



WAGENINGEN  
UNIVERSITY & RESEARCH

## Multiparental QTL analysis: can we do it in polyploids?

Acta Horticulturae

Thérèse Navarro, A.; Tumino, G.; Visser, R.G.F.; Voorrips, R.E.; Weg, W.E. et al

<https://doi.org/10.17660/ActaHortic.2020.1283.9>

This article is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this article please contact [openscience.library@wur.nl](mailto:openscience.library@wur.nl)

# Multiparental QTL analysis: can we do it in polyploids?

A. Thérèse Navarro<sup>a</sup>, G. Tumino, R.G.F. Visser, R.E. Voorrips, W.E. van de Weg and C. Maliepaard  
Plant Breeding, Wageningen University & Research, Wageningen, The Netherlands.

## Abstract

Many ornamental crops are polyploid or even exist at different ploidy levels. Polyploid QTL analysis tools have been developed in recent years, yet they are limited in the population types they accept. Biparental populations are nowadays being regarded as a limited tool for QTL discovery, as only a limited number of QTLs occurs in an experimental cross and their effects might not be stable across genetic backgrounds. Genome-Wide Association Studies include more genetic diversity but suffer from (hidden) genetic structure and low frequency of QTL alleles. Both factors influence QTL detection and effect estimation, decreasing the sensitivity of QTL analysis. Alternatively, multiparental populations (MPP) can be used, potentially combining multiple QTLs and QTL alleles with known population structure and balanced allele frequencies. Breeding populations of interconnected crosses also constitute a form of MPP and QTLs identified in them might be more applicable to commercial cultivars. To perform QTL analysis in polyploids, mixed models or Bayesian approaches that consider pedigree information are recommended. During the analysis, QTL effects are ideally estimated using Identity by Descent (IBD) alleles (genomic regions that originate from the same ancestor) which can be obtained through haplotype estimation. Although MPPs could thus be a powerful set-up to estimate polyploid haplotypes, a software gap was identified as no current polyploid haplotyping tools are able to utilize MPP pedigree information to obtain haplotypes across an MPP. In order to utilize MPPs to their full extent and expand polyploid QTL analyses to encompass typical breeding populations, new haplotyping tools must be developed.

**Keywords:** breeding populations, IBD, GWAS, family-based analysis, pedigree-based analysis

## INTRODUCTION

Polyploidy, the multiplicity of genome copies within a cell, is an important evolutionary phenomenon that has played a crucial role in plant evolution (Comai, 2005; Soltis and Soltis, 2012). This genetic condition has also been utilized in breeding, particularly in the ornamental field, due to its direct effects on organ size and morphology, and its ability to restore fertility in interspecific hybrids. In fact, in the recent book *Ornamental Crops* (van Huylenbroeck, 2018), in which molecular breeding techniques in ornamentals are reviewed, virtually all crops mentioned deal with polyploidy in one or more of these cases: i) in natural polyploid or mixed ploidy populations, ii) in cultivars that had been unconsciously selected for polyploidy, iii) in plants with induced polyploidy to alter morphology or bridge interspecific fertility barriers. Given the interest of moving from classical to molecular breeding approaches, it is essential to develop and expand methodologies that allow breeders and researchers to analyze organisms that differ from the diploid standard.

One of those techniques is quantitative trait loci (QTL) mapping. The term QTL arose almost accidentally in a mathematical article by Geldermann (1975), in which he described a marker-based method to associate variation in a *quantitative trait* with genetic *loci* (in a population of segregating individuals). Although Geldermann did not pioneer the idea, his acronym was rapidly adopted and has nowadays become an essential tool in breeding and research. Polyploid QTL models were proposed early on (Kempthorne 1957), but their application has lagged in comparison to diploids until genotyping technologies and computational resources were good enough to handle polyploid genetic complexity (Doerge

<sup>a</sup>E-mail: alejandro.theresenavarro@wur.nl



and Craig, 2000; Xie and Xu, 2000).

Interpretation of a QTL analysis and its results depends directly on the population type, the genotyping platform and the statistical method used to detect QTLs. For instance, a QTL found in an F<sub>2</sub> biparental population identifies those genomic regions *for which the parents are polymorphic* and whose variation in the F<sub>2</sub> can be associated with phenotypic variation, suggesting a link between the genes in that area and the trait in question. Although the usefulness of this method has allowed a great variety of functional genes to be uncovered, the limitations of this approach are well known: QTLs detected in a biparental population might not be functional in other genetic backgrounds and not all causative loci can be detected due to the limited genetic diversity of the population's parents.

Alternatively, genome-wide association studies (GWAS) can be used, where a group of genetically diverse individuals (generally with unknown relationship between them) are used. In these populations, linkage between marker and QTL allele is due to evolutionary linkage disequilibrium (LD) rather than the LD caused by recent relatedness in biparental populations. Despite their great usefulness, GWAS designs remain controversial (Tam et al. 2019). Although they allow to study a wider range of genetic diversity, effects from rare or weak QTL alleles cannot be adequately estimated, and thus are generally missed. Additionally, population structure and allelic diversity act as confounding factors that must be taken into account in order to avoid statistical artifacts (Yu et al., 2006; Korte and Farlow, 2013).

Nevertheless, both approaches represent the two extremes of a gradient of population diversity (Würschum, 2012). Alternatively, multiple biparental populations that share parents can be used. We will refer to them as multiparental populations (MPP), although in literature they are also known as pedigreed populations, connected populations or families. MPPs harbour a higher level of genetic diversity than a single biparental population, allow testing parental genes in multiple genetic backgrounds, and can be expected to have a more balanced genetic structure compared to GWAS panels composed of a random sample of diverse individuals. Moreover, MPPs are particularly suited for the type of exploratory crossing that is common in breeding efforts: a few interesting parents are selected, intercrossed and a small population is raised from each cross. While traditionally each cross is analyzed separately, the MPP approach proposes joining all crosses in a single analysis.

In this article, we consider the existing MPP populations in plant breeding and discuss the statistical implications of using MPPs in QTL analysis, with special attention to the analytical issues that arise with polyploid genetic analysis.

## **MULTIPARENTAL POPULATION TYPES**

### **Experimental populations**

In plants, experimental MPPs have been developed for a long time. Diallel crosses, populations where a set of parents are crossed in all possible combinations (full diallel) or omitting reciprocal crosses (half diallel), were and are still used in breeding since the definition of general and specific combining abilities were laid down by Sprague and Tatum (1942). However, they were not developed for QTL analysis, but as a form of evaluating parental contributions to hybrids, obtaining an evaluation of the quality of a parent as a source for breeding (Griffing, 1956).

More recently, complex MPP schemes have been developed. Multi-parent Advanced Generation Intercross (MAGIC) populations (Cavanagh et al., 2008) have already been developed in a variety of crops, both diploid (e.g. maize, rice, tomato) and (allo-)polyploid (e.g. wheat, peanut) (Huang et al., 2015). The principle of MAGIC is to combine alleles from different founders in a single genome, and thus evaluate each QTL allele in a variety of different genetic backgrounds. Another MPP scheme is termed nested association mapping (NAM), in which a central parent is crossed with a set of peripheral parents, followed by a series of backcrosses and selfings. This population design has been adopted in fewer crops, but nevertheless with great success. The maize NAM population (McMullen et al., 2009) is undoubtedly its most famous example, but other NAMs have also been developed in sorghum (Bouchet et al., 2017), soybean (Song et al., 2017), barley (Maurer et al., 2015) and a single

polyploid: wheat (Bajgain et al., 2016).

### Breeding populations

Although the experimental MPP schemes mentioned above are useful innovations, they represent only a small fraction of existing MPPs. In breeding programmes, it is common practice to select phenotypically interesting parents and cross them together or with other cultivars, in order to explore new trait combinations. Thus, each parent contributes to many biparental populations, and these sets of connected crosses can be regarded as an MPP.

Breeding populations are regularly screened for interesting phenotypes, and genotyping these populations is becoming an increasingly common practice. Thus, both genotypic and phenotypic data are likely to already be available for these ad-hoc MPPs. Additionally, it has been suggested that utilizing such populations for QTL detection provides results that are more readily translatable into breeding application due to their direct connection to the final cultivars (Jansen et al., 2003; Verhoeven et al., 2006; Würschum, 2012; Bink et al., 2012; Bardol et al., 2013; Han et al., 2016).

For these reasons, development of analytical methods that allow QTL analysis in polyploid MPPs (Figure 1) is especially relevant for breeding efforts, as it will allow breeders to describe and explore more deeply the breeding material present in a program.

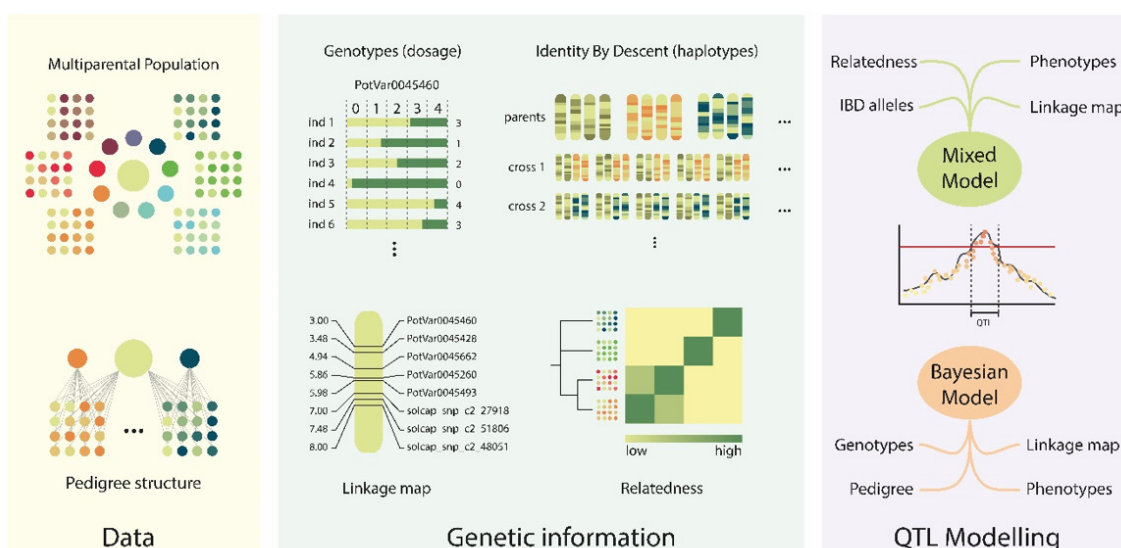


Figure 1. Required elements for polyploid multiparental population (MPP) analysis. Left panel: a multiparental population is obtained. Its pedigree structure can be used in later analysis. Middle panel: raw genetic data can be used to estimate: i) genotypes (dosages) to determine heterozygote classes, ii) haplotypes, to determine which parents and offspring are identical by descent along chromosomes, iii) linkage map, to determine marker distance and order, and iv) genetic structure, as a relatedness matrix, from pedigree and or genotype information. Right panel: phenotypes and linkage maps are used to predict QTL positions. With mixed models, genetic structure is accounted for by the relatedness matrix and IBD alleles, while in Bayesian models that is achieved by incorporating pedigree information in the QTL models.

### STATISTICAL CONSIDERATIONS

#### Genetic structure

QTL analyses rely on association between marker and QTL alleles (also known as allelic disequilibrium, linkage disequilibrium, gametic phase disequilibrium or gametic

disequilibrium), which occurs when alleles of two loci are found together more often than by chance (Flint-Garcia et al., 2003). Physical linkage is not the only possible cause of LD: bottlenecks, genetic drift, natural selection, domestication, breeding history or recent relatedness can generate long-range LD across a population, even on different chromosomes, a property generally called *population structure* (Yu et al., 2006).

The presence of population structure is one of the major differences between biparental populations and MPPs. In biparental populations all individuals are “equally” related (i.e. that any two individuals have the same probability of having a level of relatedness  $k$ ), but in MPPs, those individuals that originate from the same cross (full siblings) are more highly related than those that originate from different crosses (half-siblings or unrelated). Importantly, unlike in GWAS panels, these patterns of relatedness are predictable and can be incorporated in the MPP design process in order to obtain a balanced genetic structure.

A direct consequence of genetic structure is a non-random distribution of alleles across the population, which may lead to an increased rate of type I errors when genetic structure correlates to the phenotype. To resolve this issue, genetic structure must be considered. Two statistical frameworks have been used so far for this purpose: mixed models, where a relatedness matrix is used (Yu et al., 2006); and Bayesian QTL models in which pedigree information is included in their likelihoods (Bink et al., 2014). Strictly speaking, only mixed models have been applied using a polyploid model (Ferrão et al., 2018) through the R package GWASpoly (Rosyara et al., 2016). However, the diploid models of FlexQTL™ (Bink et al., 2014) have been used in allo-octoploid strawberry by treating each subgenome separately (Mangandi et al., 2017; Verma et al., 2017; Anciro et al., 2018). Expansion of this Bayesian framework to polyploidy would thus be an interesting development for autopolyploids and allopolyploids for which genotyping tools cannot obtain subgenome-specific genotypes.

### **Modeling QTLs in multiparental populations**

QTL detection requires estimation of QTL effects. The number and properties of these effects are determined by the type of situation we expect to encounter. In general, we can summarize MPP QTL modeling in four categories (Han et al., 2016):

#### **1. Each effect is specific for each cross and parent genotype (full model in Jannink and Jansen, 2001; disconnected model in Blanc et al., 2006).**

It is assumed that there is no shared information between crosses. We know that protein and gene functions might vary depending on the context in which they are expressed (i.e. genetic background, environmental factors), but completely disregarding shared information between crosses is an over-conservative approach that ignores the potential to increase statistical power by using MPPs.

#### **2. Parental alleles are unique and effects are shared between crosses (reduced model in Jannink and Jansen, 2001; connected model in Blanc et al., 2006; parental model in Garin et al., 2017).**

QTLs are estimated across families, but we assume that each parent contributes different alleles to the MPP. While that might be realistic when all parents originate from completely divergent genepools, it does not reflect most breeding populations, where parents have some degree of relatedness and thus might share alleles at certain loci.

#### **3. Identity-by-descent (IBD) segments among parents harbour the same alleles, with identical effects between crosses (HaploMQM- model in Jansen et al., 2003; LDLA model in Bardol et al., 2013; ancestral model in Garin et al., 2017).**

Two alleles are said to be identical by descent if they have originated from a common ancestor. This method requires the identification of parental alleles that are IBD and estimation of the effects for each unique allele. If IBD can be estimated (see below) and QTL effects are stable between crosses, this is the most powerful method in MPPs (Jansen et al., 2003).

#### **4. Effects are estimated per marker and identical between crosses (model-B in Würschum, 2012; biallelic model in Garin et al., 2017).**

In this case it is assumed that a marker allele indicates QTL allele, i.e. that *each* QTL allele is in linkage with a different marker allele.

Since we expect multiple QTL alleles, this would require a multiallelic marker system densely spread across the genome. These requirements are not yet met by any modern marker system: high-density SNPs are mostly biallelic, and multiallelic markers lack the high-density (and cost effectiveness) of SNPs. Whole genome sequencing could meet these criteria, but fully reconstructing all chromosomes (haplotyping of the whole genome) is still impossible in polyploids (see below).

#### **CONSEQUENCES OF POLYPLOIDY IN MPP ANALYSIS**

Polyploid and diploid MPPs do not differ qualitatively. Although some population parameters such as genetic drift, heterozygosity and allele frequencies are different in polyploids, these do not impact significantly the design or properties of MPPs. Complications arise on the analytical side during genotyping, linkage mapping and haplotyping.

#### **The polyploid genotyping problem**

Unlike diploids, polyploids present multiple heterozygote classes which must be distinguished. When two alleles (e.g. A and C) are detected, they must be quantified in order to estimate their *dosage* (ACCC, AACC or AAAC). Various tools have been developed for this using both fluorescence intensities of SNP arrays (Voorrips et al., 2011; Serang et al., 2012; Schmitz Carley et al., 2017; Zych et al., 2019) and read data from genotyping by sequencing (Gerard et al., 2018). These tools have helped not only in genotyping polyploids, but also in understanding the types of uncertainties that arise with each measurement technique (e.g. background fluorescence or allele bias in SNP arrays, sequencing error or overdispersion in sequence counts of GBS). In order to improve genotype estimation, allele frequency expectations in a population are usually included. There are two common frequency assumptions: a biparental  $F_1$  population, where frequencies depend directly on parental genotypes; and Hardy-Weinberg equilibrium, which is useful in randomly sampled natural populations (e.g. Voorrips et al., 2011; Gerard et al., 2018). To accommodate MPPs, frequency expectations must be adapted to reflect the structure of an MPP, i.e. to model multiple  $F_1$ s. Such work has already been performed for fitTetra (Zych et al., 2019) and there exist programs initially developed with MPPs in mind (Schmitz Carley et al., 2017). These are positive improvements, yet more complex MPPs (e.g. combining pedigreed individuals and  $F_1$ s) might still require further developments.

#### **The polyploid mapping problem**

Linkage mapping is an important tool as it allows to characterize the recombination behavior along the chromosomes of a species. More importantly, in order to detect QTLs, marker order must be known. Allopolyploids, that segregate as diploids, can use diploid mapping software to obtain a linkage map. However, autopolyploids, that generally segregate polysomically require dedicated models. Few programs are available for mapping in autopolyploids (Hackett et al., 2014; Bourke et al., 2018), the most flexible being polymapR (Bourke et al., 2018), as it can estimate integrated maps under different ploidies and segregation models (bivalent, preferential or multivalent pairing of chromosomes). Nevertheless, the package can only estimate maps in  $F_1$  populations. In order to generate a map for an MPP either multiple  $F_1$  maps should be generated and integrated, or polymapR should be adapted to accept other population structures. In any case, current methodologies do not allow to generate linkage maps for (auto)polyploid MPPs.

#### **The polyploid haplotyping problem**

More specifically relevant to MPPs is estimation of IBD. One can speak of two “kinds” of IBD: on the one hand family-IBD, e.g. regions of chromosomes from two offspring individuals that originate from the same *parental chromosome*; and on the other, ancestral-IBD, those

chromosomal regions that originate from the same *common ancestor* and that are broken down by recombination events (Browning and Browning, 2011a). The latter might even span across closely related species.

Generally, IBD is estimated using haplotypes (e.g. Meuwissen et al., 2001), the concatenation of adjacent polymorphisms, most commonly SNPs. Finding the haplotypes is simple when the number of possible combinations is low, (with high homozygosity and low ploidy). In heterozygous polyploids the issue of finding the underlying haplotypes becomes increasingly complex due to the great number of possible haplotypes. Combining polymorphisms to form haplotypes is a process known as phasing or haplotyping and can be performed in different ways depending on the type of data one uses.

Firstly, genotypes can be obtained as independent polymorphism scores (e.g. from a SNP array), or as sequence reads, where each read might contain information about multiple polymorphic sites, thus providing short-range SNP phase information. Secondly, we might wish to resolve the haplotypes of a single individual, or of a group of individuals (related or unrelated).

Haplotypes of a single individual can currently only be resolved using sequence reads, as independent SNP scores do not allow us to decide between the multiple haplotype possibilities. There already exist multiple tools that can phase polyploid haplotypes using next-generation sequencing (NGS) reads (Aguiar and Istrail, 2012; Berger et al., 2014; Das and Vikalo, 2015; He et al., 2018), and for short reads these will provide accurate results if sequencing is performed at adequate depth (Motazediz et al., 2018a). In the future, haplotype library methods based on inferring the most likely haplotype given a set of previously identified haplotypes (e.g. Pook et al., 2019), might provide haplotyping solutions for single individuals if SNP-arrays are used. These methods, however, are still in development even for diploids and thus might take some time to reach polyploids.

Phasing using populations has, comparatively, received less attention. Browning and Browning (2011b) reviewed existing methods for diploids, and divided them in two main groups: i) phasing methods for unrelated individuals, which use either coalescent theory to haplotype likelihood via Hidden Markov Models (Scheet and Stephens, 2006; Li et al., 2010; Browning and Browning, 2011a) or a parsimony principle (Neigenfind et al., 2008); and ii) phasing models for related individuals, in which pedigree information and Mendelian constraints allow to determine likely haplotypes (Abecasis et al., 2002). Since 2011, other population haplotyping tools have been released for polyploids for independent SNPs in  $F_1$ s (Zheng et al., 2016), for long reads in pedigreed individuals (Garg et al., 2016) and for short reads in parent-offspring trios (Motazediz et al., 2018b).

None of the above-mentioned methods can currently exploit information across MPPs. Thus, there is no available methodology that is able to transform unphased SNP genotypes in polyploid MPPs into the multiallelic markers that are required to apply the modeling strategy 3 described above. This gap does not allow to fully utilize multiparental population QTL detection methods in polyploids and represents a lag of polyploid methodology with respect to diploids.

## CONCLUSIONS

Multiparental populations (MPPs) are an interesting prospect that could allow to identify and utilize QTLs with more relevance for breeding applications. In that regard, we must consider also MPPs beyond experimental populations and realize that the breeding populations of interconnected crosses that are regularly generated as a form of exploratory cultivar evaluation also constitute useful MPPs.

QTL modeling of MPPs is more challenging than in biparental crosses due to genetic structure and higher allelic diversity, but mixed models have shown their usefulness in analyzing polyploid MPPs and Bayesian models, if adapted to polyploid organisms, could also prove a useful tool. Regarding the models of QTL effects, IBD-based (haplotype) estimates are the most theoretically consistent method to perform QTL analysis in MPPs, as they account both for family-based linkage and possible sharing of alleles between parents. However, estimation of IBD in polyploids is a challenging task and no method has yet been developed



that can adequately obtain haplotypes across MPPs fully capitalizing on the shared information between crosses, either from sequencing reads or from unphased SNPs.

Polyloid genetic tools are usually developed as extensions from less general diploid models. Similarly, MPP polyloid tools must be developed as generalizations of methodologies that were developed for application in different population types. To that end it will be useful to look both at polyloid tools for biparental and GWAS populations and at diploid tools for MPP analysis, harvesting developments from both fields.

## ACKNOWLEDGEMENTS

This paper was written as part of the Polyloids project “Novel genetic and genomic tools for polyloid crops” (project numbers BO-26.03-009-004 and BO-50-002-022), funded by the Dutch Top Sector Alliance for Knowledge and Innovation (TKI Topsector): horticulture and propagation materials. The author would also like to acknowledge the contribution of Paul Arens, Peter Bourke and René Smulders for their useful guidance on writing this manuscript.

## Literature cited

- Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* *30* (1), 97–101 <https://doi.org/10.1038/ng786>. PubMed
- Aguiar, D., and Istrail, S. (2012). HapCompass: a fast cycle basis algorithm for accurate haplotype assembly of sequence data. *J. Comput. Biol.* *19* (6), 577–590 <https://doi.org/10.1089/cmb.2012.0084>. PubMed
- Anciro, A., Mangandi, J., Verma, S., Peres, N., Whitaker, V.M., and Lee, S. (2018). *FaRCg1*: a quantitative trait locus conferring resistance to *Colletotrichum* crown rot caused by *Colletotrichum gloeosporioides* in octoploid strawberry. *Theor. Appl. Genet.* *131* (10), 2167–2177 <https://doi.org/10.1007/s00122-018-3145-z>. PubMed
- Bajgain, P., Rouse, M.N., Tsilo, T.J., Macharia, G.K., Bhavani, S., Jin, Y., and Anderson, J.A. (2016). Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLoS One* *11* (5), e0155760 <https://doi.org/10.1371/journal.pone.0155760>. PubMed
- Bardol, N., Ventelon, M., Mangin, B., Jasson, S., Loywick, V., Couton, F., Derue, C., Blanchard, P., Charcosset, A., and Moreau, L. (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* (11), 2717–2736 <https://doi.org/10.1007/s00122-013-2167-9>. PubMed
- Berger, E., Yorukoglu, D., Peng, J., and Berger, B. (2014). HapTree: a novel Bayesian framework for single individual polyplotting using NGS data. *PLOS Comput. Biol.* *10* (3), e1003502 <https://doi.org/10.1371/journal.pcbi.1003502>. PubMed
- Bink, M.C.A.M., Totir, L.R., ter Braak, C.J.F., Winkler, C.R., Boer, M.P., and Smith, O.S. (2012). QTL linkage analysis of connected populations using ancestral marker and pedigree information. *Theor. Appl. Genet.* *124* (6), 1097–1113 <https://doi.org/10.1007/s00122-011-1772-8>. PubMed
- Bink, M.C.A.M., Jansen, J., Madduri, M., Voorrips, R.E., Durel, C.-E., Kouassi, A.B., Laurens, F., Mathis, F., Gessler, C., Gobbin, D., et al. (2014). Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theor. Appl. Genet.* *127* (5), 1073–1090 <https://doi.org/10.1007/s00122-014-2281-3>. PubMed
- Blanc, G., Charcosset, A., Mangin, B., Gallais, A., and Moreau, L. (2006). Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theor. Appl. Genet.* *113* (2), 206–224 <https://doi.org/10.1007/s00122-006-0287-1>. PubMed
- Bouchet, S., Olatoye, M.O., Marla, S.R., Perumal, R., Tesso, T., Yu, J., Tuinstra, M., and Morris, G.P. (2017). Increased power to dissect adaptive traits in global Sorghum diversity using a nested association mapping population. *Genetics* *206* (2), 573–585 <https://doi.org/10.1534/genetics.116.198499>. PubMed
- Bourke, P.M., van Geest, G., Voorrips, R.E., Jansen, J., Kranenburg, T., Shahin, A., Visser, R.G.F., Arens, P., Smulders, M.J.M., and Maliepaard, C. (2018). polymapR-linkage analysis and genetic map construction from F1 populations of outcrossing polyloids. *Bioinformatics* *34* (20), 3496–3502 <https://doi.org/10.1093/bioinformatics/bty371>. PubMed
- Browning, B.L., and Browning, S.R. (2011a). A fast, powerful method for detecting identity by descent. *Am. J. Hum. Genet.* *88* (2), 173–182 <https://doi.org/10.1016/j.ajhg.2011.01.010>. PubMed
- Browning, S.R., and Browning, B.L. (2011b). Haplotype phasing: existing methods and new developments. *Nat. Rev. Genet.* *12* (10), 703–714 <https://doi.org/10.1038/nrg3054>. PubMed



- Cavanagh, C., Morell, M., Mackay, I., and Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* *11* (2), 215–221 <https://doi.org/10.1016/j.pbi.2008.01.002>. PubMed
- Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* *6* (11), 836–846 <https://doi.org/10.1038/nrg1711>. PubMed
- Das, S., and Vikalo, H. (2015). SDhaP: haplotype assembly for diploids and polyploids via semi-definite programming. *BMC Genomics* *16* (1), 260 <https://doi.org/10.1186/s12864-015-1408-5>. PubMed
- Doerge, R.W., and Craig, B.A. (2000). Model selection for quantitative trait locus analysis in polyploids. *Proc. Natl. Acad. Sci. USA* *97* (14), 7951–7956 <https://doi.org/10.1073/pnas.97.14.7951>. PubMed
- Ferrão, L.F.V., Benevenuto, J., Oliveira, I.B., Cellon, C., Olmstead, J., Kirst, M., Resende, M.F.R., Jr., and Munoz, P. (2018). Insights into the genetic basis of blueberry fruit-related traits using diploid and polyploid models in a GWAS context. *Front. Ecol. Evol.* *6*, 107 <https://doi.org/10.3389/fevo.2018.00107>.
- Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S., 4th. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* *54* (1), 357–374 <https://doi.org/10.1146/annurev.arplant.54.031902.134907>. PubMed
- Garg, S., Martin, M., and Marschall, T. (2016). Read-based phasing of related individuals. *Bioinformatics* *32* (12), i234–i242 <https://doi.org/10.1093/bioinformatics/btw276>. PubMed
- Garin, V., Wimmer, V., Mezouk, S., Malosetti, M., and van Eeuwijk, F. (2017). How do the type of QTL effect and the form of the residual term influence QTL detection in multi-parent populations? A case study in the maize EU-NAM population. *Theor. Appl. Genet.* *130* (8), 1753–1764 <https://doi.org/10.1007/s00122-017-2923-3>. PubMed
- Geldermann, H. (1975). Investigations on inheritance of quantitative characters in animals by gene markers I. *Methods. Theor. Appl. Genet.* *46* (7), 319–330 <https://doi.org/10.1007/BF00281673>. PubMed
- Gerard, D., Ferrão, L.F.V., Garcia, A.A.F., and Stephens, M. (2018). Genotyping polyploids from messy sequencing data. *Genetics* *210* (3), 789–807 <https://doi.org/10.1534/genetics.118.301468>. PubMed
- Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* *9* (4), 463–493 <https://doi.org/10.1071/BI9560463>.
- Hackett, C.A., Bradshaw, J.E., and Bryan, G.J. (2014). QTL mapping in autotetraploids using SNP dosage information. *Theor. Appl. Genet.* *127* (9), 1885–1904 <https://doi.org/10.1007/s00122-014-2347-2>. PubMed
- Han, S., Utz, H.F., Liu, W., Schrag, T.A., Stange, M., Würschum, T., Miedaner, T., Bauer, E., Schön, C.-C., and Melchinger, A.E. (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* (2), 431–444 <https://doi.org/10.1007/s00122-015-2637-3>. PubMed
- He, D., Saha, S., Finkers, R., and Parida, L. (2018). Efficient algorithms for polyploid haplotype phasing. *BMC Genomics* *19* (S2, Suppl 2), 110 <https://doi.org/10.1186/s12864-018-4464-9>. PubMed
- Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P., Leung, H., Varshney, R.K., and Cavanagh, C.R. (2015). MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.* *128* (6), 999–1017 <https://doi.org/10.1007/s00122-015-2506-0>. PubMed
- Jannink, J.L., and Jansen, R. (2001). Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* *157* (1), 445–454. PubMed
- Jansen, R.C., Jannink, J.-L., and Beavis, W.D. (2003). Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. *Crop Sci.* *43* (3), 829–834 <https://doi.org/10.2135/cropsci2003.8290>.
- Kempthorne, O. (1957). *An introduction to genetic statistics* (New York: John Wiley & Sons, Inc.).
- Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* *9* (1), 29 <https://doi.org/10.1186/1746-4811-9-29>. PubMed
- Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* *34* (8), 816–834 <https://doi.org/10.1002/gepi.20533>. PubMed
- Mangandi, J., Verma, S., Osorio, L., Peres, N.A., van de Weg, E., and Whitaker, V.M. (2017). Pedigree-based analysis in a multiparental population of octoploid strawberry reveals qtl alleles conferring resistance to *Phytophthora cactorum*. *G3 Genes, Genomes. G3* (Bethesda) *7* (6), 1707–1719 <https://doi.org/10.1534/g3.117.042119>. PubMed
- Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E., Kilian, B., Reif, J.C., and Pillen, K. (2015). Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* *16* (1), 290 <https://doi.org/10.1186/s12864-015-1459-7>. PubMed
- McMullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C.,

- Bottoms, C., et al. (2009). Genetic properties of the maize nested association mapping population. *Science* 325 (5941), 737–740 <https://doi.org/10.1126/science.1174320>. PubMed
- Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4), 1819–1829. PubMed
- Motazed, E., Finkers, R., Maliepaard, C., and de Ridder, D. (2018a). Exploiting next-generation sequencing to solve the haplotyping puzzle in polyploids: a simulation study. *Brief. Bioinform.* 19 (3), 387–403. PubMed
- Motazed, E., de Ridder, D., Finkers, R., Baldwin, S., Thomson, S., Monaghan, K., and Maliepaard, C. (2018b). TriPoly: haplotype estimation for polyploids using sequencing data of related individuals. *Bioinformatics* 34 (22), 3864–3872 <https://doi.org/10.1093/bioinformatics/bty442>. PubMed
- Neigenfind, J., Gyetvai, G., Basekow, R., Diehl, S., Achenbach, U., Gebhardt, C., Selbig, J., and Kersten, B. (2008). Haplotype inference from unphased SNP data in heterozygous polyploids based on SAT. *BMC Genomics* 9 (1), 356 <https://doi.org/10.1186/1471-2164-9-356>. PubMed
- Pook, T., Schlather, M., de Los Campos, G., Mayer, M., Schoen, C.C., and Simianer, H. (2019). HaploBlocker: creation of subgroup specific haplotype blocks and libraries. *Genetics* 212 (4), 1045–1061 <https://doi.org/10.1534/genetics.119.302283>. PubMed
- Rosyara, U.R., De Jong, W.S., Douches, D.S., and Endelman, J.B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome* 9 (2), 9 <https://doi.org/10.3835/plantgenome2015.08.0073>. PubMed
- Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78 (4), 629–644 <https://doi.org/10.1086/502802>. PubMed
- Schmitz Carley, C.A., Coombs, J.J., Douches, D.S., Bethke, P.C., Palta, J.P., Novy, R.G., and Endelman, J.B. (2017). Automated tetraploid genotype calling by hierarchical clustering. *Theor. Appl. Genet.* 130 (4), 717–726 <https://doi.org/10.1007/s00122-016-2845-5>. PubMed
- Serang, O., Mollinari, M., and Garcia, A.A.F. (2012). Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. *PLoS One* 7 (2), e30906 <https://doi.org/10.1371/journal.pone.0030906>. PubMed
- Soltis, P.S., and Soltis, D.E. (2012). *Polyploidy and Genome Evolution* (Berlin, Heidelberg: Springer).
- Song, Q., Yan, L., Quigley, C., Jordan, B.D., Fickus, E., Schroeder, S., Song, B.-H., Charles An, Y.-Q., Hyten, D., Nelson, R., et al. (2017). Genetic characterization of the soybean nested association mapping population. *Plant Genome* 10 (2), 2 <https://doi.org/10.3835/plantgenome2016.10.0109>. PubMed
- Sprague, G.F., and Tatum, L.A. (1942). General vs. specific combining ability in single crosses of Corn. *Agron. J.* 34 (10), 923–932 <https://doi.org/10.2134/agronj1942.00021962003400100008x>.
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20 (8), 467–484 <https://doi.org/10.1038/s41576-019-0127-1>. PubMed
- Van Huylenbroeck, J. (2018). *Ornamental Crops* (Springer International Publishing).
- Verhoeven, K.J.F., Jannink, J.-L., and McIntyre, L.M. (2006). Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity* 96 (2), 139–149 <https://doi.org/10.1038/sj.hdy.6800763>. PubMed
- Verma, S., Zurn, J.D., Salinas, N., Mathey, M.M., Denoyes, B., Hancock, J.F., Finn, C.E., Bassil, N.V., and Whitaker, V.M. (2017). Clarifying sub-genomic positions of QTLs for flowering habit and fruit quality in U.S. strawberry (*Fragaria×ananassa*) breeding populations using pedigree-based QTL analysis. *Hortic. Res.* 4 (1), 17062 <https://doi.org/10.1038/hortres.2017.62>. PubMed
- Voorrips, R.E., Gort, G., and Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* 12 (1), 172 <https://doi.org/10.1186/1471-2105-12-172>. PubMed
- Würschum, T. (2012). Mapping QTL for agronomic traits in breeding populations. *Theor. Appl. Genet.* 125 (2), 201–210 <https://doi.org/10.1007/s00122-012-1887-6>. PubMed
- Xie, C., and Xu, S. (2000). Mapping quantitative trait loci in tetraploid populations. *Genet. Res.* 76 (1), 105–115 <https://doi.org/10.1017/S0016672399004395>. PubMed
- Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38 (2), 203–208 <https://doi.org/10.1038/ng1702>. PubMed
- Zheng, C., Voorrips, R.E., Jansen, J., Hackett, C.A., Ho, J., and Bink, M.C.A.M. (2016). Probabilistic multilocus haplotype reconstruction in outcrossing tetraploids. *Genetics* 203 (1), 119–131 <https://doi.org/10.1534/genetics.115.185579>. PubMed

Zych, K., Gort, G., Maliepaard, C.A., Jansen, R.C., and Voorrips, R.E. (2019). FitTetra 2.0 - improved genotype calling for tetraploids with multiple population and parental data support. *BMC Bioinformatics* 20 (1), 148 <https://doi.org/10.1186/s12859-019-2703-y>. PubMed