Theor Appl Genet (2012) 125:201–210
DOI 10.1007/s00122-012-1887-6

REVIEW

Mapping QTL for agronomic traits in breeding populations

Tobias Würschum

Received: 9 January 2012 / Accepted: 27 April 2012 / Published online: 22 May 2012

Abstract Detection of quantitative trait loci (QTL) in breeding populations offers the advantage that these QTL are of direct relevance for the improvement of crops via knowledge-based breeding. As phenotypic data are routinely generated in breeding programs and the costs for genotyping are constantly decreasing, it is tempting to exploit this information to unravel the genetic architecture underlying important agronomic traits in crops. This review characterizes the germplasm from breeding populations available for QTL detection, provides a classification of the different QTL mapping approaches that are available, and highlights important considerations concerning study design and biometrical models suitable for QTL analysis.

Introduction

Most important agronomic traits are quantitatively inherited as opposed to traits that are controlled by a single gene (monogenic) or by a few genes (oligogenic) (Lynch and Walsh 1998; Falconer and Mackay 1996). The nature of quantitative traits is that their expression is controlled by tens, hundreds, or even thousands of quantitative trait loci (QTL), most of them having only a small effect on the trait (Mackay et al. 2009). Since the advent of molecular markers, researchers and breeders have aimed to identify functional markers associated with these QTL for implementation of marker-assisted selection (Dekkers and Hospital 2002). Historically QTL detection started with linkage mapping in biparental populations. As the results were often not transferable to other populations, researchers opted for meta-QTL studies, and more recently for the joint analysis of multiple segregating populations. In addition, technological advances in recent years leading to an abundance of markers even for polyploid species and the option for full-genome sequencing for crops (Ganal et al. 2009; Varshney et al. 2009) have driven the development of novel QTL mapping approaches. Different aspects of QTL mapping have been addressed in recent reviews (e.g., Jansen 2007; Nordborg and Weigel 2008; Myles et al. 2009; Van Eeuwijk et al. 2010b) but these have not focused on the specific requirements of QTL detection in breeding populations, i.e. populations derived from applied breeding programs.

This review aims to bridge this gap and provide an overview of the current status of methods used for QTL detection in breeding populations. In particular the objectives are to (1) characterize the germplasm available for QTL detection in breeding programs, (2) classify the different mapping approaches with regard to the composition of the mapping population, the use of identity-by-descent information, the exploitation of linkage disequilibrium, and the biometrical model used for analysis, (3) outline important considerations resulting from the choice of a QTL mapping approach and (4) discuss the contribution of epistasis to the expression of complex traits in breeding populations.

Properties of breeding populations

Breeding populations are different from natural populations in that instead of undergoing natural selection they

Communicated by R. Varshney.

T. Würschum (✉)
State Plant Breeding Institute, University of Hohenheim,
70593 Stuttgart, Germany
e-mail: tobias.wuerschum@uni-hohenheim.de

are subject to selection by the breeder. In addition, the mating process is also controlled by the breeder. Thus, some of the forces affecting natural populations are also active in breeding populations (e.g., genetic drift), but their composition and genetic properties are tightly controlled. As breeding populations represent elite material their genetic basis is generally much smaller than that of natural populations. Hence, QTL affecting the trait in natural populations may not be segregating in breeding populations due to fixation. Consequently, QTL mapping experiments based on crosses between elite and exotic material have a high chance of detecting QTL, but these are often only of relevance for applied plant breeding for the introgression of traits from exotic material, but not for the selection within elite material (Jannink et al. 2001). By contrast, QTL mapping in elite breeding material offers the possibility to dissect the genetic architecture underlying quantitative traits in breeding populations and to identify QTL which are of direct relevance for breeders.

All breeding programs follow a certain general scheme (Fig. 1a). New genotypic variation is generated by crosses between promising lines. The resulting families will, in early generations (e.g. F_2), be rather large and consist of unselected material. These plants are only phenotyped at few locations and the heritability is low or moderate. The selection intensity is high and consequently mid-generation lines are already selected and family sizes reduced. These plants are phenotyped more intensively and further selection results in highly selected lines in late generations, which are phenotyped most intensively resulting in the highest heritabilities. This iterative process is initiated every year such that at each time-point, plant material from all different stages is available representing a mixture of breeding populations with related individuals.

Classical QTL mapping is achieved using biparental populations. In breeding programs these early-generation families are often of insufficient population size to warrant a high QTL detection power and are phenotyped with low heritability. Thus, single families from breeding programs will in most cases be inappropriate for QTL detection. QTL mapping experiments can rather be carried out on multiple early- or mid-generation families, or with a diverse panel of late-generation lines that at the same time represent the parents of future crosses. Both the selection of promising parental lines from the breeding pool and the selection of superior progenies within families can be facilitated by the detection of QTL. The biometrical approaches available for QTL detection, even though not specifically designed for that purpose, reflect these two scenarios.

It must be noted here that with the exception of families directly derived from F_1 plants (F_2 or DH), the lines in breeding populations that are used for QTL analysis are phenotypically preselected resulting in unequal allele frequencies at the loci under selection. This unidirectional

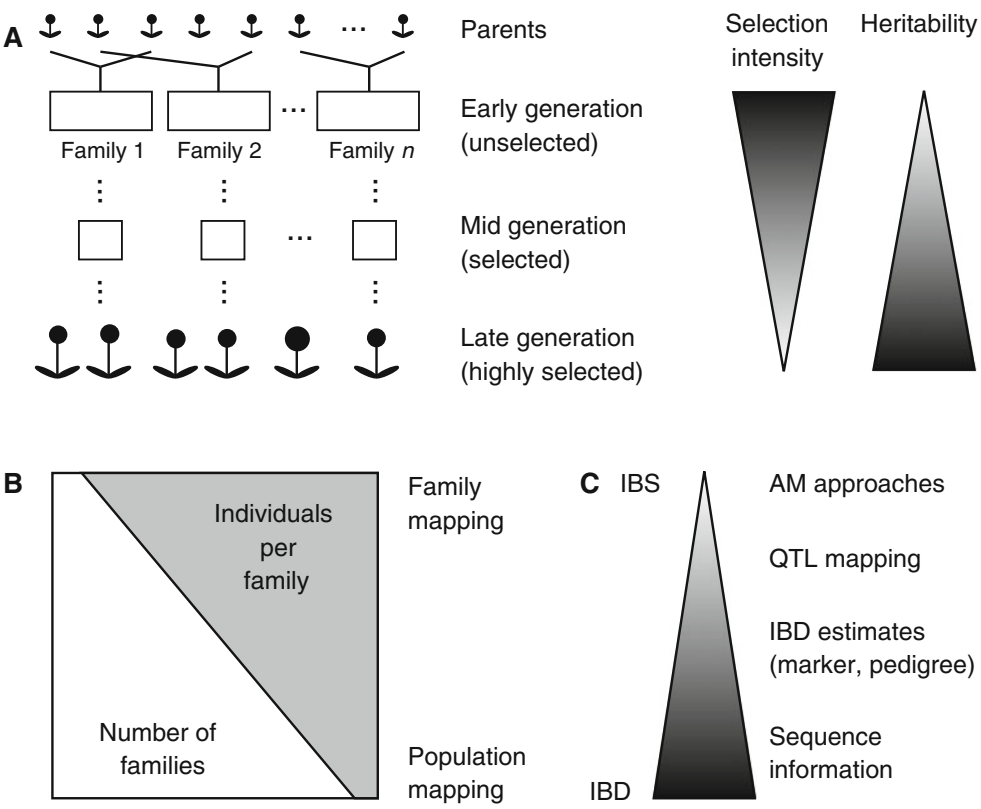

Fig. 1 Schematic representation of **a** the germplasm available from breeding programs, **b** the composition of population and family mapping populations, and **c** the use of identity-by-state (IBS) and identity-by-descent (IBD) information by the different mapping approaches

selection process can have a significant effect on the QTL detection power and the estimation of QTL effects. Biometrical approaches to account for this problem in the analysis have been devised for biparental populations (e.g., Melchinger et al. 2012), but still need to be developed for populations comprising multiple families.

Classification of QTL mapping approaches

Since the first QTL mapping experiments in biparental populations, many QTL mapping approaches have been implemented which are in part similar but also possess conceptual differences. The following section provides a classification of the different QTL mapping approaches.

Family and population mapping

One distinguishing feature is whether QTL detection is done based on segregating families or on a diverse panel of lines. Following Myles et al. (2009) these are further referred to as family mapping and population mapping, respectively (Fig. 1b). The latter represents a diverse panel of lines but can also be regarded as many families with very small family size (i.e. only one individual per family in the most extreme case). Nevertheless, there is no clear distinction but rather a smooth transition between family and population mapping (Fig. 1b). While family mapping investigates only few alleles per locus at a time (the alleles present in the parents of the families), population mapping includes many more alleles in the analysis (potentially as many as there are lines) and therefore represents a wider sample of germplasm and genetic backgrounds.

Identity-by-state and identity-by-descent

Another way of characterizing the different mapping approaches is by their utilization of identity-by-state (IBS) or identity-by-descent (IBD) information (Fig. 1c). IBS refers to alleles that are identical irrespective of whether they are inherited from a common ancestor. IBD estimates on the other hand provide a measure of whether alleles that are IBS are copies of the same ancestral allele; that is whether they are identical by descent. Conceptually IBD estimates can also be interpreted as predictions of IBS at unobserved loci (Powell et al. 2010).

IBS approaches have the disadvantage that plants carrying identical marker alleles are grouped together, but in case linkage disequilibrium (LD) between the marker and the QTL is not complete, not all plants with this marker allele will carry the positive QTL allele (Fig. 2). This is especially true for biallelic marker types such as commonly used SNPs. The grouping of markers to haplotypes (Fig. 2b) or the estimation of IBD probabilities helps to alleviate this problem (Jansen et al. 2003). Association mapping approaches as used nowadays operate with identity-by-state information of each marker. By contrast, linkage mapping methodology is an identity-by-descent approach that traces the parental origin of the alleles. However, information beyond the parents is not considered and all parents are regarded as unrelated, an assumption that is certainly not valid in breeding populations. In order to incorporate relationship information above the parent's level, more sophisticated identity-by-descent approaches have been suggested (e.g., Wu and Zheng 2001; Meuwissen et al. 2002; Van Eeuwijk et al. 2010a). These use pedigree and/or marker information to obtain IBD estimates at each tested locus which can be integrated into the model in the form of a symmetric matrix of IBD probabilities among all lines.

The availability of full-genome sequence data will enable a further step toward the true relationships between the genotypes of the mapping population. On the other hand, sequence data will also identify all causal polymorphisms such that the mapping approaches will no longer rely on LD between a marker and the QTL, rendering IBD

Fig. 2 **a** Linkage disequilibrium (LD) as the basis for association mapping approaches. Three markers are depicted and the LD between these markers and the QTL. Markers 2 and 3, though at a similar distance from the QTL, may have a different extent of LD with the QTL, illustrating that LD is variable along chromosomes. **b** Disadvantage of identity-by-state (IBS) approaches and the effect of haplotype formation

estimates unnecessary. The contribution of intragenic epistasis, i.e. favorable haplotypes, to the heritability of complex traits (Würschum et al. 2012a), however, may make the use of haplotypes still worthwhile. One difference between these IBS and IBD approaches is the theoretical reference population. The following reference populations are conceivable: the current population under study, the parental population, intermediate founder individuals some generations before the parents, or the ultimate founder lines even further back in the pedigree. It must be noted that the QTL effects that are estimated are specific for the chosen reference population. Thus, if an ancient base population is used, the genetic variance and the effects are specific for this ancient base and are more difficult to interpret. It is therefore convenient to use approaches with the current population as the base (Powell et al. 2010).

Exploitation of linkage disequilibrium

As implied by the name, linkage mapping utilizes linkage whereas association mapping, also referred to as linkage disequilibrium mapping, is based on linkage disequilibrium between markers and the QTL (Fig. 2a). Association mapping approaches are advantageous as they facilitate a higher mapping resolution. The drawback of association mapping is its reduced QTL detection power, which is linearly dependent on the LD between the QTL and the marker. QTL with small effects can only be detected if they are in strong LD with a marker (Van Inghelandt et al. 2011). In addition, incomplete LD between a marker and a QTL will lead to an underestimation of the variance explained by the QTL. Again, the availability of sequence data and consequently of the causal polymorphisms (LD = 1) will eliminate this disadvantage of association approaches. The methodology is applicable to both population and family mapping. Applied to the latter, the approach exploits the LD generated by linkage in addition to the historical LD present between the parental lines. The inclusion of an IBD matrix or of a variance–covariance matrix for the QTL effect calculated based on pedigree and/or marker information can also be interpreted as the addition of linkage disequilibrium information to a linkage analysis (Van Eeuwijk et al. 2010a). Such approaches exploiting linkage and LD thus potentially offer a high QTL detection power combined with a good mapping resolution.

Biometrical models for QTL detection

Classical linkage mapping methodology operates with estimated conditional probabilities instead of marker data alone and uses background markers as covariates to control for QTL outside the genomic region of interest (Jansen and Stam 1994; Zeng 1994). This approach is also applicable to family mapping as it can be extended toward multiple segregating families (Blanc et al. 2006). Association mapping, which has been developed by human geneticists, has recently been adopted by plant geneticists and applied to population mapping and to family mapping (Thornsberry et al. 2001; Breseghello and Sorrells 2006; Zhao et al. 2007; Harjes et al. 2008). To detect QTL in populations with complex pedigree, mixed models (e.g. Parisseaux and Bernardo 2004; Yu et al. 2006; Van Eeuwijk et al. 2010a) and Bayesian approaches (e.g. Bink et al. 2002, 2008; Bauer et al. 2009; Gasbarra et al. 2009) have been proposed. An alternative to the joint analysis of experimental data are meta-QTL analyses, which compare QTL results from different experiments and can identify commonly detected QTL.

In the context of a mixed model analysis, the QTL can be modeled as fixed or as random effect. Generally, the objective of a fixed effect approach is to estimate the effects of specific alleles, whereas a variance due to the effect is estimated if QTL are modeled as random. An advantage of fixed QTL effects is that the test for significance is less stringent, and in addition fixed QTL effects readily allow the estimation of an effect of each allele and thus the identification of the favorable allele as well as lines carrying that allele. However, to impose a variance–covariance structure on the QTL effect (e.g., IBD matrix) it must be modeled as random. Incorporating kinship information by mixed models has recently also been shown to be advantageous for QTL detection in populations derived from genetically complex crosses that have been subject to selection (Malosetti et al. 2011).

Design of QTL mapping experiments and important considerations

The above-mentioned examples show that different approaches are available for QTL detection in breeding populations. There are, however, conceptual differences between the approaches and the advantages and limitations that need to be considered are discussed here.

Choice of the mapping population

The choice between population mapping (parents or late generations) or family mapping (early or mid generations), and consequently the composition of the mapping population, greatly depends on the intention of the experimenter. Population mapping evaluates individual lines from the breeding pool while family mapping investigates the effects of QTL within families. If the experimenter is interested in an analysis of QTL present in the potential

parents of future crosses then a population mapping approach should be taken. Subsequent to the selection of promising parental lines, new variation in breeding programs is accomplished by crosses between these lines. Consequently, breeders are also interested in QTL that facilitate the identification of superior lines within such crosses and to this end family mapping is the appropriate choice.

For population mapping, the composition of the data set is crucial as the effect of favorable alleles can only be detected if they are present in the population at a certain frequency. If the experimenter is interested in QTL for a resistance trait, for example, then enough plants carrying that resistance must be included in the population since QTL with low minor allele frequency will not be detectable (Myles et al. 2009). The experimenter should also be aware that such lines often have a lower performance for other traits, e.g., yield, which means that yield QTL will be identified that distinguish the resistance carrying lines from the elite lines. These QTL are only of use for the breeder to clean the genetic background when the resistance is introduced into elite material, but will not be informative within elite material. In breeding programs this situation is often encountered when plant material from pre-breeding programs is included in the mapping population. These genotypes enrich the population with new alleles and can increase the genotypic variance but a careful interpretation is required for the implementation of the detected QTL (e.g., von der Ohe et al. 2010; Miedaner et al. 2009). In addition, the mapping population must have a certain size, as small population sizes typically lead to the detection of only large effect QTL and increase the chance of false positives (Wang et al. 2011).

The design of family mapping experiments is also an important issue. As mentioned later, the LD generated by the design is minimized by a high number of parental lines (Würschum et al. 2012b). Mating designs including many parents with a balanced contribution are, for example, single round robin or double round robin designs. Populations derived from breeding programs will generally not follow such a specific design and will include families of varying size. This alone, however, will not have a strong negative impact on the QTL analysis. Nevertheless, when compiling a breeding population for a family mapping experiment the aim should be to balance the contribution of parental alleles (Liu et al. 2012a). If a particular parent would only contribute to a single small family, the QTL detection power for these alleles would be low. An advantage of family mapping over population mapping is that it enables the detection of QTL that are only present in the breeding population at low frequency, as the controlled crosses allow the frequency of such QTL to be artificially increased (Myles et al. 2009). The QTL detection power in family mapping will be largely determined by the population size (Verhoeven et al. 2006). Large population sizes (>500 individuals) are required to detect QTL with medium effect size and to enable the detection of QTL with small effects the population size must be increased even further.

Collection of phenotypic data

An important issue for QTL detection in breeding populations is that the phenotypic data from breeding programs is often generated by combining multiple trials thus resulting in unbalanced designs. For population mapping, Wang et al. (2011) have recently compared balanced with unbalanced data sets and found that balanced data sets may be advantageous in reducing the number of false-positive QTL. Despite the potential problem of an inflated false-positive rate, these results demonstrate that phenotypic data from breeding programs can be used for QTL detection without the need for specialized balanced experimental designs. In line with this, approaches have been suggested to mine existing phenotypic and genotypic data that are routinely generated in breeding programs for the detection of QTL (Parisseaux and Bernardo 2004; Yu et al. 2005).

Another important consideration is that a statistically sound joint analysis of the phenotypic data requires overlapping genotypes between different trials, locations and years (breeding cycles). As the genotypes within the breeding programs are constantly changing, the checks included in each field trial appear predestined for this purpose. In breeding programs these are, however, often chosen by the local breeder and thus not overlapping. In order to link different phenotypic data in a joint analysis, it therefore appears advisable to use common checks throughout the breeding program. If checks must be changed over time, then not all of them should be replaced at once, to allow an overlap with previous years.

Another crucial factor that strongly determines the success of a QTL mapping experiment is the phenotyping intensity. High heritabilities are a prerequisite for reliable QTL results and a high predictive power of the detected QTL i.e. a low bias in the estimation of the proportion of genotypic variance explained by these QTL (Bradbury et al. 2011; Liu et al. 2012a). In addition, if breeding germplasm is selected for a QTL mapping approach, it may be sensible to replace the casual measurements commonly used in breeding trials with more careful measurements to fully exploit the potential of QTL detection approaches (Wang et al. 2011). To take full advantage of the vast amount of phenotypic data generated in breeding programs, a comprehensive database management system that allows the integration of phenotypic and genotypic data is certainly advantageous (Heckenberger et al. 2008).

Confounding effects of genetic relatedness

A potential problem of population mapping is the inherent population stratification present in plant populations. Population stratification can be divided into population structure, that is the presence of two or more major sub-populations, and family structure which refers to the different degrees of relatedness among the lines. As for "true" QTL, any non-functional associations between the variation of the trait and the genetic relatedness will also be detected as QTL (Zhao et al. 2007). Such genotype–phenotype covariance is especially apparent for traits involved in adaptation as phenotypic variation between populations is highly correlated with genotypic differences between them. It has long been recognized as a potential problem that can result in spurious associations and thus in a high number of false-positive QTL (Lander and Schork 1994), and different approaches have been developed to correct for genetic relatedness. Different methods have been suggested to correct for population structure (Q or P matrix) and family structure (K matrix) (Price et al. 2006; Yu et al. 2006; Malosetti et al. 2007; Zhang et al. 2010). All are based on marker data and their simultaneous use may result in an over-correction and consequently in a reduced QTL detection power if Q (P) and K explain a major part of the phenotypic variation already. The application of the K matrix by mixed models has recently been shown to be sufficient for the analysis of breeding populations (Bradbury et al. 2011; Wang et al. 2011; Würschum et al. 2011a, b). However, as the appropriate correction depends on the extent of the genotype–phenotype covariance this must be determined separately for each data set.

Family mapping, by contrast, limits these problems encountered by population mapping as the controlled crosses allow to break up the covariance between genotype and phenotype (Myles et al. 2009). Thus, for traits showing a high correlation between genetic relatedness and phenotypic similarity, family mapping is a promising method to detect true QTL and minimize the detection of false-positive QTL. However, even family mapping approaches require a correction for population structure; i.e. for the different segregating families. A recent comparison of biometrical models for family mapping has shown that if marker effects are not modeled as nested within families, an effect for the segregating family should be included in the model to correct for differences in family means (Fig. 3) (Würschum et al. 2012b).

Variation in QTL effect estimates

A major limitation of QTL mapping in biparental populations is that the estimated effects are specific to that population and QTL results are often not transferable to other

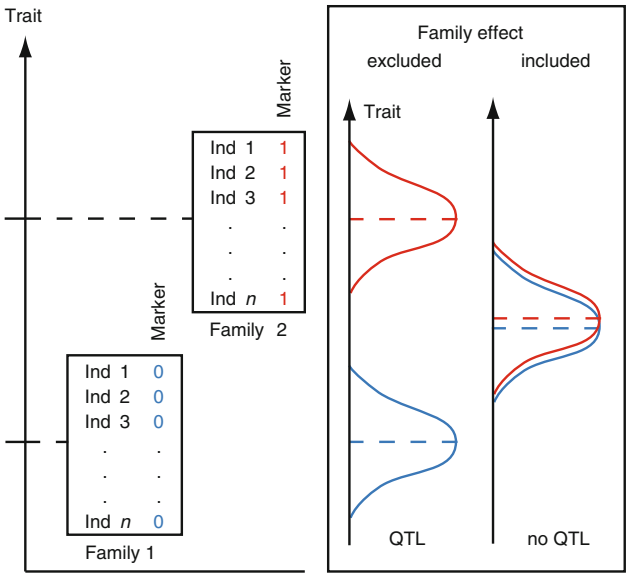

Fig. 3 Schematic representation illustrating that if different alleles are fixed in families with different family means this can result in the detection of QTL explaining only among-family variance. The detection of such QTL can be reduced by a biometrical model incorporating an effect for the segregating family

populations thus limiting their use for marker-assisted selection programs (Holland 2007; Bernardo 2008). The advantage of family mapping is that QTL are detected based on multiple segregating families which should allow more robust estimates of QTL effects across populations. Based on a diallel design, Steinhoff et al. (2011) recently partitioned the effect of each detected QTL into a part that is variable between families and a part that is common to all families. Their results confirm that for the majority of the QTL the effect estimates are specific for a particular family and can consequently not easily be transferred to other families.

This finding is further substantiated by several mapping experiments based on multiple families (Blanc et al. 2006; Coles et al. 2010; Liu et al. 2011) which showed that there is tremendous variation in allele substitution effects if these are estimated separately in each family. Allele substitution effects are a population-specific measure and often this variation is interpreted as a dependency of the QTL effect on the genetic background; i.e. epistasis. Steinhoff et al. (2012) showed that such differences in allele substitution effects are more likely to arise due to multiple alleles at a QTL locus or due to differences in allele frequencies between families. Allele frequencies are affected by segregation distortion that can arise due to selection (McMullen et al. 2009; Alheit et al. 2011), but are also subject to the sampling process especially if family sizes are small (Fig. 4). Taken together, even when QTL are mapped in multiple families the effects of the QTL in independent populations remain difficult to predict.

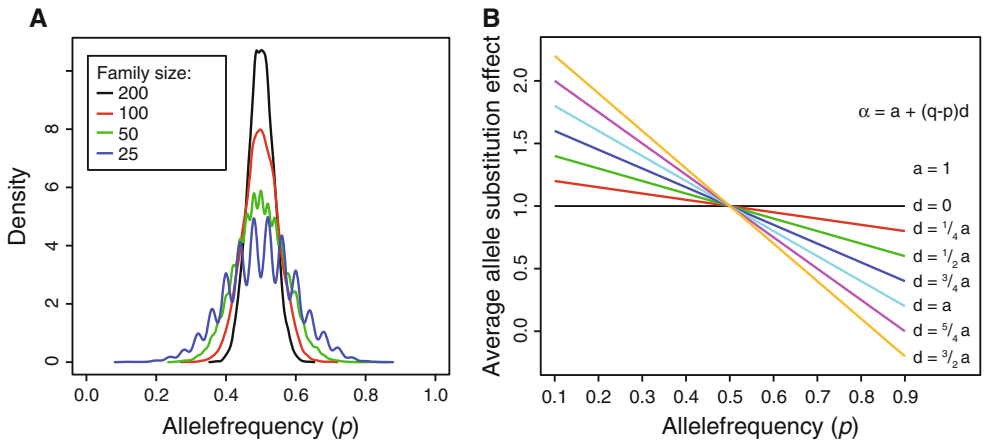

Fig. 4 **a** Effect of genotypic sampling on the allele frequency. Density distributions of the results are shown for different family sizes based on a simulation with 10,000 runs. **b** Dependency of the allele substitution (α) effect on the allele frequency in the population and on the degree of dominance, assuming a population in Hardy–Weinberg equilibrium

If QTL are detected by population mapping, the estimated allele substitution effects are specific for this population of diverse lines, and based on this the effect of the QTL within families cannot be predicted. Thus, population mapping can identify lines carrying certain favorable alleles as well as estimates of the position of the QTL, but does not provide estimates of the QTL effects within families. These have to be determined in independent experiments.

Mapping resolution

Both linkage mapping and linkage disequilibrium mapping approaches identify associations between the genotype and the phenotype. The difference lies in the number of recombinations that are exploited and consequently in the mapping resolution that can be achieved. QTL mapping approaches based on genetic linkage within a segregating family exploit only the comparably few recombinations that have occurred during the establishment of that family. Owing to the short history of that population, recombination has had little time to shuffle the genome to smaller fragments surrounding the QTL and consequently QTL can only be localized to large chromosomal regions.

By contrast, mapping approaches that exploit linkage disequilibrium make use of all recombinations that have occurred during the history of the sampled population and consequently warrant a much higher mapping resolution. LD, the correlation of alleles at separate loci, is subject to different forces (Flint-Garcia et al. 2003) and is therefore always population specific. The decay of LD with genetic map distance is variable across the genome and consequently so is the mapping resolution that can be achieved (Fig. 2a) (Van Inghelandt et al. 2011; Würschum et al. 2011a). In addition, LD is variable among, but also within species. In elite lines of the outbreeding species maize, examples showed LD to decay within an extremely short distance (Van Inghelandt et al. 2011) and to stretch long distances in more closely related lines (Rafalski 2002). By contrast in breeding material of self-pollinating species like wheat, LD decays more slowly (Reif et al. 2011). The exploited LD stems from unrecorded forces during the history of that population and must be investigated as a first step in linkage disequilibrium mapping to obtain an idea of the mapping resolution that can be realized and the required marker density. In addition, it must be noted that LD in family mapping is also generated by the mating design and the number of parental lines (Verhoeven et al. 2006; Würschum et al. 2012b). A high mapping resolution in family mapping can thus only be realized by including a high number of parental lines.

Predictive power of detected QTL and validation

Recent family mapping results based on elite maize breeding material have shown that QTL for complex traits such as grain yield and grain moisture could be detected (Blanc et al. 2006; Liu et al. 2011, 2012b; Steinhoff et al. 2011). Applying a cross-validation approach revealed a relative bias in the estimates of genotypic variance explained by the detected QTL of 10–60 % depending on the trait and the experiment. This highlights that irrespective of the biometrical model used for the analysis, a cross-validation approach should be applied to obtain unbiased estimates of the predictive power of the detected QTL. A major advantage of breeding programs is that there will always be independent populations available, even from subsequent breeding cycles. These enable a rapid validation of the

identified QTL and the elimination of false-positive QTL. Thus, valuable QTL can readily be identified and incorporated in marker-assisted selection programs.

An important consideration when analyzing multiple families is that the genotypic variance explained by the detected QTL can be attributable to within-family variance and/or to among-family variance (Würschum et al. 2012b; Liu et al. 2011). QTL due to among-family variance can arise if different alleles are fixed in families with different family means (Fig. 3). These QTL are, however, prone to an enhanced false-positive rate. An efficient way to detect QTL explaining mainly within-family variance is the choice of an appropriate biometrical model (Fig. 3) (Würschum et al. 2012b).

Taken together, even for highly complex traits, QTL mapping experiments in breeding populations can identify QTL that explain a considerable proportion of the genotypic variance and these can be validated in available independent populations. This illustrates that a three-step analysis may be most appropriate for the detection of QTL by family mapping and validation in breeding programs. In a first step, QTL are detected in single families and can in the next generation(s) be used for the identification of plants within this family that are homozygous for the detected QTL. Interestingly, these plants that have been identified as being homozygous for important QTL can already be used for crosses with complementary plants from the same or from other crosses to recombine positive alleles from different QTL into a single genotype. The separate analysis of single families has the additional advantage that it can increase QTL detection power as compared to a joint analysis, particularly if the families are very heterogeneous with regard to QTL that are segregating in them (in that case the non-segregating families will attenuate the QTL signal) (Li et al. 2011). Second, a joint analysis of multiple families will yield additional information concerning the variation in QTL effects and the transferability of the detected QTL. In the third step, families from subsequent cycles can be used for validation of the identified QTL. Conversely, if a population mapping approach is applied, the identified QTL can be used to select parents based on a gene stacking approach (i.e., combining favorable alleles from different parents). The same QTL can then be used for the selection of the desired plants in early generations.

Contribution of epistasis

Epistasis refers to interactions between alleles from two or more genetic loci in the genome (Carlborg and Haley 2004; Phillips 2008). The consequence of epistasis is that the phenotype of an individual cannot be predicted simply by the sum of the single-locus effects, but rather depends on the specific combinations of loci (Lynch and Walsh 1998). In germplasm that has experienced selection, epistasis has been shown to contribute to the expression of complex traits (Dudley and Johnson 2009). It is thus of importance for plant breeders to obtain an estimate of the genetic architecture of the trait, that is of the contribution of main effects and of epistatic interactions to the genotypic variance. It must be noted that the detection of epistatic QTL will rely even more on large population sizes than the detection of main effects. The most promising approach to detect epistatic QTL appears to be a full two-dimensional scan for all possible pairwise interactions. Such scans are nowadays computationally feasible and have successfully been used to detect epistatic interactions in family mapping (e.g., Buckler et al. 2009; Liu et al. 2011; Steinhoff et al. 2012; Würschum et al. 2012c), and in population mapping (e.g., Li et al. 2010; Massman et al. 2011; Reif et al. 2011; Würschum et al. 2011a, b; Yu et al. 2011).

In summary, epistasis appears to be of minor importance in breeding populations. For most crops and traits, epistasis could be detected but the proportion of genotypic variance explained by these epistatic QTL was small compared to that of the main effect QTL. There are, however, exceptions where individual epistatic QTL have been identified which explain a proportion of genotypic variance comparable to that of the main effects (Miedaner et al. 2011; Reif et al. 2011). As the forces active in natural populations are not effective in breeding populations, epistatic interactions may be selected and maintained, thus contributing to the expression of the trait (Steinhoff et al. 2012). In addition, some results suggest the presence of epistatic master regulators; i.e. loci that appear to be involved in a large number of interactions (Reif et al. 2011; Würschum et al. 2011a). In conclusion, the contribution of epistasis to the genetic architecture of agronomic traits in breeding populations appears to be small. Nevertheless, given the effort required to establish, phenotype and genotype these populations, an epistasis scan seems advisable as single epistatic QTL may have large effects and thus may improve knowledge-based breeding.

Conclusions

The availability of vast amounts of phenotypic data from breeding programs and the decreasing costs for genotyping or even re-sequencing make it attractive to exploit this information for QTL detection. Depending on the intention of the experiment, population mapping can be used to identify potential parents in the breeding pool carrying favorable alleles. Alternatively, family mapping can be used to detect QTL for the selection of superior lines within

crosses. Both approaches will profit from a further increase in marker density or the availability of sequence data. Irrespective of the mapping approach, the selection of plants to be included in the mapping population is of utmost importance as is the choice of an appropriate biometrical model. Challenging research questions remain, such as how to incorporate sequence data or biometrical approaches to account for the influence of selection within families on the QTL detection power in family mapping. Taken together, the mapping approaches available today represent a powerful tool to dissect the genetic architecture of complex agronomic traits in breeding populations for an improved knowledge-based breeding of crops.

Acknowledgments The author thanks Jochen Reif, H. Friedrich Utz, and Matthew R. Tucker for helpful discussions and critically reading the manuscript and greatly appreciates the helpful comments and suggestions of two anonymous reviewers.

References

Alheit KV, Reif JC, Maurer HP, Hahn V, Weissmann EA et al (2011) Detection of segregation distortion loci in triticale (× *Triticosecale Wittmack*) based on a high-density DArT marker consensus genetic linkage map. BMC Genomics 12:380

Bauer AM, Hoti F, von Korff M, Pillen K, Léon J et al (2009) Advanced backcross-QTL analysis in spring barley (*H. vulgare* ssp. *spontaneum*) comparing a REML versus a Bayesian model in multi-environmental field trials. Theor Appl Genet 119: 105–123

Bernardo R (2008) Molecular marker and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet 104:751–762

Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE et al (2008) Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161:85–96

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224

Bradbury P, Parker T, Hamblin MT, Jannink JL (2011) Assessment of power and false discovery rate in genome-wide association studies using the BarleyCAP germplasm. Crop Sci 51:52–59

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5:618–625

Coles ND, McMullen MD, Balint-Kurti PJ, Pratt RC, Holland JB (2010) Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. Genetics 184: 799–822

Dekkers JCM, Hospital F (2002) Multifactorial genetics: the use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3:22–32

Dudley JW, Johnson GR (2009) Epistatic models improve prediction of performance in corn. Crop Sci 49:763–770

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Addison Wesley Longman, Harlow

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Ann Rev Plant Biol 54:357–374

Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Gasbarra D, Pirinen M, Sillanpää MJ, Arjas E (2009) Bayesian quantitative trait locus mapping based on reconstruction of recent genetic histories. Genetics 183:709–721

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Bermudez Kandianis C et al (2008) Natural genetic variation in epsilon lycopene cyclase tapped for maize biofortification. Science 319:300–333

Heckenberger M, Maurer HP, Melchinger AE, Frisch M (2008) The plabsoft database: a comprehensive database management system for integrating phenotypic and genomic data in academic and commercial plant breeding programs. Euphytica 161: 173–179

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161

Jannink JL, Bink M, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6:337–342

Jansen RC (2007) Quantitative trait loci in inbred lines. In: Handbook of Statistical Genetics, 3rd edn. Wiley, New York. ISBN: 978-0-470-05830-5

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455

Jansen RC, Jannink JL, Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. Crop Sci 43:829–834

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

Li L, Paulo MJ, Van Eeuwijk F, Gebhardt C (2010) Statistical epistasis between candidate gene alleles for complex tuber traits in an association mapping population of tetraploid potato. Theor Appl Genet 121:1303–1310

Li H, Bradbury P, Ersoz E, Buckler ES, Wang J (2011) Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. PLoS ONE 6(3):e17573. doi:10.1371/journal.pone.0017573

Liu W, Gowda M, Steinhoff J, Maurer HP, Würschum T et al (2011) Association mapping in an elite maize breeding population. Theor Appl Genet 123:847–858

Liu W, Maurer HP, Reif JC, Cossic F, Würschum T (2012a) Optimum design of family structure and allocation of resources in association mapping with lines from multiple crosses (in review)

Liu W, Reif JC, Cossic F, Würschum T (2012b) Comparison of biometrical approaches for QTL detection in multiple segregating populations. Theor Appl Genet (accepted)

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Assoc, Sunderland

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577

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

Malosetti M, Van Eeuwijk FA, Boer MP, Casas AM, Elía M 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

Massman J, Cooper B, Horsley R, Neate S, Dill-Macky R et al (2011) Genome-wide association mapping of fusarium head blight

resistance in contemporary barley breeding germplasm. Mol Breeding 27:439–454
McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740
Melchinger AE, Orsini E, Schön CC (2012) QTL mapping under truncation selection in homozygous lines derived from biparental crosses. Theor Appl Genet 124:543–553
Meuwissen THE, Karlsen A, Lien S, Olsaker I, Goddard ME (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 161:373–379
Miedaner T, Wilde F, Korzun V, Ebmeyer E, Schmolke M et al (2009) Marker selection for Fusarium head blight resistance based on quantitative trait loci (QTL) from two European sources compared to phenotypic selection in winter wheat. Euphytica 166:219–227
Miedaner T, Würschum T, Maurer HP, Korzun V, Ebmeyer E, Reif JC (2011) Association mapping for Fusarium head blight resistance in soft European winter wheat. Mol Breed 28:647–655
Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z et al (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202
Nordborg M, Weigel D (2008) Next-generation genetics in plants. Nature 456:720–723
Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet 109:508–514
Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9:855–867
Powell JE, Visscher PM, Goddard ME (2010) Reconciling the analysis of IBD abd IBS in complex trait studies. Nat Rev Genet 11:800–805
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100
Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T et al (2011) Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theor Appl Genet 123:283–292
Steinhoff J, Liu W, Maurer HP, Würschum T, Longin FH et al (2011) Multiple-line cross QTL-mapping in European elite maize. Crop Sci 51:2505–2516
Steinhoff J, Liu W, Reif JC, Ranc N, Würschum T (2012) Detection of QTL for flowering time in multiple families of elite maize (in review)
Thornsberry JM, Goodman MM, Doebley H, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289
Van Eeuwijk FA, Boer M, Totir LR, Bink M, Wright D et al (2010a) Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. Theor Appl Genet 120:429–440
Van Eeuwijk FA, Bink MCAM, Chenu K, Chapman SC (2010b) Detection and use of QTL for complex traits in multiple environments. Cur Opin Plant Biol 13:193–205
Van Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor Appl Genet 123:11–20
Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530
Verhoeven KJF, Jannink JL, McIntyre LM (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96:139–149
Von der Ohe C, Ebmeyer E, Korzun V, Miedaner T (2010) Agronomic and quality performance of winter wheat backcross populations carrying non-adapted Fusarium head blight resistance QTL. Crop Sci 50:2283–2290
Wang H, Smith KP, Combs E, Blake T, Horsley RD et al (2011) Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. Theor Appl Genet. doi:10.1007/s00122-011-1691-8
Wu R, Zheng ZB (2001) Joint linkage and linkage disequilibrium mapping in natural populations. Genetics 157:899–909
Würschum T, Maurer HP, Kraft T, Janssen G, Nilsson C et al (2011a) Genome-wide association mapping of agronomic traits in sugar beet. Theor Appl Genet 123:1121–1131
Würschum T, Maurer HP, Schulz B, Möhring J, Reif JC (2011b) Genome-wide association mapping reveals epistasis and genetic interaction networks in sugar beet. Theor Appl Genet 123:109–118
Würschum T, Maurer HP, Dreyer F, Reif JC (2012a) Effect of inter- and intragenic epistasis on the heritability of complex traits (in review)
Würschum T, Liu W, Gowda M, Maurer HP, Fischer S et al (2012b) Comparison of biometrical models for joint linkage association mapping. Heredity 108:332–340
Würschum T, Liu W, Maurer HP, Abel S, Reif JC (2012c) Dissecting the genetic architecture of agronomic traits in multiple segregating populations in rapeseed (*Brassica napus* L.). Theor Appl Genet 124:153–161
Yu J, Arbelbide M, Bernardo R (2005) Power of in silico QTL mapping from phenotypic, pedigree, and marker data in a hybrid breeding program. Theor Appl Genet 110:1061–1067
Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208
Yu LX, Lorenz A, Rutkoski J, Singh RP, Bhavani S et al (2011) Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor Appl Genet. doi:10.1007/s00122-011-1664-y
Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468
Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360
Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C et al (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4