fit and parsimony to pick up the pertinent type of QTL effect.

To sum up, in the context of MPPs, such as diallel and NAM designs, we can generally detect any form of inter-locus epistasis as a form of QTL-by-genetic-background interaction. Based on this idea, we developed a 1D IBD-based mixed model approach — cP&F — to estimate QTLs with effects that are either family-specific or parent-specific. Our approach improved QTL mapping power in simulated data sets and increased the number of identified QTLs and the proportion of explained genetic variance in empirical MPP designs. The IBD information underlying QTLs with family-specific effects can be used in subsequent 1D or 2D genome scans to identify the second QTL in two-locus interactions.

Supplementary material

A. Genetic variance calculation in simulated MPPs considering epistasis

The genetic effect contributed by the main additive effect of the three *simQTLs* and the epistasis between *simQTL1* and *simQTL3* per family is expressed as: $Q_{family} = z_1 a_1 + z_2 a_2 + z_3 a_3 + z_1 z_3 \delta_{1,3}$. In each family of *simDiallel* and *simNAM* designs, we assume the three *simQTLs* from different chromosomes are independent, i.e., $cov(z_2, z_3) = cov(z_1, z_3) = cov(z_2, z_3) = cov(z_2, z_1 z_3) = 0$, so we can deduce the unified formula of genetic variance in each family as:

$$var(Q_{family}) = a_1^2 var(z_1) + a_2^2 var(z_2) + a_3^2 var(z_3) + \delta_{1,3}^2 var(z_1 z_3) + 2a_1 \delta_{1,3} cov(z_1, z_1 z_3) + 2a_3 \delta_{1,3} cov(z_3, z_1 z_3).$$

The variance and covariance of those genotype indicators z_1 , z_2 , and z_3 is determined by their genetic frequencies f_{z_x} or f_{z_y} (x or $y = 1, 2, 3$ for the three *simQTLs*) in each family as:

$$\begin{aligned} var(z_x) &= E(z_x^2) - \{E(z_x)\}^2 \\ &= 1 - (f_{z_x=1} - f_{z_x=-1})^2 \end{aligned}$$

and

$$\begin{aligned} cov(z_x, z_x z_y) &= E(z_x^2 z_y) - E(z_x) E(z_x z_y) \\ &= E(z_y) - E(z_x) E(z_x z_y) \\ &= f_{z_y=1} - f_{z_y=-1} - (f_{z_x=1} - f_{z_x=-1}) (f_{z_x z_y=1} - f_{z_x z_y=-1}). \end{aligned}$$

Without considering residual heterozygotes, expected genotype frequencies are listed in the table:

MPP	Genotypes at simQTLs		Segregation	Expected genotype frequencies					
	Parent1: simQTL1 2 3	Parent2: simQTL1 2 3		f_z		f_z		f_z	
				z_1 = 1	z_1 = -1	z_2 = 1	z_2 = -1	z_3 = 1	z_3 = -1
simDiallel	A: $M_1 M_1 M_2 M_2 M_3 M_3$	B: $m_1 m_1 M_2 M_2 m_3 m_3$	seg $M_2 M_2$ seg	1/2	1/2	1	0	1/2	1/2
	A: $M_1 M_1 M_2 M_2 M_3 M_3$	C: $m_1 m_1 m_2 m_2 M_3 M_3$	seg seg $M_3 M_3$	1/2	1/2	1/2	1/2	1	0
	B: $m_1 m_1 M_2 M_2 m_3 m_3$	C: $m_1 m_1 m_2 m_2 M_3 M_3$	$m_1 m_1$ seg seg	0	1	1/2	1/2	1/2	1/2
simNAM	D: $M_1 M_1 m_2 m_2 m_3 m_3$	A: $M_1 M_1 M_2 M_2 M_3 M_3$	$M_1 M_1$ seg seg	1	0	1/2	1/2	1/2	1/2
	D: $M_1 M_1 m_2 m_2 m_3 m_3$	B: $m_1 m_1 M_2 M_2 m_3 m_3$	seg seg $m_3 m_3$	1/2	1/2	1/2	1/2	0	1
	D: $M_1 M_1 m_2 m_2 m_3 m_3$	C: $m_1 m_1 m_2 m_2 M_3 M_3$	seg $m_2 m_2$ seg	1/2	1/2	0	1	1/2	1/2

Therefore, in *simDiallel*, we obtain the formula for the calculation of expected genetic variance per family as:

$$\begin{aligned} \text{var}(Q_{A \times B}) &= a_1^2 + a_3^2 + \delta_{1,3}^2 \\ \text{var}(Q_{A \times C}) &= a_1^2 + a_2^2 + \delta_{1,3}^2 + 2a_1\delta_{1,3} = (a_1 + \delta_{1,3})^2 + a_2^2 \\ \text{var}(Q_{B \times C}) &= a_2^2 + a_3^2 + \delta_{1,3}^2 - 2a_3\delta_{1,3} = a_2^2 + (a_3 - \delta_{1,3})^2. \end{aligned}$$

In *simNAM*, we have:

$$\begin{aligned} \text{var}(Q_{D \times A}) &= a_2^2 + a_3^2 + \delta_{1,3}^2 + 2a_3\delta_{1,3} = a_2^2 + (a_3 + \delta_{1,3})^2 \\ \text{var}(Q_{D \times B}) &= a_1^2 + a_2^2 + \delta_{1,3}^2 - 2a_1\delta_{1,3} = (a_1 - \delta_{1,3})^2 + a_2^2 \\ \text{var}(Q_{D \times C}) &= a_1^2 + a_3^2 + \delta_{1,3}^2. \end{aligned}$$

The above formulas of genetic variance per family demonstrate the genetic component associated with each *simQTL*, and thus we can calculate the expected heritability (h^2) for each *simQTL* per family given the simulated residual error e .

In the *simDiallel* design,

for the family $A \times B$:

$$h_{\text{simQTL1}_{A \times B}}^2 = \frac{a_1^2 + \delta_{1,3}^2}{a_1^2 + a_3^2 + \delta_{1,3}^2 + \text{var}(e)}; h_{\text{simQTL2}_{A \times B}}^2 = 0; h_{\text{simQTL3}_{A \times B}}^2 = \frac{a_3^2 + \delta_{1,3}^2}{a_1^2 + a_3^2 + \delta_{1,3}^2 + \text{var}(e)}.$$

for the family $A \times C$:

$$h_{\text{simQTL1}_{A \times C}}^2 = \frac{(a_1 + \delta_{1,3})^2}{(a_1 + \delta_{1,3})^2 + a_2^2 + \text{var}(e)}; h_{\text{simQTL2}_{A \times C}}^2 = \frac{a_2^2}{(a_1 + \delta_{1,3})^2 + a_2^2 + \text{var}(e)}; h_{\text{simQTL3}_{A \times C}}^2 = 0.$$

for the family $B \times C$:

$$h_{\text{simQTL1}_{B \times C}}^2 = 0; h_{\text{simQTL2}_{B \times C}}^2 = \frac{a_2^2}{a_2^2 + (a_3 - \delta_{1,3})^2 + \text{var}(e)}; h_{\text{simQTL3}_{B \times C}}^2 = \frac{(a_3 - \delta_{1,3})^2}{a_2^2 + (a_3 - \delta_{1,3})^2 + \text{var}(e)}.$$

In the *simNAM* design,

for the family $D \times A$:

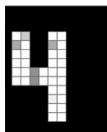
$$h_{\text{simQTL1}_{D \times A}}^2 = 0; h_{\text{simQTL2}_{D \times A}}^2 = \frac{a_2^2}{a_2^2 + (a_3 + \delta_{1,3})^2 + \text{var}(e)}; h_{\text{simQTL3}_{D \times A}}^2 = \frac{(a_3 + \delta_{1,3})^2}{a_2^2 + (a_3 + \delta_{1,3})^2 + \text{var}(e)}.$$

for the family $D \times B$:

$$h_{\text{simQTL1}_{D \times B}}^2 = \frac{(a_1 - \delta_{1,3})^2}{(a_1 - \delta_{1,3})^2 + a_2^2 + \text{var}(e)}; h_{\text{simQTL2}_{D \times B}}^2 = \frac{a_2^2}{(a_1 - \delta_{1,3})^2 + a_2^2 + \text{var}(e)}; h_{\text{simQTL3}_{D \times B}}^2 = 0.$$

for the family $D \times C$:

$$h_{simQTL1D \times C}^2 = \frac{a_1^2 + \delta_{1,}^2}{a_1^2 + a_3^2 + \delta_{1,3}^2 + var(e)}; h_{simQTL2D \times C}^2 = 0; h_{simQTL3D \times C}^2 = \frac{a_3^2 + \delta_{1,}^2}{a_1^2 + a_3^2 + \delta_{1,3}^2 + var(e)}$$



B. QTL profiles of empirical data sets for the cP&F approach and the reference model rP

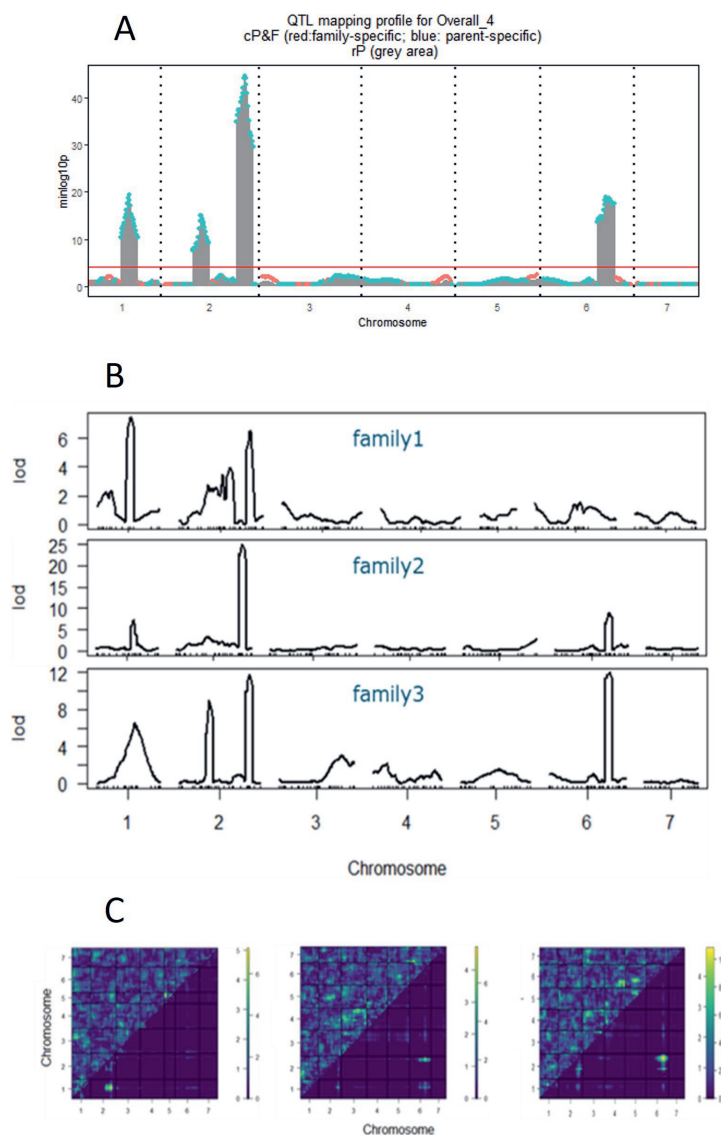


Figure S1. **A.** QTL profiles for the cucumber NAM design using the cP&F approach, compared with the mapping profiles using the reference model rP. **B.** the QTL profiles of composite interval mapping for each bi-parental family by r/qtl. **C.** The heat map of LOD signals from the two-dimensional scan by r/qtl. The LOD signal of testing epistasis is displayed in the upper left triangle; The LOD signal of testing additive effects is displayed in the lower right triangle. In the color scale on the right, numbers to the left and right correspond to LOD signals for interactive and additive effects, respectively.

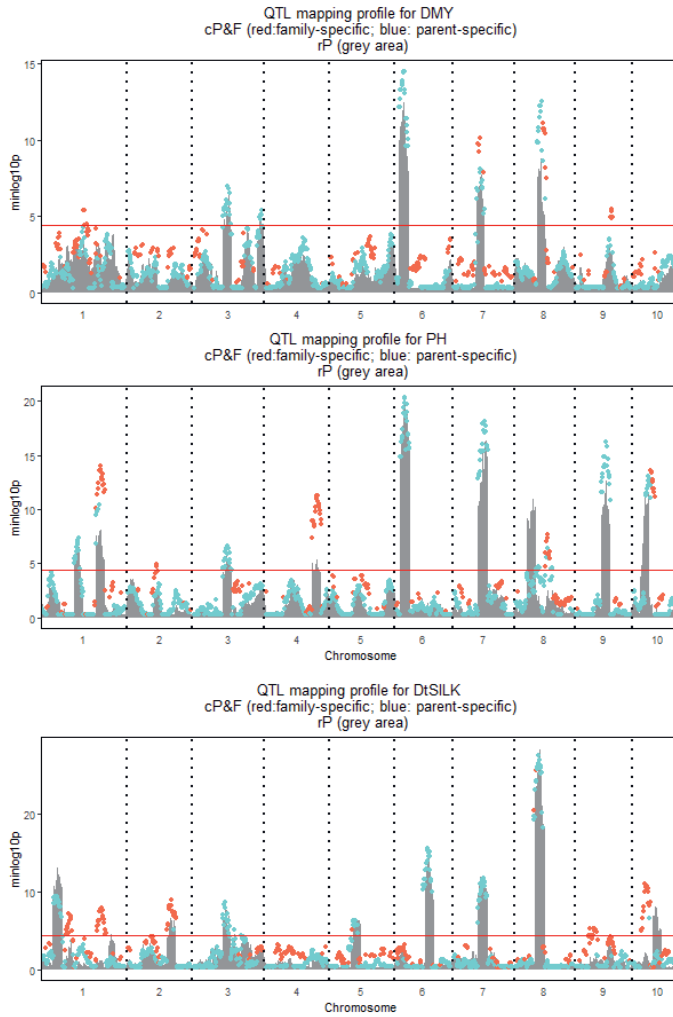


Figure S2. QTL profiles for the maize NAM design using the cP&F approach, compared with the mapping profiles using the reference model rP.

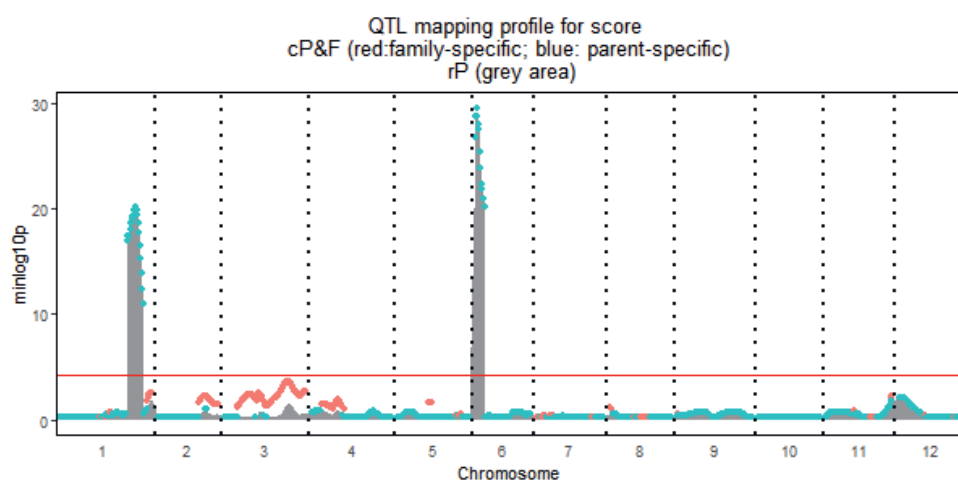


Figure S3. QTL profiles for the tomato NAM design using the cP&F approach, compared with the mapping profiles using the reference model rP.

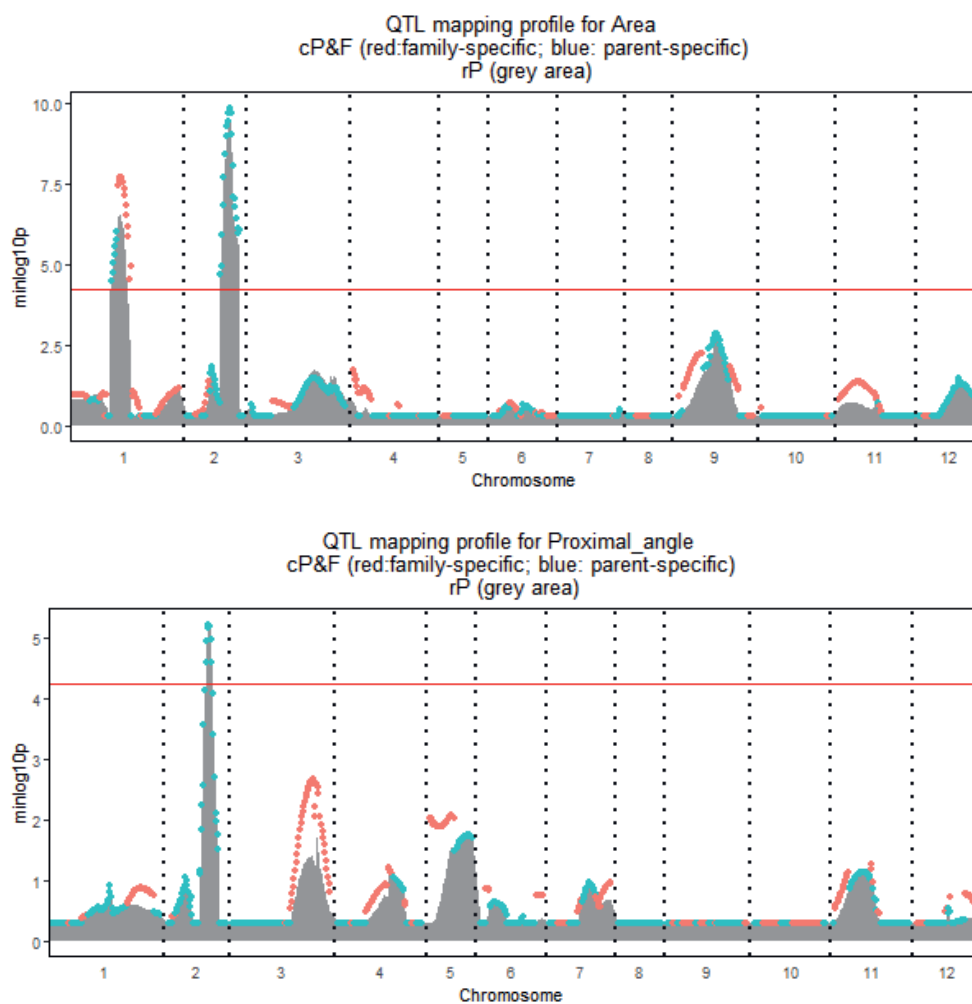


Figure S4. QTL profiles for the tomato diallel using the cP&F approach, compared with the mapping profiles using the reference model rP.

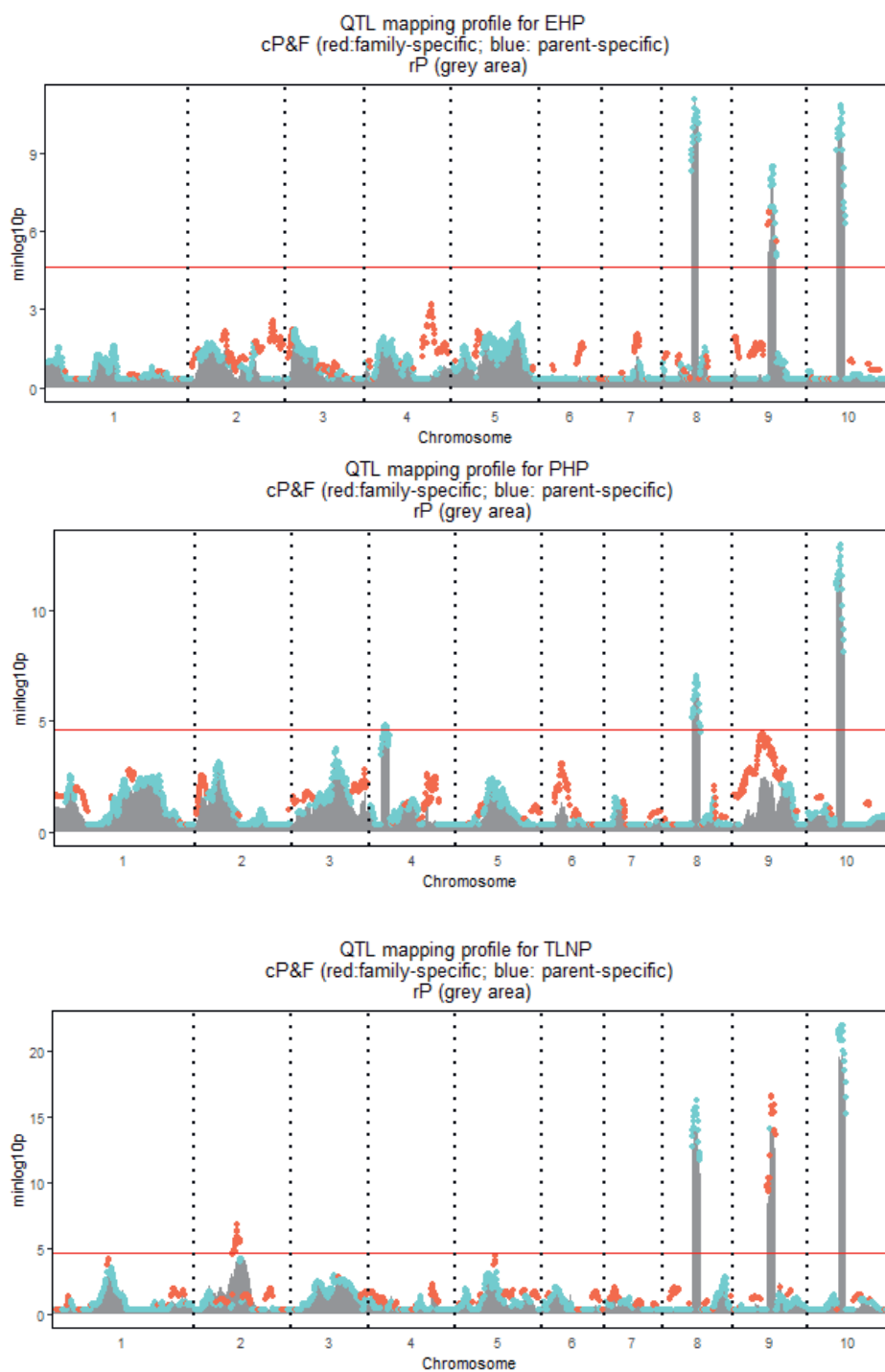


Figure S5. QTL profiles for the maize diallel using the cP&F approach, compared with the mapping profiles using the reference model rP.

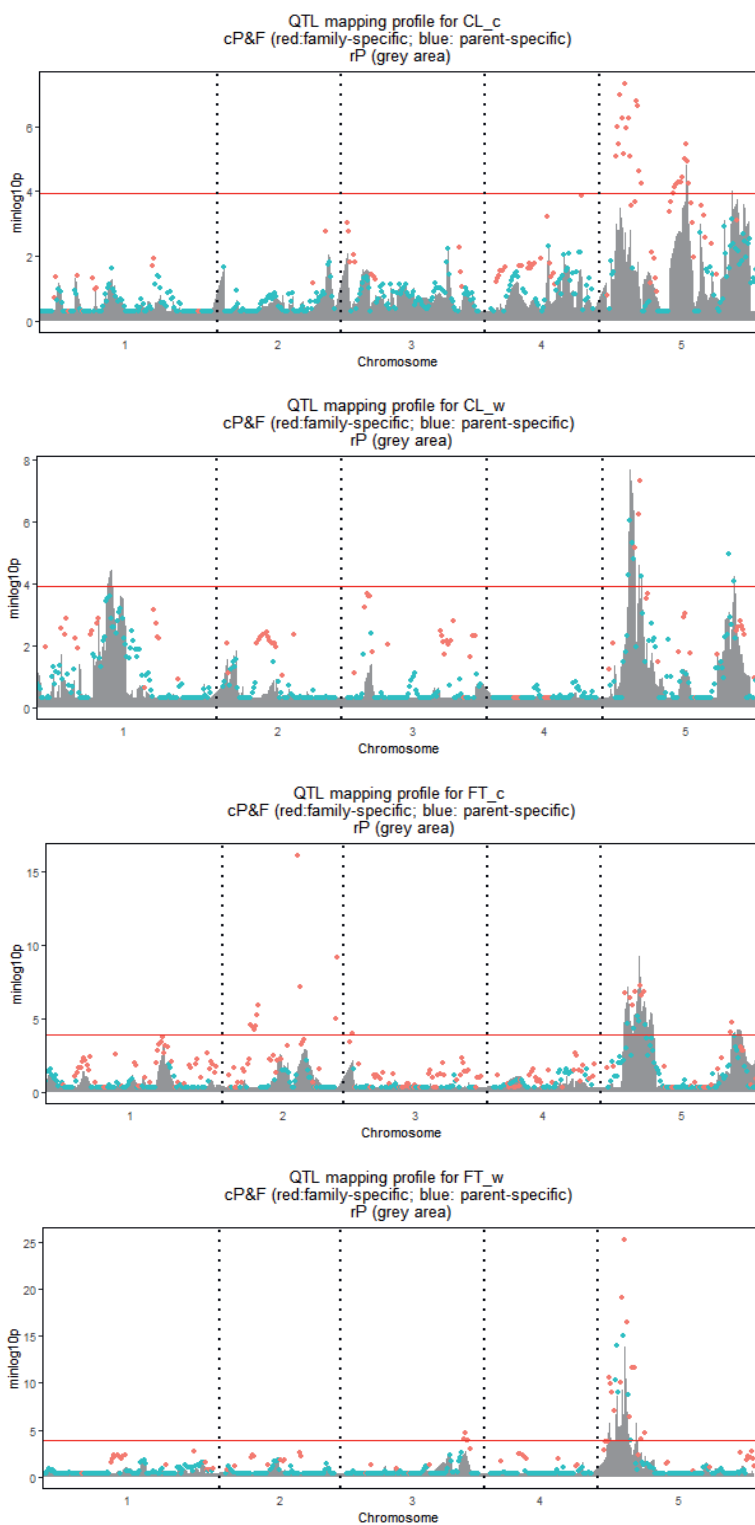
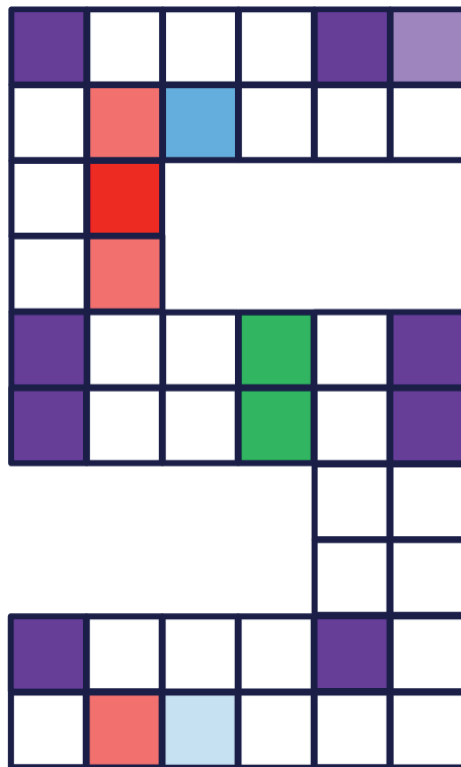


Figure S6. QTL profiles for the *Arabidopsis* connected diallel using the cP&F approach, compared with the mapping profiles using the reference model rP.

Chapter 5

An IBD-based mixed model approach to detection of QTL-by-Environment interactions with multi-parent populations



Abstract:

Key message A mixed model approach is presented for the detection and modeling of QTL-by-Environment Interactions (QEI) for multi-parent populations that are evaluated in multi-environment trials. Four models for QTL effects are presented with varying forms of consistency and dependence in relation to environments and families. An essential part of the approach is the calculation of design matrices for QTL effects that are based on Identity-by-Descent probabilities between parents and offspring.

Multi-parent populations (MPPs) are attractive for genetic and breeding studies because they combine genetic diversity with easy-to-control population structure. Most methods for mapping QTLs in MPPs focus on the detection of QTLs in single environments. Little attention has been given to mapping QTLs in multi-environment trials (METs) and to detecting and modelling QTL-by-environment interactions (QEIs). We present mixed model approaches for the detection and modelling of consistent versus environment dependent QTLs, i.e., QTL-by-Environment Interaction (QEI). QTL effects are assumed to be normally distributed with variances expressing consistency or dependence on environments and families. The entries of the corresponding design matrices are functions of identity-by-descent (IBD) probabilities between parents and offspring and follow from the parental origin of offspring DNA. A polygenic effect is added to the models to account for background genetic variation. We illustrate the wide applicability of our method by analyzing several public MPP data sets with observations from METs. The examples include diallel, nested association mapping (NAM), and multi-parent advanced inter-cross (MAGIC) populations. The results of our approach compare favorably to those of previous studies that used tailored methods.

Keywords: cowpea, diallel, identity-by-descent, MAGIC, maize, MET, multi-environment trials, multi-parent populations, NAM, QTL-by-environment interaction

1. Introduction

Genotype-by-environment interaction (GEI) implies the differential behavior of genotypes across a range of environmental conditions. Broadly adapted genotypes usually perform stably across a wide set of environmental conditions, while narrowly adapted genotypes do well under specific conditions. A visual check on GEI inspects genotypic responses over an environmental gradient, i.e., looks at reaction norms. Nonparallelism of reaction norms in the form of divergence, convergence, and cross-over interaction points to GEI (Crews 2005). Studying GEI is important for the optimal allocation of resources in the design of breeding programs and enhances our understanding of the genetic mechanism underlying diverse phenotypic responses across environments (Van Eeuwijk et al. 2016; Bustos-Korts et al. 2022).

From another perspective, GEI expresses itself in the form of heterogeneity of genetic variances and correlations across environments (Przystalski et al. 2008). Therefore, for analyzing multi-environment trials (MET), linear mixed model approaches (Smith et al. 2005; van Eeuwijk et al. 2010a; Malosetti et al. 2013) are recommended (Malosetti et al. 2004; Boer et al. 2007; Eeuwijk et al. 2007; van Eeuwijk et al. 2010a). The genetic basis of GEI can be resolved by including explicit marker information to model and identify QTL-by-environment interactions (QEI) (Boer et al. 2007; Malosetti et al. 2013).

The favorable properties of MPPs, such as the inclusion of wide genetic diversity in the parents and the easy-to-control population structure in a QTL analysis will benefit the investigation of QEI. Compared with the use of bi-parental populations in METs, the higher genetic diversity of MPPs will increase the chance of displaying polymorphisms at QTLs that interact with environments. In comparison to diversity panels, the known pedigree of MPPs alleviates the problems of population structure and minor frequencies. A few papers have attempted to analyze multi-environment trials for multi-parent populations (MET&MPP) with typically QTL analyses per environment and subsequent comparison of QTL test statistics or $-\log_{10}(p)$ -value profiles. Such an approach was used for a cowpea eight-way MAGIC design by Huynh et al. (2018). Alternatively, in a maize diallel design, QTLs for the photoperiod sensitivity were mapped using the difference between long-day and short-day observations (Coles et al. 2010). In a maize NAM design, QTL effects for yield were assessed between two weakly correlated locations with consideration of the genetic covariance between environments (Garin et al. 2020). In other cases of maize NAM and diallel MET&MPPs (Buckler et al. 2009; Giraud et al. 2014), QTL detection was limited to the mapping of the average response across environments, ignoring QEI.

A generic statistical approach is needed for MET&MPP analysis to investigate QEI. In this study, we propose IBD-based mixed model approaches for studying QEI in various MET&MPPs. In QTL models, random QTL effects are defined with different effect types as stable or unstable across environments as well as families. A background polygenic effect is added that can be structured by family and environment. This approach is applied to several MET&MPP data sets, including the cowpea MAGIC (Huynh et al. 2018), a maize diallel (Coles et al. 2010), and two maize NAM designs (Bauer et al. 2013; Giraud et al. 2014; Garin et al. 2020).



2. Methodology

2.1. General framework

Figure 1 demonstrates the framework of MET&MPP analysis for studying QEI. First, IBD probabilities are computed with pedigree and genome information of parents and progenies. The best linear unbiased estimates (BLUEs) of phenotypes can be obtained by single trial phenotypic analysis with correction for block and spatial effects. Subsequently, a MET&MPP analysis is employed using mixed model approaches to study QEI. An important issue in the building of mixed models for METs is the formulation of the variance-covariance structure between trials. An informal way of studying this structure is via a genotype-by-environment interaction (GGE) biplot analysis. Details of each step are described below and details of the proposed IBD-based MET&MPP analysis using mixed models for QEI study are elaborated in the section “Mixed models for IBD-based QTL analysis of METs for MPPs”.

Entries of design matrices, i.e., genetic predictors, to test for QTL effects and QEI are derived from marker information and have the form of expected numbers of alleles originating from individual partners. These expected numbers of alleles are therefore simple functions of IBD probabilities. IBD probabilities are calculated by the R package *statgenIBD* (Boer and van Rossum 2021a) for a wide range of MPP designs and by a Mathematica tool called *RABBIT* for any other complicated design (Zheng 2019). Both *RABBIT* and *statgenIBD* adopt the framework of Hidden Markov Models (HMM) and inheritance vectors to calculate IBD probabilities using the pedigree and genome information of parents and offspring (Zheng et al. 2014, 2015; Boer and van Rossum 2021a). In the current paper, we chose to compute IBDs at a grid of 5cM along the genome for the empirical MET&MPP data sets as a compromise between mapping resolution and computation time. Other choices for the grid resolution are possible. Multi-environment phenotypes that act as inputs to our models are typically genotypic means for offspring of families that were measured across multiple trials. The genotypic means are best linear unbiased estimators (BLUEs) calculated per trial, i.e., outputs of single trial analyses that take into account block, row, column, and spatial effects. These BLUEs and corresponding standard errors can be calculated by the R package *statgenSTA* (Rossum et al. 2021b).

To investigate the heterogeneity of genotypic variances and correlations across trials, we performed a GGE biplot analysis (Yan et al. 2000). The GGE biplot is a rank-two principal component fitting to the genotype-by-trial table of BLUEs in which the first axis is closely related to the genotypic main effects, while the second axis shows GEI effects. The R package *statgenGxE* (Rossum et al. 2021a) was used to create the GGE biplots. These plots can be helpful to investigate the contributions of individual QTLs to the genetic correlations between trials.

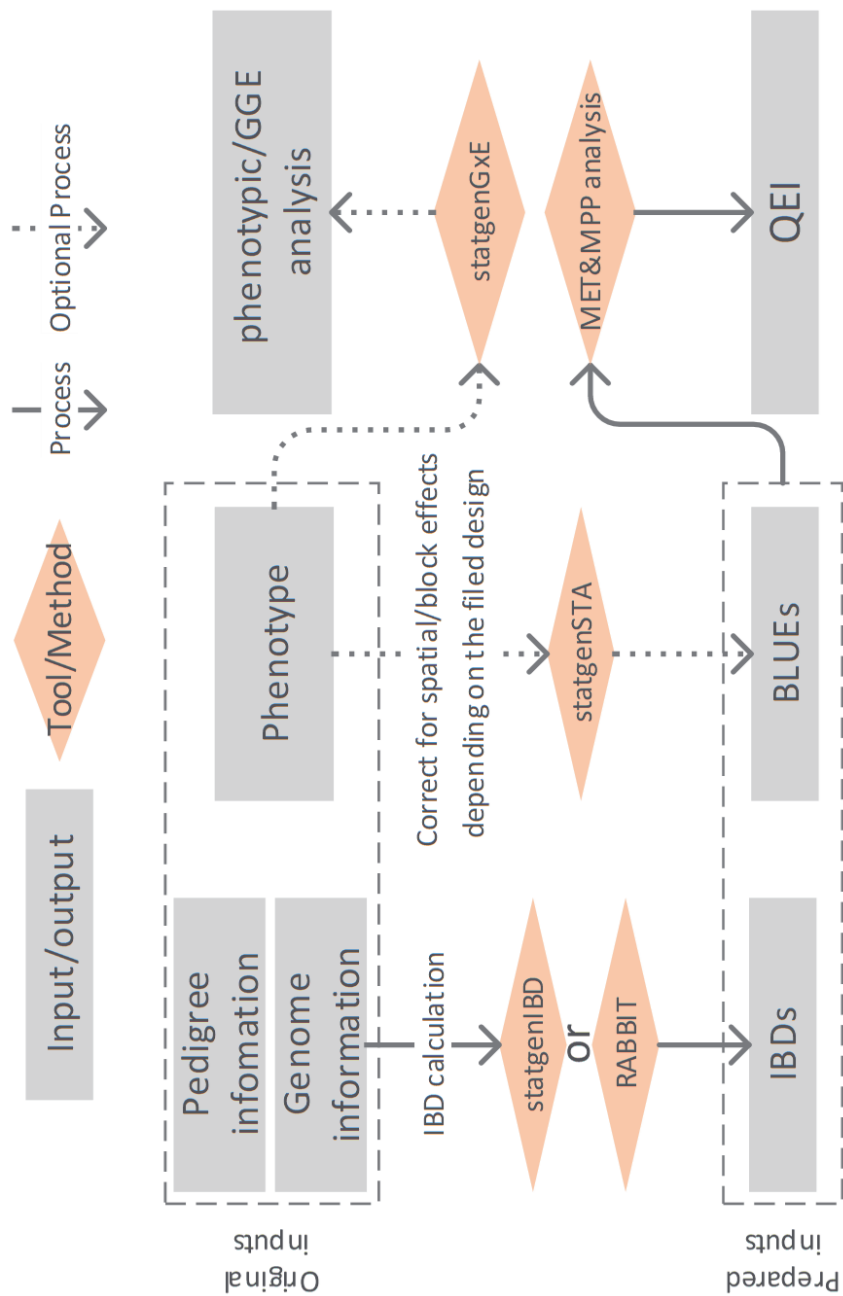


Figure 1. The framework of IBD-based MET&MPP analysis for QEI study.

2.2. Mixed models for IBD-based QTL analysis of METs for MPPs

2.2.1. Models for QTL effects

Four QTL models are constructed differing in the effect types at a putative QTL in terms of stability across families and environments: consistent across both environments and families (environment-consistent and family-consistent, EC&FC QTL), environment-specific and family-consistent (ES&FC QTL), environment-consistent and family-specific (EC&FS QTL), or environment-specific and family-specific (ES&FS QTL). For MAGIC populations, we consider the consistency and specificity of QTL effects only in relation to the environments.

We elaborate on these four QTL models in the case of a MET&MPP with C crosses (or families) derived from P parents across J environments or trials. n_{cj} denotes the number of observations (genotypic BLUEs) for the c -th family in the j -th environment; the total number of observations from the c -th family across all J environments is $\sum_j n_{cj} = n_{c.}$; the total number of observations in the j -th environment across all C families is $\sum_c n_{cj} = n_{.j}$; so the total number of observations in the MET&MPP is $\sum_j \sum_c n_{cj} = \sum_c n_{c.} = \sum_j n_{.j} = N$. In all model equations, we present later, **matrices** are presented in bold font and random terms are underlined. The general mixed model to map a single QTL can be expressed as:

$$\underline{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_q \underline{\mathbf{u}}_q + \underline{\mathbf{g}} + \underline{\boldsymbol{\varepsilon}}.$$

\underline{Y} is the $N \times 1$ column vector for all N observations in a MET&MPP. The systematic part, $\mathbf{X}\boldsymbol{\beta}$, models effects for families and environments, and their interactions. The structure of the design matrix \mathbf{Z}_q and the vector of QTL effects $\underline{\mathbf{u}}_q$ is determined by the effect type at the QTL. The four QTL models regarding \mathbf{Z}_q and $\underline{\mathbf{u}}_q$ are described with superscripts to indicate the matrix dimensions to help distinguish the structures.

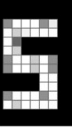
	Environment-consistent (EC)	Environment-specific (ES)
Family-consistent (FC)	$\mathbf{Z}_q^{(N \times P)} = [\boldsymbol{\pi}_{q,1}^{(N \times 1)} \dots \boldsymbol{\pi}_{q,k}^{(N \times 1)} \dots \boldsymbol{\pi}_{q,P}^{(N \times 1)}],$ <p style="text-align: center;">and</p> $\underline{\mathbf{u}}_q^{(P \times 1)} = [\underline{a}_{q,1} \dots \underline{a}_{q,k} \dots \underline{a}_{q,P}]^T$ $\sim MVN(0, \mathbf{I}_P \sigma_q^2)$	$\mathbf{Z}_q^{(N \times PJ)} = \bigoplus_{j=1}^J \mathbf{Z}_{q,j}^{(n_j \times P)},$ <p style="text-align: center;">and</p> $\underline{\mathbf{u}}_q^{(PJ \times 1)} = \left[\left(\underline{\mathbf{u}}_{q,1}^{(P \times 1)} \right)^T \dots \left(\underline{\mathbf{u}}_{q,j}^{(P \times 1)} \right)^T \dots \left(\underline{\mathbf{u}}_{q,J}^{(P \times 1)} \right)^T \right]^T$ $\sim MVN(0, \bigoplus_{j=1}^J \mathbf{I}_P \sigma_{q,j}^2)$
Family-specific (FS)	$\mathbf{Z}_q^{(N \times 2C)} = \bigoplus_{c=1}^C [\boldsymbol{\pi}_{q,P1_c}^{(n_c \times 1)} \boldsymbol{\pi}_{q,P2_c}^{(n_c \times 1)}],$ <p style="text-align: center;">and</p> $\underline{\mathbf{u}}_q^{(2C \times 1)} = [\underline{a}_{q,P1_1} \underline{a}_{q,P2_1} \dots \underline{a}_{q,P1_c} \underline{a}_{q,P2_c} \dots \underline{a}_{q,P1_C} \underline{a}_{q,P2_C}]^T$ $\sim MVN(0, \bigoplus_{c=1}^C \mathbf{I}_{2C} \sigma_{q,c}^2)$	$\mathbf{Z}_q^{(N \times 2CJ)} = \bigoplus_{j=1}^J \bigoplus_{c=1}^C [\boldsymbol{\pi}_{q,P1_c}^{(n_{cj} \times 1)} \boldsymbol{\pi}_{q,P2_c}^{(n_{cj} \times 1)}],$ <p style="text-align: center;">and</p> $\underline{\mathbf{u}}_q^{(2JC \times 1)} = \left[\left(\underline{\mathbf{u}}_{q,1}^{(2C \times 1)} \right)^T \dots \left(\underline{\mathbf{u}}_{q,j}^{(2C \times 1)} \right)^T \dots \left(\underline{\mathbf{u}}_{q,J}^{(2C \times 1)} \right)^T \right]^T$ $\sim MVN(0, \bigoplus_{j=1}^J \bigoplus_{c=1}^C \mathbf{I}_{2C} \sigma_{q,j,c}^2)$

EC&FC QTL

In the EC&FC QTL model, the QTL effect at the q -th position is assumed to be stable across environments and families, which is comparable to the generic IBD-based mixed model approach to map additive QTLs in previous studies (Li et al. 2021, 2022). $\mathbf{Z}_q^{(N \times P)}$ is a $N \times P$ design matrix that can be partitioned into P column vectors. Each of the P column vectors is a $N \times 1$ column vector, $\boldsymbol{\pi}_{q,k}^{(N \times 1)}$, containing expected numbers of allelic copies associated with the k -th ($k = 1, 2, \dots, P$) parent for all N observations. The $P \times 1$ column vector $\underline{\mathbf{u}}_q^{(P \times 1)}$ contains element $\underline{a}_{q,k}$ representing the QTL effect associated with the k -th parent. The variance of $\underline{\mathbf{u}}_q^{(P \times 1)}$ is $\mathbf{I}_P \sigma_q^2$, where \mathbf{I}_P is the P -dimensional identity matrix and σ_q^2 is the genetic variance for the putative QTL, implying a homogeneous VCOV structure for the QTL effect across all environments and families.

ES&FC QTL

In the ES&FC QTL model, the QTL effect is defined as being unstable across J environments due to QEI but consistent across C families. $\mathbf{Z}_q^{(N \times PJ)}$ is a $N \times PJ$ design matrix that can be split into J diagonal components. Each of the J components is a $n_j \times P$ design matrix, $\mathbf{Z}_{q,j}^{(n_j \times P)}$, whose structure is comparable to that defined in $\mathbf{Z}_q^{(N \times P)}$, but $\mathbf{Z}_{q,j}^{(n_j \times P)}$ is designated for the j -th environment with n_j being the number of observations in this environment. The vector of effects $\underline{\mathbf{u}}_q^{(PJ \times 1)}$ is a $PJ \times 1$ column vector that can be partitioned into J column sub vectors. Each of the J sub column vectors is a $P \times 1$ column vector $\underline{\mathbf{u}}_{q,j}^{(P \times 1)}$ that contains the



QTL effects of P parents in the j -th environment with $I_P \sigma_{q,j}^2$ being the environment-specific QTL variance.

EC&FS QTL

In the EC&FS QTL model, parental QTL effects are specified within each of C bi-parental families and therefore vary between families, while being stable across all J environments. $\mathbf{Z}_q^{(N \times 2C)}$ is a $N \times 2C$ design matrix that can be partitioned into C diagonal components. Each of these components contains two $n_c \times 1$ column vectors, $\boldsymbol{\pi}_{q,P1_c}^{(n_c \times 1)}$ and $\boldsymbol{\pi}_{q,P2_c}^{(n_c \times 1)}$, with as elements the expected numbers of allele copies associated with the two parents of that family, $P1_c$ and $P2_c$. The random QTL effects $\mathbf{u}_q^{(2C \times 1)}$ form a $2C \times 1$ column vector associated with the pairs of parents and have variance is $I_2 \sigma_{q,c}^2$ in the c -th family.

ES&FS QTL

In the ES&FS QTL model, the QTL effects are neither stable across environments nor families, but both environment and family specific. The ES&FS QTL model is established by merging the ES&FC and EC&FS QTL models. $\mathbf{Z}_q^{(N \times 2CJ)}$ is a $N \times 2CJ$ design matrix with JC diagonal components. Each of these components possesses a pair of $n_{cj} \times 1$ column vectors, $\boldsymbol{\pi}_{q,P1_c}^{(n_{cj} \times 1)}$ and $\boldsymbol{\pi}_{q,P2_c}^{(n_{cj} \times 1)}$, whose structure resembles that in the EC&FS QTL model but is designated for the j -th environment, where n_{cj} is the number of observations in the c -th family and j -th environment. Remember that $\sum_{j=1}^J \sum_{c=1}^C n_{cj} = N$. The $2JC \times 1$ column vector $\mathbf{u}_q^{(2JC \times 1)}$ can be partitioned into J column vectors with each column vector having the form $\mathbf{u}_{q,j}^{(2C \times 1)}$, comparable to that in the EC&FS QTL model. For the j -th environment, the variance of the QTL effects across the C families can be written as $\oplus_{c=1}^C I_2 \sigma_{q,j,c}^2$. QTL variances are heterogeneous and depend on environment and family simultaneously.

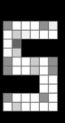
2.2.2. Models for polygenic effect

The QTL models above can be combined in single or multi-QTL models with a structured polygenic effect, \mathbf{g} . We write for the VCOV of \mathbf{g} , $\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_{MET} \otimes \boldsymbol{\Sigma}_{MPP}$, with $\boldsymbol{\Sigma}_{MPP}$ defined by the relations between the families and $\boldsymbol{\Sigma}_{MET}$ by the relations between the environments. For $\boldsymbol{\Sigma}_{MPP}$, we allowed for identity or homogeneity ($\boldsymbol{\Sigma}_{MPP_{id}}$), heterogeneity ($\boldsymbol{\Sigma}_{MPP_{idh}}$), or a population structure as a marker-based kinship matrix ($\boldsymbol{\Sigma}_{MPP_{kin}}$). For $\boldsymbol{\Sigma}_{MET}$, we assumed identity or homogeneity ($\boldsymbol{\Sigma}_{MET_{id}}$), heterogeneity ($\boldsymbol{\Sigma}_{MET_{idh}}$), and an unstructured model ($\boldsymbol{\Sigma}_{MET_{us}}$). The combination of three structure models for $\boldsymbol{\Sigma}_{MPP}$ and three structure models for $\boldsymbol{\Sigma}_{MET}$ generate nine VCOV structures for the polygenic background effect. For each of those nine VCOV models, the number of variances and correlation parameters is given. Because the models are fitted on tables of genotype by environment BLUEs, the ‘genetic’ variances and correlations in the upper two rows also contain a non-genetic contribution stemming from plot error. The VCOV model that we used as a default for genome scans was $\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_{MET_{us}} \otimes \boldsymbol{\Sigma}_{MPP_{idh}}$.

	$\Sigma_{MET_{id}}$	$\Sigma_{MET_{idh}}$	$\Sigma_{MET_{us}}$
$\Sigma_{MPP_{id}}$	1 variance σ_g^2	J variances $\sigma_{g_j}^2$	J variances and J(J-1)/2 correlations $\sigma_{g_j}^2$ and $\sigma_{g_{jj^*}}^2$
$\Sigma_{MPP_{idh}}$	J variances $\sigma_{g_c}^2$	CJ variances $\sigma_{g_{jc}}^2$	CJ variances C(J-1)/2 correlations $\sigma_{g_{jc}}^2$ and $\sigma_{g_{jj^*c}}^2$
$\Sigma_{MPP_{kin}}$	2 variances σ_g^2 and σ_e^2	J+1 variances σ_g^2 and $\sigma_{e_j}^2$	J(J+1)/2+1 parameters σ_g^2 , $\sigma_{e_j}^2$, and $\sigma_{e_{jj^*}}^2$

2.2.3. Model selection and genome-wide QTL scans

To perform genome-wide QTL scans, various strategies can be followed. A common strategy is to first identify the best VCOV model for the polygenic effect by fitting different options for the VCOV in a model without QTLs and use an information criterion like the Akaike information criterion (AIC) (Akaike 1974) or the Bayesian information criterion (BIC) to select the most suitable model. See for example, Boer et al. (2007), Verbyla (2019), and Boer et al. (2020). After the selection of a VCOV for the polygenic effect, genome scans can be performed for the identification of QTLs. Options now are to fit the most complex QTL model (ES&FS) everywhere. Upon detection of a significant QTL, one could try to simplify the QTL model to see whether an EC model can be used in place of an ES model, and similarly one can replace an FS model with an FC model. We did indeed compare various VCOV models for the polygenic effect, but chose in the end for a pragmatic approach by using the following model for the VCOV of the polygenic effect, $\Sigma_g = \Sigma_{MET_{us}} \otimes \Sigma_{MPP_{idh}}$, and then produced four full genome scans, one for each of the four QTL models (EC&FC, ES&FC, EC&FS, and ES&FS). Within each scan, we tested for significance of the QTL with a likelihood ratio test (LRT). The corresponding p -value followed from an approximation of the test-statistic by a mixture of a χ^2 distributions with 0 and 1 degrees of freedom (Self and Liang 1987). We used a conservative Bonferroni-corrected threshold at a genome-wide significance level of 0.05 to test for QTLs. From a first scan, we retained the most significant QTL with the highest $-\log_{10}(p)$ value as a co-factor in a second scan, with a window of 20 cM around the first detected QTL within which no additional QTLs will be fitted. After the second scan, this procedure to identify co-factors will be repeated until no further QTLs are found anymore and the $-\log_{10}(p)$ profile stabilizes. The final model is then a multi-QTL model consisting of all QTLs identified in earlier scans. The multi-QTL models produced by the four scans were then compared and a synthesis model was created from the union of these four models. When a QTL was present in more than one multi-QTL model the nature of the QTL was assessed by comparing the AICs between the models with different formulations for the QTL effects. The AIC was calculated as $AIC = -2\ln(L_{MAX}) + 2\tau$, where

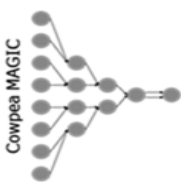
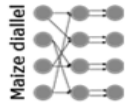
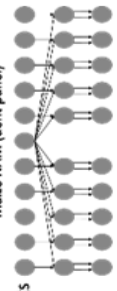
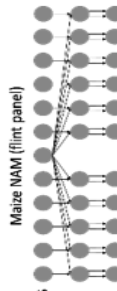


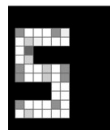
L_{MAX} is the restricted maximum likelihood (REML) and τ is the number of estimated variance parameters. The model with the smallest AIC value was selected to define the nature of the QTL effect.

3. Materials

To test our approach, we collected several empirical MET&MPP data sets (Table 1) including a cowpea MAGIC, a maize diallel, and two maize NAM designs. The cowpea MAGIC and maize diallel designs were screened in managed stress trials, while the two maize NAM designs were screened across several geographic locations in the EU. In this paper, we did not distinguish the two types of trials and use the term ‘environments’ to describe the sets of trial conditions.

Table 1. Summary of empirical MET&MPPs.

MPPs	Pop. size	Traits of interest	METs	Reference
<p>Cowpea MAGIC</p>  <p>Parents 2-way 4-way 8-way ⋮ F8</p>	305	Flowering time (FT)	Long-day and short-day environments combined with full and restricted irrigation conditions	(Huynh et al. 2018)
<p>Maize diallel</p>  <p>Parents F1 ⋮ RIL</p>	569	Growing degree days to silking (GDDTs) and anthesis (GDDTA)	Summer (long-day) and winter (short-day) seasons	(Coles et al. 2010)
<p>Maize NAM (dent panel)</p>  <p>Parents F1 DH</p>	841	Dry matter yield (DMY), plant height (PH), days to silking (DSILK)	Four locations across EU	(Giraud et al. 2014)
<p>Maize NAM (flint panel)</p>  <p>Parents F1 DH</p>	811	Dry matter yield (DMY), plant height (PH), days to silking (DSILK)	Six locations across EU	(Giraud et al. 2014; Garin et al. 2020)



3.1. MAGIC design

The cowpea MAGIC design was derived from eight parents that generated an F8 generation (Huynh et al. 2018). Each genotype was screened under the combination of two day-length conditions (short-day and long-day) with two irrigation conditions (full and restricted irrigation) denoted by fi_l (full irrigation under long-day), fi_s (full irrigation under short-day), ri_l (restricted irrigation under long-day), and ri_s (restricted irrigation under short-day). The trait of interest was the flowering time (FT).

In the previous study (Huynh et al. 2018), QTL mapping was performed for each environment condition by simple interval mapping with founder haplotype probabilities fitted in a linear model using the package R/mpMap (Huang and George 2011).

3.2. Maize diallel design

In the maize diallel design, four bi-parental families were created by crossing each of two temperate inbred parents to each of two photoperiod-sensitive tropical parents (Coles et al. 2010). Recombinant inbred lines (RILs) were screened under long-day (summer seasons) and short-day (winter seasons) conditions for photoperiod-related traits, e.g., days to silking (DTS) and anthesis (DTA), whose measurements were converted to growing degree days (GDDTS and GDDTA) to account for the influence of temperature.

In the previous study (Coles et al. 2010), QTL mapping for photoperiodic responses was performed by calculating the difference in response between short-day and long-day conditions. In addition, QTL mapping for the separate conditions was performed by the tool MCQTL 4.0 (Jourjon et al. 2005).

3.3. Maize NAM designs: dent and flint panels

Two maize NAM designs, a flint and a dent panel, were taken from the maize EU-NAM project (Bauer et al. 2013; Lehermeier et al. 2014; Giraud et al. 2014). In the flint panel, 11 doubled haploid (DH) families were derived from the central parent UH007 crossed with 11 peripheral parents. Multiple traits such as dry matter yield (DMY), days to silking (DtSILK), and plant height (PH) were measured across six locations in the EU, namely Wadersloh (Germany), Ploudaniel (France), La Coruña (Spain), Einbeck (Germany), Roggenstein (Germany), and Eckartseier (Germany). In the dent panel, 10 families were derived from the central parent F353 which was crossed with 10 peripheral parents. The same traits as measured in the flint panel were measured again, but across four locations only, namely Wadersloh (Germany), Mons (Germany), Einbeck (Germany), and Roggenstein (Germany).

In previous work by Giraud (2014), combined linkage and linkage disequilibrium mapping was performed for both NAM panels, while Garin (2020) used the NAM flint panel as an example for a QEI study using only the Roggenstein and La Coruña locations.

4. Results

4.1. VCOV structure for polygenic effects

First, we compared different VCOV structures for the polygenic effect. Table 2 compares the BIC and AIC values of the nine VCOV models. With respect to the cowpea MAGIC design with only one family, there is no difference between the $\Sigma_{MPP_{id}}$ and $\Sigma_{MPP_{idh}}$ structure models (so denoted by $\Sigma_{MPP_{id(h)}}$). The smallest BIC and AIC values were found for the $\Sigma_{MET_{us}} \otimes \Sigma_{MPP_{id(h)}}$ model. The analysis of trait GDDTA and GDDTS in the maize diallel MET&MPP also revealed small BIC or AIC values for $\Sigma_{MET_{us}} \otimes \Sigma_{MPP_{idh}}$ models. For the two maize NAM MET&MPPs that involved relatively larger numbers of parents and families, we assessed the VCOV models on AIC values. No big differences occurred between $\Sigma_{MET_{us}} \otimes \Sigma_{MPP_{idh}}$ and $\Sigma_{MET_{us}} \otimes \Sigma_{MPP_{kin}}$, with similar AIC values, but solving for $\Sigma_{MET_{us}} \otimes \Sigma_{MPP_{kin}}$ is computationally intensive in a later QTL model. Therefore, for practical reasons, we adopted the $\Sigma_{MET_{us}} \otimes \Sigma_{MPP_{idh}}$ model to describe the VCOV of the background polygenic effects.

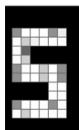


Table 2. Assessing the VCOV structure for the polygenic effect.

MPP	# Family	# Env.	Trait	Structure model for polygenic effects Σ_g		# Variance and/or covariance parameters	BIC	AIC				
cowpeaMAGIC	1	4	FT	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id(h)}}$	1	6115.3	6110.2				
					$\Sigma_{MPP_{kin}}$	2	6116.2	6106.1				
				$\Sigma_{MET_{idh}}$	$\Sigma_{MPP_{id(h)}}$	4	5637.1	5617.0				
					$\Sigma_{MPP_{kin}}$	5	5637.0	5611.9				
				$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id(h)}}$	10	4881.2	4830.9				
					$\Sigma_{MPP_{kin}}$	11	4888.2	4832.9				
maizeDiallel	4	2	GDDTA	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$	1	9410.1	9405.1				
					$\Sigma_{MPP_{idh}}$	4	9360.0	9339.9				
					$\Sigma_{MPP_{kin}}$	2	9417.1	9407.1				
				$\Sigma_{MET_{idh}}$	$\Sigma_{MPP_{id}}$	2	9194.5	9184.4				
					$\Sigma_{MPP_{idh}}$	8	9148.3	9108.1				
					$\Sigma_{MPP_{kin}}$	3	9201.5	9186.4				
				$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	3	8967.9	8952.8				
					$\Sigma_{MPP_{idh}}$	12	8934.2	8873.9				
					$\Sigma_{MPP_{kin}}$	4	8974.9	8954.8				
				GDDTS	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$	1	9932.1	9927.1			
						$\Sigma_{MPP_{idh}}$	4	9929.3	9909.2			
						$\Sigma_{MPP_{kin}}$	2	9939.2	9929.1			
			$\Sigma_{MET_{idh}}$		$\Sigma_{MPP_{id}}$	2	9711.9	9701.8				
					$\Sigma_{MPP_{idh}}$	8	9699.8	9659.6				
					$\Sigma_{MPP_{kin}}$	3	9718.9	9703.8				
			$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	3	9549.0	9533.9					
				$\Sigma_{MPP_{idh}}$	12	9549.9	9489.5					
				$\Sigma_{MPP_{kin}}$	4	9556.0	9535.9					
				MaizeNAM (dent panel)	10	4	DMY	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$	1	21159.4	21153.3
									$\Sigma_{MPP_{idh}}$	10	21180.3	21119.5
									$\Sigma_{MPP_{kin}}$	2	21133.9	21121.7
			$\Sigma_{MET_{idh}}$					$\Sigma_{MPP_{id}}$	4	20977.5	20953.1	
								$\Sigma_{MPP_{idh}}$	40	21184.6	20941.4	
								$\Sigma_{MPP_{kin}}$	5	20940.1	20909.7	
$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	10	20604.1				20543.2					
	$\Sigma_{MPP_{idh}}$	100	21175.4				20567.3					
	$\Sigma_{MPP_{kin}}$	11	20612.2				20545.3					
	DtSILK	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$				1	8696.5	8690.4			
			$\Sigma_{MPP_{idh}}$				10	8687.2	8626.4			
			$\Sigma_{MPP_{kin}}$				2	8684.5	8672.3			
$\Sigma_{MET_{idh}}$		$\Sigma_{MPP_{id}}$	4				8581.7	8557.4				
		$\Sigma_{MPP_{idh}}$	40				8734.2	8490.9				
		$\Sigma_{MPP_{kin}}$	5				8569.9	8539.5				
$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	10	7536.7				7475.8					
	$\Sigma_{MPP_{idh}}$	100	8057.2				7448.9					
	$\Sigma_{MPP_{kin}}$	11	7544.8				7477.9					

maizeNAM (flint panel)	11	6	PH	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$	1	19241.0	19234.9
					$\Sigma_{MPP_{idh}}$	10	19282.2	19221.3
					$\Sigma_{MPP_{kin}}$	2	19214.9	19202.8
				$\Sigma_{MET_{idh}}$	$\Sigma_{MPP_{id}}$	4	19072.1	19047.8
					$\Sigma_{MPP_{idh}}$	40	19303.7	19060.3
					$\Sigma_{MPP_{kin}}$	5	19055.2	19024.8
				$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	10	18239.7	18178.8
					$\Sigma_{MPP_{idh}}$	100	18838.8	18230.5
					$\Sigma_{MPP_{kin}}$	11	18247.8	18180.9
			DMY	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$	1	34086.6	34080.2
					$\Sigma_{MPP_{idh}}$	11	34094.0	34022.8
					$\Sigma_{MPP_{kin}}$	2	34030.2	34017.2
				$\Sigma_{MET_{idh}}$	$\Sigma_{MPP_{id}}$	6	33698.5	33659.7
					$\Sigma_{MPP_{idh}}$	66	34009.6	33582.5
					$\Sigma_{MPP_{kin}}$	7	33632.4	33587.1
				$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	21	32953.8	32817.9
					$\Sigma_{MPP_{idh}}$	231	34311.5	32816.5
					$\Sigma_{MPP_{kin}}$	22	32962.3	32819.9
			DtSILK	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$	1	13406.7	13400.2
					$\Sigma_{MPP_{idh}}$	11	13076.1	13004.9
					$\Sigma_{MPP_{kin}}$	2	13254.1	13241.2
				$\Sigma_{MET_{idh}}$	$\Sigma_{MPP_{id}}$	6	13328.1	13289.3
					$\Sigma_{MPP_{idh}}$	66	13204.9	12777.6
					$\Sigma_{MPP_{kin}}$	7	13184.2	13138.9
				$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	21	10731.9	10595.9
					$\Sigma_{MPP_{idh}}$	231	11897.3	10401.6
					$\Sigma_{MPP_{kin}}$	22	10740.5	10598.0
			PH	$\Sigma_{MET_{id}}$	$\Sigma_{MPP_{id}}$	1	30117.3	30110.8
					$\Sigma_{MPP_{idh}}$	11	30085.4	30014.2
					$\Sigma_{MPP_{kin}}$	2	30031.3	30018.4
				$\Sigma_{MET_{idh}}$	$\Sigma_{MPP_{id}}$	6	29789.0	29750.2
					$\Sigma_{MPP_{idh}}$	66	30085.4	29658.0
					$\Sigma_{MPP_{kin}}$	7	29691.7	29646.4
				$\Sigma_{MET_{us}}$	$\Sigma_{MPP_{id}}$	21	28666.0	28530.0
					$\Sigma_{MPP_{idh}}$	231	30096.7	28600.9
					$\Sigma_{MPP_{kin}}$	22	28674.5	28532.0



4.2. Detecting QEI from QTL models

We present the results of QEI analysis for empirical cowpea MAGIC, maize diallel, and two maize NAM MET&MPPs from three perspectives. First, we checked the GGE biplot to explore the genotypic correlations between the environments. We expect that weakly correlated environments might reveal potential QEI. We looked at identified environment-specific QTLs and checked whether their effect profiles align with the genotypic correlations between environments. We compared our results with previous analyses that used different methods to investigate QEI.

4.2.1. Cowpea MAGIC for flowering time (FT)

In the cowpea MAGIC MET&MPP for trait FT, the GGE analysis (Figure 2) shows that short-day environments under both full and restricted irrigations (fi_s and ri_s) were highly correlated. Long-day environments under both irrigation conditions (fi_l and ri_l) showed a positive correlation. The length of the two vectors for the long-day environments points to a relatively high genetic variance.

Our QEI analysis results (Figure 2) were in agreement with previous analyses that mapped QTLs in separate long-day and short-day environments (Huynh et al. 2018). We identified two environment-specific QTLs on chromosome 9 and 11 with contrasting parental effects under long-day environments (fi_l and ri_l). Huynh's work (2018) detected the same two QTLs on chromosome 9 and 11 with much stronger signals under the long-day condition than the short-day condition. Other QTLs (on chromosome 4 and 5) were detected under both conditions in the previous study, and these two QTLs were identified as environment-consistent QTLs using our approach. A weak QTL identified on chromosome 1 in the short-day environment by Huynh (2018) was not identified anymore, while our method detected a new environment-consistent QTL on chromosome 2.

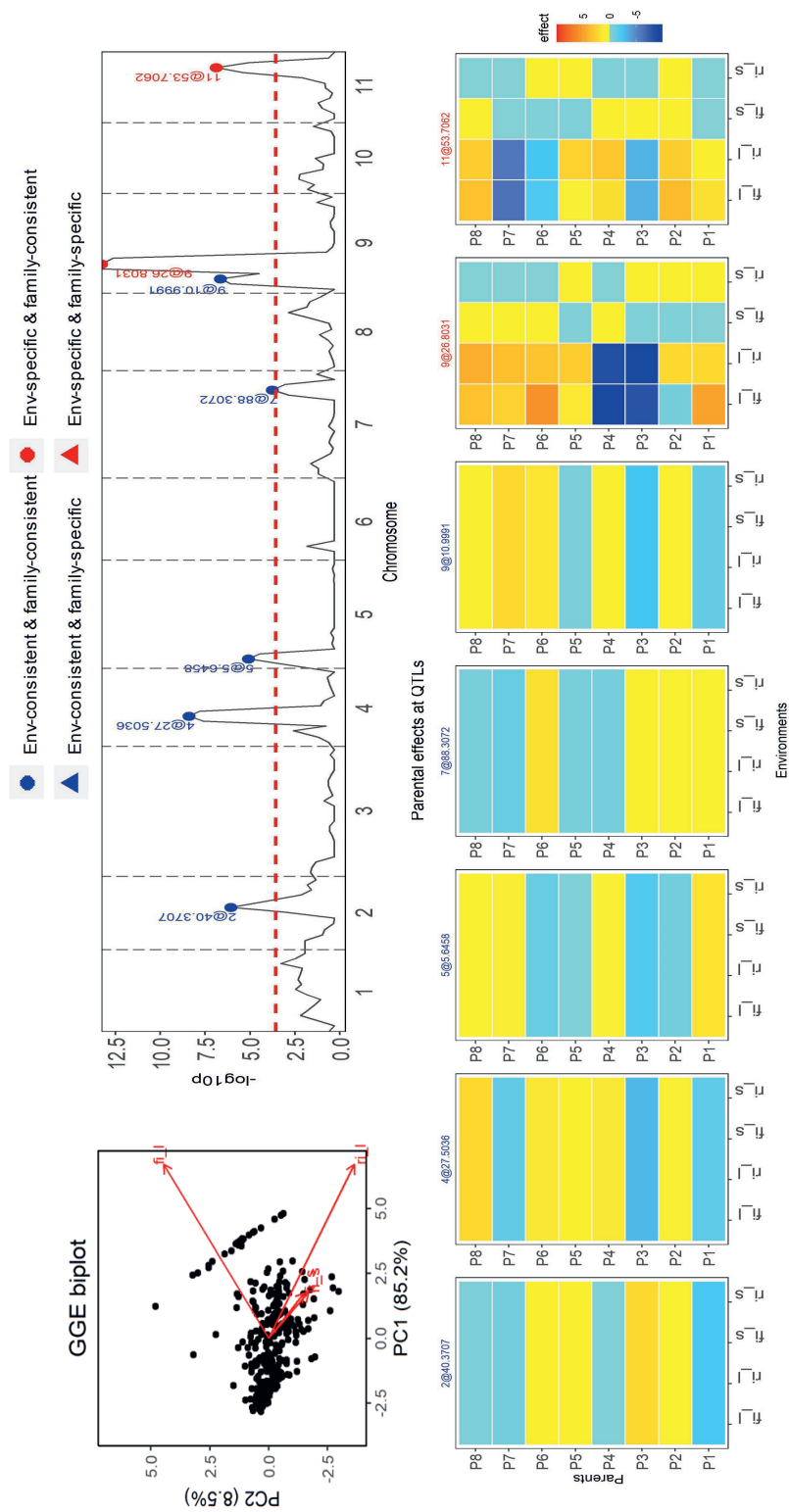


Figure 2. Results of QEI analysis of cowpea MAGIC population for flowering time (FT). **Top Left** GGE biplot for genotype/environment relationships. **Top right** mapping profile for QTLs with different effect types corresponding to families and environments. **Bottom** effect profiles at QTLs with different effect types corresponding to families and environment.



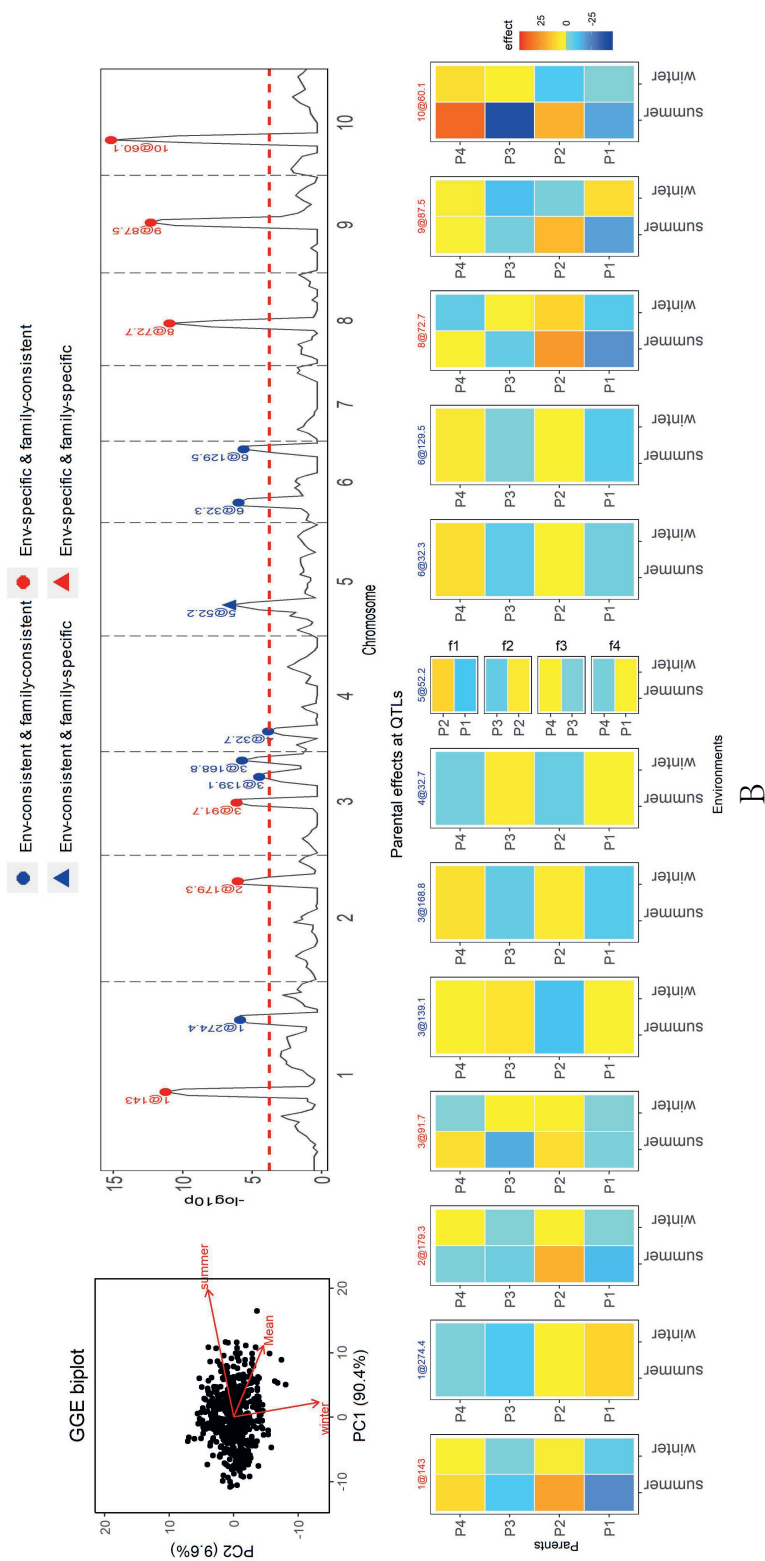
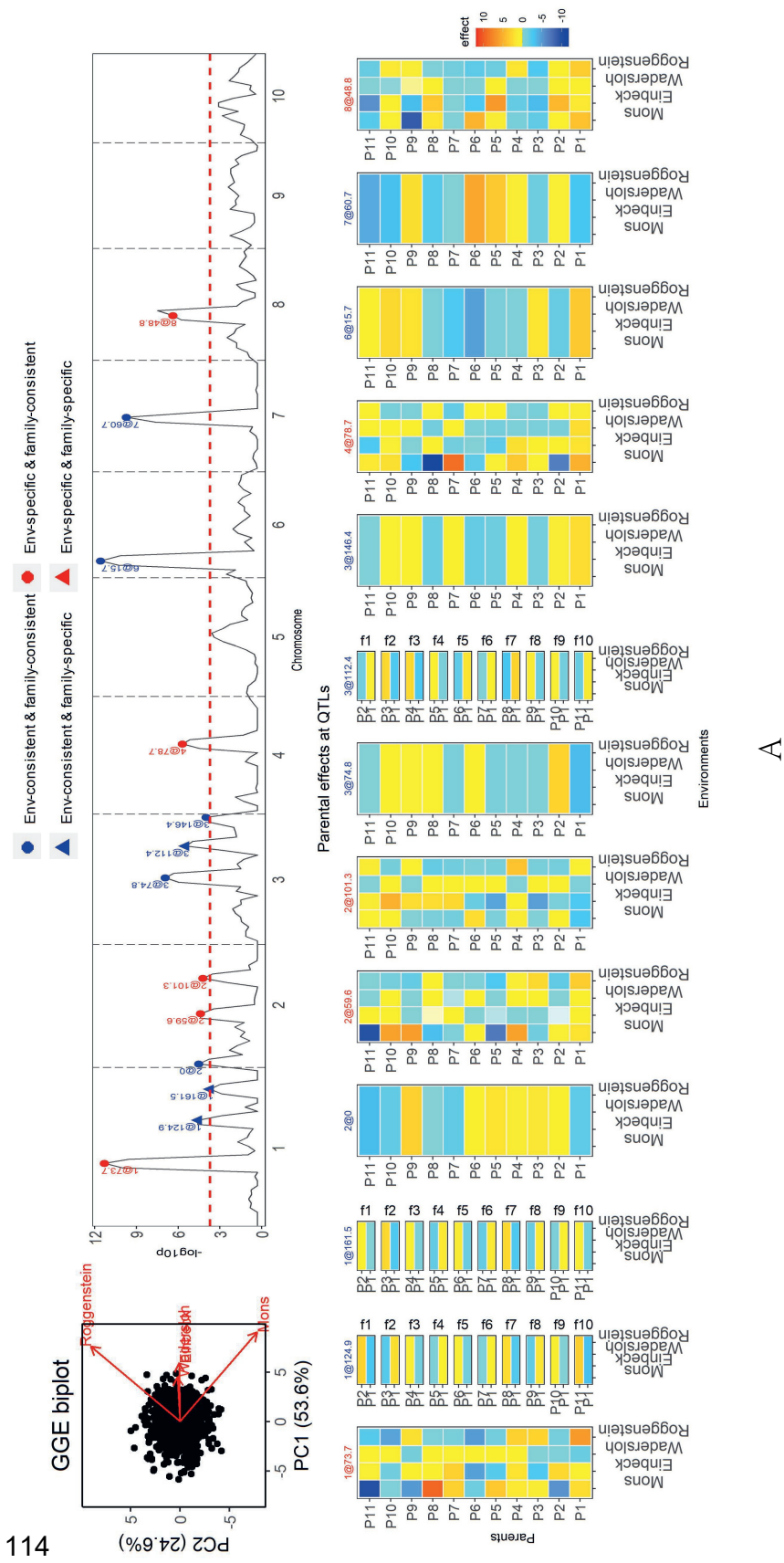
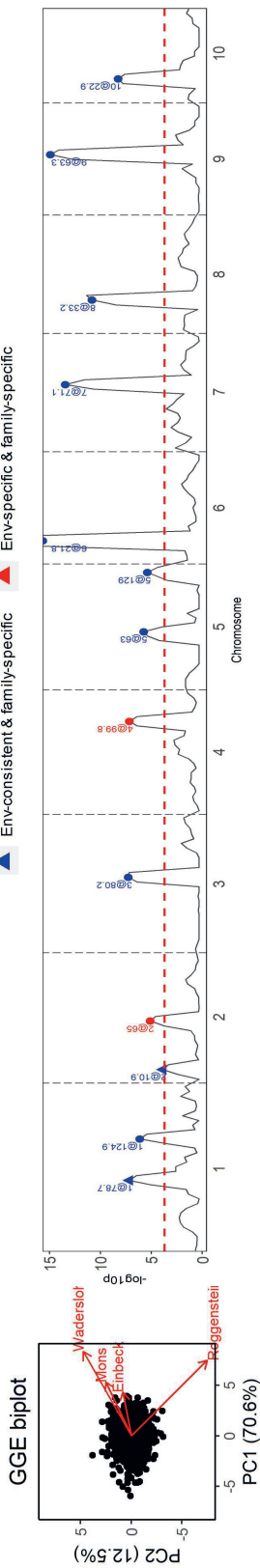
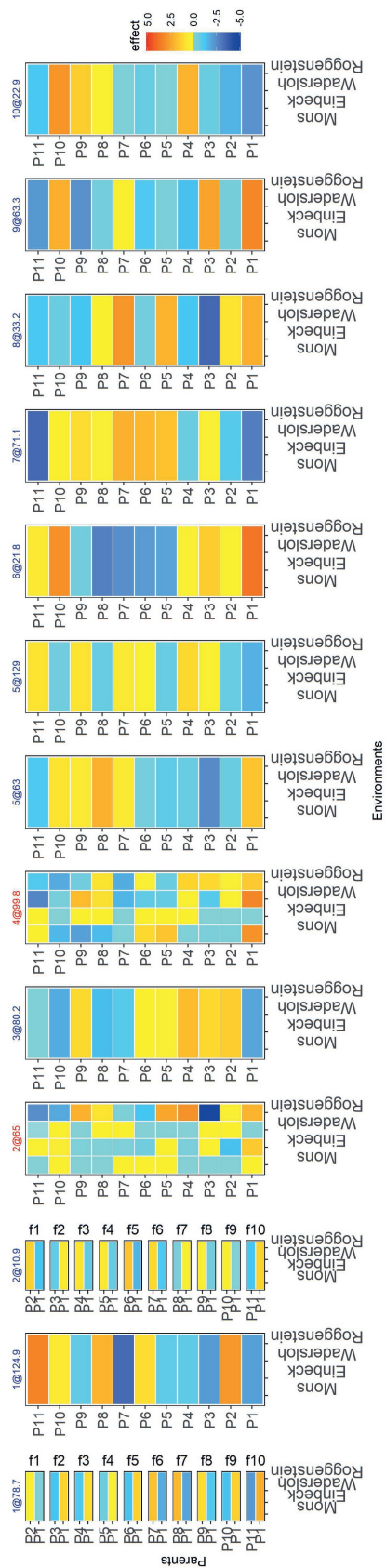


Figure 3. Results of QEI analysis of maize diallel design for **A**, growing degree days to anthesis (GDDTA) and **B**, growing degree days to silk (GDDTS). **Top Left** GGE biplot for genotype/environment relationships. **Top right** mapping profile for QTLs with different effect types corresponding to families and environments. **Bottom** effect profiles at QTLs with different effect types corresponding to families and environment.





Parental effects at QTLs



B

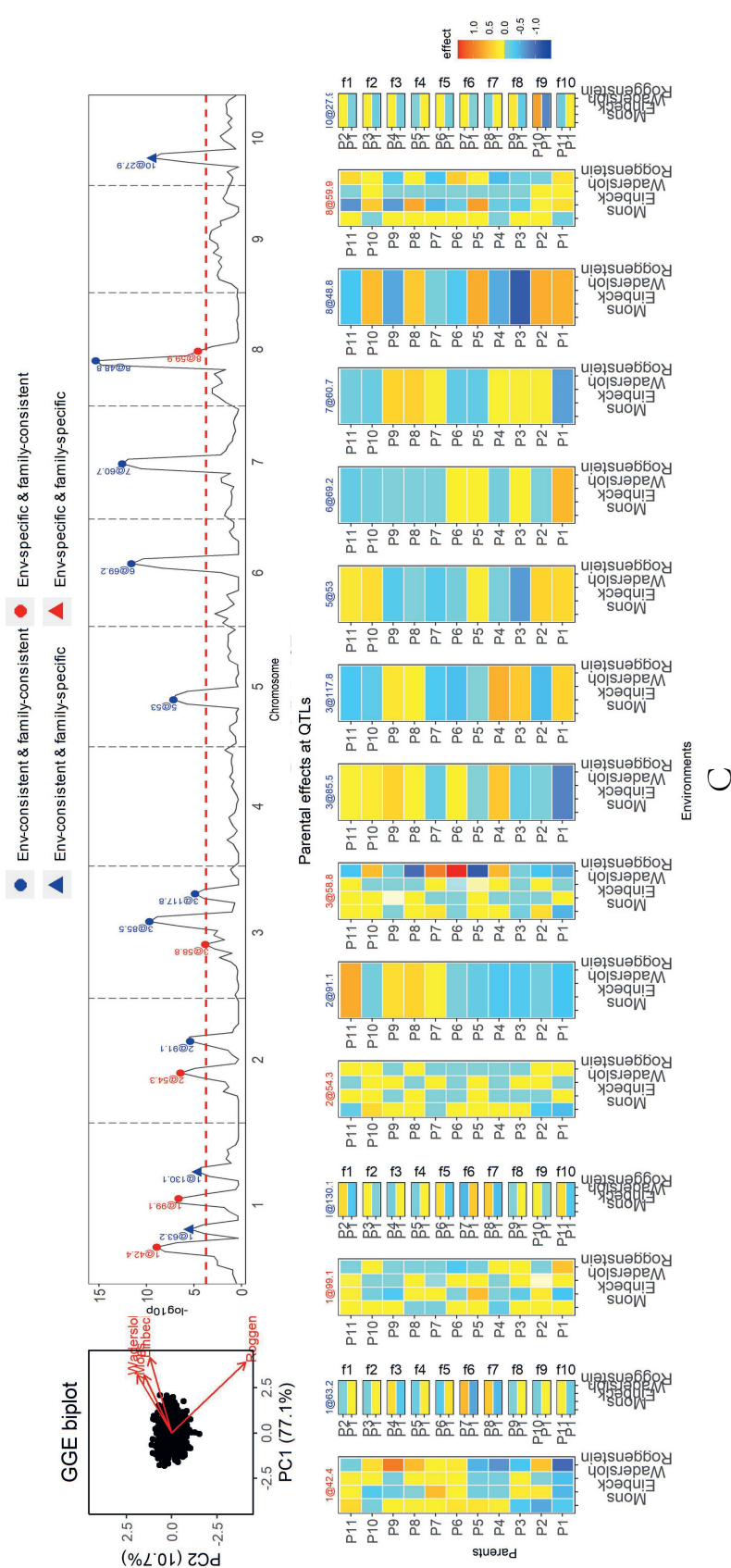
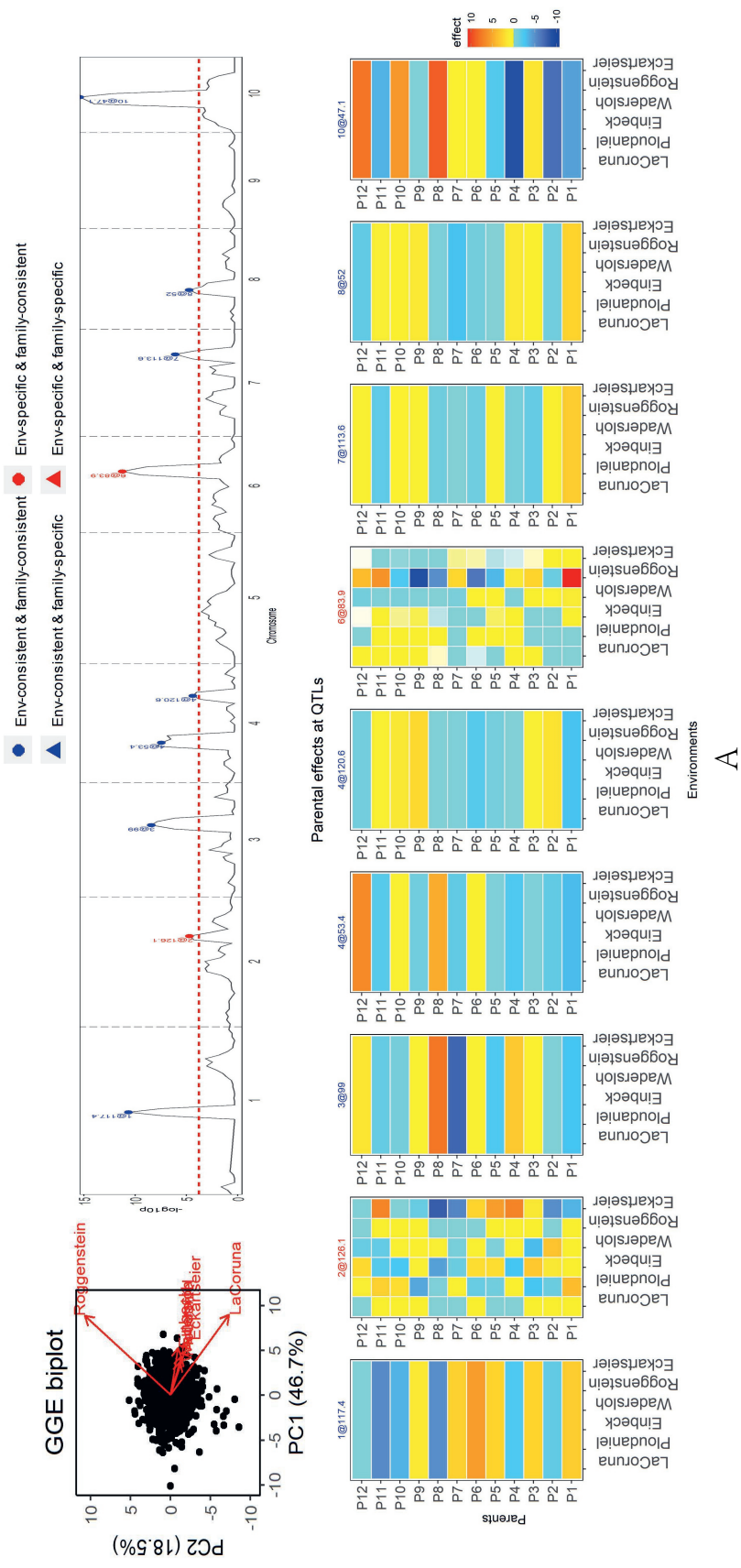
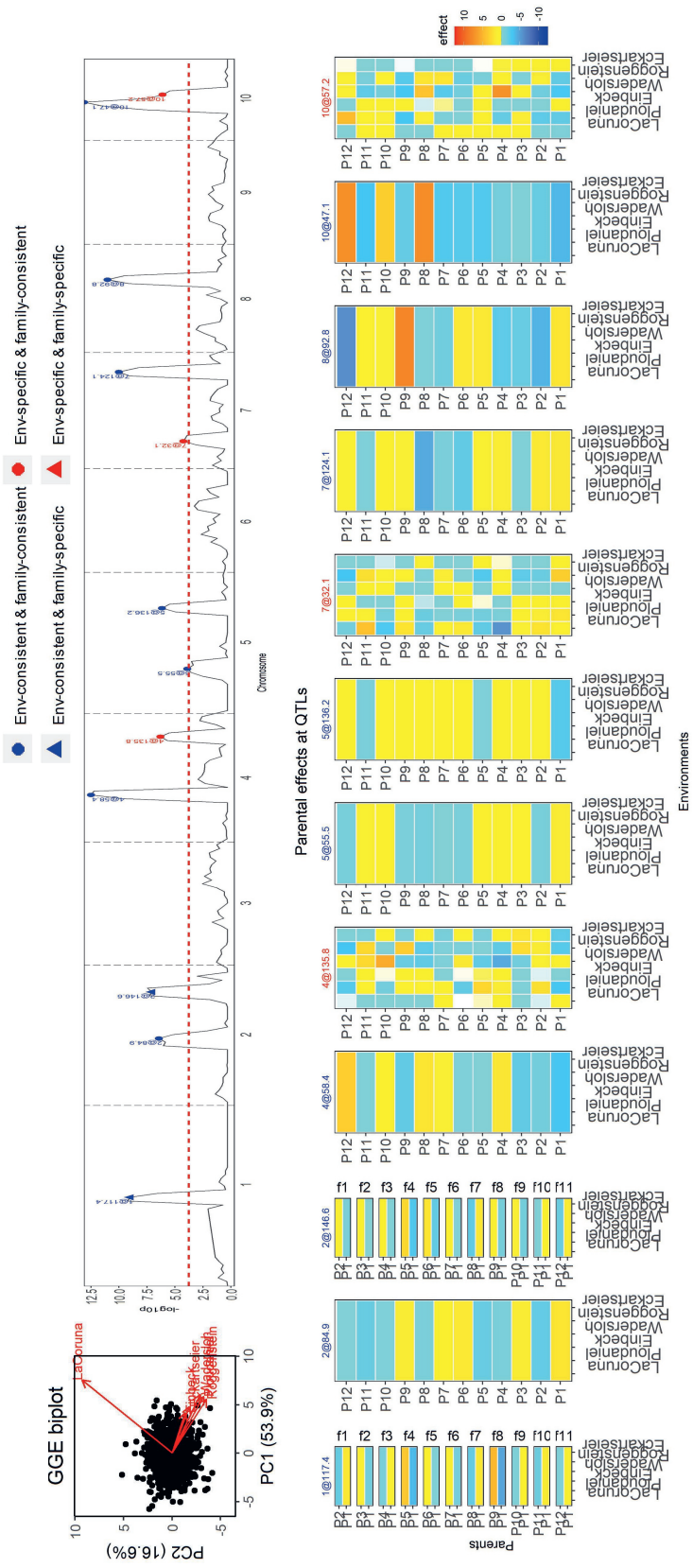


Figure 4. Results of QEI analysis of maize NAM dent design for **A.** dry matter yield (DMY), **B.** plant height (PH), and **C.** days to silk (DSILK). **Top Left** GGE biplot for genotype/environment relationships. **Top right** mapping profile for QTLs with different effect types corresponding to families and environments. **Bottom** effect profiles at QTLs with different effect types corresponding to families and environment.





B

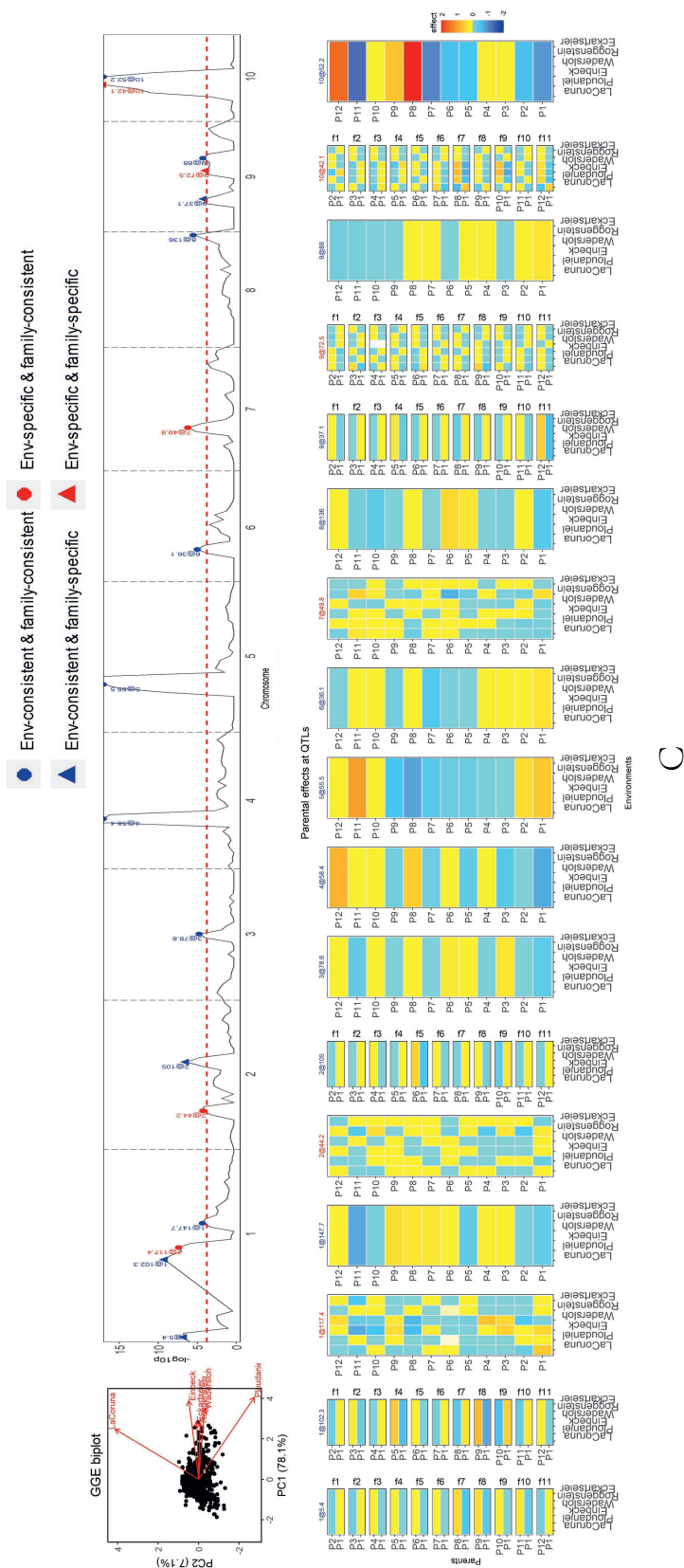


Figure 5. Results of QEI analysis of maize NAM flint design for **A**. dry matter yield (DMY), **B**. plant height (PH), and **C**. days to silk (DSILK). **Top Left** GGE biplot for genotype/environment relationships. **Top right** mapping profile for QTLs with different effect types corresponding to families and environments. **Bottom** effect profiles at QTLs with different effect types corresponding to families and environment.

4.2.2. Maize diallel design for days to silking (GDDTS) and anthesis (GDDTA)

Two flowering-related traits, GDDTS and GDDTA, were measured in the summer and winter seasons to evaluate the photoperiod sensitivity under long-day and short-day conditions. For both traits, the correlations between summer and winter conditions were weak (Figure 3).

In the QEI analysis, we found that nearly half of the identified QTLs were environment-specific QTLs for both GDDTS and GDDTA. For the trait GDDTA (Figure 3a), in total 10 QTLs were detected, among which 6 QTLs were environment specific. Five of the 6 environment-specific QTLs were comparable to QTLs for photoperiodic responses (calculated as differences of responses between the two seasons) from a previous study (Coles et al. 2010). Particularly strong environment-specific QTLs on chromosome 8, 9, and 10 were identified with contrasting parental effects in especially the summer season (GDDTAS). The same summer season QTLs were identified by Coles et al. (2010) using single environment analysis. As for the trait GDDTS (Figure 3b), we identified 6 environment-specific QTLs out of a total of 13 QTLs, which were in accordance with the 6 QTLs for photoperiodic responses identified by Coles et al. (2010) with strong effects under summer conditions (GDDTSS).

For both traits GDDTA and GDDTS, a few environment-consistent QTLs were estimated with specific effects within families (EC&FS QTLs), which implied QTL-by-family background interactions. We did not identify ES&FS QTLs for GDDTA and GDDTS, but analysis of other traits (Supplementary material), such as plant height (PH), ear height (EH), and total leaf number (TLN) in this maize diallel MET&MPP, revealed some ES&FS QTLs.

4.2.3. Maize NAM (dent panel) design for dry matter yield (DMY), day to silking (DtSILK), and plant height (PH)

The maize NAM dent panel was screened for important traits such as DMY, DtSILK, and PH, under four locations across the EU. The GGE plots (Figure 4) illustrate different levels of environment correlations. Regarding trait DMY, environment Roggenstein and INR were weakly correlated, and the other two environments Einbeck and Wadersloh were highly positively correlated; as for trait DtSILK, three environments Wadersloh, Mons, and Einbeck were clustered together and they were weakly correlated with Roggenstein. For trait PH, environment Roggenstein was separated, weakly correlated, with the rest of the environments.

We identified several environment-specific QTLs in the QEI analysis for each trait. For example, five environment-specific QTLs out of a total of 13 QTLs for trait DMY were detected using our methodology (Figure 4a). The effect profiles at these five environment-specific QTLs display the most differential effect profiles between the two weakly correlated environments INR and Roggenstein. For the same data, Giraud et al. (2014) performed combined linkage disequilibrium linkage analysis (LDLA) for adjusted means over the environments. The environment-consistent QTLs identified by our QEI analysis were comparable to QTLs on chromosome 1, 3, 6, and 7 by Giraud et al (2014). Notably, environment-specific QTLs on chromosome 1, 2, and 4 identified in our study were not reported by Giraud et al. (2014). Concerning DtSILK, five environment-specific QTLs out of

a total of 16 QTLs rendered differential effect profiles between Roggenstein and other environments. For PH, two environment-specific QTLs out of a total of 13 QTLs were found, and the differential effect profiles were between the two weakly correlated environments Roggenstein and Wadersloh.

For all traits, we detected QTLs with consistent effects across environments that could be either consistent across families as well as specific to families (EC&FC and EC&FS QTLs), but no environment-specific and family-specific (ES&FS) QTLs were found for any of the traits.

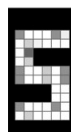
4.2.4. Maize NAM (flint panel) design for dry matter yield (DMY), days to silking (DtSILK), and plant height (PH)

The maize NAM design of the flint panel was screened across six locations in the EU for DMY, DtSILK, and PH. With respect to trait DMY (Figure 5a), the GGE biplot shows three clusters (La Coruña, Roggenstein, and another five environments). We detected nine QTLs among which two QTLs on chromosome 2 and 6 are environment-specific. For these two environment-specific QTLs, we refer to the work of Garin (2020) where QEI was assessed between the two weakly correlated environments La Coruña and Roggenstein. A QEI signal on chromosome 6 identified by Garin et al.(2020) agrees with our results that this QTL is environment-specific with strong effects in environment Roggenstein. We found an extra environment-specific QTL on chromosome 2 with differential effect profiles between Eckartseier and other environments, which was not present in the work by Garin et al. (2020) using only Roggenstein and La Coruña. The QEI identified on chromosome 5 by Garin et al.(2020) with a relatively weak signal was not identified anymore by our method over six environments. Later, we selected the same two environments Roggenstein and La Coruña to reanalyze DMY, and then the QTL on chromosome 5 was identified again as being environment-specific.

5. Discussion

5.1. QEI analysis by assessing the stability of QTL effects across families and environments

Through the MET&MPP analysis, QEI is detected by testing different types of QTL effects across diverse environments and families by employing IBD-based mixed model approaches. Previous studies have discussed the advantages of using multi-allelic IBDs versus bi-allelic identity-by-state (IBS) markers (Jurcic et al. 2021; Li et al. 2021) and modeling QTL terms as random versus fixed (Boer et al. 2007; Wang et al. 2022b). In the current study for MET&MPP analysis, we propose to fit functions of IBD probabilities as genetic predictors to indicate the expected numbers of allele copies originating from each parent, which allows the estimation of multi-allelic QTL effects in relation to parent origins (Wei and Xu 2016; Li et al. 2021). By contrast to modeling fixed QTL effects in the MET&MPP analysis, treating QTL effects as random term avoids the issue of overfitting many genetic parameters, particularly for complex MET&MPPs involving large numbers of parents,



families, and environments. It is also convenient to define the nature of random QTL effects (EC&FC, ES&FC, EC&FS, and ES&FS) by modeling the combinations of homogeneity or heterogeneity of genetic structures corresponding to both environments and families.

EC&FC QTLs can be regarded as the most stable QTLs in the MET&MPP analysis. EC&FC QTLs are comparable to those QTLs identified by the approach using adjusted genotype means over diverse environments (Giraud et al. 2014; Garin et al. 2020). EC&FC QTLs identified in MET&MPP analysis agree with QTLs identified in single environment analyses, like those in Coles et al. (2010) and Huynh et al. (2018). EC&FC QTLs tend to be detected with higher power and resolution in MET&MPP analysis than when analyzing separate environments or a genotypic main effect across environments, as the joint analysis of all environments increases the sample size for mapping stable QTLs.

QTLs detected in separate environments with differential mapping profiles are likely to be detected as ES&FC QTLs by MET&MPP analysis. Effect profiles at ES&FC QTLs associated with parent origins across environments tend to be differential between weakly correlated environments, while highly and positively correlated environments rarely convey QEI signals. When using the difference between two environments as a response variable for QTL mapping, such as in the maize diallel MPP (Coles et al. 2010), ES&FC QTLs in our MET&MPP analysis were comparable to those QTLs mapped in single environment analyses. However, we notice that with small population sizes, weak QEI signals are likely to be overlooked. For example, in the maize NAM flint panel (811 genotypes across 6 environments) the QEI signal on chromosome 5 related to environments La Coruña and Roggenstein (Garin et al. 2020) disappeared in the MET&MPP analysis using all 6 environments, but this QTL could be identified again under the two selected environments Coruña and Roggenstein.

From the perspective of genetics, defining QTL effects within or across families allows evaluating the potential QTL-by-family interaction (QFI) (Jannink and Jansen 2001; Blanc et al. 2006; Han et al. 2016), as the rich germplasm resource from multiple parents increases the chances to combine interacting QTLs across the MPP and diversify the segregation configurations across families. For FS QTLs that are stable across environments (EC&FS QTLs), simple digenic interaction can be further investigated for pairs of markers by two-dimensional genome scans in families where interacting QTLs segregate. Such bi- or multi-dimensional QTL scans for epistatic interactions require larger family sizes for sufficient detection power. For the detection of ES&FS QTLs whose family-specific effects are unstable across environments, again large families are required, which may explain why we could detect very few ES&FS QTLs in the maize diallel design, with around 140 genotypes per family over 2 environments, and hardly any ES&FS QTL in the maize NAM design, with around 80 genotypes per family.

5.2. Suitable VCOV structures corresponding to background family and environment effects

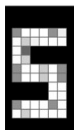
GEI analysis in MET&MPP designs cannot ignore the family or population structure and genetic correlations between environments. An additional objective of the current study was

to determine and verify the appropriate VCOV model for the polygenic effect. We have tested different combinations of family structures and genotypic variances and correlations for sets of trials and showed that the unstructured VCOV model for between environment correlations combined with the heterogeneity of family variances ($\Sigma_{MPP_{idh}} \otimes \Sigma_{MET_{us}}$) is a suitable model for most of the MET&MPP data sets.

With respect to modelling heterogeneity of family variances ($\Sigma_{MPP_{idh}}$), the heterogeneity of family structure can be a consequence of different segregation scenarios of QTLs and potential non-additive effects (epistasis) in the specific family background (Li et al. 2021). Alternatively, we also modelled individual relatedness by imposing a marker-based kinship matrix, with the unstructured VCOV model corresponding to between environment correlations ($\Sigma_{MPP_{kin}} \otimes \Sigma_{MET_{us}}$). Modelling the polygenic effect with a $\Sigma_{MPP_{kin}} \otimes \Sigma_{MET_{us}}$ structure is not always better than modelling with a $\Sigma_{MPP_{idh}} \otimes \Sigma_{MET_{us}}$ structure. In the analysis of our empirical MET&MPPs, we used standard MAGIC, NAM, and diallel designs with balanced family sizes and controlled population structures. However, for more complex MPP designs, the modelling of only heterogeneous family variance ignoring family correlations might be over-optimistic and thus adopting the $\Sigma_{MPP_{kin}} \otimes \Sigma_{MET_{us}}$ VCOV structure model might be better to describe the background polygenic and between environment genotypic effects. A mixed model with population structure, $\Sigma_{MPP_{kin}}$, combined with an unstructured VCOV model, $\Sigma_{MET_{us}}$, corresponding to between environment correlations, $\Sigma_{MPP_{kin}} \otimes \Sigma_{MET_{us}}$, may be so computationally intensive that a later whole-genome QTL scan may become extremely time-consuming.

With respect to environment variance and correlations, the unstructured VCOV model $\Sigma_{MET_{us}}$ was proven to be a suitable model. For different traits in various MET&MPP designs analysed in the current study, different levels of similarities between pairs of environments can be observed from the GGE analysis, and thus it is more reasonable to adopt the unstructured VCOV model. Using an unstructured VCOV model is also suggested in previous studies from both theoretical and practical perspectives (Boer et al. 2007; van Eeuwijk et al. 2010a; Malosetti et al. 2013). We also tried other VCOV structures with fewer parameters, such as the factor analytic model, $\Sigma_{MET_{fa}}$ (Piepho 1997, 1998; Smith et al. 2001). However, in several cases, solving the $\Sigma_{MPP_{idh}} \otimes \Sigma_{MET_{fa}}$ was not successful due to singularity issues. An alternative way of reducing the number of estimated variance parameters is to impose an unstructured VCOV model on environmental clusters in place of trials (Malosetti et al. 2013).

In the final QTL model after fitting all QTL candidates with their selected effect types, the VCOV structure model for the background family and environment effects tends to become simpler, or the difference between the complex structure model and the simple model become smaller in terms of AIC or BIC values. This indicates that QTLs explain part of the initial heterogeneity in genotypic variances and correlations.



Supplementary material

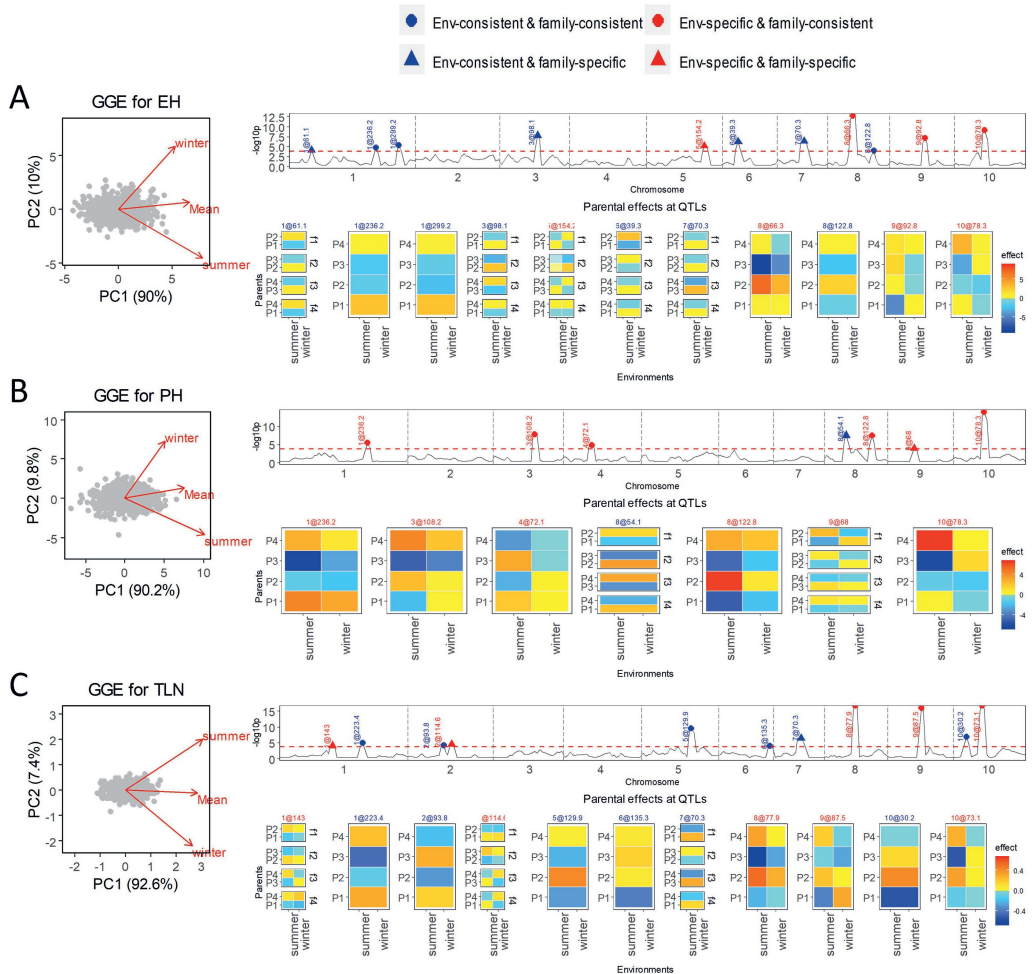
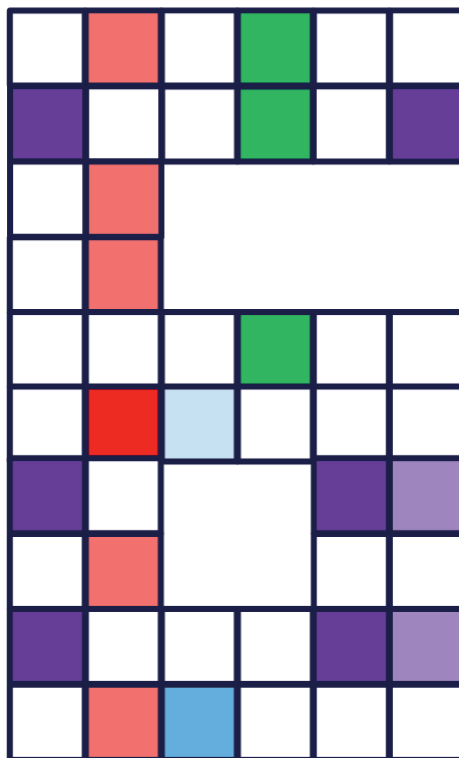


Figure S1. Results of QEI analysis of maize diallel design for **A.** ear height (EH) and **B.** plant height (PH), and **C.** total leaf number (TLN). **Left** GGE biplot for genotype/environment relationships. **Top right** mapping profile for QTLs with different effect types corresponding to families and environments. **Bottom right** effect profiles at QTLs with different effect types corresponding to families and environment.

Chapter 6

General discussion



1. Prologue

The development of good statistical methodologies is one of the key elements of successful QTL mapping. Particularly, innovative and complex multi-parent mapping populations, (i.e., MPPs), drive the need to develop suitable statistical tools for QTL mapping. The popularity and success of using MPPs for genetic studies can be explained by the fact that they provide a bridge between linkage mapping and association mapping approaches (Arrones et al. 2020). However, a complication in QTL mapping for MPP designs is how to deal with the multitude of design options and possible genetic architectures. This thesis describes the development of IBD-based mixed model approaches for QTL mapping in a wide range of MPP designs. We also extend this generic IBD-based mixed model approach to more advanced models for unravelling epistasis and QTL-by-environment interaction (QEI) using MPPs containing multiple crosses (multi-cross population) and MPPs that are evaluated in multi-environment trials (MET&MPPs). The statistical models developed in this thesis were successfully applied and verified on a wide range of simulated and empirical MPPs including NAM, diallel, MAGIC, and sophisticated MPPs with known pedigree information in diverse species like maize, cowpea, soybean, barley, tomato, cucumber, and *Arabidopsis*.

2. Mapping population

2.1. MPPs are novel populations that address traditional challenges

QTL mapping in MPPs solves some of the recurring problems in the application of marker-assist selection/breeding (MAS) (Liao et al. 2001; Li 2008). Efficient MAS depends on a reliable marker-phenotype relationship that can successfully predict phenotypic performance. The marker-phenotype relation can be split into marker-QTL and QTL-phenotype correlations which are both affected by many factors (Collard and Mackill 2008; Cobb et al. 2019). MPPs can help to reveal reliable marker-QTL and QTL-phenotype correlations.

From the perspective of a QTL-phenotype correlation, even when the phenotyping strategy is very reliable, an identified QTL in a single bi-parental background may still require confirmation in other genetic backgrounds. It is often observed that QTL effects estimated within one population background are unstable or disappear in another genetic background (Melchinger et al. 1998; Liu and Zeng 2000; Crepieux et al. 2005; Pascual et al. 2016). A proper QTL validation procedure should confirm QTLs in different genetic backgrounds (Collard and Mackill 2008). The issue that genetic effects cannot always be stably transferred into other genetic backgrounds can be alleviated by identifying QTLs in a wider population as when performing joint analysis of diverse data sets in meta-QTL studies (Griffiths et al. 2009; Courtois et al. 2009; Kumar et al. 2020), and a good alternative is the use of multi-cross populations with diverse genetic backgrounds (Tsaih et al. 2005) such as in half-diallel (Tsaih et al. 2005; Paulo et al. 2008) and NAM (Yu et al. 2008) populations. The issue of untransferable QTLs between families is touched upon, among other things, in **Chapter 4** from the view of epistasis. Here we show that interacting QTLs can display different effects from one family to another family. When breeders want to use QTL alleles with observed QTL by background

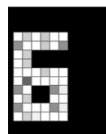
interactions for introgression into new germplasm they should take into account possible limitations to the generality of the introgressed QTL alleles.

From the perspective of QTL-marker correlation, a marker as a proxy for a QTL relies on the distance on the genetic map. The tight linkage between a marker and the true QTL in one mapping population can be contrasting with a lower linkage in a population representing a broader diversity (Sharp et al. 2001; Thomas 2003). Such poor QTL-marker correlation may result in a disappointing application of identified QTLs in a MAS program. The confidence in QTL-marker correlations can be enhanced in MPPs because high levels of polymorphisms in MPPs benefit the construction of a reliable consensus genetic map and increase the chance to find an accurate QTL.

2.2. Choosing between MPPs

Although an MPP facilitates QTL mapping and the MAS process, it remains a challenge to choose the most appropriate type of MPPs for a specific breeding program design or genetic study. A study by Bernardo (2021) investigated the impact of different diallel and n-way cross MPPs on genetic gain and doubted whether higher diversity included in MPPs leads to improved genetic gain. Another study (Garin et al. 2021) addresses the influence of allelic diversity on QTL detection in different multi-cross MPPs, including diallel, chessboard (incomplete diallel with crosses along multiple diagonals), factorial, and NAM populations, and shows that fewer but larger crosses from a reduced number of parents produce larger detection power. In **Chapter 2** of the thesis, we simulated diallel, NAM, and four-way cross MPPs from four real *Arabidopsis* parents to evaluate the performance of various QTL detection methodologies. From those simulations, it appeared that a single best MPP does not exist in terms of mapping power and resolution, as QTL mapping results depend on many factors including the choice of parents, crossing design, family sizes, and allele segregation configurations and frequencies.

It is important to exploit the unique potential of MPP designs to achieve specific breeding objectives. To test the usefulness of donor genes in elite genetic backgrounds, the NAM population is a promising approach because each donor parent can be crossed as a peripheral parent to a central elite parent (Cobb et al. 2019). Diallel populations consist of multiple connected families and allow the assessment of general combining ability (GCA) and specific combining ability (SCA) in hybrid groups (Verhoeven et al. 2006; Karademir and Gençer 2010; Sadaiah et al. 2013; Giraud et al. 2017). A MAGIC population represents a mosaic of parental genomes and can accumulate multiple QTLs in a homogeneous genetic background, i.e., without population structure, providing high QTL mapping power and good opportunities to pyramid favourable QTLs from diverse parents (Mackay et al. 2014; Dell'Acqua et al. 2015; Scott et al. 2020). One recent example of a MAGIC is a tomato four-way cross population to study yield component QTLs in relation to parental origin (Tsutsumi-Morita et al. 2021). In this thesis, we explored the potential of using different MPPs in QTL mapping (**Chapter 4** and **5**). It has been shown that the studies of complex genetic architectures for non-additive QTLs require specific types of MPPs, such as multi-cross MPPs to study epistasis and MET&MPPs for QTL×Environment interaction analysis.



3. Using IBDs as genetic predictors in MPPs

The basis for our IBD-based approaches is the creation of design matrices along the genome representing the allelic information in offspring in relation to parental origins. The entries in these design matrices are derived from conditional QTL genotypes given the full sets of flanking markers on either side of the QTL. In this thesis, we applied a general Hidden Markov Model (HMM) framework called reconstructing ancestry blocks bit by bit (RABBIT) to calculate the IBD probabilities for both standard and complex MPP designs (Zheng et al. 2014, 2015, 2018). The difference of this general HMM framework from other strategies is that RABBIT constructs a general continuous time Markov model for parent origin processes in any kind of MPPs, where the rate matrix is deduced from the expected densities of various types of junctions (observable recombination breakpoints). RABBIT also accounts for missing genotypic data and typing errors of both parents and progenies, and thus it is robust to low-quality genotyping.

IBD probabilities which are calculated from the pedigree and whole genome marker information are in general more informative to represent the genome composition than IBS for most of the simulated and empirical data in this thesis. To illustrate our point, we simulated a four-way cross $(A \times B) \times (C \times D)$ (Figure 1). When genotypes at a simulated genomic position are polymorphic at the IBS level between parent A on the one hand and the remaining set of parents at the other end, an IBS approach will only test a contrast between parent A and the remaining parents, while in an IBD approach contrasts between all parents are possible (provided the IBD probabilities are different for all parents). Design matrices based on IBD probabilities capture more of the allelic variation between multiple parents and this may explain the improvement of mapping power and resolution of such QTLs in IBD approaches versus the benchmark IBS-based GWAS approach shown in **Chapter 2**.

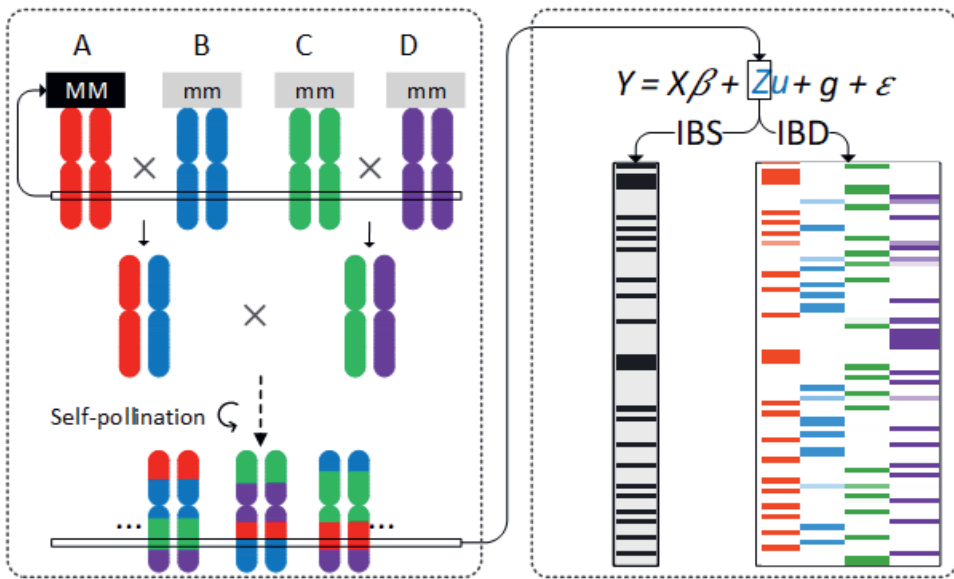


Figure 1. Comparing the construction of design matrices for QTL effects using IBS and IBD probabilities. One genomic position is shown for a simulated four-way cross with parent A carrying a different allele from the other parents. For the IBS-based design matrix, the black and grey entries in the design matrix Z indicate the bi-allelic states for all individuals, in the rows of the matrix. For the IBD-based design matrix, colours correspond with parents and the intensity of colour reflects IBD probabilities between parents and offspring.

Regarding the use of IBD probabilities, our methods are related to so-called haplotype-GWAS (hGWAS) or random-model approaches in cases of MAGIC populations (Wei and Xu 2016; Wang et al. 2022a) and the connected model for joint analysis for multi-cross populations (Blanc et al. 2006; Giraud et al. 2014; Han et al. 2016). In the hGWAS approach, an IBD block is assigned to the parent with the highest conditional QTL probability or IBD probability in that DNA segment but it is limited to RIL or DH individuals. In the connected model for joint cross analysis, multi-allelic QTL effects are associated with the parents and are estimated across the whole offspring population, so design matrices Z along the genome are of dimension number of offspring \times number of parents, $N \times P$, with entries indicating the expected number of parental alleles at a genomic position. As an extension, QTL effects can be defined at ancestral levels defined above parents, which is known as linkage disequilibrium and linkage (LDLA) analysis (Bardol et al. 2013; Giraud et al. 2014; Han et al. 2016).

Another way to utilize IBD information in QTL mapping is to use IBD information to structure the random genotype effects by imposing a variance-covariance structure on the genotypes, reflecting their genetic relatedness, and test for a significant variance component (George et al. 2000; Crepieux et al. 2004, 2005). Such an approach was applied to analyse a maize hybrid program (van Eeuwijk et al. 2010b) and a complex pedigree in palm (Tisné et al. 2015) to test the presence of QTLs in heterotic groups.

4. Mixed model approaches

Throughout this thesis, we propose the use of mixed model approaches for QTL mapping in MPPs. Mixed model approaches are prevalent for GWAS in diversity panels. In GWAS approaches, corrections for population structure were incorporated by imposing a relationship matrix on the polygenic effects (Yu et al. 2006; Zhao et al. 2007; Hyun et al. 2008; Zhang et al. 2010). Our MPP mixed model approaches can be seen as an extension and generalization of the GWAS mixed model. We have shown that mixed model approaches provide versatile and flexible solutions to handle complex MPPs by modelling random QTL effects and accounting for polygenic backgrounds.

4.1. Random parental effects at QTLs

Modelling QTL effects as random or fixed is a recurring question in QTL mapping for diversity panels and MPPs. In this thesis, QTL effects are recommended to be modelled as random. Genetic effects are routinely defined as fixed for easy interpretation of the allelic effects and statistical tests in standard and balanced MPPs. For example, in a NAM design where a central parent is crossed with multiple peripheral parents, it is convenient to model the fixed genetic effect relative to the central parent as the reference. However, complex MPPs present analytical challenges and demand specialized methodologies to take care of a large number of estimated parameters in relation to parents and/or families. A solution to the difficulties of estimation and testing for large numbers of fixed QTL effects is to choose QTL effects as random effects with a common variance (Wei and Xu 2016). An additional advantage of modelling QTL effects as random is that it may be easier to deal with unbalanced data sets. An illustration of a QTL modelling approach for MPPs with random parental effects with a common variance is given in the multi-parent whole-genome average interval mapping procedure (WGAIM) by Verbyla et al. (2012, 2014b).

Beyond detecting QTLs, we also want to estimate QTL effects and variances. For fixed QTL alleles, allelic substitution effects at significant QTLs, as well as the contribution of the allelic variation to the genetic and phenotypic variance, are often overestimated due to under-sampling of offspring individuals (known as Beavis effect) of the mapping population (Xu 2003; Beavis 2019; Wang et al. 2022b). First, the Beavis effect describes the phenomenon that the variance associated with QTLs is significantly inflated if the offspring population size for identifying those QTLs is not large enough (Xu 2003; Beavis 2019). Secondly, the commonly used estimator for genetic variance obtained by squaring the estimated QTL effect is biased because this estimator does not consider the estimation error for the QTL effects (Broman 2001; Luo et al. 2003; Xu 2003). Incorporating the estimation errors into the estimator of genetic variance is one way to adjust the bias. The alternative of directly estimating the genetic variance from the application of a mixed or random model (Wang et al. 2022b) is more attractive. Treating QTL effects as random allows the estimation of variance components related to QTLs by restricted maximum likelihood (REML), and it benefits the estimation of unbiased genetic variances.

In the context of MPPs, we simulated an eight-way MAGIC RIL population to demonstrate the differences in estimated QTL effects, variance, and prediction abilities between fixed-QTL

and random-QTL models. Box 1 shows the details of our simulations and their results. We observed the following. 1) The realized distributions of estimated fixed QTL effects, $\hat{Q}_{P_{fix}}$, were more dispersed than those of their random counterparts, $\hat{Q}_{P_{random}}$, which subscribes to the view that QTL effects tend to be overestimated when the population size is small (Xu 2003; Beavis 2019); 2) the estimated variance for random QTL effects, $\hat{\sigma}_{QTL_{random}}^2$, was close to variance that was used to simulate those effects, $\sigma_{QTL_{sim}}^2 = 0.5$, while the estimated variance of fixed QTL effects, $\hat{\sigma}_{QTL_{fixed}}^2$, was biased to a larger value, which agrees with the assumption that the random model produces an unbiased estimate of the genetic variance (Wang et al. 2022b).

Box 1 Comparison of random-QTL model and fixed-QTL model regarding estimated effects, variance, and prediction accuracy in the case of an eight-way MAGIC.

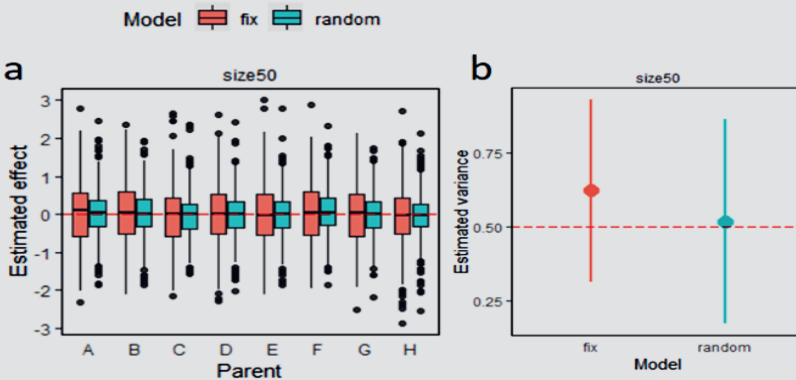
Genetic model: $Y = \mu + Q + \varepsilon$. Phenotypes are explained by overall mean μ , genetic effects Q at a QTL, and residual error $\varepsilon \sim N(0, 1)$.

Simulation and calculation:

- QTL effects contributed by each parent $Q_P \sim N(0, \sigma_{QTL_{sim}}^2)$ where $P \in \{A, B, C, D, E, F, G, H\}$
- Estimated parent effect from the fixed and random model: $\hat{Q}_{P_{fix}}$ and $\hat{Q}_{P_{random}}$
- In the fixed-QTL model, the estimator of genetic variance is derived from estimated fixed effects as: $\hat{\sigma}_{QTL_{fixed}}^2 = \frac{1}{8} \sum \hat{Q}_{P_{fixed}}^2 - \left(\frac{1}{8} \sum \hat{Q}_{P_{fixed}} \right)^2$, assuming that each offspring individual has an equal chance of 1/8 to be IBD with a parent at any position.
- In the random-QTL model, genetic variance can be directly estimated from the corresponding variance component: $\hat{\sigma}_{QTL_{random}}^2$.

Simulate $\sigma_{QTL_{sim}}^2 = 0.5$ with a small population size $N=50$ (500 replications)

- Figure a: $\hat{Q}_{P_{fix}}$ vs $\hat{Q}_{P_{random}}$ Figure b: $\hat{\sigma}_{QTL_{fixed}}^2$ vs $\hat{\sigma}_{QTL_{random}}^2$



4.2. Polygenic effects and individual relatedness

The phenotypic variation of a complex trait contains contributions of major and minor QTLs. Especially the detection of QTLs with small effects will be hindered by QTLs with larger effects still segregating and adding variation to the error. When testing for the presence of a QTL at a genomic position, it is therefore recommended to control for the segregation of QTLs at other genomic positions. This control can be incorporated via one or more co-factors or via the inclusion of a random polygenic effect with a kinship structuration based on IBS or IBD relations between genotypes. Both approaches alone or combined will reduce the residual error. In this thesis, we often perform multiple rounds of genome-wide QTL identification scans with sets of cofactors that consisted of QTLs identified in earlier rounds of scanning.

We proposed to impose kinship to structure the random polygenic effect when testing for IBD-based random QTLs. Multi-level relatedness in MPPs follows from relations between offspring that can be full-sib, half-sib, or non-related in NAM and diallel designs. More complicated relations are possible in MAGIC designs depending on the way of intercrossing the parents, i.e., the definition of funnels for two- to n-way crosses. Comparable to standard GWAS models, the generic model we adopted in Chapter 2 included multiple multi-allelic QTLs with a polygenic term structured by kinship as IBS relations between genotypes and a residual term with family-specific variances. In Chapter 3 we show that the kinship relations for the polygenic term can also be derived from IBD information and then lead to similar results as the use of an IBS-based kinship matrix does.

In Chapters 4 and 5, we included epistasis and/or QEI in our MPP models, which led to strong computational demands when we wanted to retain a polygenic term with kinship structure on its VCOV. Therefore, we omitted the polygenic term with kinship structure but retained a multi-QTL model whose residual contained mainly genetic and a little bit of non-genetic (plot) variation with a family-specific variance. For most standard and balanced MPPs, the use of multiple cofactors and a family-specific genetic background should be adequate to account for the QTLs elsewhere. We expected to be able to identify most of the QTLs in our iterative QTL identification procedure for each of the individual families, after which a relatively simple residual should remain in the individual families. A proof of this last statement can be found in the observation that for many of our empirical data sets the final multi-QTL model with and without the marker(IBS)-based kinship matrix yielded roughly equal residual errors and similar AIC or BIC values.

5. Genetic architecture in MPPs: additive and non-additive effects

The versatility and flexibility of the IBD-based mixed model approach permit the investigation of complex genetic architectures. Based on the general IBD-based mixed model approach, we modified the design matrices Z and VCOV structures of parental effects G associated with a QTL to differentiate consistent, family-specific, and environment-specific effects (Figure 3) for investigations of additive, epistatic, and QTL-by-environment interacting effects, respectively.

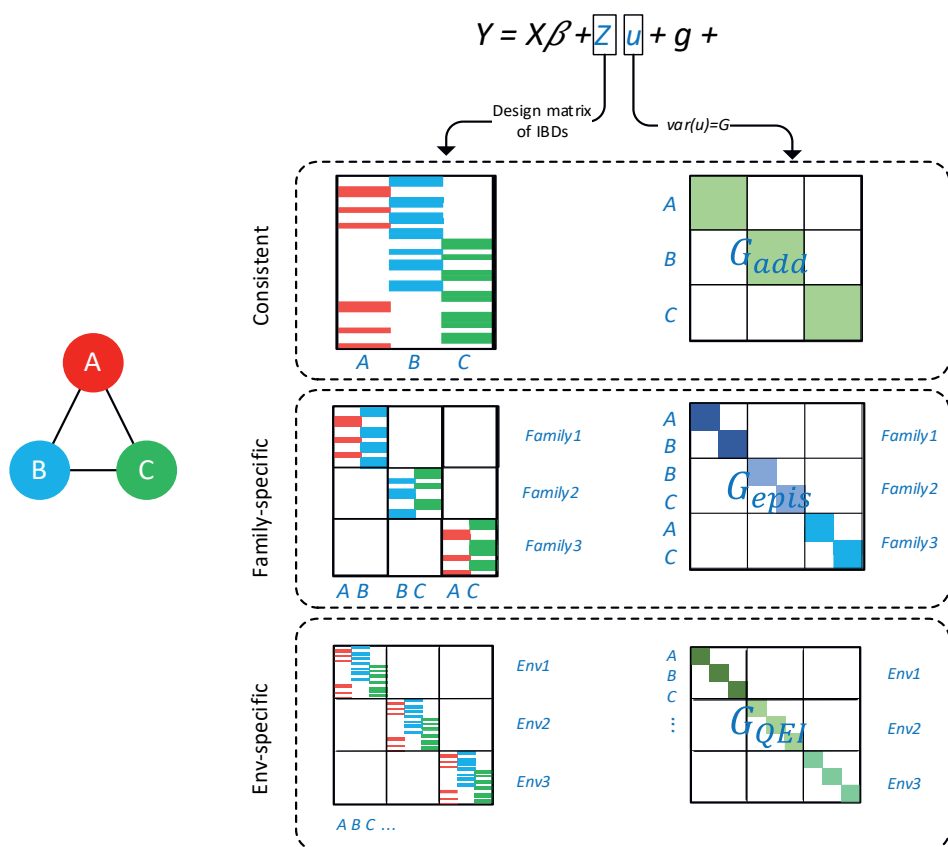


Figure 2. Schematic representation of proposed mixed model approaches in this thesis regarding various IBD-based design matrices and VCOV structures of parental effects corresponding to additive effects, epistasis (in the form of family-specific genetic effects), and QTL-by-environment interaction (in the form of environment-specific effects). We consider a diallel population consisting of three RIL families: A×B, B×C, and A×C. In the Z design matrices, each entry is a function of IBD probabilities in relation to parent A (red), B (blue), and C (green); white denotes 0 probabilities. Vector u indicates the QTL effects related to parents, whose VCOV structure G depends on the types of QTL effects. Entries in G with the same colour represent the same genetic variance, and the intensity of the colour represents the magnitude of the variance. For example, in the consistent model, all parent effects share the same genetic variance; in the family-specific model, family-specific genetic variances are modelled; and in the environment-specific model, genetic variances are different across all three environments.

5.1. Interactions between QTLs (epistasis) and between alleles within QTLs (dominance) in MPPs

The importance of epistasis to understand the biological mechanism of complex traits is well recognized (Causse et al. 2007; Alves et al. 2012; Hansen 2013). For example, prevalent epistatic interactions were identified for fiber quality traits in a cotton MAGIC population using SNP-

based GWAS and IBD-based haplotype GWAS (Wang et al. 2022a); clear interactions between loci for drought tolerance traits were identified by a network-guided approach in a cowpea MAGIC population (Ravelombola et al. 2021). A two-dimensional locus by locus scan is the most popular way to test two-way epistasis. This approach suffers from a multiple-testing problem and is not easily extended to higher-order interactions. An alternative way of testing for epistasis was introduced as a one-dimensional test procedure using multi-cross MPPs to assess epistasis as QTL by family interaction for families that are connected by common parents (Jannink and Jansen 2001; Blanc et al. 2006; Han et al. 2016). In Chapter 4 we generalized the one-dimensional approach with fixed and bi-allelic QTL effects to a similar one-dimensional approach with random and multi-allelic epistatic effects. Our IBD-based one-dimensional scan circumvents the requirements of testing each pair of loci across the genome. Instead, it tests for a family-specific QTL effect at each genomic position, which expresses interaction with the genetic background. Subsequently, locus-by-locus interaction tests can be performed between the earlier identified family-specific QTLs and the remainder of the genome.

The IBD-based one-dimensional scan for testing epistasis is developed for general multi-cross MPPs (NAM and diallel populations) but its efficiency may rely on the specific MPPs. In Chapter 4, we used the simulated NAM and diallel designs to assess the performance of the methodology. We noticed that simulated epistatic QTLs were detected as family-specific QTLs with a higher chance in the NAM population compared to the diallel population. The reason might be that a QTL associated with the reference central parent in the NAM population can be assessed in all families, so the effect of the QTL is estimated in a large population, while a QTL associated with a particular parent in a diallel population can only be evaluated in families involving this parent, reducing the sample size for estimating epistatic QTL effect for the shared parents is limited.

For n-way crosses like MAGIC MPPs, the IBD-based one-dimensional scan for epistasis cannot be used because a standard MAGIC design has only one family. An alternative approach is to add an extra term, $M_q\delta_q$, which represents the interaction of a QTL with any other QTL in the genome. This gives the model $Y = X\beta + Z_q u_q + M_q\delta_q + g + \varepsilon$, where M is the concatenation of the products of the design matrix Z_q at the location q with all the design matrices at a grid along the genome elsewhere. The vector δ_q denotes the epistatic effects of the QTL at the location q with the QTLs elsewhere in the genome. The rest of the terms are as in Figure 3. A one-dimensional test can be adopted with the imposition of a penalty on the estimates for the epistatic effects to avoid overfitting and over-parameterization issues (Boer et al. 2002; Zhang and Xu 2005). The testing of a QTL being epistatic with QTLs elsewhere in the genome can be done by comparing two models. A first model with an appropriate penalty parameter for δ_q that allows epistasis, the alternative model, in contrast to a null model without epistasis in which a large penalty is imposed on the estimates for δ_q (Boer et al. 2002).

This thesis focused on the investigation of inter-locus interactions (epistasis) and ignored the intra-locus interaction (dominance) in non-inbred populations. We return to our *simDiallel* example in Chapter 4 which consisted of exclusively RIL populations, but now introduce a third and fourth scenario where we simulate F2 lines in the A×C family while A×B and B×C families remain RILs (Box 2). Only when family types are identical, we can assume that the heterogeneous VCOV structure of a QTL is caused by epistasis (scenario a vs. b); otherwise, we

cannot distinguish whether it is epistasis or dominance that provokes the heterogeneity of VCOV structure at a QTL (scenario a vs. d). Therefore, when assuming no epistasis, an IBD-based one-dimensional scan can be applied to MPPs with a mixture of family types to assess whether a family-specific QTL is due to dominance or not (scenario a vs. c).

Box 2 Extension of the one-dimensional scan approach from testing epistasis to testing dominance by introducing a mixture of families in a multi-cross MPP.

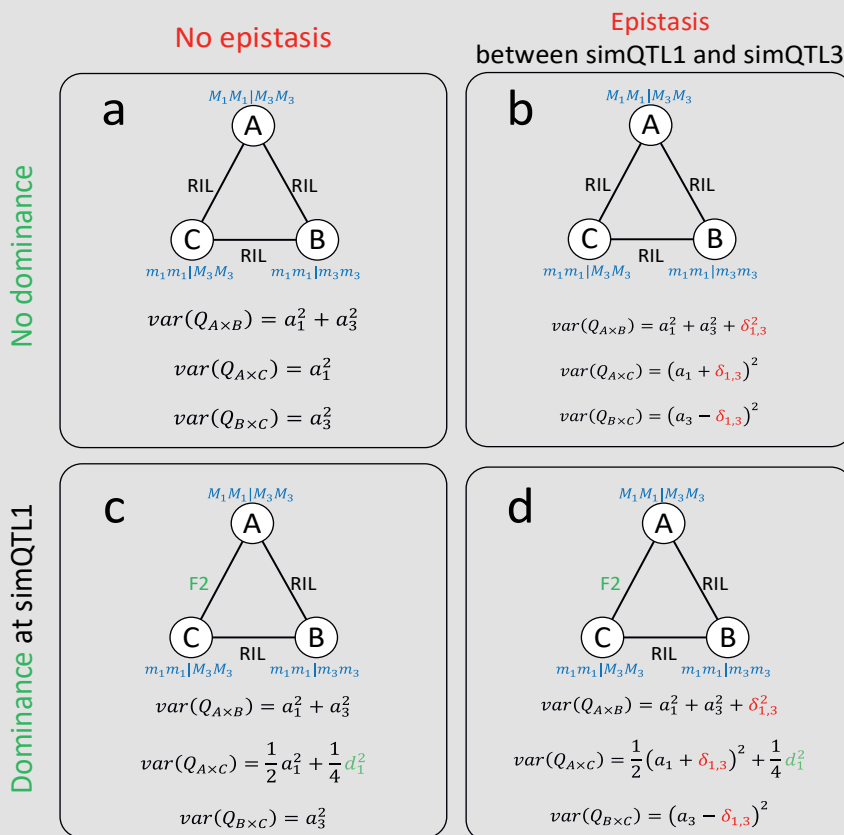
The genetic variance in each family is associated with allele frequencies at QTLs and effect sizes of additive (a_1 or a_3), epistatic ($\delta_{1,3}$), and dominant (d_1) effects.

For example, in the F2 family A×C in scenarios c and d:

$$\text{var}(QTL_{A \times C}) = \text{var}(Z_{a1}a_1 + Z_{a1}Z_{a3}\delta_{1,3} + Z_{d1}d_1)$$

$$\text{Genetic indicators are } Z_{a1} = \begin{cases} -1 (mm) & 1/4 \\ 0 (Mm) & 1/2 \\ 1 (MM) & 1/4 \end{cases}; Z_{a3} = 1 (MM); Z_{d1} = \begin{cases} 0 (mm/MM) & 1/2 \\ 1 (Mm) & 1/2 \end{cases}$$

Four scenarios are shown below depending on whether epistatic and dominance effects are present.



5.2. Multi-trial and multi-trait analysis in MPPs

Multi-environment trials (METs) become popular for MPPs, but the challenge is to develop appropriate statistical tools to assess the stability of QTL effects across both environments and families, simultaneously. The mixed model approaches developed in Chapter 5 offer a good solution to estimate the differential or consistent effect at QTLs over multiple environments in a wide range of empirical MET&MPPs. Different VCOV structures of QTL effects are modelled and compared to select the nature of the QTL effects as one out of four effect types as being consistent or specific across families and environments. Meanwhile, the genetic background effects are considered to correspond to the family or population structure as well as to reflect genetic correlations between environments.

The observation that environment-specific QTLs show differential genetic effects over diverse environments implies QEI, but the exact genetic basis needs to be further investigated. A previous study (Boer et al. 2007) in a maize BPP design regresses the effects of environment-specific QTLs on temperature covariables and it is suggested to use additional environmental information in populations with wider genetic diversity like MPPs. Regressing QTL effects on environmental covariates also benefits the prediction of genotypic behaviours in other environments (Heslot et al. 2014).

The mixed model approach to MET&MPP analysis for QEI detection can be referred to as multivariate mixed models, whose philosophy can be extended to multi-trait (MT) analysis. Multiple traits can occur as multiple physiologically correlated traits, or as traits measured at different growth phases, i.e., longitudinal traits. The multivariate mixed model approach is expected to distinguish between linked QTLs or shared (pleiotropic) QTLs for multiple traits (Verbyla et al. 2014a; Boehm et al. 2019). The analysis becomes more complex when multiple traits are obtained in multiple trials for an MPP design. For instance, a four-way cross was used to map yield and yield components over different conditions (Meng et al. 2016). For a RIL population (Alimi et al. 2013), a mixed model for QTL mapping approach was presented that combines multiple trait and multi-environment aspects. A generalization of this approach to MPPs would be welcome.

6. Other extensions

6.1. Regularization regression approaches

QTL testing is performed through the model selection/comparison techniques between null and alternative models. We performed multiple rounds of genome QTL scans to reach a final multi-QTL model containing a subset of the genetic predictors as QTL candidates. In the generic model (Chapter 2) to map additive QTLs, we decided to adopt a forward selection strategy, so a QTL under testing depends on the identified QTL cofactors from previous genome scans. Other subset selection strategies like backward or stepwise selection could also have been used but are computationally more intensive.

When our initial generic model approach is extended to more advanced models where QTLs are assumed to interact with family genetic background (Chapter 4) and environments

(Chapter 5), variances of QTL effects are not global but become heterogeneous to follow the crossing scheme of an MPP and the diverse environments in METs. The number of model comparisons and tests rapidly increases when models with different QTL definitions need to be compared. Instead of using a forward, backward, or stepwise approach to model building, we may employ a penalized regularization approach, like LASSO, in which the automatic selection of a QTL takes place conditional on the choice of penalty.

When regularized regression is used, the genetic predictors are mostly a function of IBS or bi-allelic information (Whittaker et al. 2000; Zhang and Xu 2005; Piepho 2009; Ogutu et al. 2012). We propose to investigate IBD probabilities as genetic predictors in regularization approaches for MPPs and give an example of such an approach in Box 3.

Box 3 Basics of regularized regression approaches using IBS information and the extension to the use of IBD information in MPPs assuming only additive effects.

- Based on IBS (bi-allelic states):

$$Y = \sum_{m=1}^M X_m a_m + \varepsilon$$

Y is a $N \times 1$ column vector of centred and adjusted genotypic means. $\sum_{m=1}^M X_m a_m$ represents the genetic contributions from all M positions along the genome, where X_m is the $N \times 1$ column vector of the design matrix with entry values 0, 1, or 2 to indicate the number of allele copies at the m -th position; and a_m is allelic effect at the m -th position. The residual term ε follows a normal distribution.

a_m is estimated by minimizing the function:

$$F_{\lambda, \gamma}(a) = \underset{a}{\operatorname{argmin}} \{ (Y - \sum_{m=1}^M X_m a_m)^T (Y - \sum_{m=1}^M X_m a_m) + \lambda \sum_{m=1}^M |a_m|^\gamma \},$$

$(Y - \sum_{m=1}^M X_m a_m)^T (Y - \sum_{m=1}^M X_m a_m)$ is the loss function; and $\lambda \sum_{m=1}^M |a_m|^\gamma$ is the penalty function where the parameter γ determines the type of penalty. If $\gamma = 0$, the estimator is equivalent to subset selection; for LASSO, $\gamma = 1$, and irrelevant markers have effects shrunk to 0; for ridge regression $\gamma = 2$.

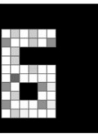
- Based on IBD probabilities:

$$Y = \sum_{m=1}^M Z_m u_m + \varepsilon$$

The $N \times 1$ column vector X_m in the IBS-based model can be replaced by the $N \times P$ (P is the number of parents) design matrix Z_m to estimate the QTL effects related to the parents in the $P \times 1$ column vector u_m at the m -th grid location. The target of the IBD-based model is to identify groups of predictors as QTLs, with each group consisting of the columns of a Z_m matrix. One penalized regression option is the general group LASSO (gLASSO) approach (Ogutu and Piepho 2014), where we would need to minimize the function:

$$gLASSO(u) = \underset{u}{\operatorname{argmin}} \{ \|Y - \sum_{m=1}^M Z_m u_m\|_2^2 + \lambda \sum_{m=1}^M \sqrt{P} \|u_m\|_2 \} \text{ (Yang and Zou 2015).}$$

Optimizing penalty parameters can be performed on the data by grid search from varying λ values based on different validation criteria like the Akaike information criterion (AIC), Bayesian information criterion (BIC), or k -fold cross-validation. Next, the regularization path is computed at the optimal λ to estimate the marker effects a (IBS-based approach) and QTL allele effects u (IBD-based approach).



6.2. Genomic prediction

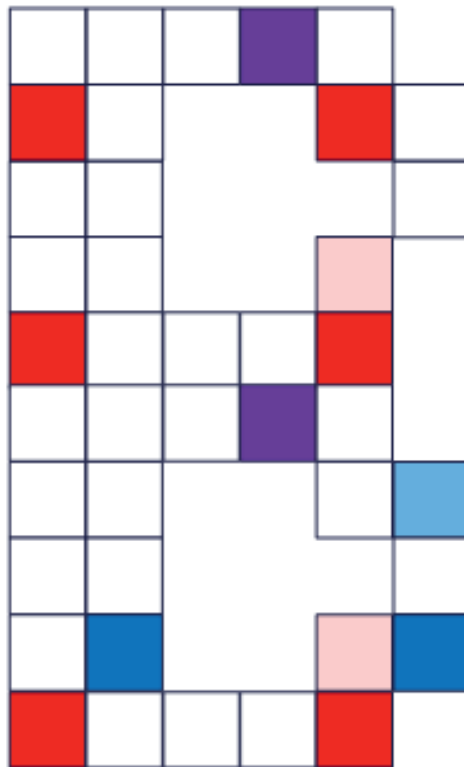
This thesis focused on QTL mapping in MPP designs. An issue arises when the target trait is driven by many QTLs with minor effects (MacKay et al. 2009) that make it challenging to identify QTLs and introgress them into breeding populations. This issue can be mitigated by genomic prediction/selection methods where molecular variation at the whole genome is used to predict or select the desirable phenotype instead of a subset of significant QTLs (Alimi 2016; Lyra et al. 2020). However, genomic prediction methods merely provide statistical prediction power rather than throwing light on the genetic mechanism underlying the target trait. In the context of MPPs, it is interesting to investigate whether using whole-genome IBD information as genetic predictors (such as the approach presented in Box 3) can yield higher predictive ability.

Even though genomic prediction models are superior to using QTL tagging markers alone for trait prediction (Rutkoski et al. 2012; Bustos-Korts et al. 2016), the inclusion of major QTLs as fixed effects alongside a polygenic term structured by kinship is shown to improve prediction accuracy (Bustos-Korts et al. 2016; Zaïm et al. 2020). In future work, the QTLs identified by our IBD-based model approaches can be incorporated into genomic prediction models to predict complex traits with the possibility to consider epistasis and QEI genetic architectures.

7. Epilogue

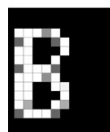
This thesis has exploited the great potential of MPP designs in QTL studies by employing IBD probabilities as genetic predictors and mixed model approaches to estimate QTLs. The IBD probabilities, which are computed using the information of pedigree, whole genome, and crossing schemes, can informatively and precisely describe the genome compositions referring to parent origins. Mixed model approaches provide versatile and flexible solutions to estimate genetic variance at a putative QTL while accounting for the genetic background effects (Chapter 2). This generic IBD-based mixed model approach is extended to more advanced models that allow dissecting genetic architectures underlining complex traits, such as epistasis in multi-cross MPPs (Chapter 4) and QEI analysis in MET&MPPs (Chapter 5). This thesis contributes to unifying the statistical methodologies for QTL mapping in a wide range of MPPs as well as to fully exploiting the potentials in specific MPPs to dissect complex genetic architectures, which will be of benefit to breeding programs and genetic studies.

Bibliography



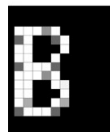
- Abdelraheem A, Thyssen GN, Fang DD, et al (2021) GWAS reveals consistent QTL for drought and salt tolerance in a MAGIC population of 550 lines derived from intermating of 11 Upland cotton (*Gossypium hirsutum*) parents. *Mol Genet Genomics* 296:119–129. <https://doi.org/10.1007/s00438-020-01733-2>
- Abdulkina LR, Kobayashi C, Lovell JT, et al (2019) Components of the ribosome biogenesis pathway underlie establishment of telomere length set point in *Arabidopsis*. *Nat Commun* 10:. <https://doi.org/10.1038/s41467-019-13448-z>
- Abiola O, Angel JM, Avner P, et al (2003) The nature and identification of quantitative trait loci: A community's view. *Nat. Rev. Genet.* 4:911–916
- Aflakparast M, Salimi H, Gerami A, et al (2014) Cuckoo search epistasis: A new method for exploring significant genetic interactions. *Heredity (Edinb)* 112:666–674. <https://doi.org/10.1038/hdy.2014.4>
- Akaike H (1974) A New Look at the Statistical Model Identification. *IEEE Trans Automat Contr* 19:716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Alimi NA (2016) Statistical methods for QTL mapping and genomic prediction of multiple traits and environments: case studies in pepper. <https://doi.org/10.18174/390205>
- Alimi NA, Bink MC, Dieleman JA, et al (2013) Multi-trait and multi-environment QTL analyses of yield and a set of physiological traits in pepper. *Theor Appl Genet* 126:2597–2625. <https://doi.org/10.1007/S00122-013-2160-3/TABLES/12>
- Altendorf KR, DeHaan LR, Larson SR, Anderson JA (2021) QTL for seed shattering and threshability in intermediate wheatgrass align closely with well-studied orthologs from wheat, barley, and rice. *Plant Genome* 14:. <https://doi.org/10.1002/tpg2.20145>
- Alves AA, Rosado CCG, Faria DA, et al (2012) Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of *Eucalyptus* to *Puccinia psidii* rust infection. *Euphytica* 183:27–38. <https://doi.org/10.1007/s10681-011-0455-5>
- Amadeu RR, Muñoz PR, Zheng C, Endelman JB (2021) QTL mapping in outbred tetraploid (and diploid) diallel populations. *Genetics* 219
- Arrones A, Vilanova S, Plazas M, et al (2020) The dawn of the age of multi-parent magic populations in plant breeding: Novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. *Biology (Basel)*. 9:1–25
- Bajgain P, Rouse MN, Tsilo TJ, et al (2016) Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLoS One* 11:e0155760. <https://doi.org/10.1371/journal.pone.0155760>
- Bardol N, Ventelon M, Mangin B, et al (2013) Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. *Theor Appl Genet* 126:2717–2736. <https://doi.org/10.1007/s00122-013-2167-9>
- Bauer E, Falque M, Walter H, et al (2013) Intraspecific variation of recombination rate in maize. *Genome Biol* 14:R103. <https://doi.org/10.1186/gb-2013-14-9-r103>

- Baxter I, Brazelton JN, Yu D, et al (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genet* 6:e1001193. <https://doi.org/10.1371/journal.pgen.1001193>
- Beavis WD (2019) QTL Analyses: Power, Precision, and Accuracy. In: *Molecular Dissection of Complex Traits*. CRC Press, pp 145–162
- Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor Appl Genet* 67:35–43. <https://doi.org/10.1007/BF00303919>
- Bernardo R (2021) Multiparental populations in line development: Genetic gain, diversity, and practical limitations. *Crop Sci* 61:4139–4150. <https://doi.org/10.1002/csc2.20632>
- Beyer P, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11:215–221
- Blanc G, Charcosset A, Mangin B, et al (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: An application in maize. *Theor Appl Genet* 113:206–224. <https://doi.org/10.1007/s00122-006-0287-1>
- Bloom JS, Ehrenreich IM, Loo WT, et al (2013) Finding the sources of missing heritability in a yeast cross. *Nature* 494:234–237. <https://doi.org/10.1038/nature11867>
- Bloom JS, Kotenko I, Sadhu MJ, et al (2015) Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. *Nat Commun* 6:1–6. <https://doi.org/10.1038/ncomms9712>
- Boehm FJ, Chesler EJ, Yandell BS, Broman KW (2019) Testing pleiotropy vs. separate QTL in multiparental populations. *G3 Genes, Genomes, Genet* 9:2317–2324. <https://doi.org/10.1534/g3.119.400098>
- Boer MP, Piepho HP, Williams ER (2020) Linear Variance, P-splines and Neighbour Differences for Spatial Adjustment in Field Trials: How are they Related? *J Agric Biol Environ Stat* 25:676–698. <https://doi.org/10.1007/s13253-020-00412-4>
- Boer MP, Ter Braak CJF, Jansen RC (2002) A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* 162:951–960. <https://doi.org/10.1093/genetics/162.2.951>
- Boer MP, van Rossum B (2021a) statgenIBD: Calculation of IBD Probabilities. <https://CRAN.R-project.org/package=statgenIBD>
- Boer MP, van Rossum B (2021b) LMMsolver: Linear Mixed Model Solver. <https://CRAN.R-project.org/package=LMMsolver>
- Boer MP, Wright D, Feng L, 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. <https://doi.org/10.1534/genetics.107.071068>
- Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Brem RB, Storey JD, Whittle J, Kruglyak L (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436:701–703. <https://doi.org/10.1038/nature03865>



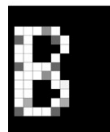
- Broman KW (2001) Review of statistical methods for QTL mapping in experimental crosses. *Lab Anim* (NY) 30:44–52
- Broman KW, Gatti DM, Simecek P, et al (2019) R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* 211:495–502. <https://doi.org/10.1534/genetics.118.301595>
- Broman KW, Gatti DM, Simecek P, et al (2018) R/qtl2: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multi-parent Populations. *Genetics* genetics.301595.2018. <https://doi.org/10.1534/genetics.118.301595>
- Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889–890. <https://doi.org/10.1093/bioinformatics/btg112>
- Browning SR (2008) Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. *Genetics* 178:2123–2132. <https://doi.org/10.1534/genetics.107.084624>
- Buckler ES, Holland JB, Bradbury PJ, et al (2009) The genetic architecture of maize flowering time. *Science* (80-) 325:714–718. <https://doi.org/10.1126/science.1174276>
- Burgos E, Belen De Luca M, Diouf I, et al (2021) Validated MAGIC and GWAS population mapping reveals the link between vitamin E content and natural variation in chorismate metabolism in tomato. *Plant J* 105:907–923. <https://doi.org/10.1111/tpj.15077>
- Bustos-Korts D, Boer MP, Layton J, et al (2022) Identification of environment types and adaptation zones with self-organizing maps; applications to sunflower multi-environment data in Europe. *Theor Appl Genet* 1:1–24. <https://doi.org/10.1007/s00122-022-04098-9>
- Bustos-Korts D, Malosetti M, Chapman S, et al (2016) Improvement of predictive ability by uniform coverage of the target genetic space. *G3 Genes, Genomes, Genet* 6:3733–3747. <https://doi.org/10.1534/g3.116.035410>
- Butler D, Cullis B, Gilmour A, Gogel B (2009) Analysis of Mixed Models for S-language Environments: ASReml–R Reference Manual. *Dep Prim Ind Fish* 160
- Butler DG, Cullis BR, Gilmour AR, et al (2018) ASReml-R Reference Manual Version 4. ASReml-R Ref Man 176
- Campanelli G, Sestili S, Acciarri N, et al (2019) Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy* 9:119. <https://doi.org/10.3390/agronomy9030119>
- Carlborg Ö, Haley CS (2004) Epistasis: Too often neglected in complex trait studies? *Nat. Rev. Genet.* 5:618–625
- Carlborg Ö, Hocking PM, Burt DW, Haley CS (2004) Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. *Genet Res* 83:197–209. <https://doi.org/10.1017/S0016672304006779>
- Carlborg Ö, Kerje S, Schütz K, et al (2003) A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Res* 13:413–421. <https://doi.org/10.1101/gr.528003>
- Causse M, Chaib J, Lecomte L, et al (2007) Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. *Theor Appl Genet* 115:429–442. <https://doi.org/10.1007/s00122-007-0578-1>

- Cheng R, Abney M, Palmer AA, Skol AD (2011) QTLRel: An R Package for Genome-wide Association Studies in which Relatedness is a Concern. *BMC Genet* 12:1–3. <https://doi.org/10.1186/1471-2156-12-66>
- Churchill GA, Airey DC, Allayee H, et al (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36:1133–1137
- Cobb JN, Biswas PS, Platten JD (2019) Back to the future: revisiting MAS as a tool for modern plant breeding. *Theor. Appl. Genet.* 132:647–667
- Coles ND, McMullen MD, Balint-Kurti PJ, et al (2010) Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. *Genetics* 184:799–812. <https://doi.org/10.1534/genetics.109.110304>
- Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. In: *Euphytica*. Springer, pp 169–196
- Collard BCY, Mackill DJ (2008) Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* 363:557–572
- Cordell HJ (2009) Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* 10:392–404
- Courtois B, Ahmadi N, Khowaja F, et al (2009) Rice root genetic architecture: Meta-analysis from a drought QTL database. *Rice* 2:115–128. <https://doi.org/10.1007/s12284-009-9028-9>
- Crainiceanu CM, Ruppert D (2004) Likelihood ratio tests in linear mixed models with one variance component. *J R Stat Soc Ser B Stat Methodol* 66:165–185. <https://doi.org/10.1111/j.1467-9868.2004.00438.x>
- Crepieux S, Lebreton C, Flament P, Charmet G (2005) Application of a new IBD-based QTL mapping method to common wheat breeding population: Analysis of kernel hardness and dough strength. *Theor Appl Genet* 111:1409–1419. <https://doi.org/10.1007/s00122-005-0073-5>
- Crepieux S, Lebreton C, Servin B, Charmet G (2004) IBD-based QTL detection in inbred pedigrees: A case study of cereal breeding programs. *Euphytica* 137:101–109. <https://doi.org/10.1023/B:EUPH.0000040507.44711.93>
- Crews DP (2005) *Phenotypic plasticity: Functional and conceptual approaches*. Oxford University Press
- Crow JF (2010) On epistasis: Why it is unimportant in polygenic directional selection. *Philos. Trans. R. Soc. B Biol. Sci.* 365:1241–1244
- Dang VH, Hill CB, Zhang XQ, et al (2020) Genetic dissection of the interactions between semi-dwarfing genes *sdw1* and *ari-e* and their effects on agronomic traits in a barley MAGIC population. *Mol Breed* 40:. <https://doi.org/10.1007/s11032-020-01145-5>
- Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet* 3:22–32. <https://doi.org/10.1038/nrg701>
- Dell'Acqua M, Gatti DM, Pea G, et al (2015) Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in *Zea mays*. *Genome Biol* 16:167. <https://doi.org/10.1186/s13059-015-0716-z>
- Diaz S, Ariza-Suarez D, Ramdeen R, et al (2021) Genetic Architecture and Genomic Prediction of Cooking Time in Common Bean (*Phaseolus vulgaris* L.). *Front Plant Sci* 11:. <https://doi.org/10.3389/fpls.2020.622213>



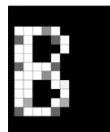
- Diouf IA, Derivot L, Bitton F, et al (2018) Water deficit and salinity stress reveal many specific QTL for plant growth and fruit quality traits in tomato. *Front Plant Sci* 9:279. <https://doi.org/10.3389/fpls.2018.00279>
- Eeuwijk FA Van, Malosetti M, Boer MP (2007) Modelling the Genetic Basis of Response Curves Underlying Genotype \times Environment Interaction. In: *Scale and Complexity in Plant Systems Research*. Springer, pp 115–126
- Ehrenreich IM (2017) Epistasis: Searching for interacting genetic variants using crosses. *Genetics* 206:531–535
- Endelman JB, Plomion C (2014) LPmerge: An R package for merging genetic maps by linear programming. *Bioinformatics* 30:1623–1624. <https://doi.org/10.1093/bioinformatics/btu091>
- Fijneman RJA, De Vries SS, Jansen RC, Demant P (1996) Complex interactions of new quantitative trait loci, Sluc1, Sluc2, Sluc3, and Sluc4, that influence the susceptibility to lung cancer in the mouse. *Nat Genet* 14:465–467. <https://doi.org/10.1038/ng1296-465>
- Flint-Garcia SA, Thornsberry JM, Edwards IV SB (2003) Structure of Linkage Disequilibrium in Plants. *Annu Rev Plant Biol* 54:357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Forsberg SKG, Bloom JS, Sadhu MJ, et al (2017) Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat Genet* 49:497–503. <https://doi.org/10.1038/ng.3800>
- Gage JL, Monier B, Giri A, Buckler ES (2020) Ten years of the maize nested association mapping population: Impact, limitations, and future directions. *Plant Cell* 32:2083–2093. <https://doi.org/10.1105/tpc.19.00951>
- Galiano-Carneiro AL, Kessel B, Presterl T, et al (2021) Multi-parent QTL mapping reveals stable QTL conferring resistance to *Gibberella* ear rot in maize. *Euphytica* 217:. <https://doi.org/10.1007/s10681-020-02748-x>
- Gardner KA, Wittern LM, Mackay IJ (2016) A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnol J* 14:1406–1417. <https://doi.org/10.1111/pbi.12504>
- Garin V, Malosetti M, van Eeuwijk F (2020) Multi-parent multi-environment QTL analysis: an illustration with the EU-NAM Flint population. *Theor Appl Genet* 133:2627–2638. <https://doi.org/10.1007/s00122-020-03621-0>
- Garin V, Wimmer V, Borchardt D, et al (2021) The influence of QTL allelic diversity on QTL detection in multi-parent populations: a simulation study in sugar beet. *BMC Genomic Data* 22:1–12. <https://doi.org/10.1186/s12863-021-00960-9>
- Garin V, Wimmer V, Mezouk S, et al (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. *Theor Appl Genet* 130:1753–1764. <https://doi.org/10.1007/s00122-017-2923-3>
- Gatti DM, Svenson KL, Shabalin A, et al (2014) Quantitative trait locus mapping methods for diversity outbred mice. *G3 Genes, Genomes, Genet* 4:1623–1633. <https://doi.org/10.1534/g3.114.013748>

- Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. *Theor Appl Genet* 46:319–330.
<https://doi.org/10.1007/BF00281673>
- George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in Complex pedigrees: A two-step variance component approach. *Genetics* 156:2081–2092.
<https://doi.org/10.1093/genetics/156.4.2081>
- Gimelfarb A, Lande R (1994) Simulation of marker assisted selection for non-additive traits. *Genet Res* 64:127–136. <https://doi.org/10.1017/S0016672300032730>
- Giraud H, Bauland C, Falque M, et al (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* 207:1167–1180.
<https://doi.org/10.1534/genetics.117.300305>
- Giraud H, Lehermeier C, Bauer E, et al (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* 198:1717–1734.
<https://doi.org/10.1534/genetics.114.169367>
- Gleeson AC, Cullis BR (1987) Residual Maximum Likelihood (REML) Estimation of a Neighbour Model for Field Experiments. *Biometrics* 43:277.
<https://doi.org/10.2307/2531812>
- Gramazio P, Yan H, Hasing T, et al (2019) Whole-Genome Resequencing of Seven Eggplant (*Solanum melongena*) and One Wild Relative (*S. incanum*) Accessions Provides New Insights and Breeding Tools for Eggplant Enhancement. *Front Plant Sci* 10:.
<https://doi.org/10.3389/fpls.2019.01220>
- Grieco M, Schmidt M, Warnemünde S, et al (2022) Dynamics and genetic regulation of leaf nutrient concentration in barley based on hyperspectral imaging and machine learning. *Plant Sci* 315:.
<https://doi.org/10.1016/j.plantsci.2021.111123>
- Griffiths S, Simmonds J, Leverington M, et al (2009) Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. *Theor Appl Genet* 119:383–395. <https://doi.org/10.1007/s00122-009-1046-x>
- Han S, Utz HF, Liu W, et al (2016) Choice of models for QTL mapping with multiple families and design of the training set for prediction of Fusarium resistance traits in maize. *Theor Appl Genet* 129:431–444. <https://doi.org/10.1007/s00122-015-2637-3>
- Hansen TF (2013) Why epistasis is important for selection and adaptation. *Evolution (N Y)* 67:3501–3511. <https://doi.org/10.1111/evo.12214>
- Hautsalo J, Novakazi F, Jalli M, et al (2021) Pyramiding of scald resistance genes in four spring barley MAGIC populations. *Theor Appl Genet* 134:3829–3843.
<https://doi.org/10.1007/s00122-021-03930-y>
- He XH, Zhang YM (2011) A complete solution for dissecting pure main and epistatic effects of QTL in triple testcross design. *PLoS One* 6:24575.
<https://doi.org/10.1371/journal.pone.0024575>
- Heslot N, Akdemir D, Sorrells ME, Jannink JL (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet* 127:463–480.
<https://doi.org/10.1007/s00122-013-2231-5>



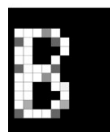
- Holland JB (2007) Genetic architecture of complex traits in plants. *Curr Opin Plant Biol* 10:156–161. <https://doi.org/10.1016/j.pbi.2007.01.003>
- Holland JB, Moser HS, O'Donoghue LS, Lee M (1997) QTLs and epistasis associated with vernalization responses in oat. *Crop Sci* 37:1306–1316. <https://doi.org/10.2135/cropsci1997.0011183X003700040047x>
- Hu J, Xiao G, Jiang P, et al (2022) QTL detection for bread wheat processing quality in a nested association mapping population of semi-wild and domesticated wheat varieties. *BMC Plant Biol* 22:. <https://doi.org/10.1186/s12870-022-03523-x>
- Huang BE, George AW (2011) R/mpMap: A computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics* 27:727–729. <https://doi.org/10.1093/bioinformatics/btq719>
- Huang BE, Verbyla KL, Verbyla AP, et al (2015) MAGIC populations in crops: current status and future prospects. *Theor Appl Genet* 128:999–1017. <https://doi.org/10.1007/s00122-015-2506-0>
- Huang M, Liu X, Zhou Y, et al (2019) BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8:. <https://doi.org/10.1093/gigascience/giy154>
- Huang X, Paulo MJ, Boer M, et al (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proc Natl Acad Sci U S A* 108:4488–4493. <https://doi.org/10.1073/pnas.1100465108>
- Huynh BL, Ehlers JD, Huang BE, et al (2018) A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *Plant J* 93:1129–1142. <https://doi.org/10.1111/tpj.13827>
- Hyun MK, Zaitlen NA, Wade CM, et al (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723. <https://doi.org/10.1534/genetics.107.080101>
- Islam MS, Thyssen GN, Jenkins JN, et al (2016) A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton. *BMC Genomics* 17:1–17. <https://doi.org/10.1186/s12864-016-3249-2>
- Jannink JL, Jansen R (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* 157:445–454. <https://doi.org/10.1093/genetics/157.1.445>
- Jansen RC (2004) Quantitative Trait Loci in Inbred Lines. In: *Handbook of Statistical Genetics*. John Wiley & Sons, Ltd
- Jourjon MF, Jasson S, Marcel J, et al (2005) MCQTL: Multi-allelic QTL mapping in multi-cross design. *Bioinformatics* 21:128–130. <https://doi.org/10.1093/bioinformatics/bth481>
- Jurcic EJ, Villalba P V., Pathauer PS, et al (2021) Single-step genomic prediction of *Eucalyptus dunnii* using different identity-by-descent and identity-by-state relationship matrices. *Heredity (Edinb)* 127:176–189. <https://doi.org/10.1038/s41437-021-00450-9>
- Karademir E, Gençer O (2010) Combining ability and heterosis for yield and fiber quality properties in cotton (*G. hirsutum* L.) obtained by half diallel mating design. *Not Bot Horti Agrobot Cluj-Napoca* 38:222–227. <https://doi.org/10.15835/NBHA3813528>
- Kearsey MJ (1998) The principles of QTL analysis (a minimal mathematics approach). *J. Exp. Bot.* 49:1619–1623

- Kitony JK, Sunohara H, Tasaki M, et al (2021) Development of an aus-derived nested association mapping (Aus-nam) population in rice. *Plants* 10: <https://doi.org/10.3390/plants10061255>
- Korstanje R, Paigen B (2002) From QTL to gene: The harvest begins [1]. *Nat. Genet.* 31:235–236
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9:1–9
- Kover PX, Valdar W, Trakalo J, et al (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 5:e1000551. <https://doi.org/10.1371/journal.pgen.1000551>
- Kumar A, Saripalli G, Jan I, et al (2020) Meta-QTL analysis and identification of candidate genes for drought tolerance in bread wheat (*Triticum aestivum* L.). *Physiol Mol Biol Plants* 26:1713–1725. <https://doi.org/10.1007/s12298-020-00847-6>
- Lander ES, Botstein S (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185. <https://doi.org/10.1093/genetics/121.1.185>
- Leal SM (2001) Genetics and Analysis of Quantitative Traits. *Am J Hum Genet* 68:548–549. <https://doi.org/10.1086/318209>
- Lehermeier C, Krämer N, Bauer E, et al (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics* 198:3–16. <https://doi.org/10.1534/genetics.114.161943>
- Li H, Bradbury P, Ersoz E, et al (2011) Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. *PLoS One* 6: <https://doi.org/10.1371/journal.pone.0017573>
- Li H, Ribaut JM, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor Appl Genet* 116:243–260. <https://doi.org/10.1007/s00122-007-0663-5>
- Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics* 175:361–374. <https://doi.org/10.1534/genetics.106.066811>
- Li W, Boer MP, van Rossum B-J, et al (2022) statgenMPP: an R package implementing an IBD-based mixed model approach for QTL mapping in a wide range of multi-parent populations. *Bioinformatics*. <https://doi.org/10.1093/BIOINFORMATICS/BTAC662>
- Li W, Boer MP, Zheng C, et al (2021) An IBD-based mixed model approach for QTL mapping in multiparental populations. *Theor Appl Genet* 1:1–18. <https://doi.org/10.1007/s00122-021-03919-7>
- Li Z-K (2008) QTL mapping in rice: a few critical considerations. pp 153–171
- Liang F, Zhan W, Hu G, et al (2022) Five plants per RIL for phenotyping traits of high or moderate heritability ensure the power of QTL mapping in a rice MAGIC population. *Mol Breed* 42:1–14. <https://doi.org/10.1007/s11032-022-01299-4>
- Liao CY, Wu P, Hu B, Yi KK (2001) Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. *Theor Appl Genet* 103:104–111. <https://doi.org/10.1007/s001220000528>
- Liller CB, Walla A, Boer MP, et al (2017) Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. *Theor Appl Genet* 130:269–281. <https://doi.org/10.1007/s00122-016-2807-y>



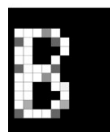
- Lipka AE, Tian F, Wang Q, et al (2012) GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Liu K, Goodman M, Muse S, et al (2003) Genetic Structure and Diversity among Maize Inbred Lines as Inferred from DNA Microsatellites. *Genetics* 165:2117–2128. <https://doi.org/10.1093/genetics/165.4.2117>
- Liu X, Huang M, Fan B, et al (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genet* 12:e1005767. <https://doi.org/10.1371/journal.pgen.1005767>
- Liu Y, Zeng ZB (2000) A general mixture model approach for mapping quantitative trait loci from diverse cross designs involving multiple inbred lines. *Genet Res* 75:345–355. <https://doi.org/10.1017/S0016672300004493>
- Long N, Gianola D, Rosa GJM, Weigel KA (2011) Marker-assisted prediction of non-additive genetic values. *Genetica* 139:843–854. <https://doi.org/10.1007/s10709-011-9588-7>
- López-Malvar A, Malvar RA, Butrón A, et al (2022) Identification of single nucleotide polymorphisms (SNPs) for maize cell wall hydroxycinnamates using a multi-parent advanced generation intercross (MAGIC) population. *Phytochemistry* 193:. <https://doi.org/10.1016/j.phytochem.2021.113002>
- Luo L, Mao Y, Xu S (2003) Correcting the Bias in Estimation of Genetic Variances Contributed by Individual QTL. *Genet* 2003 1192 119:107–114. <https://doi.org/10.1023/A:1026028928003>
- Lyra DH, Virlet N, Sadeghi-Tehran P, et al (2020) Functional QTL mapping and genomic prediction of canopy height in wheat measured using a robotic field phenotyping platform. *J Exp Bot* 71:1885–1898. <https://doi.org/10.1093/JXB/ERZ545>
- Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 12:57–63
- Mackay IJ, Bansept-Basler P, Bentley AR, et al (2014) An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: Creation, properties, and validation. *G3 Genes, Genomes, Genet* 4:1603–1610. <https://doi.org/10.1534/g3.114.012963>
- MacKay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: Challenges and prospects. *Nat Rev Genet* 10:565–577. <https://doi.org/10.1038/nrg2612>
- Malosetti M, Ribaut JM, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Front Physiol* 4 MAR:44. <https://doi.org/10.3389/fphys.2013.00044>
- Malosetti M, Van Der Linden CG, Vosman B, Van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175:879–889. <https://doi.org/10.1534/genetics.105.054932>
- Malosetti M, van Eeuwijk FA, Boer MP, et al (2011) Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. *Theor Appl Genet* 122:1605–1616. <https://doi.org/10.1007/s00122-011-1558-z>

- Malosetti M, Voltas J, Romagosa I, et al (2004) Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137:139–145. <https://doi.org/10.1023/B:EUPH.0000040511.46388.ef>
- Manolio TA, Collins FS, Cox NJ, et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753. <https://doi.org/10.1038/nature08494>
- McMullen MD, Kresovich S, Villeda HS, et al (2009) Genetic properties of the maize nested association mapping population. *Science* (80-) 325:737–740. <https://doi.org/10.1126/science.1174320>
- Melchinger AE, Friedrich Utz H, Schör CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149:383–403. <https://doi.org/10.1093/genetics/149.1.383>
- Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J* 3:269–283. <https://doi.org/10.1016/j.cj.2015.01.001>
- Meng L, Zhao X, Ponce K, et al (2016) QTL mapping for agronomic traits using multi-parent advanced generation inter-cross (MAGIC) populations derived from diverse elite indica rice lines. *F Crop Res* 189:19–42. <https://doi.org/10.1016/j.fcr.2016.02.004>
- Michel KJ, Lima DC, Hundley H, et al (2022) Genetic mapping and prediction of flowering time and plant height in a maize Stiff Stalk MAGIC population. *Genetics*. <https://doi.org/10.1093/genetics/iyac063>
- Mitchell-Olds T (2010) Complex-trait analysis in plants. *Genome Biol.* 11:1–3
- Molenberghs G, Verbeke G (2007) Likelihood ratio, score, and wald tests in a constrained parameter space. *Am Stat* 61:22–27. <https://doi.org/10.1198/000313007X171322>
- Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol* 147:969–977. <https://doi.org/10.1104/pp.108.118232>
- Mott R, Talbot CJ, Turri MG, et al (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci U S A* 97:12649–12654. <https://doi.org/10.1073/pnas.230304397>
- Myles S, Peiffer J, Brown PJ, et al (2009) Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell* 21:2194–2202. <https://doi.org/10.1105/tpc.109.068437>
- Nadeau JH, Frankel WN (2000) The roads from phenotypic variation to gene discovery: Mutagenesis versus QTLs. *Nat. Genet.* 25:381–384
- Odell SG, Hudson AI, Praud S, et al (2022) Modeling allelic diversity of multiparent mapping populations affects detection of quantitative trait loci. *G3 Genes, Genomes, Genet* 12:. <https://doi.org/10.1093/g3journal/jkac011>
- Ogut F, Bian Y, Bradbury PJ, Holland JB (2015) Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. *Heredity (Edinb)* 114:552–563. <https://doi.org/10.1038/hdy.2014.123>
- Ogut JO, Piepho HP (2014) Regularized group regression methods for genomic prediction: Bridge, MCP, SCAD, group bridge, group lasso, sparse group lasso, group MCP and group SCAD. In: *BMC Proceedings*. BioMed Central Ltd., pp 1–9



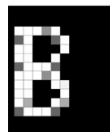
- Ogututu JO, Schulz-Streeck T, Piepho HP (2012) Genomic selection using regularized linear regression models: Ridge regression, lasso, elastic net and their extensions. *BMC Proc* 6:S10. <https://doi.org/10.1186/1753-6561-6-S2-S10>
- Olatoye MO, Marla SR, Hu Z, et al (2020) Dissecting adaptive traits with nested association mapping: Genetic architecture of inflorescence morphology in sorghum. *G3 Genes, Genomes, Genet* 10:1785–1796. <https://doi.org/10.1534/g3.119.400658>
- Ongom PO, Ejeta G (2018) mating design and genetic structure of a multi-parent advanced generation intercross (MAGIC) population of sorghum (*Sorghum bicolor* (L.) moench). *G3 Genes, Genomes, Genet* 8:331–341. <https://doi.org/10.1534/g3.117.300248>
- Paccapelo MV, Kelly AM, Christopher JT, Verbyla AP (2022) WGNAM: whole-genome nested association mapping. *Theor Appl Genet* 1:1–20. <https://doi.org/10.1007/s00122-022-04107-x>
- Papa R, Kapan DD, Counterman BA, et al (2013) Multi-Allelic Major Effect Genes Interact with Minor Effect QTLs to Control Adaptive Color Pattern Variation in *Heliconius erato*. *PLoS One* 8:e57033. <https://doi.org/10.1371/journal.pone.0057033>
- Pascual L, Albert E, Sauvage C, et al (2016) Dissecting quantitative trait variation in the resequencing era: Complementarity of bi-parental, multi-parental and association panels. *Plant Sci* 242:120–130. <https://doi.org/10.1016/j.plantsci.2015.06.017>
- Pascual L, Desplat N, Huang BE, et al (2015) Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol J* 13:565–577. <https://doi.org/10.1111/pbi.12282>
- Pascual L, Xu J, Biais B, et al (2013) Deciphering genetic diversity and inheritance of tomato fruit weight and composition through a systems biology approach. *J Exp Bot* 64:5737–5752. <https://doi.org/10.1093/jxb/ert349>
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. <https://doi.org/10.1093/biomet/58.3.545>
- Paulo MJ, Boer M, Huang X, et al (2008) A mixed model QTL analysis for a complex cross population consisting of a half diallel of two-way hybrids in *Arabidopsis thaliana*: Analysis of simulated data. *Euphytica* 161:107–114. <https://doi.org/10.1007/s10681-008-9665-x>
- Pettersson M, Besnier F, Siegel PB, Carlborg Ö (2011) Replication and explorations of High-Order epistasis using a large advanced intercross line pedigree. *PLoS Genet* 7:e1002180. <https://doi.org/10.1371/journal.pgen.1002180>
- Piepho H-P (1997) Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. *Biometrics* 53:761. <https://doi.org/10.2307/2533976>
- Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. *Crop Sci* 49:1165–1176. <https://doi.org/10.2135/cropsci2008.10.0595>
- Piepho HP (2019) A coefficient of determination (R^2) for generalized linear mixed models. *Biometrical J* 61:860–872. <https://doi.org/10.1002/bimj.201800270>
- Piepho HP (1998) Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theor Appl Genet* 1998 971 97:195–201. <https://doi.org/10.1007/S001220050885>

- Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 11:459–463. <https://doi.org/10.1038/nrg2813>
- Przystalski M, Osman A, Thiemt EM, et al (2008) Comparing the performance of cereal varieties in organic and non-organic cropping systems in different European countries. *Euphytica* 163:417–433. <https://doi.org/10.1007/s10681-008-9715-4>
- R Core Team (2022) R: A language and environment for statistical computing
- Ravelombola W, Shi A, Huynh BL, et al (2022) Genetic architecture of salt tolerance in a Multi-Parent Advanced Generation Inter-Cross (MAGIC) cowpea population. *BMC Genomics* 23:1–22. <https://doi.org/10.1186/s12864-022-08332-y>
- Ravelombola W, Shi A, Huynh BL (2021) Loci discovery, network-guided approach, and genomic prediction for drought tolerance index in a multi-parent advanced generation intercross (MAGIC) cowpea population. *Hortic Res* 8:1–13. <https://doi.org/10.1038/s41438-021-00462-w>
- Riaz A, Kockappelgren P, Hehir JG, et al (2020) Genetic analysis using a multi-parent wheat population identifies novel sources of septoria tritici blotch resistance. *Genes (Basel)* 11:1–27. <https://doi.org/10.3390/genes11080887>
- Ribaut JM, Hoisington D (1998) Marker-assisted selection: New tools and strategies. *Trends Plant Sci* 3:236–239. [https://doi.org/10.1016/S1360-1385\(98\)01240-0](https://doi.org/10.1016/S1360-1385(98)01240-0)
- Rida S, Maafi O, López-malvar A, et al (2021) Genetics of germination and seedling traits under drought stress in a magic population of maize. *Plants* 10:. <https://doi.org/10.3390/plants10091786>
- Rollar S, Geyer M, Hartl L, et al (2021a) Quantitative Trait Loci Mapping of Adult Plant and Seedling Resistance to Stripe Rust (*Puccinia striiformis* Westend.) in a Multiparent Advanced Generation Intercross Wheat Population. *Front Plant Sci* 12:. <https://doi.org/10.3389/fpls.2021.684671>
- Rollar S, Serfling A, Geyer M, et al (2021b) QTL mapping of adult plant and seedling resistance to leaf rust (*Puccinia triticina* Eriks.) in a multiparent advanced generation intercross (MAGIC) wheat population. *Theor Appl Genet* 134:37–51. <https://doi.org/10.1007/s00122-020-03657-2>
- Rossum B-J van, Eeuwijk F van, Boer M, et al (2021a) statgenGxE: Genotype by Environment (GxE) Analysis
- Rossum B-J van, Eeuwijk F van, Boer MP, et al (2021b) statgenSTA: Single Trial Analysis (STA) of Field Trials
- Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* 20:1199–1216. <https://doi.org/10.1105/tpc.108.058131>
- Rutkoski J, Benson J, Jia Y, et al (2012) Evaluation of Genomic Prediction Methods for Fusarium Head Blight Resistance in Wheat. *Plant Genome* 5:51–61. <https://doi.org/10.3835/plantgenome2012.02.0001>
- Sadaiah K, Reddy VN, Kumar SS (2013) Heterosis and combining ability studies for sugar content in sweet corn (*Zea mays saccharata* L.). *IJSRP* 3:1–5



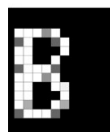
- Sandhu KS, Mihalyov PD, Lewien MJ, et al (2021) Genomic selection and genome-wide association studies for grain protein content stability in a nested association mapping population of wheat. *Agronomy* 11:. <https://doi.org/10.3390/agronomy11122528>
- Satturu V, Vattikuti JL, Durga Sai J, et al (2020) Multiple genome wide association mapping models identify quantitative trait nucleotides for brown planthopper (*Nilaparvata lugens*) resistance in MAGIC Indica population of rice. *Vaccines* 8:1–16. <https://doi.org/10.3390/vaccines8040608>
- Scott MF, Ladejobi O, Amer S, et al (2020) Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity* (Edinb). 125:396–416
- Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82:605–610. <https://doi.org/10.1080/01621459.1987.10478472>
- Seye AI, Bauland C, Giraud H, et al (2019) Quantitative trait loci mapping in hybrids between Dent and Flint maize multiparental populations reveals group-specific QTL for silage quality traits with variable pleiotropic effects on yield. *Theor Appl Genet* 132:1523–1542. <https://doi.org/10.1007/s00122-019-03296-2>
- Sham PC, Purcell SM (2014) Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15:335–346
- Sharp PJ, Johnston S, Brown G, et al (2001) Validation of molecular markers for wheat breeding. *Aust J Agric Res* 52:1357–1366. <https://doi.org/10.1071/ar01052>
- Shi J, Wang J, Zhang L (2019) Genetic mapping with background control for quantitative trait locus (QTL) in 8-parental pure-line populations. *J Hered* 110:880–891. <https://doi.org/10.1093/jhered/esz050>
- Smith A, Cullis B, Thompson R (2001) Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics* 57:1138–1147. <https://doi.org/10.1111/J.0006-341X.2001.01138.X>
- Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *J. Agric. Sci.* 143:449–462
- Song Q, Yan L, Quigley C, et al (2017) Genetic Characterization of the Soybean Nested Association Mapping Population. *Plant Genome* 10:plantgenome2016.10.0109. <https://doi.org/10.3835/plantgenome2016.10.0109>
- Sticca EL, Belbin GM, Gignoux CR (2021) Current Developments in Detection of Identity-by-Descent Methods and Applications. *Front. Genet.* 12:1725
- Stich B, Möhring J, Piepho HP, et al (2008) Comparison of mixed-model approaches for association mapping. *Genetics* 178:1745–1754. <https://doi.org/10.1534/genetics.107.079707>
- Stratz P, Baes C, Rückert C, et al (2013) A two-step approach to map quantitative trait loci for meat quality in connected porcine F2 crosses considering main and epistatic effects. *Anim Genet* 44:14–23. <https://doi.org/10.1111/j.1365-2052.2012.02360.x>
- Sul JH, Martin LS, Eskin E (2018) Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genet.* 14

- Swarts K, Bauer E, Glaubitz JC, et al (2021) Joint analysis of days to flowering reveals independent temperate adaptations in maize. *Heredity* (Edinb) 126:929–941. <https://doi.org/10.1038/s41437-021-00422-z>
- Tao Y, Zhao X, Wang X, et al (2020) Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals. *Plant Biotechnol J* 18:1093–1105. <https://doi.org/10.1111/pbi.13284>
- Taylor J (2018) Efficient linkage map construction using R/ASMap
- Taylor MB, Ehrenreich IM (2014) Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast. *PLoS Genet* 10:e1004324. <https://doi.org/10.1371/journal.pgen.1004324>
- Taylor MB, Ehrenreich IM (2015) Higher-order genetic interactions and their contribution to complex traits. *Trends Genet.* 31:34–40
- Thérèse Navarro A, Tumino G, Voorrips RE, et al (2022) Multiallelic models for QTL mapping in diverse polyploid populations. *BMC Bioinformatics* 23:. <https://doi.org/10.1186/s12859-022-04607-z>
- Thomas WTB (2003) Prospects for molecular breeding of barley. *Ann Appl Biol* 142:1–12. <https://doi.org/10.1111/j.1744-7348.2003.tb00223.x>
- Thompson EA (2013) Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics* 194:301–326. <https://doi.org/10.1534/genetics.112.148825>
- Thudi M, Gaur PM, Krishnamurthy L, et al (2014) Genomics-assisted breeding for drought tolerance in chickpea. *Funct Plant Biol* 41:1178–1190. <https://doi.org/10.1071/FP13318>
- Tisné S, Denis M, Cros D, et al (2015) Mixed model approach for IBD-based QTL mapping in a complex oil palm pedigree. *BMC Genomics* 16:1–12. <https://doi.org/10.1186/s12864-015-1985-3>
- Tsaih SW, Lu L, Airey DC, et al (2005) Quantitative trait mapping in a diallel cross of recombinant inbred lines. *Mamm Genome* 16:344–355. <https://doi.org/10.1007/s00335-004-2466-1>
- Tsutsumi-Morita Y, Heuvelink E, Khaleghi S, et al (2021) Yield dissection models to improve yield: A case study in tomato. In *Silico Plants* 3:1–18. <https://doi.org/10.1093/insilicoplants/diab012>
- Turner SD, Maurizio PL, Valdar W, et al (2018) Dissecting the genetic architecture of shoot growth in carrot (*Daucus carota* L.) using a diallel mating design. *G3 Genes, Genomes, Genet* 8:411–426. <https://doi.org/10.1534/g3.117.300235>
- Utz HF, Melchinger AE, Hohenheim U (2014) PLABQTL : A program for composite interval mapping of. 1–2
- van Eeuwijk FA, Bink MC, Chenu K, Chapman SC (2010a) Detection and use of QTL for complex traits in multiple environments. *Curr. Opin. Plant Biol.* 13:193–205
- van Eeuwijk FA, Boer M, Totir LR, et al (2010b) Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. *Theor Appl Genet* 120:429–440. <https://doi.org/10.1007/s00122-009-1205-0>
- Van Eeuwijk FA, Bustos-Korts D V., Malosetti M (2016) What should students in plant breeding know about the statistical aspects of genotype × Environment interactions? *Crop Sci* 56:2119–2140. <https://doi.org/10.2135/cropsci2015.06.0375>



- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Verbyla AP (2019) A note on model selection using information criteria for general linear models estimated using REML. *Aust New Zeal J Stat* 61:39–50. <https://doi.org/10.1111/anzs.12254>
- Verbyla AP, Cavanagh CR, Verbyla KL (2014a) Whole-genome analysis of multi-environment or multitrait QTL in MAGIC. *G3 Genes, Genomes, Genet* 4:1569–1584. <https://doi.org/10.1534/g3.114.012971>
- Verbyla AP, George AW, Cavanagh CR, Verbyla KL (2014b) Whole-genome QTL analysis for MAGIC. *Theor Appl Genet* 127:1753–1770. <https://doi.org/10.1007/s00122-014-2337-4>
- Verbyla AP, Taylor JD, Verbyla KL (2012) RWGAIM: An efficient high-dimensional random whole genome average (QTL) interval mapping approach. *Genet Res (Camb)* 94:291–306. <https://doi.org/10.1017/S0016672312000493>
- Verhoeven KJF, Jannink JL, McIntyre LM (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity (Edinb)* 96:139–149. <https://doi.org/10.1038/sj.hdy.6800763>
- Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *Am J Hum Genet* 90:7–24. <https://doi.org/10.1016/j.ajhg.2011.11.029>
- Visscher PM, Wray NR, Zhang Q, et al (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* 101:5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Wang J (2009) Inclusive Composite Interval Mapping of Quantitative Trait Genes. *Acta Agron Sin* 35:239–245. <https://doi.org/10.3724/sp.j.1006.2009.00239>
- Wang J, Crossa J, Gai J (2020) Quantitative genetic studies with applications in plant breeding in the omics era. *Crop J* 8:683–687
- Wang M, Qi Z, Thyssen GN, et al (2022a) Genomic interrogation of a MAGIC population highlights genetic factors controlling fiber quality traits in cotton. *Commun Biol* 5:1–12. <https://doi.org/10.1038/s42003-022-03022-7>
- Wang S, Xie F, Xu S (2022b) Estimating genetic variance contributed by a quantitative trait locus: A random model approach. *PLoS Comput Biol* 18:e1009923. <https://doi.org/10.1371/journal.pcbi.1009923>
- Wei J, Xu S (2016) A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. *Genetics* 202:471–486. <https://doi.org/10.1534/genetics.115.179945>
- Wei WH, Knott S, Haley CS, De Koning DJ (2010) Controlling false positives in the mapping of epistatic QTL. *Heredity (Edinb)* 104:401–409. <https://doi.org/10.1038/hdy.2009.129>
- Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. *Genet Res* 75:249–252. <https://doi.org/10.1017/S0016672399004462>
- Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* 4:e1000212. <https://doi.org/10.1371/journal.pgen.1000212>
- Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* 125:201–210

- Würschum T, Liu W, Gowda M, et al (2012) Comparison of biometrical models for joint linkage association mapping. *Heredity* (Edinb) 108:332–340. <https://doi.org/10.1038/hdy.2011.78>
- Würschum T, Maurer HP, Schulz B, et al (2011) Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. *Theor Appl Genet* 123:109–118. <https://doi.org/10.1007/s00122-011-1570-3>
- Xavier A, Jarquin D, Howard R, et al (2018) Genome-wide analysis of grain yield stability and environmental interactions in a multiparental soybean population. *G3 Genes, Genomes, Genet* 8:519–529. <https://doi.org/10.1534/g3.117.300300>
- Xiao Y, Tong H, Yang X, et al (2016) Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytol* 210:1095–1106. <https://doi.org/10.1111/nph.13814>
- Xu S (2007) An empirical Bayes method for estimating epistatic effects of quantitative trait loci. *Biometrics* 63:513–521. <https://doi.org/10.1111/j.1541-0420.2006.00711.x>
- Xu S (2003) Theoretical Basis of the Beavis Effect. *Genetics* 165:2259–2268. <https://doi.org/10.1093/genetics/165.4.2259>
- Xu Y, Li P, Yang Z, Xu C (2017) Genetic mapping of quantitative trait loci in crops. *Crop J* 5:175–184. <https://doi.org/10.1016/j.cj.2016.06.003>
- Yan W, Hunt LA, Sheng Q, Szlavnics Z (2000) Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Sci* 40:597–605. <https://doi.org/10.2135/cropsci2000.403597x>
- Yang J, Zaitlen NA, Goddard ME, et al (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 46:100–106. <https://doi.org/10.1038/ng.2876>
- Yang Y, Zou H (2015) A fast unified algorithm for solving group-lasso penalize learning problems. *Stat Comput* 25:1129–1141. <https://doi.org/10.1007/s11222-014-9498-5>
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551. <https://doi.org/10.1534/genetics.107.074245>
- Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208. <https://doi.org/10.1038/ng1702>
- Zaïm M, Kabbaj H, Kehel Z, et al (2020) Combining QTL Analysis and Genomic Predictions for Four Durum Wheat Populations Under Drought Conditions. *Front Genet* 11:316. <https://doi.org/10.3389/fgene.2020.00316>
- Zhang L, Meng L, Wang J (2019) Linkage analysis and integrated software GAPL for pure-line populations derived from four-way and eight-way crosses. *Crop J* 7:283–293. <https://doi.org/10.1016/j.cj.2018.10.006>
- Zhang S, Meng L, Wang J, Zhang L (2017) Background controlled QTL mapping in pure-line genetic populations derived from four-way crosses. *Heredity* (Edinb) 119:256–264. <https://doi.org/10.1038/hdy.2017.42>
- Zhang YM, Xu S (2005) A penalized maximum likelihood method for estimating epistatic effects of QTL. *Heredity* (Edinb) 95:96–104. <https://doi.org/10.1038/sj.hdy.6800702>



- Zhang Z, Ersoz E, Lai CQ, et al (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42:355–360. <https://doi.org/10.1038/ng.546>
- Zhao K, Aranzana MJ, Kim S, et al (2007) An Arabidopsis example of association mapping in structured samples. *PLoS Genet* 3:0071–0082. <https://doi.org/10.1371/journal.pgen.0030004>
- Zhao S, Li X, Song J, et al (2022) Genetic dissection of maize plant architecture using a novel nested association mapping population. *Plant Genome* 15:. <https://doi.org/10.1002/tpg2.20179>
- Zheng C (2019) RABBIT (v3.2) manual book
- Zheng C, Boer MP, van Eeuwijk FA (2015) Reconstruction of genome ancestry blocks in multiparental populations. *Genetics* 200:1073–1087. <https://doi.org/10.1534/genetics.115.177873>
- Zheng C, Boer MP, Van Eeuwijk FA (2014) A general modeling framework for genome ancestral origins in multiparental populations. *Genetics* 198:87–101. <https://doi.org/10.1534/genetics.114.163006>
- Zheng C, Boer MP, Van Eeuwijk FA (2018) Recursive algorithms for modeling genomic ancestral origins in a fixed pedigree. *G3 Genes, Genomes, Genet* 8:3231–3245. <https://doi.org/10.1534/g3.118.200340>
- Zheng C, Boer MP, Van Eeuwijk FA (2019) Construction of genetic linkage maps in multiparental populations. *Genetics* 212:1031–1044. <https://doi.org/10.1534/genetics.119.302229>
- Zheng D, Yamei W, Yongchao L, et al (2022) Genome-wide Association Analysis of Cadmium Content in Rice Based on MAGIC Population. *Chinese J Rice Sci* 36:35–42. <https://doi.org/10.16819/j.1001-7216.2022.210504>
- Zhu X, Leiser WL, Hahn V, Würschum T (2021a) Identification of QTL for seed yield and agronomic traits in 944 soybean (*Glycine max*) RILs from a diallel cross of early-maturing varieties. *Plant Breed* 140:254–266. <https://doi.org/10.1111/pbr.12900>
- Zhu X, Leiser WL, Hahn V, Würschum T (2021b) Identification of seed protein and oil related QTL in 944 RILs from a diallel of early-maturing European soybean. *Crop J* 9:238–247. <https://doi.org/10.1016/j.cj.2020.06.006>

Summary

The main topic of this thesis is the development of methodologies for QTL mapping in multi-parent populations (MPPs). QTL mapping or analysis is the attempt to discover the co-segregations between molecular markers (genotypes) and observations or measurements of a trait (phenotypes), which is essential for plant breeding and genetic studies. A diverse mapping population that can catch sufficient genetic variants corresponding to phenotypic variations is the fundamental requirement for successful QTL mapping. The desirable properties of MPPs, such as the inclusion of wide genetic diversity and easy-to-control population structure, make MPPs good resources for QTL mapping.

QTL mapping in MPPs is not easy. A first complication arises in how to deal with the multitude of design options in various types of MPPs, such as diallel, nested association mapping (NAM), and multi-parent advanced generation inter-cross (MAGIC) populations. A second complication is how to dissect complex genetic architectures with QTLs that not merely have additive effects, but also non-additive effects like epistasis and QTL-by-Environment interactions (QEIs). In this thesis, statistical methodologies and tools were developed for QTL mapping in various MPPs to deal with the multitude of design options and complex genetic architectures.

The first objective was to develop a generic model for mapping additive QTLs that can be applied to a wide range of MPPs. In **Chapter 2**, we develop an IBD-based mixed model approach, where additive QTL effects were modelled as random effects with a global genetic variance across the population. Design matrices for QTL effects were based on identity-by-descent (IBD) probabilities between parents and offspring that were computed by a general framework of a Hidden Markov Model (HMM). This approach also took care of individual relatedness between and across families by modelling family-specific residual variances and a marker-based kinship matrix for polygenic effect. This generic IBD-based mixed model approach was proven to increase the power and resolution of QTLs by contrast to a benchmark genome-wide association study (GWAS) model in simulated MPPs, and successfully applied to various empirical MPPs.

In **Chapter 3**, an open-source and freely available R package tool called *statgenMPP* was developed. *statgenMPP* incorporates algorithms from *statgenIBD* and *LMMsolver* or *RABBIT* into easy-to-use functionalities of the IBD probability calculation, mixed model solving, and output visualization. *statgenMPP* also provides an efficient way for calculating kinship matrices using IBD probabilities and parallel analysis for fast genome scans for QTL mapping. The efficiency of *statgenMPP* was tested on public data sets from previous studies and empirical data sets from the breeding industry.

The generic model approach developed in **Chapter 2** and **Chapter 3** for mapping additive QTLs in general MPPs was extended to models that could dissect complex genetic architectures. **Chapter 4** presented the one-dimensional scan approach to study epistasis in NAM and diallel populations. Because epistasis can lead to difficult-to-study situations when multi-QTL and high-order interactions are involved, the effects of epistatic QTLs were translated into QTL-by-family interactions or family-specific QTLs in the mixed model approach. A one-dimensional scan assessed the QTL-by-family interaction by testing the parent-specific and family-specific QTLs. This approach was proven to improve the mapping power of simulated digenic (QTL-by-QTL) interactions when contrasted to the generic model of mapping only additive (parent-specific) QTLs. The analysis of complex traits in empirical NAM and diallel populations showed the increased number of identified QTLs attributed to family-specific QTLs.

Chapter 5 extended the generic IBD-based mixed model approach to the detection of QEI for MPPs evaluated in multi-environment trials (METs). The stabilities of QTL effects were assessed across environments and families. QTLs identified in every single environment from previous studies could be detected as stable QTLs across families and environments by our approach. QTLs that were identified in certain environments, estimated with differential effects across environments, or detected for the difference in responses between environments from previous studies were detected as environment-specific QTLs or QEIs by our approach.

Chapter 6 summarized the contributions of MPPs to QTL mapping and plant breeding, addressed the advantages of proposed IBD-based mixed approaches, and discussed the future perspectives of methodological development to further explore the beneficial properties of MPPs.

Acknowledgement

Completing a PhD is a journey of hard work, dedication, and a great deal of support from others. Throughout my journey, I have been blessed to encounter many individuals who have made a profound impact on my life, and I would like to take this opportunity to express my gratitude to them.

First and foremost, I would like to extend my heartfelt thanks to my supervisors **Ronny Joosen, Martin Boer, and Fred van Eeuwijk**.

I am deeply appreciative of the opportunity to meet **Ronny**, my internship supervisor, back in 2016 at Rijk Zwaan. I would also like to extend my sincerest gratitude to **Evert** for giving me the chance to intern at Rijk Zwaan through his interview and selection process at that time. My first day at Rijk Zwaan remains fresh in my memory, particularly because of the unwavering support and guidance I received from Ronny. You imparted your knowledge and expertise to an inexperienced student like myself, and spend the entire afternoon giving me hands-on coaching on the in-house tools. This internship was truly a turning point in my academic and professional journey. As the internship project was about to end, you and Evert initiated this PhD project and offered me the opportunity. At the time, I was struggling with my future prospects and lacked the confidence to undertake such a demanding project, but with your encouragement and belief in me, I found the courage and motivation to take on this challenge. I am also grateful for the continued support and patience that you showed me throughout my PhD journey. Your professional attitude, cheerful demeanor, and unwavering commitment have made my experience as a PhD student an incredibly enriching and enjoyable one. I would like to express my heartfelt gratitude for everything you have done for me, and your invaluable contributions to my growth as a researcher and a scholar.

As my co-promotor and daily supervisor, **Martin** made a tremendous impact on my PhD journey. Throughout my research, you provided unwavering support and guidance that were a constant source of inspiration for me. I knew you during my master's thesis presentation in 2016, where you were in attendance and offered positive feedback on my work. This encounter left a lasting impression on me, and I was fortunate enough to collaborate with you in my internship after the master's project. Later I was also lucky to have you as my PhD supervisor. Your expertise and insight were invaluable in shaping my research. You cared about my personal development and provided plenty of suggestions to enhance my theoretical foundations and practical abilities. Whenever I needed, you gave me prompt responses by email or I could easily drop by your office, and you were always available and patient to discuss various perspectives. Before each presentation in Biometris and Rijk Zwaan, you would review my presentation slides in every detail and particular, and you would keep notes for me in the presentation and provide guidance on the next steps. We had the opportunity to attend the MAGIC workshop in the UK in 2019, where we had a successful collaboration, and we also had a memorable tour in Cambridge. At the end of my PhD journey, you were very responsible and supportive to help me keep my graduation process organized. Your dedication and time cannot be thanked enough, and I am grateful for all that you have

done for me. I am confident that my PhD journey would not have been as successful without your unwavering support, guidance, and encouragement.

Fred, as my promotor in this PhD project, has played a critical role in shaping my research and academic goals. You have been an important source of guidance and support throughout my journey in the field. After my internship in 2016, when I was uncertain about my future in academia and doubted whether I should take the opportunity for the PhD project, you invited me to the Biometris PhD day, from where I was inspired by the diverse and interesting topics discussed by other PhD students of yours and the supportive environment provided by you. You created a great platform for students in Biometris to access a wide range of topics. I am grateful for the opportunity to be a part of such a talented and dedicated group under your guidance. You provided me with the freedom to develop my ideas and improve myself during the first two years of my PhD journey. At the end of my PhD, you initiated a sufficient amount of meetings to provide critical feedback, which helped me to refine my research and bring it to a higher level of quality. I especially appreciate the time you took to review each chapter of my thesis and provide constructive feedback on my writing. The lessons and experiences I have gained during my PhD journey will stay with me for a lifetime. I cannot thank you enough for your unwavering support and guidance, which have been instrumental in my academic growth and success. I am proud to have had you as my promotor and I am confident that your guidance in the PhD journey will continue to shape my future in this field.

I would like to extend my heartfelt thanks to my colleagues at Biometris and Rijk Zwaan for their unwavering support and friendship throughout my PhD journey. The welcoming and friendly atmosphere they provided made my experience at both organizations truly unforgettable. I first met my officemate **Barder**. You invited me to watch movies on the first day in Biometris. Later we met **Georgios**. We shared many memorable lunches and online gaming nights after both of you left Biometris for work. This memory will always hold a special place in my heart. **Yutaka**, you conducted data analysis at Biometris and asked me questions on QTL mapping. I am grateful for your questions and academic talk with me, which helped me to enhance my confidence. As my paranymp, your support and guidance during the graduation process were invaluable. I would also like to extend my thanks to **Chaozhi** for imparting his knowledge of IBD calculations, and to **Bart-Jan** for his invaluable contributions to the R package statgenMPP building. **Jip, James, Patrick, Kevin, Henry, Fredrik, and Emma**, thank you all for creating an enriching platform for discussions and fun activities. My part-time teaching experience in Biomtris in the last two years of my PhD journey was made possible and enjoyable thanks to the organizations of **Saskia, Carolien, Evert-Jan, Eric, Nienke**, and all teachers. From Rijk Zwaan, the experience allowed me to meet wonderful colleagues such as **Lishia, Ram, Hamed, Nathelie, Vahid, and Paula**, who always made my visits to Rijk Zwaan enjoyable and memorable. Special thanks to Lisha, Vahid, and Paula for sometimes providing car-pooling support that made my travel easier. As I embark on the next phase of my journey, I will always treasure the memories and experiences I shared with my colleagues and miss them deeply.

I would like to express my deepest gratitude to my friends who have been an integral part of my journey through the challenging and demanding process of obtaining a PhD degree. **Liyu, Zhengchong, Wei, Shuo, Ziwen, Yuan, Lirong, and Mingzhen** have been my close friends

since 2014 when we arrived in the Netherlands for our master's program. Our bond strengthened, becoming like that of siblings. After some of you left the Netherlands, this community continuously generated happiness when we met **Hao, Lei, Weijia, and Miaoying**. Together, we shared countless hotpot dinners, travel experiences, and gaming nights. Your laughter and unwavering support helped me to stay positive and motivated. I will always cherish our friendship and the memories we have created together. I extend my sincerest gratitude to the amazing couple **Ying & Yang**, for their unwavering generosity in helping me with both my life and studies since my Master's program. Ying, your understanding ear and wise counsel have always been a source of comfort to me. I am grateful for your kindness and consideration. The well-chosen tie you gave me for my defense is deeply appreciated. I also extend my gratitude to another couple, **Zilin & Yan**, who have invited me to many happy events, introduced me to new friends like **Mang, Yifeng, and Qiaofeng**, and kindly accommodated my friend, Huaiyu. **Huaiyu**, we have been friends since our undergraduate program in China and I am so proud of you for pursuing your PhD dreams here in the Netherlands. I would like to extend my appreciation to the final amazing couple, **Xi & Pepijn**, who have provided me with your companionship and treated me like family. I am honored to have a room named after me in your new house. I am also grateful for the opportunity to play an important role in your wedding ceremony as a translator for Xi's family. The countless hours spent on dinners, travels, and good times with you two have been invaluable. Pepijn, as my paronym, thank you for your excellent organization and support during my PhD defense. Your unwavering presence in my life has brought me joy and happiness and I am truly fortunate to call you my family.

For my housemates and friends, **Pol, Eduward, Jildau, Jorian**, and all others who lived in the student house, I extend my sincere gratitude for creating a warm atmosphere in our shared home. Your companionship during our countless happy evenings, cooking and dining together, provided a source of joy and comfort. I am truly grateful for the support, encouragement, and love you have all shown me.

In closing, I would like to dedicate this thesis to my family, as a small token of my appreciation for their unwavering support and encouragement. 感谢我的爸爸妈妈. I would like to express my heartfelt gratitude to my parents for their love, support, and encouragement throughout my life. Your hard work, which took you away from my hometown, has provided me with a great life. Despite the challenges you faced in your marriage when I was a kid, your love and support for me never wavered. Despite limited education opportunities, you always prioritized my education, making wise decisions to ensure I attended better schools. I am deeply grateful to my mother for respecting my choices and being thrilled when I shared my plans to study in the Netherlands. Your emotional and financial support during my studies and life in the Netherlands has been invaluable. I will always be grateful to you for being there to listen and celebrate my achievements. My father, a man of few words, has shown love and support through actions. I cherish the memories of you secretly checking on me when I was homesick as a resident student in middle school, packing my monthly bag with fruits and snacks, and proudly displaying my certificates of merit. Although it is difficult in my culture to express love directly to my parents, I want to say that I love you both and thank you for

everything you have done for me. 感谢我的爷爷奶奶. I would also like to extend my gratitude to my grandparents. Your tireless care and love for me when my parents were away from home for work will always be deeply appreciated. Your unwavering patience and understanding towards a mischievous child will never be forgotten. Although you are unable to witness this milestone in my life today, I carry your love and bravely continue my journey ahead.

In conclusion, I am filled with gratitude as I pen these final words of acknowledgment. To all of you who have been a part of my journey, I extend my heartfelt thanks. Your presence in my life shaped me into the person I am today. Thank you for being there, and for making this journey that I will cherish forever!

About the author

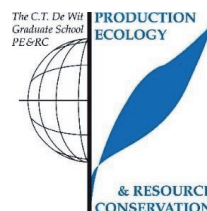
Wenhao Li was born on June 9, 1991, in Sichuan, China, a city known for its panda population and reputation as a hub of leisure and culinary arts. Growing up in this environment shaped Wenhao into an easy-going person, but he didn't leave his hometown for the first 23 years despite China's vast size. One summer during his undergraduate studies at Sichuan Agricultural University (SAU), he suddenly had the idea to study abroad. He picked up English, packed his bags, and prepared for a journey to Wageningen University and Research (WUR) for his master's program. In 2014, he left his hometown for the first time and flew to the Netherlands with excitement. During his time in the EU, Wenhao made up for the lost time in China by traveling to many countries through the EU. At WUR, he focused on plant breeding and wrote his master's thesis on QTL mapping for the resistance of leaf rust in wheat MAGIC populations. This combination of statistics, genetics, and plant pathology sparked his interest, leading him to an internship at Rijk Zwaan and eventually a PhD project in Biometris at WUR in collaboration with Rijk Zwaan in 2017. He enjoyed both his PhD and teaching but eventually had to graduate and move on. On February 28, 2023, he received his PhD diploma and began preparing for the next journey in his life.



Contact me via: wenhao.li1991@outlook.com

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)



Title of review/title project proposal (4.5 ECTS)

- Development of statistical methodologies for QTL mapping in multiparent populations

Post-graduate courses (5.3 ECTS)

- Linear and generalized linear model and linear algebra; Leiden university (2017)
- Julia programming and JVAS; Technische Universität München (2019)
- Bayesian statistics; University of Washington (2020)
- Multivariate analysis; University of Washington (2020)

Deficiency, refresh, brush-up courses (2 ECTS)

- Julia programming; Coursera (2018)

Competence strengthening/skills courses (1.9 ECTS)

- The essentials of scientific writing and presenting; Wageningen into Languages (2019)
- Critical thinking and argumentation; Wageningen Graduate Schools (2020)
- Career assessment; Wageningen Graduate Schools (2022)
- Short workshop introduction to LaTeX; PE&RC (2022)

Scientific integrity/ethics in science activities (0.6 ECTS)

- Scientific integrity; Social science Group (2022)

PE&RC Annual meetings, seminars and the PE&RC weekend (2.1 ECTS)

- PE&RC First years weekend (2018)
- PE&RC Midterm weekend (2019)
- PE&RC Last years weekend (2021)

Discussion groups / local seminars or scientific meetings (10.5 ECTS)

- Whealbi workshop (2017)
- VVSOR annual meeting (2018)
- Booking reading and discussion (2019-2021)
- ARRI forum (2020-2022)

International symposia, workshops and conferences (5.2 ECTS)

- Eucarpia biometrics; poster presentation; Ghent university (2017)
- MAGIC workshop; oral presentation; University College London (2019)
- Eucarpia biometrics; oral presentation; Paris-Saclay University (2022)

Lecturing/supervision of practicals/tutorials (29.4 ECTS)

- Linear model and linear algebra (2019)
- Advanced statistics (2020-2022)
- Statistics 1 (2020-2022)
- Statistics 2 (2020-2022)
- R for statistics (2022)

The research in this thesis has been financially supported by Rijk Zwaan Breeding B.V.



ABCDABCDABCDABCDABCD