

Chapter 16

Prospects and Limitations for Development and Application of Functional Markers in Plants

Everton A. Brenner, William D. Beavis, Jeppe R. Andersen,
and Thomas Lübberstedt

Introduction

As genotyping becomes more accessible with faster and cheaper DNA sequencing technologies and single nucleotide polymorphism (SNP) platforms, an ever-increasing number of sequence polymorphisms are revealed in various plant species. In maize, sequence comparison of six recently sequenced inbred lines revealed more than 1,000,000 SNPs and 30,000 insertions/deletions (Indels) in the maize genome (Lai et al. 2010). Knowledge about the effect of polymorphisms on trait variation has also been increasing with results from association and nested association mapping studies. Respective quantitative trait polymorphisms (QTP), sequence polymorphisms associated with phenotypic trait variation, can be converted into Functional Markers (FMs) (Andersen and Lübberstedt 2003). Contrary to linked markers, FMs are derived from polymorphisms causing phenotypic variation.

The development of FMs, as described by Andersen and Lübberstedt (2003), requires the knowledge of functionally characterized loci. Once polymorphic sites are identified within those functional loci, statistical models can be used to test for genotype-phenotype associations. Such association studies provide inferential, i.e., statistical, evidence for correlations, not necessarily reflecting biological causality. Validation of trait-associated polymorphisms, through gene introgression provides biological evidence of functionality. Importantly, functional polymorphisms can be converted into technical assays using, e.g., any of the SNP or insertion/deletion (Indel) detection technologies (Appleby et al. 2009; Gupta et al. 2008).

E.A. Brenner (✉) • W.D. Beavis • T. Lübberstedt
Department of Agronomy, Iowa State University, Agronomy Hall, Ames, IA 50011, USA
e-mail: thomasl@iastate.edu

J.R. Andersen
Department of Genetics and Biotechnology, Research Center Flakkebjerg, University of Aarhus,
Aarhus, Denmark

Examples of QTP discovery through association studies include starch biosynthesis (Wilson et al. 2004), cell wall digestibility (Andersen et al. 2008; Brenner et al. 2010; Guillet-Claude et al. 2004), flowering time (Salvi et al. 2007; Thornsberry et al. 2001), carotenoid biosynthesis (Harjes et al. 2008; Palaisa et al. 2003), inflorescence architecture (Bortiri et al. 2006), kernel properties (He et al. 2008; Shi et al. 2008), resistance to bacterial blight (Iyer-Pascuzzi and McCouch 2007), and fruit quality (Costa et al. 2008; Ogundiwin et al. 2008). Thornsberry et al. (2001) pioneered association mapping in plants by developing linkage disequilibrium (LD) mapping and employing this to identify associations between polymorphisms within the Dwarf8 (D8) gene affecting flowering time and plant height. The study was based on 92 diverse maize inbred lines from four populations: Stiff Stalk, non-Stiff Stalks, tropical, and semi-tropical. The association analysis, which was correcting for population structure, identified nine polymorphisms significantly associated with flowering time. Andersen et al. (2005) investigated the applicability of these nine polymorphisms as FMs in an independent set of 71 elite European inbred lines. Ignoring population structure, six of the nine polymorphisms were significantly associated with flowering time, and none with plant height. However, when population structure was considered, only one association between a 2-bp Indel in the promoter region and plant height remained significant, while no association was observed for flowering time. Camus-Kulandaivelu et al. (2006) evaluated a 6 bp Indel identified by Thornsberry et al. (2001) in a larger population consisting of 375 inbred lines and 275 landraces from United States and Europe. This QTP was confirmed to be associated with flowering time under long-day conditions, with different estimated allelic effects for inbreds and landraces.

This example illustrates that availability of qualified candidate genes can facilitate development of informative molecular markers by means of association studies. QTPs may, however, not be consistent across genetic backgrounds and environments. In this review, the challenges in development, estimation of genetic effects for, and application of FMs in plants are discussed.

Power, Precision and Accuracy in QTP Detection

The goal of genetic mapping studies is to identify genomic regions associated with observed phenotypic variation. In plants, linkage mapping started as a great promise to reveal chromosome fragments with higher-than-expected associations with phenotypic variation observed in segregating bi-parental populations. Today, thousands of linkage mapping experiments have been reported (Behn et al. 2004; Blanc et al. 2006; Byrne et al. 1998; Buerstmayr et al. 2003; Pinson et al. 2005; Tang et al. 2000). Identified QTLs and estimated QTL effects, however, have been rarely consistent across and even within populations, and only a minority has been used for cultivar improvement (Bernardo 2008). The reasons for lack of repeatability and application of QTL identified from linkage mapping experiments have been extensively discussed (Beavis 1994; Bernardo 2008; Scho et al. 2004). Most linkage

mapping experiments have resulted in inconsistent QTL with overestimated effects mainly due to small population sizes, stringent significance levels, and interactions with different genetic backgrounds and environments (Beavis 1994; Bernardo 2008; Xu 2003a). In addition, the limited number of recombination events accumulated in populations commonly used in linkage mapping experiments (i.e., F2 and backcross) makes it difficult to narrow the associated regions to fewer than several megabases. As a consequence, the identification of causative genes usually requires development of further recombinants of at least 500 individuals for adequate power (Beavis 1994; Lee et al. 2002).

Compared to linkage mapping in families, association mapping in populations can potentially reveal the genetic basis of phenotypic variation with much greater genetic resolution and even identify QTPs. Contrary to linkage mapping, association mapping does not rely on a controlled bi-parental segregating population, but on a collection of lines not necessarily sharing a pedigree, and therefore takes advantage of historical recombination events accumulated among lines. Smaller LD blocks due to accumulated recombination events allow greater genetic resolution, and require, as a consequence, a much higher marker density as compared to linkage mapping within families. In some plant species, like maize, reduced LD combined with large genomes may require hundreds of thousands of molecular markers to adequately cover the genome (Brown et al. 2004; Ching et al. 2002; Flint-Garcia et al. 2005; Hyten et al. 2007; Yu et al. 2008). Evaluating a massive number of markers requires multiple testing corrections to control for false positives, thus decreasing the power of identifying markers associated with phenotypic variation. Reduced power is even more problematic for quantitative traits governed by multiple genes with modest or small phenotypic effects, or alleles with strong phenotypic effect but at low frequencies in the association panel.

A third approach based on the concept of combining LD with linkage mapping has been referred to as Nested Association Mapping (NAM) (Yu et al. 2008). Several NAM populations have been developed in plant species (Guo et al. 2010). Typically these consist of families of Recombinant Inbred Lines (RILs) derived from a sample of inbred lines crossed to a reference inbred line. The relationships among progeny within families are inbred full sibs, while relationships among progeny from different families are half sibs (Bernardo 2002). NAM populations consisting of doubled haploid lines or RILs are “immortalized”, meaning that homozygous lines within each family can be evaluated in numerous locations and years without confounding effects of genetic segregation (Nordborg and Weigel 2010).

Yu et al. (2008) developed and released a maize NAM population consisting of 25 families with 200 RILs for each family. Simulations from Guo et al. (2010) suggests that NAM populations similar to the one developed from Yu et al. (2008) have adequate power to accurately and precisely identify additive polymorphisms contributing at least 5% of the variation in the phenotype. Guo et al. (2010) also observed that the resolution and power to detect QTP is maintained even if non-functional alleles are in LD with the causal variant. Two recent studies in the maize NAM populations identified alleles with small effects in association with southern leaf blight resistance and leaf architecture in maize (Kump et al. 2011; Tian et al. 2011). These results demonstrate the potential of nested designs to identify QTPs.

Challenges in Functional Marker Development: LD, Epistasis, Environmental and GxE Effects

Depending upon the sample size, LD, and genetic architecture, the mapping approaches discussed above usually identify genomic intervals associated with phenotypic variation, and FM development requires the identification of the functional variants (SNPs/Indels) within these intervals. The discrimination of functional vs. non-functional variants is often complicated by LD within candidate loci, where non-functional alleles may be associated with phenotypic variation when in LD with functional ones. Varying levels of LD have previously been observed between genes of the phenylpropanoid pathway, decaying within few hundred bps for *CCoAOMT2* and *COMT* (Guillet-Claude et al. 2004; Zein et al. 2007) while spanning more than 3.5 kb at the PAL locus (Andersen et al. 2007). Even in populations with substantial intragenic decay of LD, adjacent polymorphic sites might still be in high or complete LD, leading to an overestimation of SNPs/Indels associated with the investigated phenotype. The identification of causal genetic polymorphisms is a difficult task, and statistical evidences and the biological nature of candidate variants may have to be analyzed mutually in order to discriminate QTPs from closely associated non-causal polymorphisms. SNPs located in coding regions causing non-synonymous non-conservative amino acid changes are more likely to be functional than non-synonymous conservative and synonymous amino acid substitution (Risch 2000). Although SNPs have received more attention in mapping studies, Indels involve larger segments of DNA, and when disrupting or causing frame shifts in coding sequences, are more likely to cause phenotypic variation (i.e. loss of function mutants). Such extreme phenotypes are more likely eliminated or fixed by (natural) selection, and as a result, Indels are usually less frequent in populations as compared to SNPs in genic sequences (Clark et al. 2007; Jones et al. 2009).

Polymorphisms in non-coding regulatory regions are potentially major sources of phenotypic variation when regulating gene expression, while variants in intronic regions may create or delete a splicing site (Talerico and Berget 1990). Salvi et al. (2007) identified a non-coding cis-acting regulatory element located 70 kb upstream of an Ap2-like transcription factor which is involved in flowering time. Clark et al. (2004) and Camus-Kulandaivelu et al. (2008) also identified cis-acting regulatory in regions 60 and 100 kb upstream of the *Tb1* and *D8* genes, respectively. In effect, the search for QTPs should not be limited to exonic regions, but ideally should also encompass regulatory and intronic regions with potential impact on the investigated trait (Polidoros et al. 2009).

Besides reducing the resolution of association mapping, another LD-related issue is the identification or development of optimal QTP haplotypes when several polymorphisms within the target locus affect the trait of interest. This is a concern especially when favorable QTP alleles for one trait are closely linked to QTP alleles with unfavorable effect on other traits (Chen et al. 2010). If not available in the characterized population, development of optimal QTP allele combinations based on intragenic recombination events might be difficult to achieve, even by use of large

populations and intragenic markers. Alternatively, exotic germplasm might provide a source for novel intragenic combinations of QTP alleles. More recently, the use of Zinc finger nucleases (ZFNs) has been proposed as a promising technology to replace alleles by homologous recombination (Shukla et al. 2009). The induction of recombination in defined genomic intervals is, therefore, a promising approach to develop optimal QTP haplotypes even within large LD blocks.

Even after true QTPs have been identified, their transferability might be affected by the composition of populations in different studies, both with regard to allele frequencies at the target locus, and structure of the respective populations. D8 is the only example in plants so far, where the same locus has been studied independently in different experimental populations of inbred lines (Andersen et al. 2005; Camus-Kulandaivelu et al. 2006; Thornsberry et al. 2001). When correcting for population structure, the QTPs identified by Thornsberry et al. (2001) were not significantly associated with flowering time in the study of Andersen et al. (2005) as haplotypes were confounded with population structure in the latter study. Other factors, apart from population structure, with potential impact on the detection of QTPs are epistasis, dominance (so far, association studies in maize were conducted at line per se level), as well as environment and genotype by environment effects.

If the effects of an allele depend on a second allele, either in the same or different loci, the power to detect associations and the accuracy of estimated allelic effects are reduced. Dominance effects cause deviations from additive effects of alleles belonging to the same loci, and simple additive models not accounting for dominance would lead to biased estimation of allelic effects. The relevance of dominance bias for any given trait is directly dependent on the ratio between dominance and additive variances, and it might be reasonably neglected if dominance effects are weak (Hill et al. 2008). In some crop species, the use of RILs or DH lines gives the opportunity to estimate allelic effects free from dominance deviations, permitting more accurate phenotypic predictions of the progeny. In crops evaluated as hybrids, additive effects are still likely the major source of genetic variance among hybrids, but non-estimated dominance effects will probably contribute to phenotypic variation, causing deviations from predicted additive values.

Similarly, epistasis, i.e., the non-additive interaction among alleles at different loci can bias estimates of allelic effects (Cheverud and Routman 1995). Epistasis estimates are often limited by the number of loci included in respective models (Carlborg and Haley 2004). If interacting alleles are not considered or are unknown, it is not possible to model epistatic effects and their consequences in association analysis and FM development. For this reason, if a candidate gene is suspected to interact with other genes, e.g., those belonging to a common genetic network, associations identified for a single gene might be inaccurate and misleading. Numerous mapping studies have detected QTL \times QTL epistasis as a statistical feature causing deviation from expected additive effects (Juenger et al. 2005; Yang et al. 2010; Zhang et al. 2008), but only a few studies have investigated gene \times gene interaction affecting the phenotypic variation in plant association mapping populations (Li et al. 2010; Manicacci et al. 2009; Stracke et al. 2009).

Mapping experiments often require large population sizes for adequate power to identify QTL and accurately estimate their effects. Collecting phenotypic data across multiple environments, years, and replications is costly and challenging, and accommodating large populations in multiple environments require more efficient experimental designs involving incomplete blocks, e.g., augmented or alpha-lattice designs. Inadequate experimental designs not controlling environmental noise lead to inaccuracy in phenotypic estimation and subsequent identification of QTL and estimation of their effects, even if population sizes are adequate. Control of environmental variation within (with number of plants/plot) and among experimental rows (replications/location) are essential for estimating environment variance within locations. Experiments in multiple locations also account for Genotype x Environment interactions (GxE). Using marginal means across locations might lead to inaccurate associations and estimations of allelic effects, if GxE is significant. In case of weak genetic correlations across environments, association analyses should be conducted on an individual location basis. Clustering environments according to their genetic correlations for all pairwise comparisons across environments (Cooper and DeLacy 1994) is an alternative to classify environments into a smaller number of mega-environments based on their influence on GxE.

In conclusion, the genetic effects of QTPs are background, population, and environment dependent (Fig. 16.1). We propose to employ the term “potential” to describe the presence of a beneficial QTP allele, since this term reflects a certain potential of trait expression and is analogous to the risk concept in human genetic diseases, depending on the genetic effect and penetrance of the respective allele.

In humans, the relative risk of an individual developing a complex disease is estimated by taking into account genetic and non-genetic (i.e.: sex, age, diet, ethnicity, and others) variables. The genetic component of risk assessment is based on odds ratio: the odds of a disease occurring in individuals with a certain allele versus the odds of this disease occurring in individuals without this allele. When more than one gene (marker) is considered, the genetic risk of an individual corresponds to the product of odds ratios of individual alleles (Risch 1990; Wray et al. 2007). The same principle may be applicable in plants. Once lines are genotyped for a FM, breeding values for each genotypic class of this FM can be estimated across lines, environments and years, leading to a normal distribution of breeding values for each genotypic class. These distributions can be further characterize for their “displacement” (Risch 2000), which is defined as the number of standard deviations of the average effect of one homozygous genotypic class in relation to the other. Mendelian alleles with strong phenotypic effects are likely to have larger displacement, while alleles from genes affecting complex inherited traits are likely to have smaller displacements (Fig. 16.2).

Even though the estimation of displacement shows the average effect of one allele in relation to the other, it does not directly measure the likelihood of a genotype to contribute to a desirable phenotype. The “potential” of an allele contributing to a phenotype of interest requires establishment of a threshold separating undesirable from desirable phenotypes (Fig. 16.3). In plant breeding, the threshold may be

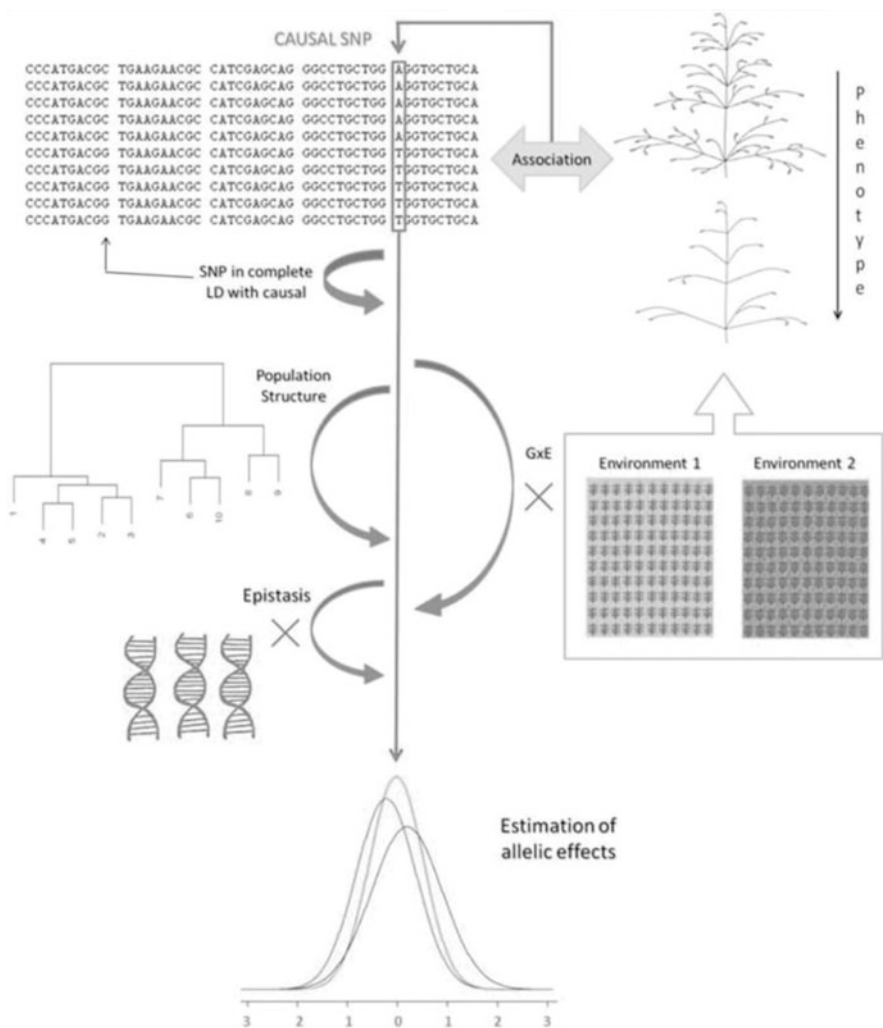


Fig. 16.1 Association between phenotype and genotype, and key components potentially impacting the identification of FMs and the estimation of their effects: *LD*, *population structure*, *GxE*, *epistasis*

defined as a value above the mean phenotype of the best commercial lines (normally used as checks in breeding experimental designs). The estimation of the potential of an allele would be defined as odds of lines passing the threshold with a certain FM genotype versus the odds of lines passing the threshold without this FM genotype.

Fig. 16.2 Potential (P) of genotypes from a FM to pass the threshold of a fictional trait. Probabilities vary according to means and distributions. Mendelian traits (A) usually display larger displacements and larger differences in probabilities, while differences are subtle in quantitative traits (C) (Based on Risch 2000)

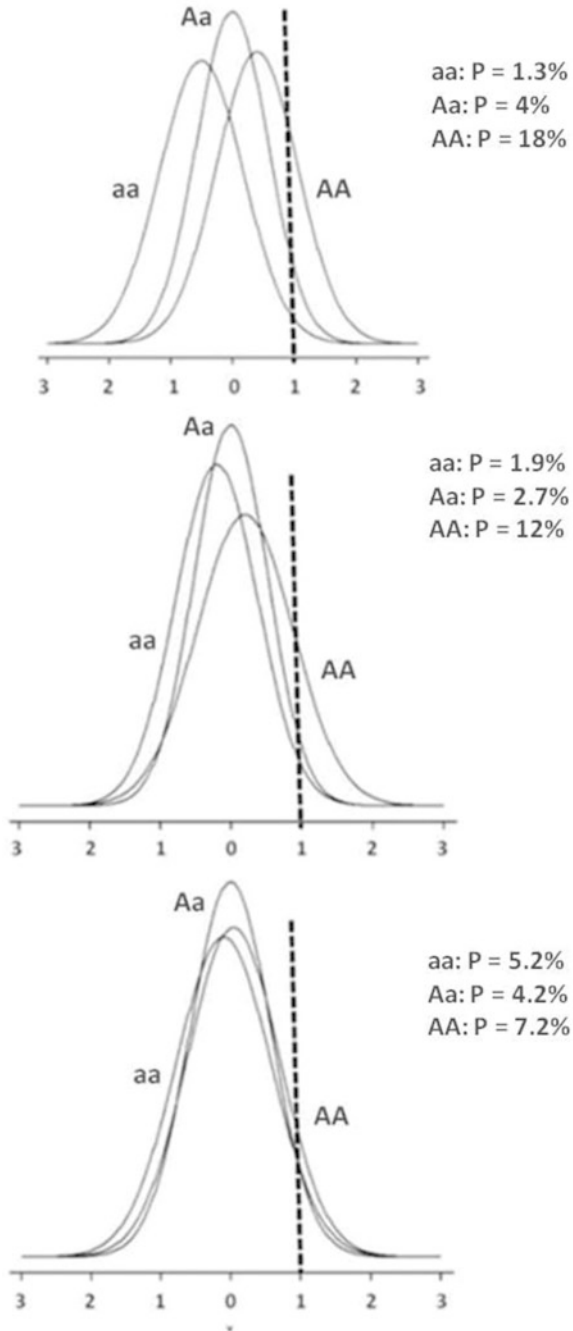
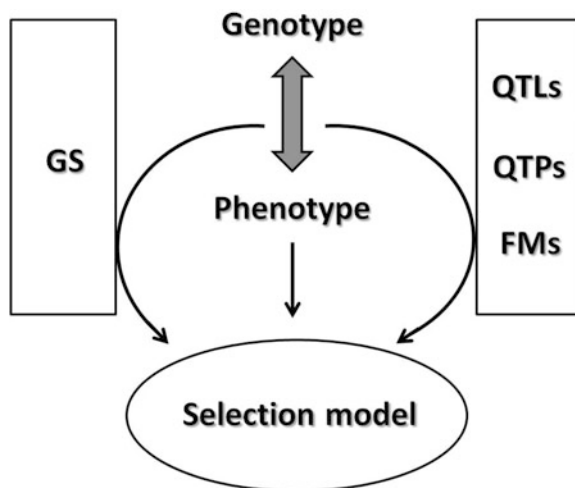


Fig. 16.3 Using the genotype-phenotype relationship to implement selection models. The level of genetic characterization of this relationship may vary from none, such as in genomic selection (*GS*), to highly characterized, such as in functional markers (*FMs*)



Systematic Collection of Genotypic and Phenotypic Information

Marker and phenotypic data accumulate as mapping experiments designed to investigate genotype-phenotype associations and/or assist breeding decisions are performed. Combining information from different mapping experiments via meta-analysis is a promising approach to enhance statistical power, reduce type 1 errors, and evaluate effects of QTL/QTP in a broader set of genetic backgrounds and environments (Heo et al. 2001). Combining data, however, is not straightforward. The definition of a phenotype and how it is measured is seldom consistent across research groups. Although standard phenotyping techniques are a common practice in the private sector, it would require dialogue among researchers in public institutions to reach a consensus. Additionally, detailed description of the experiment including information on germplasm (i.e. maturity), locations, number of replications, check lines, and statistical design, would be required for any researcher to access if an experiment should be considered for meta-analysis or not. Locations and years are not only relevant for estimation of interactions between FM and environments, but the detailed description of an environment (such as maximum/minimum daily temperatures and precipitation) might be important for specific research goals. In drought tolerance studies, for example, temperatures, and amount/distribution of precipitation during different plant development stages is essential information to map drought tolerance genes, given that maize responds differently to water stress in different developmental stages (Barker et al. 2005). With this knowledge, breeders would be able to cluster environments according to relevant climate parameters, and evaluate FM potential in different lines (backgrounds) growing in environments with stress occurring in specific developmental stages.

Besides meta-analysis, comparing mapping outcomes across independent studies is a valuable approach for accessing QTL consistency, refining estimations of QTLs/QTPs effects and narrowing QTL intervals. Pooling results from different mapping experiments is a popular practice in human genetics, where different research groups combine and compared outcomes from large genome wide association studies (GWAS) for common complex diseases, such as type 2 diabetes, coronary disease and breast cancer (McPherson et al. 2007; Scott et al. 2007; Stacey et al. 2007).

In plants, most mapping experiments have consisted of single experiments designed for QTL detection, while less attention has been given to meta and post-hoc analysis. In SoyBase, the USDA-ARS soybean genetics and genomics database, does not routinely archive raw experimental data from QTL experiments (David Grant, pers. comm.). Although the availability of such data could be used to improve QTL mapping, the soybean community has not traditionally done these analyses and so has not made the raw data available. As a consequence the genetic maps in SoyBase are constructed post-hoc by placing the published QTL positions onto a reference genetic map framework using linear scaling between this framework and the reported results. The interpretation of this composite genetic map is complicated by the facts that (1) many of the reported QTL were identified only by analysis of variance (ANOVA) based on a subset of the markers, and (2) the methods and nomenclature used for phenotypic measurements in different experiments are inconsistent. In addition, the choice of QTL mapping procedures has several important ramifications. First, QTL controlled by the same underlying gene can often show different positions due to variation in marker numbers and locations across experiments. Second, the position of the underlying gene cannot be determined relative to the reported QTL. And third, it is not possible to determine the effect of a QTL since the effect and QTL position are confounded if no composite interval mapping is used.

MaizeGDB, the USDA-ARS genetics and genomics database, contains archives for a subset of the raw data for QTL mapping experiments (Carolyn Lawrence, pers. comm.). However, because there is no community agreement on the necessity for submitting such data, it is not possible to do any comprehensive re-analysis of the data due to its incompleteness. As is the case for SoyBase, inconsistencies in trait measurement methodologies and nomenclature along with often imprecise QTL positions impair the ability to compare results between studies.

The current constraints on cross-population comparisons are being addressed by both databases. MIQAS (Minimum Information for QTL and Association Studies, <http://miqas.sourceforge.net/>) will be adopted to ensure that all QTL studies report a critical minimum of information about a given QTL. In particular, researchers will be encouraged to use interval mapping to identify and position QTL rather than simple ANOVA. Standard ontologies for traits and, where possible, accepted methods used to measure them are being developed.

The Buckler lab has developed standardized phenotyping tools in maize (<http://www.maizegenetics.net/phenotyping-tools>) which could develop into community standards. Also, all phenotypic data from the NAM population will be made

publicly available. This together with the NAM GWAS (<http://cbsuapps.tc.cornell.edu/namgwas.aspx>) will facilitate unprecedented in silico mapping opportunities. Together these improvements to the public databases will facilitate the re-analysis of combined trait and mapping data from multiple populations. This should produce refined genetic positions for QTL which are needed to identify candidate genes.

Application of Functional Markers

Resulting from the rapid progress in sequencing technology, the genomic sequence of additional maize inbreds beyond B73 is already reality (Lai et al. 2010). Projects like the NAM community approach (Kump et al. 2011; Tian et al. 2011; Yu et al. 2008) will lead to accumulation of further characterized genes and QTPs of agronomic relevance. Thus, the number of functionally characterized polymorphisms in maize as prerequisite for functional marker development will substantially increase over the next decade. FMs might be useful for various steps along the process of cultivar development. These include (1) identification of novel or better alleles (QTPs haplotypes) for characterized genes in exotic germplasm collections, (2) identification of complementary parents for development of new inbreds, (3) description of the “genetic potential” of new inbreds, and (4) variety registration and description. FMs will also be essential to test for negative pleiotropic side-effects. This will in addition lead to a better understanding of the nature of trait correlations, or “pleiotropic” effects described for major genes (Chen and Lübberstedt 2010). Various studies found close genetic correlations between plant height and flowering time. Interestingly, flowering time associated polymorphisms in D8, a gene initially identified by its mutant allele leading to dwarfing, had no effects on plant height (Thornsberry et al. 2001). Similarly, mutant alleles of brown midrib genes in maize were found to affect other agronomic characters, including plant height and biomass yield (Pedersen et al. 2005). However, none of the polymorphisms within the Bm3 gene affecting forage quality affected any of these agronomic traits (Chen et al. 2010). In conclusion, for composition of optimal haplotypes for genes shown to affect one or more traits of interest, multiple traits need to be considered.

It remains to be seen, how FMs will contribute to marker-assisted (recurrent) selection, in particular as compared to genomic selection (GS) procedures based on low cost markers without requirements on their functional characterization (Bernardo and Yu 2007). Although most empirical studies of GS are still limited, accurate estimates of breeding values combined with the possibility for selection of kernels before planting (by seed-chipping) and selection in off-season winter nurseries makes GS very promising for maximizing genetic gain in breeding programs, especially when compared to marker assisted selection based only on markers with statistically significant trait associations (Bernardo and Yu 2007; Heffner et al. 2009; Mayor and Bernardo 2009). GS has been described as brute-force and black box procedure to increase genetic gain (Bernardo and Yu 2007), as selection is based on a large number of markers without prior knowledge of QTL positions or genetic mechanisms involved in phenotypic variation. Markers

in LD with favorable QTL receive a large estimated breeding value, even if the QTL is unknown. In GS, lines are selected based on the sum of estimated breeding values of markers across the whole genome, rather than site specific introgression of significant QTL.

Current research on GS is focused on developing statistical methods that incrementally improve the accuracy, i.e., the correlation between predicted and observed breeding values of individuals in a breeding population (de los Campos et al. 2009; Gianola et al. 2006; Habier et al. 2007; Heffner et al. 2009; Kizilkaya et al. 2010; Xu 2003b; Zhong et al. 2009). Alternatively, as functional genomic knowledge increases it seems reasonable to hypothesize that the concept of gene pyramiding could be extended to genome assembly (GA) for polygenic traits. To our knowledge GS has not been compared with gene pyramiding, much less GA. The question is: what criteria should be used to make such a comparison? While genetic gain, or its accuracy component, is a simple criterion, it is not realistic. Actual breeding decisions are based upon multiple breeding objectives, such as maximizing genetic gain, while maintaining genetic diversity throughout the genomes of the breeding population.

Xu et al. (2011) used an operations research approach to address the challenges imposed by varying degrees of LD among favorable functional alleles to assemble a desired phenotype in minimal time while avoiding loss of genetic diversity for other loci in a population. Importantly, using an optimization approach changes the framework for evaluation from a simple criterion of accuracy to the more realistic situation of meeting multiple breeding objectives simultaneously. Hypothetically, GA, based on knowledge of FMs, LD, and genomic diversity should outperform GS for realistic breeding objectives. The likely outcome will be conditional, i.e., depending upon the structure of the breeding population, genetic architecture of the trait, and genome structure we will likely find Pareto Frontiers describing when the hypothesis is true and when it is false.

The question remaining is how GS would benefit from an increasing number of characterized functional genes. Calus et al. (2008) showed that haplotype versus random marker-based GS is more efficient to predict breeding values. It therefore appears likely that marker-multiplexes employed in GS procedures based on previously characterized QTPs are at least superior to random markers in populations with low LD. For populations with high LD, where markers are more likely to be in LD with favorable QTL, prior knowledge of FM might not improve genetic gain in GS (Ødegård et al. 2009). The contribution of FM to the genetic gain in this case, however, will come from an increasing knowledge of allele effects, distributions, and environment/genetic interactions.

Perspective: Future Opportunities

New sequencing platforms have motivated genome sequencing projects in larger populations in different species. In humans, the 1000 Genome project was launched in 2008 as a consortium involving more than 75 universities and companies

worldwide. The goal is to sequence genomes and reveal sequence polymorphisms in more than a thousand individuals from different ethnic groups. Another large sequencing initiative is the Genome 10K Project, aiming to sequence the genomes of 10,000 vertebrate species by 2015 (<http://genome10k.soe.ucsc.edu/>). In plants, the 1001 Genomes Project was initiated in 2008, with the objective of revealing whole-genome sequence variants in 1,001 accessions of *Arabidopsis thaliana*. In maize, seven inbred lines have been resequenced by public institutions in United States and China (Lai et al. 2010; Schnable et al. 2009).

The challenge will be translating this huge amount of genomic information into QTL, QTPs and FMs for crop improvement. In plant breeding, the importance of a marker normally depends on how it predicts the phenotype, and accurate predictions depend on accurate estimation of marker effects based on phenotypic evaluations. Although phenotyping has become more efficient over the years with larger and automated field machinery and hand held computers, field characterization of breeding lines normally requires large allocation of land and labor work. As genotyping costs reduce, phenotyping becomes the major bottleneck in marker assisted breeding. More recently “phenomics”, which is using instruments arrays that allows high through-put screening of thousands of lines consistently in short periods of time, has been suggested as the approach that will make phenotyping “catch up” with genomics (Finkel 2009). The use of phenomics, however, will not surrogate field experimentation, and allocation of land and phenotyping labor will still be necessary for major plant breeding traits.

Another challenge associated with FM development is the biological validation of statistically inferred QTPs. Transgenic constructions require time consuming regulations for field evaluations, and are usually vulnerable to position effects, which substantially affect the expression of genes depending on the (random) introgression site in the genome. Backcrossing has been a traditional approach for introgression of moderate number of alleles, but it has the drawback of introgressing unwanted genome from the donor parent by linkage drag. The magnitude of linkage drag can be minimized by selection of recurrent markers flanking the target region. This approach, however, requires larger populations as flanking markers are closer to the target region (Hospital 2001). Recently, ZFN was introduced as a promising technology to assist allele introgression without some of the drawbacks from transgenic and backcross approaches. ZFN promotes recombination in defined chromosome segments, permitting allele introgression without linkage drag with smaller population sizes (Shukla et al. 2009).

Even though phenotyping, validation, and introgression of favorable QTPs are still major drawbacks, identification of candidate QTPs and subsequent FM development are increasingly reported. A number of FMs have already been developed in different plant species (Fan et al. 2009; Ji et al. 2010; Iyer-Pascuzzi and McCouch 2007; Shi et al. 2008; Su et al. 2010; Tommasini et al. 2006). Developing optimal strategies to integrate this increasing knowledge of functionality of genomic regions, and combining this information with phenotypic and GS will be essential to maximize genetic gain. Most likely, FMs will have to be evaluated on a case by cases basis, where their significance to the genetic gain will depend on the populations and environments of individual breeding programs.

References

- Andersen JR, Lübberstedt T (2003) Functional markers in plants. *Trends Plant Sci* 8:554–560
- Andersen JR, Schrag T, Melchinger AE, Imad Z, Lübberstedt T (2005) Validation of Dwarf8 polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). *Theor Appl Genet* 111:206–217
- Andersen JR, Zein I, Wenzel G, Krützfeldt B, Eder J, Ouzunova M, Lübberstedt T (2007) High levels of linkage disequilibrium and associations with forage quality at a phenylalanine ammonia-lyase locus in European maize (*Zea mays* L.) inbreds. *Theor Appl Genet* 114:307–319
- Andersen JR, Zein I, Wenzel G, Darnhofer B, Eder J, Ouzunova M, Lübberstedt T (2008) Characterization of phenylpropanoid pathway genes within European maize (*Zea mays* L.) inbreds. *BMC Plant Biol* 8:2
- Appleby N, Edwards D, Batley J (2009) New technologies for ultra-high throughput genotyping in plants. In: Somers D, Langridge P, Gustafson J (eds) *Plant genomics methods and protocols*. Humana Press, New York, pp 19–39
- Barker T, Campos H, Cooper M, Dolan D, Edmeades GO, Habben J, Schussler J, Wright D, Zinselmeier C (2005) Improving drought tolerance in maize. *Plant Breed Rev* 25:173–253
- Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. 49th annual corn and sorghum industry research conference, ASTA, Washington, DC, pp 250–266
- Behn A, Hartl L, Schweizer G, Wenzel G, Baumer M (2004) QTL mapping for resistance against non-parasitic leaf spots in a spring barley doubled haploid population. *Theor Appl Genet* 108:1229–1235
- Bernardo R (2002) *Breeding for quantitative traits in plants*. Stemma Press, Woodbury
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci* 48:1649–1664
- Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. *Crop Sci* 47:1082–1090
- 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
- Bortiri E, Chuck G, Vollbrecht E, Rocheford T, Martienssen R, Hake S (2006) *ramosa2* encodes a lateral organ boundary domain protein that determines the fate of stem cells in branch meristems of maize. *Plant Cell* 18:574–585
- Brenner EA, Zein I, Chen Y, Andersen JR, Wenzel G, Ouzunova M, Eder J, Darnhofer B, Frei U, Barrière Y, Lübberstedt T (2010) Polymorphisms in O-methyltransferase genes are associated with stover cell wall digestibility in European maize (*Zea mays* L.). *BMC Plant Biol* 10:27
- Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Nat Acad Sci USA* 101:15255–15260
- Buerstmayr H, Steiner B, Hartl L, Griesser M, Angerer N, Lengauer D, Miedaner T, Schneider B, Lemmens M (2003) Molecular mapping of QTLs for fusarium head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. *Theor Appl Genet* 107:503–508
- Byrne PF, McMullen MD, Wiseman BR, Snook ME, Musket TA, Theuri JM, Widstrom NW, Coe EH (1998) Maize silk maysin concentration and corn earworm antibiosis: QTLs and genetic mechanisms. *Crop Sci* 38:461–471
- Calus MPL, Meuwissen THE, Roos APW, Veerkamp RF (2008) Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178:553–561
- Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. *Genetics* 172:2449–2463

- Camus-Kulandaivelu L, Chevin LM, Tollon-Cordet C, Charcosset A, Manicacci D, Tenaillon MI (2008) Patterns of molecular evolution associated with two selective sweeps in the Tb1-Dwarf8 region in maize. *Genetics* 180:1107–1121
- Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Genetics* 5:618–625
- Chen Y, Lübberstedt T (2010) Molecular basis of trait correlations. *Trends Plant Sci* 15:454–461
- Chen Y, Zein I, Brenner EA, Andersen JR, Landbeck M, Ouzunova M, Lübberstedt T (2010) Polymorphisms in monolignol biosynthetic genes are associated with biomass yield and agronomic traits in European maize (*Zea mays* L.). *BMC Plant Biol* 10(12)
- Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. *Genetics* 139:1455–1461
- Ching A, Caldwell KS, Jung M, Dolan MSO, Smith H, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 14:1–14
- Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene tb1. *Proc Nat Acad Sci* 101:700–707
- Clark TG, Andrew T, Cooper GM, Margulies EH, Mullikin JC, Balding DJ (2007) Functional constraint and small insertions and deletions in the ENCODE regions of the human genome. *Genome Biol* 8:180
- Cooper M, DeLacy IH (1994) Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theor Appl Genet* 88:561–572
- Costa F, Weg WE, Stella S, Dondini L, Pratesi D, Musacchi S, Sansavini S (2008) Map position and functional allelic diversity of Md-Exp7, a new putative expansin gene associated with fruit softening in apple (*Malus × domestica* Borkh.) and pear (*Pyrus communis*). *Tree Genet Genomes* 4:575–586
- de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375–385
- Fan C, Yu S, Wang C, Xing Y (2009) A causal C-A mutation in the second exon of GS3 highly associated with rice grain length and validated as a functional marker. *Theor Appl Genet* 118:465–472
- Finkel E (2009) With “phenomics” plant scientists hope to shift breeding into overdrive. *Science* 325:380–381
- Flint-Garcia SA, Thuitet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J* 44:1054–1064
- Gianola D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173:1761–1776
- Guillet-Claude C, Birolleau-Touchard C, Manicacci D, Fourmann M, Barraud S, Carret V, Martinant JP, Barrière Y (2004) Genetic diversity associated with variation in silage corn digestibility for three O-methyltransferase genes involved in lignin biosynthesis. *Theor Appl Genet* 110:126–135
- Guo B, Slepner DA, Beavis WD (2010) Nested association mapping for identification of functional markers. *Genetics* 186:373–383
- Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101:5–18
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397
- Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wrutzel ET, Yan J, Buckler ES (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319:330–333

- He XY, Zhang YL, He ZH, Wu YP, XiaoYG MCX, Xia XC (2008) Characterization of phytoene synthase 1 gene (Psy1) located on common wheat chromosome 7A and development of a functional marker. *Theor Appl Genet* 116:213–221
- Heffner EL, Sorrells ME, Jannink J (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12
- Heo M, Leibel RL, Boyer BB et al (2001) Pooling analysis of genetic data: the association of leptin receptor (LEPR) polymorphisms with variables related to human adiposity. *Genetics* 159:1163–1178
- Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4:2
- Hospital F (2001) Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics* 158:1363–1379
- Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175:1937–1944
- Iyer-Pascuzzi AS, McCouch SR (2007) Functional markers for xa5-mediated resistance in rice (*Oryza sativa*, L.). *Mol Breed* 19:291–296
- Ji Q, Lu J, Chao Q, Zhang Y, Zhang M, Gu M, Xu M (2010) Two sequence alterations, a 136 bp InDel and an A/C polymorphic site, in the S5 locus are associated with spikelet fertility of indica-japonica hybrid in rice. *J Genet Genomics* 37:57–68
- Jones E, Chu WC, Ayele M, Ho J, Bruggeman E, Yourstone K, Rafalski R, Smith OS, McMullen MD, Bezawada C, Warren J, Babayev J, Basu S, Smith S (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.) germplasm. *Mol Breed* 24:165–176
- Juenger TE, Sen S, Stowe KA, Simms EL (2005) Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica* 123:87–105
- Kizilkaya K, Fernando RL, Garrick DJ (2010) Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J Anim Sci* 88:544–551
- Kump KL, Bradbury PJ, Wissner RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat Genet* 43:2
- Lai J, Li R, Xu X et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature* 42:1027–1030
- Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Mol Biol* 48:453–461
- 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 Gen* 121:1303–1310
- Manicacci D, Camus-Kulandaivelu L, Fourmann M et al (2009) Epistatic interactions between *Opaqe2* transcriptional activator and its target gene *CyPPDK1* control kernel trait variation in maize. *Plant Physiol* 150:506–520
- Mayor PJ, Bernardo R (2009) Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F2 populations. *Crop Sci* 49:1719–1725
- McPherson R, Persemlidis A, Kavaslar N et al (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science* 316:1488–1491
- Nordborg M, Weigel D (2010) Next-generation genetics in plants. *Nature* 456:10–13
- Ødegård J, Sonesson AK, Yazdi MH, Meuwissen THE (2009) Introgression of a major QTL from an inferior into a superior population using genomic selection. *Genet Sel Evol* 41:38
- Ogundiwin EA, Peace CP, Nicolet CM, Rashbrook VK, Gradziel TM, Bliss FA, Parfitt D, Crisosto CH (2008) Leucoanthocyanidin dioxygenase gene (PpLDOX): a potential functional marker for cold storage browning in peach. *Tree Genet Genomes* 4:543–554

- Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* 15:1795–1806
- Pedersen JF, Vogel KP, Funnell DL (2005) Impact of reduced lignin on plant fitness. *Crop Sci* 45:812–819
- Pinson SRM, Capdevielle FM, Oard JH (2005) Blight resistance in rice using recombinant inbred lines. *Crop Sci* 45:503–510
- Polidoros AN, Mylona PV, Arnholdt-Schmitt B (2009) Aox gene structure, transcript variation and expression in plants. *Physiol Plant* 137:342–353
- Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 46:229–241
- Risch N (2000) Searching for genetic determinants in the new millennium. *Nature* 405:847–856
- Salvi S, Sponza G, Morgante M et al (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Nat Acad Sci* 104:11376–11381
- Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115
- Scho CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167:485–498
- Scott LJ, Mohlke KL, Bonnycastle LL et al (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345
- Shi WYY, Chen S, Xu M (2008) Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. *Mol Breed* 22:185–192
- Shukla VK, Doyon Y, Miller JC et al (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459:437–441
- Stacey SN, Manolescu A, Sulem P et al (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 39:865–869
- Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A, Piepho HP (2009) Association mapping reveals gene action and interactions in the determination of flowering time in barley. *Theor Appl Genet* 118:259–273
- Su Z, Hao C, Wang L, Dong Y, Zhang X (2010) Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 122:211–223
- Talerico M, Berget SM (1990) Effect of 5' splice site mutations on splicing of the preceding intron. *Mol Cel Biol* 10:6299–6305
- Tang D, Wu W, Li W, Lu H, Worland AJ (2000) Mapping of QTLs conferring resistance to bacterial leaf streak in rice. *Theor Appl Genet* 101:286–291
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* 28:286–289
- Tian F, Dradbury PJ, Brown PJ et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43:159–162
- Tommasini L, Yahiaoui N, Srichumpa P, Keller B (2006) Development of functional markers specific for seven *Pm3* resistance alleles and their validation in the bread wheat gene pool. *Theor Appl Gen* 114:165–175
- Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16:2719–2733
- Wray NR, Goddard ME, Visscher PM (2007) Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 17:1520–1528
- Xu S (2003a) Estimating polygenic effects using markers of the entire genome. *Genetics* 163:789–801
- Xu S (2003b) Theoretical basis of the Beavis effect. *Genetics* 165:2259–2268
- Xu P, Wang L, Beavis WD (2011) An optimization approach to gene stacking. *Eur J Oper Res* 214(1):168–178

- Yang X, Guo Y, Yan J, Zhang J, Song T, Rocheford T, Li JS (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. *Theor Appl Genet* 120:665–678
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551
- Zein I, Wenzel G, Andersen JR, Lübberstedt T (2007) Low level of linkage disequilibrium at the *COMT* (caffeic acid O-methyl transferase) locus in European maize (*Zea mays* L.). *Genet Res Crop Evol* 54:139–148
- Zhang K, Tian J, Zhao L, Wang S (2008) Mapping QTLs with epistatic effects and QTL x environment interactions for plant height using a doubled haploid population in cultivated wheat. *J Genet Genomics* 35:119–127
- Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182:355–364