Association Mapping for Enhancing Maize (*Zea mays* L.) Genetic Improvement

Jianbing Yan,* Marilyn Warburton, and Jonathan Crouch

ABSTRACT

Association mapping through linkage disequilibrium (LD) analysis is a powerful tool for the dissection of complex agronomic traits and for the identification of alleles that can contribute to the enhancement of a target trait. With the developments of high throughput genotyping techniques and advanced statistical approaches as well as the assembling and characterization of multiple association mapping panels, maize has become the model crop for association analysis. In this paper, we summarize progress in maize association mapping and the impacts of genetic diversity, rate of LD decay, population size, and population structure. We also review the use of candidate genes and gene-based markers in maize association mapping studies that has generated particularly promising results. In addition, we examine recent developments in genome-wide genotyping techniques that promise to improve the power of association mapping and significantly refine our understanding of the genetic architecture of complex quantitative traits. The new challenges and opportunities associated with genome-wide analysis studies are discussed. In conclusion, we review the current and future impacts of association mapping on maize improvement along with the potential benefits for poor people in developing countries who are dependent on this crop for their food security and livelihoods.

J. Yan, National Maize Improvement Center of China, China Agricultural Univ., Beijing 100193, China; J. Yan and J. Crouch, International Maize and Wheat Improvement Center (CIMMYT), Apartado Postal 6-640, 06600 Mexico, DF, Mexico; M. Warburton, USDA-ARS, Corn Host Plant Resistance Research Unit, Box 9555, Mississippi State, MS 39762. Received 27 Apr. 2010. *Corresponding author (yjianbing@gmail.com).

Abbreviations: CGIAR, Consultative Group on International Agricultural Research; *crtRB1*, β-*carotene hydroxylase* gene; GWA, genome-wide association; GWS, genome-wide selection; LD, linkage disequilibrium; MAF, minor allelic frequency; MAGIC, multiparent advanced generation intercross; MARS, marker-assisted recurrent selection; MAS, marker-assisted selection; NAM, nested association mapping; nMDS, nonmetric multidimensional scaling; PCA, principal components analysis; QTL, quantitative trait locus or loci; SNP, single nucleotide polymorphism.

Maize (*Zea mays* L.) is the world's most widely grown crop with an annual global production of 826 million t in 2008 (available at http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor [verified 8 Dec. 2010]). There are 14 countries where maize is estimated to provide 25 to 50% of the total human energy consumption and a further 27 countries where maize provides 10 to 25% of the total energy consumption (FAO, 2009a). Maize is also an important source of cooking oil, biofuel, and animal feed. By 2050, the predicted 9 billion people in the world will require 70% more food than today's population, and a large proportion of the increased demand will come from developing countries (FAO, 2009b). It is estimated that more than half of the increased demand for cereals as a whole will come from maize farmers and consumers. The necessary increase in maize production will require substantial changes in agronomic practices

Published in Crop Sci. 51:433–449 (2011).
doi: 10.2135/cropsci2010.04.0233
Published online 29 Dec. 2010.

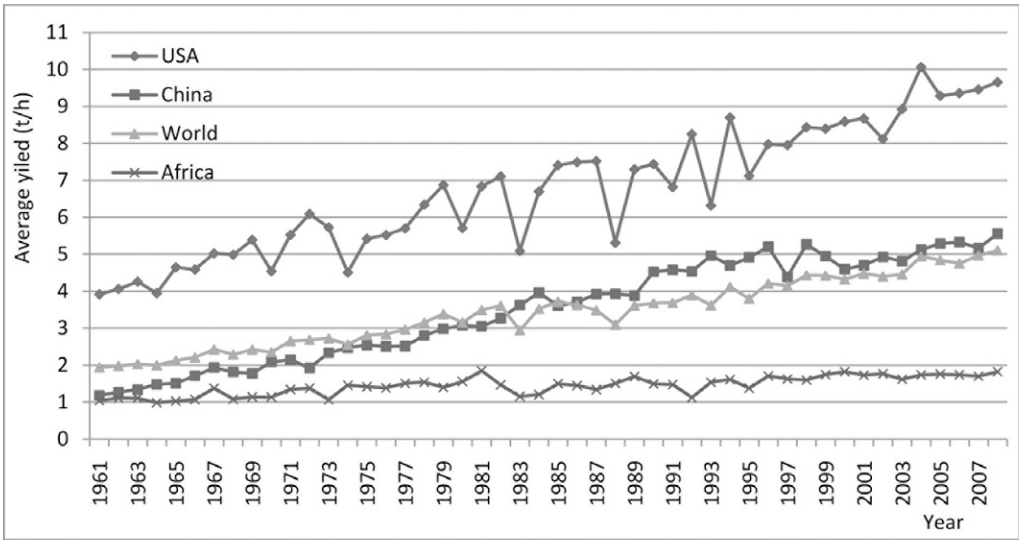

Figure 1. Average yield of maize during 1961 through 2008 for the United States and China compared to averages across Africa and the whole world (data from FAO in 2010; http://faostat.fao.org [verified 6 Dec. 2010]).

and methods of genetic improvement. However, there is a danger that these improved yields will come at a high environmental cost due to overapplication of synthetic fertilizers, which cannot be sustained (Robertson and Vitousek, 2009). Fortunately, maize is also an important model organism for cytogenetics, genetics, genomics, and functional genomics studies (Strable and Scanlon, 2009). Thus, there is a tremendous innovation stream for maize breeders to utilize in their attempts to substantially increase maize productivity in an environmentally sensitive way.

The technology developments of the "Green Revolution" (including irrigation, fertilizer, and new cultivars) led to more than a doubling of global maize, wheat (*Triticum aestivum* L.), and rice (*Oryza sativa* L.) production between 1966 and 2000 (Khush, 2001). Maize exceeded this trend in most areas (except Africa) largely due to the additional benefits of harnessing heterosis in hybrids and the improved performance and adaptation of the inbred parents (Fig. 1). However, the rate of yield improvement is not in line with current and predicted increases in demand. The situation is especially severe in Africa, which did not benefit from yield increases from Green Revolution varieties (Ejeta, 2010). Consequently, average maize yields in Africa have only increased by about 0.5 t ha^{-1} over the past half century compared to a 6 t ha^{-1} increase during the same period in the United States (Fig. 1). As maize is a primary staple food in many African countries, increasing maize productivity is a key priority for African agricultural development to reduce poverty and hunger in this region and thus a cornerstone of the proposed African Green Revolution (Ejeta, 2010). However, the challenge to successfully replicate the Asian Green Revolution in Africa is confounded by a multitude of environmental stresses that are becoming more dynamic due to climate change and thus create a highly difficult target environment for maize breeders and farmers (Collier et al., 2008). Thus, where the Asian Green Revolution was driven by a few oligogenic traits, the African Green Revolution will require products that effectively combine many complex traits into easily disseminated new maize varieties. Fortunately, the new techniques of applied genomics research and molecular breeding are ready to meet these demands through a knowledge-led approach to maize breeding.

Quantitative trait locus (QTL) mapping is a powerful and well-established tool for studying the genetic basis of complex quantitative traits in plants and animals. More than 10,000 articles published during the last three decades on QTL mapping in different species are listed in the Pubmed database (available at http://www.ncbi.nlm.nih.gov/pubmed [verified 6 Dec. 2010]). Of these, more than 360 articles relate to reports of over 1000 QTLs associated with various traits in maize. Despite the surfeit of mapping publications, to date only a few QTLs have been identified at the gene level through cloning (Moose and Mumm, 2008). This is mainly because map-based cloning of QTLs is a very time consuming and expensive process in maize and other crop species. Association mapping has been widely used to study the genetic basis of complex traits in human and animal systems and is a very efficient and effective method for confirming candidate genes or for identifying new genes (Altshuler et al., 2008; Hunter and Crawford, 2008). Association mapping is now being increasingly used in a wide range of plants (Rafalski, 2010), where it appears to be more powerful than in humans or animals (Zhu et al., 2008). Unlike linkage mapping, association mapping can explore all the recombination events and mutations in a given population and with a higher resolution (Yu and Buckler, 2006). However, association mapping has a lower power to detect rare alleles in a population, even those with large effects, than linkage mapping (Visscher, 2008). In this review we discuss recent progress and the particular strengths of association mapping in

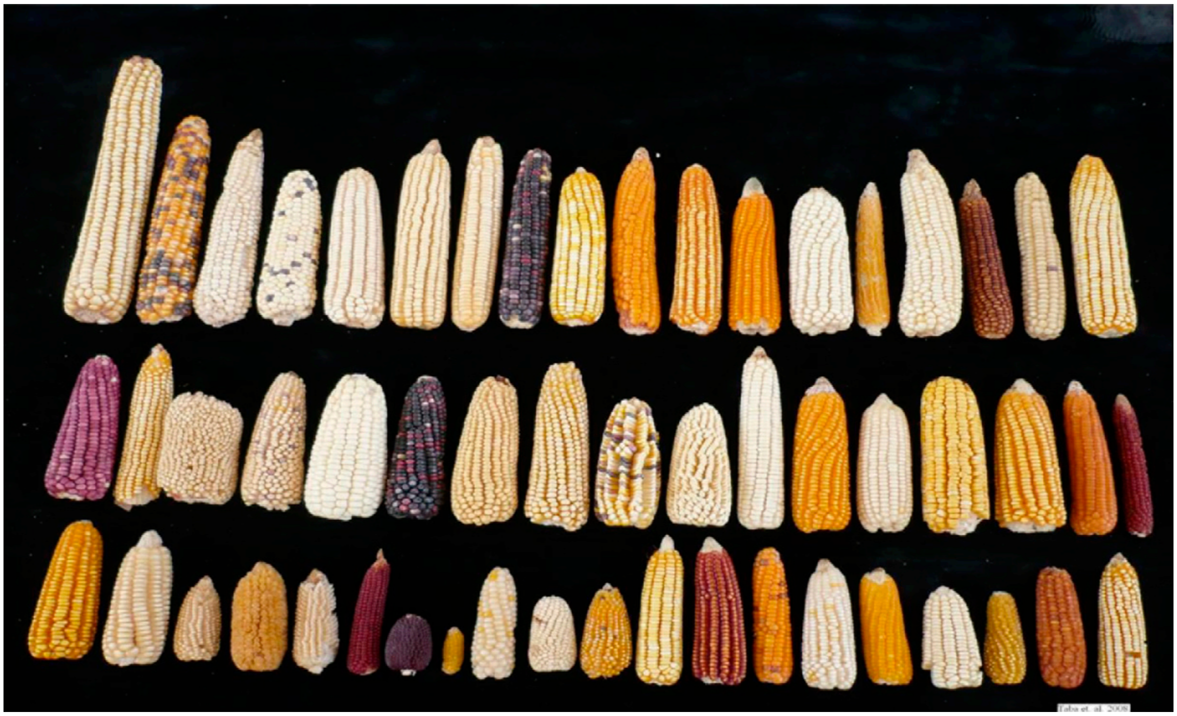

Figure 2. Examples of the range of phenotypic variation in maize germplasm held in the CIMMYT genebank (photo provided by Dr. Suketoshi Taba, CIMMYT).

maize and the requirements for its effective use in enhancing maize genetic improvement.

Abundant Genetic Diversity in the Maize Gene Pool Improves the Power of Association Mapping

Genetic mapping via linkage or association analyses cannot be performed in the absence of measurable polymorphisms, so abundant differences at the phenotypic level and a high density of polymorphisms at the DNA sequence level are essential. Maize shows an amazing degree of phenotypic diversity: plant height can range from 0.5 to 5 m at maturity; flowering dates vary from 2 to 11 mo after planting; the ear and kernels vary in color, length, size, shape, etc. (Sprague and Dudley, 1988; Fig. 2). For nearly every trait of economic or agronomic importance, there are measurable phenotypic differences within the global maize germplasm pool. The tremendous allelic diversity underlying this astonishing phenotypic variation in maize has been exploited and used throughout history, first via farmer selection and more recently in breeding programs by geneticists and breeders. During the past 10,000 yr since domestication from its wild relative teosinte (*Zea mays* subsp. *parviglumis*), maize has retained and further generated vast quantities of allelic diversity and genes via an active system of transposable elements. In addition, gene flow from teosinte and between maize populations (enhanced by the outcrossing nature of maize), farmer and natural selection (especially following introduction into new growing regions), recombination, drift, and mutation have all contributed to the diversity seen in maize germplasm (Walbot, 2009).

It has been estimated that there is a polymorphism between two diverse lines every 44 bp throughout the maize genomic sequence (Gore et al., 2009) and that the divergence between two maize inbred lines is even greater than between human and chimpanzees, which diverged as independent species 3.5 million years ago (Buckler and Stevens, 2005). Maize has a moderately sized genome of 2300 Mbp, predicted to contain over 32,000 genes (Schnable et al., 2009). Several million polymorphisms including single nucleotide polymorphisms (SNPs) and indels have been identified through comparison of the sequences of 27 diverse inbred lines (Gore et al., 2009). There are typically multiple polymorphisms within each gene, leading to a higher frequency of amino acid differences than in most plants, which translates to the high levels of phenotypic differences observed at the whole plant level (Ching et al., 2002; Rafalski, 2010). Thus, there are multiple haplotypes (combinations of alleles and/or SNPs within a gene) within each gene that allow association mapping of almost any trait, providing the germplasm panel captures a large proportion of the total diversity available for that trait (Li and Jiang, 2005; Zhu et al., 2008).

Linkage Disequilibrium is an Important Factor in Association Mapping

Linkage disequilibrium (LD) is the nonrandom association in a population of alleles at two or more loci. The term was originally defined in relation to the population of alleles that reside on the same chromosome. Although LD is a population-based phenomenon (rather than an individual genome-based phenomenon), it is generally observed that there tends to be a higher LD between alleles that are located more closely together (because of their lower probability of being separated from one another by recombination). Thus, the random association between alleles might be reduced by linkage thereby creating the so-called disequilibrium. However, with the increasing use of genome-wide association (GWA) studies the term is now being used in a broader context by some researchers to also encompass alleles across chromosomes that show some association. In collections of commercial maize germplasm the rate of LD decay (the rate of return to random association between two given alleles) is relatively slow. However, decay of LD between two polymorphic sites in diverse maize germplasm collections occurs very rapidly within a few kilobase pairs due to the high rate of recombination in this material (Tenaillon et al., 2001; Gore et al., 2009; Yan et al., 2009). Both linkage mapping and association mapping studies aim to identify functional sequence variants (alleles) encoding changes in phenotype or markers sufficiently closely linked to them to allow breeders to routinely select and manipulate these alleles in diverse populations. In traditional linkage mapping studies, segregating individuals are genotyped with hundreds or thousands of random markers, and there is a low probability that these markers will include the functional DNA variants themselves or even markers closely linked to them.

Besides physical distance on the chromosome, many factors affect the breakdown of LD, including genetic drift, natural and artificial selection, mating system, and admixture of different populations (Flint-Garcia et al., 2003; Gaut and Long, 2003; Yu and Buckler, 2006). Several statistical parameters can be used to estimate the extent of LD (Hedrick, 1987), most commonly r^2, which estimates the correlation between allelic states of two given polymorphic loci. Based on multiple case studies in maize, LD decay ranges from less than 1 kbp (Tenaillon et al., 2001) in landraces to more than 100 kbp in elite (more closely related) breeding lines (Ching et al., 2002). Given this relationship, association analysis is particularly powerful in maize, as the resolution can be controlled by choice of association mapping panel: more elite germplasm for higher LD or more diverse and/or exotic germplasm for less LD. For example, significant marker–trait associations can be identified using elite lines with higher LD that will then require fewer markers, whereas more diverse lines with smaller linkage blocks (and thus lower LD) will require more markers but will get closer to the gene of interest.

Linkage disequilibrium can be greatly overestimated (especially at larger genomic distances) when sample sizes smaller than 50 individuals are used (Yan et al., 2009). Decay in LD also varies widely in different chromosomal regions (Yan et al., 2009). This may be due to the great variation in recombination rates along the chromosomes, including a low recombination rate in centromeric regions and a high recombination rate within genic regions due to retrotransposon insertions (Dooner and He, 2008). Very extensive LD has been found in regions that have experienced strong selective sweeps (Jung et al., 2004; Tian et al., 2009), such as is found around the *Y1* gene, which controls carotenoid (and thus color) production in maize grains (Palaisa et al., 2003). High levels of nucleotide diversity in and around this gene have been identified in white maize germplasm but not in modern yellow maize cultivars (Palaisa et al., 2003; Fu et al., 2010), which have been strongly selected for the health benefits of carotenoids for humans and animals (Mangelsdorf and Fraps, 1931). This selection pressure has caused the LD around the *Y1* locus of yellow maize to span hundreds of kilobase pairs. Another sequenced region of chromosome 10 contains a long LD region covering 1 Mbp, indicating a lack of recent recombination or a lack of sequence diversity perhaps due to selection, but the underlying genes have not been identified (Tian et al., 2009). Based on genome-wide sequencing of 27 diverse inbred lines, more than 100 LD blocks of different sizes (from thousands to millions of base pairs in length) have been identified in the maize genome (Gore et al., 2009). Within these regions for some sets of germplasm, it may not be possible to identify markers very closely linked to the functional mutation of target genes. For this reason, choice of appropriate germplasm to maximize the number of historical recombinations and mutation events (and thus reduce LD) within and around the gene of interest is critical for the success of association analysis.

In general, genetic linkage mapping studies identify linkage between a marker and the more distant functional DNA sequence by creating biparental mapping populations that have experienced only had a few generations of recombination since their creation, thereby increasing the probability that random markers will still be in LD with functional variants. However, linked markers identified in this manner may not be suitable for marker-assisted selection involving unrelated maize genotypes, since the linkage between the markers and the useful functional variants may have been broken during the recombination history of these unrelated genotypes. Random markers used for association mapping in maize must be much closer to the functional variants for a statistically significant association to be detected due to LD breakdown in a diverse association mapping panel. Many generations of recombination separate unrelated lines in a diverse association mapping panel (starting from their most recent common ancestor) compared to the lines in a genetic linkage mapping population (starting from the two lines crossed to generate them). The level of LD among the markers used to genotype the individuals in any given association mapping panel is an important index for a successful association study, as it will help to estimate the resolution and minimum number of markers needed for detecting significant associations (Yan et al., 2009).

Population Structure Matters for Association Mapping

Population structure can cause some allele frequencies to differ significantly between subpopulations, which can create unexpected LD between unlinked loci across the genome (Ersoz et al., 2009). For example, the *d8* gene has been proposed to affect flowering time and has been analyzed in three different sets of germplasm (Thornsberry et al., 2001; Andersen et al., 2005; Camus-Kulandaivelu et al., 2006). When population structure was ignored, significant associations were identified in all three independent studies. Flowering time is an important component of adaptation that has been under high selection pressure during domestication and migration, and allele frequencies of genes related to this trait can thus vary in different subpopulations adapted to contrasting latitudes. For example, 33 to 35% of the phenotypic variation in a diverse maize panel for male flowering time and female flowering time was found to be explained by population structure (Flint-Garcia et al., 2005). Therefore, if subpopulation structure is not accounted for, spurious (i.e., noncausative) associations may be detected between flowering time and alleles at other loci that are all differentially distributed among subpopulations. Associations between *d8* and flowering time disappeared in some cases when the analysis was adjusted for population structure.

Nearly all traits of agronomic or economic importance have been intensively selected during extensive breeding efforts over the past century. This has led to significant population structure issues for all association analyses using modern maize germplasm that must be quantified and the analysis adjusted so that results are not seriously compromised (Ersoz et al., 2009). Neutral markers can be used to estimate population structure using traditional fingerprinting and diversity analyses. Several statistical methods have been used to control the effect of population structure in association analyses including genomic control (Devlin and Roeder, 1999; Mackay, and Powell, 2007), structured association (Pritchard et al., 2000; Falush et al., 2003), principal components analysis (PCA) (Patterson et al., 2006; Price et al., 2006), nonmetric multidimensional scaling (nMDS) (Zhu and Yu, 2009), and the unified mixed-model approach (Flint-Garcia et al., 2005; Yu et al., 2006; Zhang et al., 2010b). A two-stage dimension determination approach using both PCA and nMDS has been demonstrated to be the best approach to capture the major structure of association panels to maximize the rejection of false positives while maximizing the statistical power to identify real associations (Zhu and Yu, 2009).

Candidate Gene Strategy

Association analyses generally have a low statistical power for correlating rare alleles with phenotypic differences, yet these may be highly beneficial variants that breeders are searching for. The candidate gene method of association analysis is a hypothesis-driven approach for complex trait dissection that aims to improve the odds of identifying the most important alleles. It involves genotyping or resequencing those genes considered to have a high probability of association with the phenotype(s) of interest within the germplasm being tested. There are a number of different approaches to implementing this strategy depending on the method used to identify the candidate gene and the level of confidence the researcher has in the likelihood that a given gene is important for the target trait. In the past it was common to sequence the gene of interest as fully as possible across a limited number of diverse lines (typically 24 to 48) to identify possible causal polymorphisms, such as SNPs causing amino acid changes or indels in untranslated or translated regions. The selected polymorphisms were then screened across a larger germplasm collection (of hundreds or thousands of genotypes) using inexpensive PCR-based SNP and/or indel genotyping assays (rather than sequencing) to confirm the associations between genotype and phenotype. In another method, the partial or entire gene is sequenced in all individuals of a germplasm panel (of several hundred genotypes) to identify significant associations, either with the causal polymorphism(s) or a polymorphism that is within LD distance to a causal polymorphism. Although this is a more expensive approach, it may identify rare polymorphisms that can be missed by the first strategy. Determining which method to use has generally been based on the level of funding and the amount of time available for each study. However, resequencing of the entire gene has the added advantage that it can directly identify the best haplotype for each target breeding purpose.

More than 20 studies of candidate gene association analysis in maize have been published to date (Table 1). These studies used candidate genes from well characterized and relatively simple metabolic pathways (Wilson et al., 2004; Harjes et al., 2008; Yan et al., 2010b) or those with extensive prior evidence for the role of the candidate gene(s) in the control of the phenotype of interest. Such evidence may include information from map-based cloning studies (Salvi et al., 2007; Ducrocq et al., 2008; Zheng et al., 2008; Buckler et al., 2009), information from closely related species (Li et al., 2010a, b), and information from QTL mapping studies and/or expression results (Krill et al., 2010). Expressed genes underlying major QTLs are used in association analysis to confirm in which gene(s) the causal polymorphisms can be found and to identify additional significant polymorphisms. *Vgt1* is a major QTL affecting flowering time that was isolated via map-based cloning and confirmed by association analysis (Salvi et al., 2007). Independent association analysis also discovered additional significant associations that explained more of the phenotypic variation of the trait (Ducrocq et al., 2008; Buckler et al., 2009).

Table 1. Summary of candidate gene association studies in maize.

Populations	Sample size	Background markers	Association method[†]	Candidate genes	Traits	References
Diverse inbred lines	92	141	LR+Q	*Dwarf8*	Flowering time and plant and ear height	Thornsberry et al., 2001
Elite inbred lines	71	55	LR+Q, GLM–Q	*Dwarf8*	Flowering time and plant height	Andersen et al., 2005
Diverse inbred lines and landraces	375 and 275	55 and 24	LR+Q, GLM+Q	*Dwarf8*	Flowering time	Camus-Kulandaivelu et al., 2006
Diverse inbred lines	95	141	LR+Q, GLM+Q	*Vgt1*	Flowering time	Salvi et al., 2007
Diverse inbred lines and landraces	375	55	MLM	*Vgt1*	Flowering time	Ducrocq et al., 2008
Diverse inbred lines	282	89 plus 553	MLM	*Vgt1, ZmRap2.7*	Flowering time	Buckler et al., 2009
Elite inbred lines	75	–	Case-control	*Y1*	Endosperm color	Palaisa et al., 2003
Diverse inbred lines	34	–	stepwise multiple linear regression	*CCoAOMT1, CCoAOMT2, AldOMT*	Cell wall digestibility	Guillet-Claude et al., 2004a
Diverse inbred lines	31	–	ANOVA	*ZmPox3*	Forage quality traits	Guillet-Claude et al., 2004b
Diverse inbred lines	97	47	LR+Q	*ae1, bt2, sh1, sh2, sugary1, waxy1*	Kernel composition and starch pasting properties	Wilson et al., 2004
Diverse inbred lines	42	101	LR+Q,GLM–Q	*bm3*	Forage quality traits	Lübberstedt et al., 2005
Diverse inbred lines	86	141	LR+Q	*a1, c2, whp1*	Maysin and chlorogenic acid content	Szalma et al., 2005
Diverse inbred lines	57	–	Haplotype tree scanning	*Sugary1*	Sweet taste	Tracy et al., 2006
Diverse inbred lines	32	101	LR+Q, GLM–Q	*PAL*	Forage quality traits	Andersen et al., 2007
Diverse inbred lines	40	101	MLM, GLM+Q	*C4H, 4CL1, 4CL2, C3H, F5H, CAD*	Forage quality traits	Andersen et al., 2008
Diverse inbred lines	282	89 plus 553	MLM	*lcyE*[§]	Carotenoid content	Harjes et al., 2008
Elite lines	71	–	unknown	*DGAT*	Oil content and composition	Zheng et al., 2008
Diverse inbred lines	281	89 plus 553	GLM+Q, MLM	*bx1*	DIMBOA[‡]	Butrón et al., 2010
Diverse inbred lines	121	82 plus 884	MLM	*GS3*	Kernel shape and weight	Li et al., 2010a
Diverse inbred lines	121	82 plus 884	MLM	*GW2*	Kernel shape and weight	Li et al., 2010b
Diverse inbred lines	375	55	MLM, GLM+Q	*opaque2, CyPPDK1*	Kernel quality traits	Manicacci et al., 2009
Diverse inbred lines	277	89 plus 553	MLM	*bif2*	Flowering time	Pressoir et al., 2009
Diverse inbred lines	281 245 155	89 plus 553 50 82 plus 884	MLM	*crtRB1*[¶]	Carotenoid content	Yan et al., 2010b
Diverse inbred lines	282	89 plus 553	MLM	*IDH*	Central carbon metabolism	Zhang et al., 2010
Diverse inbred lines	282	89 plus 553	MLM,GLM+Q MLM+C	21 genes[#]	Aluminum tolerance	Krill et al., 2010

[†] +, considering population structure; –, not considering population structure; GLM, generalized linear model; LR, logistic regression; MLM, mixed linear model (Q+K model); Q, population structure.

[‡] DIMBOA, 2-4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one.

[§] *lcye, lcyopene epsilon cyclase* gene.

[¶] *crtRB1, β-carotene hydroxylase* gene.

[#] Malic enzyme (ME), iron-responsive transporter-like (FE), major facilitator superfamily antiporter (ANT1), ABC transporter-like protein (ABC), isocitrate lyase (ISL), amino acid permease AUX1 (AUX1), SAH hydrolase (SAHH), cytochrome P450 (P450), pectin methylesterase (PME), phosphatidylinositol 3-kinase (PI3K), germin2 (oxalate oxidase) (OO2), isocitrate dehydrogenase (IDH), fumerase (FUM), ZmALMT1 (AL1), ZmALMT2 (AL2), ZmALMT3 (AL3), ZmALMT5 (AL5), ZmALMT8 (AL8), ZmALMT9 (AL9), ZmALMT16 (AL16), ZmASL (ASL).

Few QTLs affecting important traits in maize have yet been fully sequenced in multiple lines due to the complexity of the maize genome and the difficulty of cloning QTLs. Using information from related species may help direct the search for the gene(s) underlying the QTL of interest. Rice has a reasonably small genome and the full sequence has been available to researchers for several years (Goff et al., 2002), which has enabled a number of rice QTLs to be cloned, including QTLs contributing to yield and yield components (Xing and Zhang, 2010). Comparative genomics has revealed extensive macrosynteny and microsynteny between rice and maize genomes (Salse et al., 2004), and QTL controlling the same or similar traits have been identified in orthologous regions of maize, rice, and sorghum [*Sorghum bicolor* (L.) Moench] (Paterson et al., 1995; Yan et al., 2004). Homologous genes can have similar functions in different species (Kojima et al., 2002; Yano et al., 2000) or can influence the same trait but through different functions (Cockram et al., 2007). A cloned gene in rice may help to extract the orthologous sequence in maize and thereby identify candidate genes for maize association analysis studies. If the candidate gene influences the target function in maize, the polymorphism(s) underlying beneficial changes in expression of the target trait can then be identified. Clearly, caution must be applied when using this approach, as the causal polymorphism(s) in the maize gene may be completely different to those found to be important for the same trait in rice (Li et al., 2010a). As long as the same gene plays an important role in the target trait in both species, this approach will still provide a valuable shortcut. However, many genes identified in mutant screens in *Arabidopsis* spp. were not useful as candidates for maize flowering time in the nested association mapping (NAM) panel of maize lines (Buckler et al., 2009).

Marker–trait associations have been identified in maize based on information from the rice gene *GS3* (Fan et al., 2006), which underlies a major QTL affecting rice grain size. This gene was cloned based on its map position in rice (Fan et al., 2006), and a common SNP occurring in the second exon resulting in a stop codon was identified as the causal polymorphism of larger rice grain size (Fan et al., 2006, 2009). The *GS3* ortholog was cloned in maize and resequenced for association mapping analysis, and several polymorphisms significantly associated with maize kernel size were identified (Li et al., 2010a). None of the causal polymorphisms in maize were the same as the premature stop codon SNP found in rice and a similar trend was also observed in another gene, *GW2* (Li et al., 2010b), which implies that the orthlogous genes have different mechanisms in the two species (Li et al., 2010a, b).

Pathway-driven identification and selection of candidate genes has proven to be a successful strategy in several association studies. One successful use of pathway information to identify candidate genes focused on the production of provitamin A carotenoids, which are converted to vitamin A when metabolized by humans and animals. Biofortification of food and feed with provitamin A is an economical approach to address the global challenge of vitamin A deficiency. Gene-based marker-assisted selection (MAS) offers an efficient and highly cost effective method for selection of high provitamin A maize breeding lines (Harjes et al., 2008). The genes for plant carotenoid synthesis have been elucidated primarily in model systems and with very few exceptions the pathway is identical in other plants, although the number of genes controlling each step in the reaction can vary (Vallabhaneni and Wurtzel, 2009). Combining this genetic knowledge with detailed information about the metabolism and catabolism of carotenoids in maize and other plant species, two studies were performed to identify candidate genes for verification via association analysis of high provitamin A levels in maize grain. One study found that the allelic variation at the *lcyopene epsilon cyclase* gene (*lcye*) could explain over half the phenotypic variation in provitamin A levels, (Harjes et al., 2008), while the second study confirmed that alleles at β-*carotene hydroxylases* gene (*crtRB1*) explained 40% of the phenotypic variation in β-carotene levels (Yan et al., 2010b). It is significant and unexpected that haplotypes in just two genes could explain such a large proportion of the phenotypic variation in a trait showing continuous variation. This clearly begs the question whether similar assays can be developed for more complex traits such as pest and disease resistance or even drought tolerance and yield components.

Association mapping provides a means of identifying marker–trait associations using panels of germplasm but does not prove the function of the gene. Association mapping can be adversely affected by many factors, including population structure, small sample size, and low frequency of specific alleles, that may increase the detection of a false positive associations. It is very difficult to say which significance level is acceptable in a given association study. Alternatives include the recalibration of the probability based on the false discovery rate estimation and the use of bonferroni tests to avoid false positives. However, these methods are rarely used in candidate gene association studies. The use of stringent probability thresholds will reduce the danger of false positives, but this must be carefully balanced to minimize the level of rejection of true positives caused by setting the thresholds too high. In a recent study, it was estimated that many genes affecting human height were not detected due to overly stringent significance tests; this may have led to reducing the estimated heritability compared to what would be expected from conventional measurements of the trait (Yang et al., 2010).

Many other methods can be also used to confirm that the identified polymorphisms are indeed significantly associated with the target trait. For example, a *p* value may be improved by adding additional individuals to the same panel or confirmed in independent panels of germplasm thereby increasing the researcher's confidence in the marker–trait association. Or a completely different approach such as linkage analysis and/or expression studies can also be used. It is difficult to define a standard criterion for validation, but multiple lines of evidence are generally considered necessary for any solid conclusion. Conclusive proof of gene function is only achieved through gene cloning and transgenic expression of that gene construct in appropriate genotypes, but for a plant breeder an acceptable level of confidence in the function of a gene marker can be established by validation in a range of target breeding populations and/or near isogenic lines.

The Evolution from Candidate Gene-Based Association Analysis to Genome-Wide Association Study

Candidate gene-based association analysis, as mentioned above, is a hypothesis-driven approach that requires detailed prior knowledge of potential candidate genes. The historical emphasis on this approach in maize resulted from the low detection power of association analyses based on random markers, which was particularly severe due to the especially rapid LD decay in maize (Tenaillon et al., 2001; Gore et al., 2009; Yan et al., 2009). The number of available maize SNP markers is dramatically increasing (http://www.panzea.org/lit/data_sets.html#SNPs [verified 8 Dec. 2010]), so it should now be possible to identify markers covering every chromosomal region in the maize genome, ensuring that there are markers within and closely linked to genes contributing to complex target traits. These so-called GWA studies have been widely used in human and animal systems (Altshuler et al., 2008; Hunter and Crawford, 2008) and are already being used in some plant studies (Chan et al., 2009; Waugh et al., 2009; Atwell et al., 2010). Genome-wide association studies start with the genotypic characterization of a collection of individuals with a sufficient number of polymorphic markers to place one or a few markers in each LD block, which usually requires several hundred thousand SNPs or more. This density of genotyping can be achieved through array-based systems that can simultaneously genotype up to 1 million SNPs (Gupta et al., 2008; Yan et al., 2010a). Alternatively, the massive increase in sequencing capability and the dramatic decrease in unit costs provided by next generation sequencing technologies (Metzker, 2009) may be harnessed for the large-scale genotyping requirements of GWA study.

The minimum number of markers needed for a successful GWA study depends on the genome size and the rate of LD decay of the target germplasm. For example, in the model species *Arabidopsis thaliana* (L.) Heynh., it is estimated that 140,000 markers should provide a good coverage of the 125 Mbp genome (Kim et al., 2007). However, in maize, it is estimated that more than 10 million markers are needed to cover the 2300 Mbp genome due to the combined effects of a much larger genome and a much more rapid LD decay in maize (Myles et al., 2009). However, there is currently no empirical validation of this target number of markers, and only resequencing via next generation sequencing may be able to provide this level of genotyping (Lupski et al., 2010) and only in species for which reference genomes are available. Moreover, the statistical analysis of resultant datasets will present a major challenge to currently available data-handling pipelines. Using elite lines to construct the association mapping panel should significantly reduce the numbers of markers required for GWA studies in maize. In one example, a gene associated with oleic acid content was identified using genome scanning with only 8590 loci in 553 elite maize inbreds (Beló et al., 2008). In this sense, the developed maize SNP50 array (a high density maize SNP array containing >56,000 SNPs from ~19,000 genes) will still be useful for GWA studies for some particular traits (Martin, 2010). However, we knew at least 50 genes affecting the oil content in maize (Laurie et al., 2004), but only one could be detected using 8500 loci, which implies more markers are required to have the highest probability to detect all loci dealing with quantitative traits, especially from diverse sources. An alternative is the use of SNP haplotypes to replace single SNPs in GWA studies, which may significantly reduce the number of markers required while also providing more QTL detection power. The construction of SNP haplotypes for maize breeding is still unclear, but the successful application of GWA studies in humans using haploytpes has already shown clear promise (Schaid, 2004).

By focusing on polymorphisms within the expressed region of all genes, marker density can be decreased significantly without substantially decreasing the QTL detection power. This is because polymorphisms in gene regions should have a higher probability of being functionally important compared to randomly selected polymorphisms. This is supported by most studies on cloned QTL (Alonso-Blanco et al., 2005; Salvi and Tuberosa, 2005). Thus, in the near term, GWA studies may focus on expressed regions using a gene-centric approach (Jorgenson and Witte, 2006; Ng et al., 2010) although this approach might miss some important functional polymorphisms such as *cis*-regulatory regions, which can be up to tens of kilobase pairs away from the target gene (Clark et al., 2006). The maize genome is predicted to have over 32,000 genes, based on analysis of the B73 genomic sequence (Schnable et al., 2009), which are being annotated with an average length of approximately 1.4 kbp in sequencing studies of complementary DNA (cDNA) (Alexandrov et al., 2009; Soderlund et al., 2009). Thus, it will now be possible to identify the vast majority of the polymorphisms within each of these genes and provide a large and powerful set of gene-based markers for GWA studies. Assuming 50,000 genes in the maize genome and 10 to 20 markers developed within the expressed regions of each gene, maize GWA studies will have sufficient QTL detection power if 500,000 to 1,000,000 well chosen markers are used. In this way, the use of markers developed from the expressed portion of the genome would allow a 10- to 20-fold reduction in the necessary marker density compared to the predicted requirement of 10 to 15 million random markers (Myles et al., 2009).

Hundreds of GWA studies have been performed in human and animal systems that have identified thousands of genes or SNP markers, most of which are associated with small effects on target traits (Altshuler et al., 2008; Hunter and Crawford, 2008; http://www.genome.gov/gwastudies [verified 6 Dec. 2010]). In addition, the aggregate heritability of genes identified by GWA studies and associated with a given trait is still quite low. In human and animal studies, GWA studies may be biased by two major interconnected factors: a dominance of SNPs with a lower minor allelic frequency (MAF) (<5%) and small population size (Manolio et al., 2009). In these cases, rare alleles (including those responsible for large effects on the target trait) are effectively hidden in the surveyed population due to the lack of statistical power to assign an

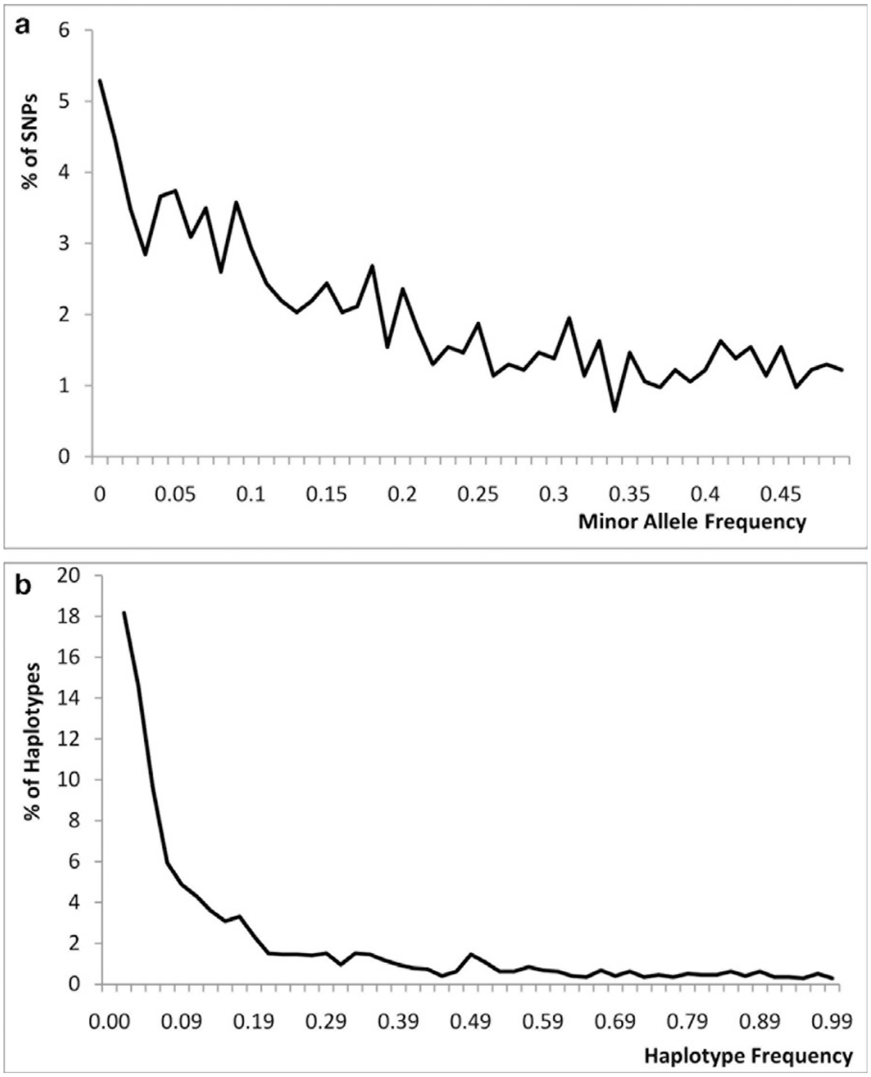

Figure 3. (a) Single nucleotide polymorphism (SNP) frequency based on screening 632 diverse maize lines with 1229 SNPs. An aggregate total of 36% of the SNPs have a minor allele frequency (MAF) of 0.1 or less (data from Yan et al., 2009). (b) Haplotype frequency based on 538 loci in 632 diverse maize lines. An aggregate total of 55% of the haplotypes have a minor allelic frequency (MAF) of 0.1 or less (data from Yan et al., 2009).

association when there are so few representatives carrying the rare alleles in that population (Manolio et al., 2009). In a recent study, even known causal SNPs have not been identified in a GWA study using an array designed specifically for them (Lusis and Pajukanta, 2008).

It is likely that these same problems will also be major limitations for GWA studies in plants (Myles et al., 2009). In one study, 632 diverse maize lines were genotyped with 1536 SNPs developed from 582 candidate genes (Yan et al., 2009). More than 36% of individual SNPs (Fig. 3a) and more than 55% of single gene-based haplotypes (Fig. 3b) had a MAF of less than 10%. In the study involving *crtRB1*, the most favorable allele was very rare (MAF < 5%) in temperate germplasm and totally absent in tropical and subtropical germplasm. This allele would not have been detected in a GWA study, and validation required the development of several linkage mapping populations segregating at this locus (Yan et al., 2010b). Thus, applying association analysis to targeted segregating biparental populations is currently the major means of identifying and validating these rare but important alleles (Manenti et al., 2009). Considering the lines commonly used for breeding may only contain a few haploytpes, association mapping should have the highest power to estimate the contribution of these haplotypes to the trait of interest.

Potential Solutions for the Major Constraints to Association Studies

One big advantage that genetic studies of plants have traditionally enjoyed is that populations of different genetic structure can be created to serve specific purposes. A powerful example is the NAM population consisting of 25 recombinant inbred line (RIL) populations created by crossing a diverse range of 25 important temperate and tropical breeding lines with one common, well characterized parent (B73) (Yu et al., 2008; Buckler et al., 2009; McMullen et al.,

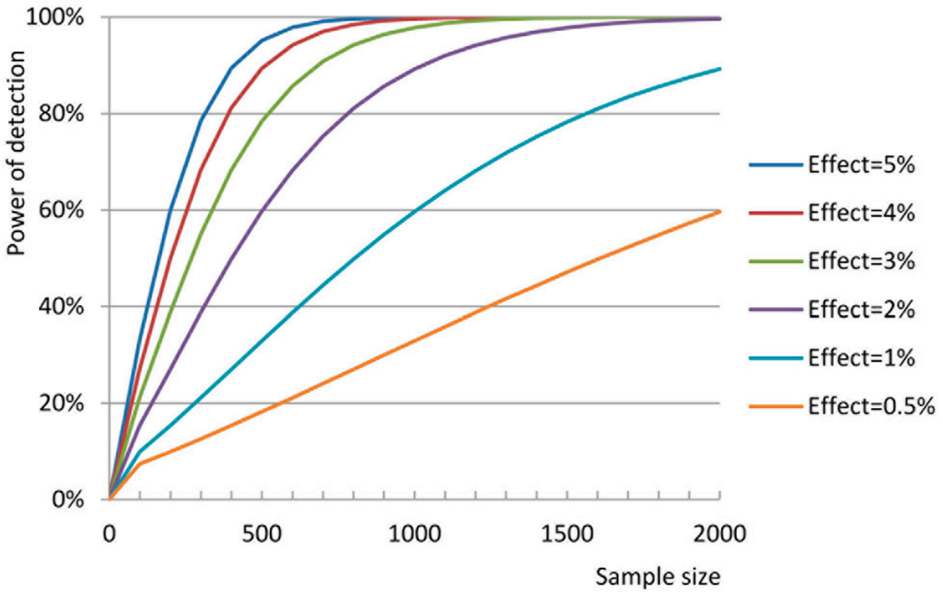

Figure 4. Effect on quantitative trait loci (QTL) detection power (proportion of real QTL detected) of increasing population size for QTL contributing 0.5 to 5.0% of the total phenotypic variation of the target trait. The simulation was performed using Genetic Power Calculator (GPC; Purcell et al., 2003) assuming the linkage disequilibrium (LD) of markers was equal to R^2 = 0.8, minor allelic frequency (MAF) = 0.1, and sibling correlations = 0.1.

2009). Another useful approach is the multiparent advanced generation intercross (MAGIC) population, originally proposed in animals (Mott et al., 2000) and now used in plants as well (Kover et al., 2009; Chintamanani et al., 2010). The NAM and MAGIC populations provide an ideal resource for gene identification and validation in maize including the identification of numerous small-effect QTL contributing to a target agronomic trait (Buckler et al., 2009; Kover et al., 2009). The NAM and MAGIC approaches may also boost the low detection power of traditional association mapping to detect rare alleles (Visscher, 2008). This may be critically important since genes having common variants with modest effects on complex traits may also have rare variants with large effects, which may be the preferred targets for plant breeders (Manolio et al., 2009).

Another challenge for GWA studies in maize is the large number of loci with small effects that contribute to most quantitative traits, including flowering time (Buckler et al., 2009), oil content (Laurie et al., 2004), and drought tolerance (Messmer et al., 2009). Although this phenomenon is commonly reported in animal species, in some plant species such as rice, sorghum, wheat and *Arabidopsis* spp. (Mackay, 2009), a smaller number of genes, each with very large effect on a quantitative trait, has been reported for flowering time and grain quality traits. The presence of large-effect QTL may be due to the inbreeding nature of some of these plant species. Animals and outcrossing plants such as maize may conform more to the infinitesimal model of quantitative inheritance (Buckler et al., 2009; Atwell et al., 2010), in which there are an infinite (or at least, very large) number of genes, each contributing a very small amount to a quantitative trait. This is a critically important issue as it may infer that maize researchers should take their lead from methodological advances in human and animal genetics rather than those in the *Arabidopsis* and rice communities, although direct comparison with specific reports is often not possible since many human genetics studies are based on case-control studies.

The power of association studies is determined by the size of the experimental population, the magnitude of the target allele effect, the density of markers used, and the rate of LD decay between marker and target allele as well as errors in phenotyping and genotyping data and the desired resultant statistical significance level (Gordon and Finch, 2005). Increasing the number of individuals phenotyped has a much more substantial effect on the power of QTL detection (especially for small-effect loci) than increasing the density of genotyping (Long and Langley, 1999). As shown in Fig. 4, using a population with 500 individuals provides an 80% probability of detecting a gene that explains 3% or more of the phenotypic variation, while 1500 individuals are needed to achieve the same probability of detection for a gene that only explains 1% of the variation of the target trait. Because only a few of the ~50 QTL identified by NAM for flowering time in maize explained more than 3% of the variation in this trait (Buckler et al., 2009), populations of 500 genotypes clearly provide insufficient QTL detection power to be of much value in GWA studies of agronomic traits in maize. Similar observations have also been made in quantitative trait studies of humans (Visscher, 2008). Much larger population sizes are therefore going to be needed for detection of most QTL in maize, although smaller population sizes may be adequate for detection of alleles of large effect in self-pollinated plant species (Rostoks et al., 2006; Atwell et al., 2010). However, as has become standard practice in human genomics research (Purcell et al., 2003), plant genomics researchers should confirm a high power of detection of their experimental design before initiating a study, preferably in the region of 99% (Gordon and Finch, 2005).

Precise phenotyping is another key constraint for any marker–trait association analysis (Montes et al., 2007; Myles et al., 2009; Rafalski, 2010). In animal systems it is very difficult to obtain replicated phenotypic measurements for each genotype, but in plants it is relative easy to generate pure breeding or homogenous lines for phenotyping in replicated trials across multiple environments and seasons. This approach has become very popular in plants as it significantly increases the heritability of the resulting data. However, the cost and time required for such phenotyping, especially as recommended population sizes have increased, has become a rate limiting factor for the overall process. This has led to heavy investment into the development of high throughput and highly precise phenotyping techniques (Finkel, 2009; Fernie and Schauer, 2009). For many important traits, efficient and cost effective phenotyping methods are still lacking. For example, it takes 1 h per sample to measure maize kernel carotenoid content using the high performance liquid chromatography (HPLC), at a cost of over US$50 (Yan et al., 2010b), and this is still the routine procedure for most provitamin A breeding programs. Other faster and/or cheaper methods are usually not sufficiently precise to measure the trait in a way that provides breeders with acceptable levels of genetic gain. This is clearly a compelling candidate for marker-assisted selection.

Although plant breeders have developed many rapid methods of scoring important agronomic traits including yield, these are often not sufficiently quantitative for fine mapping. Thus, traits of low heritability that are difficult, time consuming, or expensive to accurately measure may be disaggregated into component traits, which are usually more highly heritable and easier to measure. Alternatively justification for measuring a more expensive or time consuming component trait may be made to improve precision or heritability of the overall mapping process. This is perhaps most extreme with metabolomic or biochemical phenotypes that are very expensive and difficult to screen but often present the most promising surrogates for accurately measuring low heritability traits such as yield. Marker–trait association studies will always benefit from a more precise measurement of the phenotype (Fernie and Schauer, 2009).

Clearly, experimental design must always be carefully optimized to maximize the power of the analysis. Nevertheless, there always remains a chance that statistically significant associations are due to chance (Gordon and Finch, 2005) and it is highly difficult to distinguish true associations from spurious associations (Atwell et al., 2010). For this reason, validation of candidate markers in independent populations remains an essential element in the process despite its time consuming and expensive nature. However, once markers have been identified that have been shown to be tightly and robustly linked to the target trait, they provide several magnitudes of return on investment through increased speed and cost efficiency of breeding programs.

How Can Association Mapping Help Crop Improvement?

Genes found to have significant associations with target traits can be resequenced in a diverse panel of germplasm to identify causal mutations and the most favorable alleles for trait improvement and to develop simple PCR-based markers for MAS (Harjes et al., 2008; Yan et al., 2010b). Gene-based markers are more accurate than linked markers for the prediction of phenotype, since the marker–trait association will not be lost during segregation in the course of recurrent breeding selection cycles. Results from association analysis can be used to predict the best haplotype across one or multiple genes for optimum expression of the target trait. Using the *crtRB1* gene as an example, six common haplotypes were identified that conferred different levels of the target trait (Yan et al., 2010b), with an eightfold difference in the phenotype between the common "best" and "worst" haplotype. In theory, the optimum haplotype can be reconstructed from any cross of two parents containing different components of the desired haplotype. However, in practice, different donors vary in their background effects, in terms of the effects of alleles at other loci that directly or indirectly influence the target trait. This significantly influences the value of parental genotypes contributing the most favorable component haplotypes with respect to overall breeding efficiency. Fortunately, association analysis can help to determine which one is the best donor, something that linkage analysis cannot. For example, in the case of the *crtRB1* gene (Yan et al., 2010b), two segregating populations were used for QTL mapping of the target trait. Major QTL were mapped that explained similar percentages of phenotypic variation for the target trait in both populations, but it was not known if the parent containing the favorable alleles in each of the two populations would have had the same effect on the target traits in a different genetic background. However, information from association analysis allowed us to determine that one of the parents was the best donor of the optimum component haplotype for future breeding programs. Clearly this is dependent on the determination of multiple allele effects in association mapping analysis that possess the chosen component haplotype in different backgrounds, enabling an estimation of context-dependent allele effect.

In the past, plant breeders have tended to focus on incremental improvement of a few key agronomic traits.

However, many breeders have dreamt of moving to an ideotype breeding approach, in which a theoretical ideal profile of characters is defined and breeding strategies are designed to reach that goal. Generally, this strategy is not possible through conventional phenotypic selection, although partial success has been seen, such as development of a new plant type in rice (Khush, 2005). Many of the genes controlling a wide range of desirable agronomic traits have now been mapped in diverse sources of germplasm. If information on plant physiology, pathology, entomology, and biochemistry is available to design the optimal plant ideotype, then theoretically molecular breeding offers a mechanism for pyramiding these genes into a single breeding line to create the ideal new cultivar (Peleman and van der Voort, 2003). This approach is already being pursued to breed higher yielding and more stress-tolerant rice (Takeda and Matsuoka, 2008; Xing and Zhang, 2010). However, the pyramiding of multiple genes with small effects from diverse genetic backgrounds into a single line still faces substantial practical challenges. For example, to find one individual in a segregating population that carries all the target beneficial alleles is increasingly difficult as more and more traits are considered simultaneously. Even with disaggregated breeding schemes this requires thousands of progeny to be generated and screened. In addition, the unpredictable implications of epistasis and genotype × environment effects further complicate the process as one attempts to consider more traits simultaneously.

An ideotype breeding scheme would need to be accomplished over several generations, but the complexity and duration of the breeding program will remain high. As with phenotypic based breeding, ideotype breeding schemes for traits controlled by a relatively small number of major genes remain more attractive and cost effective, but for many of the most agronomically important traits in maize such as yield and drought tolerance, the huge number of genes involved means that a design-led approach is still out of reach (Buckler et al., 2009). Nevertheless, marker-assisted recurrent selection (MARS) and genome-wide selection (GWS) may provide a compromise approach for effective molecular breeding of these complex traits. In MARS the best lines in a segregating population are identified based on phenotypic evaluation and then an index of marker profiles associated with those lines is constructed. The markers are chosen based on linkage with traits of interest in the segregating population, and the index is used in (typically three) subsequent cycles of marker-based selection (Johnson, 2004). It has been proposed that only a small number of markers (20–40) associated with the target trait(s) would be needed (Koebner, 2003). This approach is currently being tested in over 20 breeding populations in the Drought Tolerant Maize for Africa project (DTMA; http://dtma.cimmyt.org [verified 6 Dec. 2010]).

While MARS is essentially a linkage mapping-based approach applied to biparental breeding populations, GWS is an LD analysis-based approach that relies on estimating the effect of each marker without testing the significance of its association with the target phenotype. Because GWS uses amalgamated populations from multiple parents for the simultaneous estimation of the phenotypic effect at many markers, it requires a much higher density of markers, especially when using a collection of germplasm with a rapid LD decay. The phenotypic effect of every marker is estimated using an appropriate model applied to phenotype data from one or more cycles of evaluating diverse germplasm. Then, in each cycle of GWS, all markers will be used to estimate the breeding value of each line, which will then determine the best lines to be selected for the next cycle (Meuwissen et al., 2001; Bernardo and Yu, 2007). Generally, GWS and MARS use random markers that are linked to the target gene rather than within the actual gene. However, it is inevitable that maize molecular breeders will want to use the most informative markers from within genes for MARS and GWS, when they are available, to reduce the total number of markers required to achieve a high level of selective power. Similarly, it is expected that GWS and GWA studies will be combined in future maize improvement strategies, using GWA studies with many markers to identify and empirically validate a subset of significant markers for GWS. Frisch et al. (2010) have used transcriptional data from a 46,000 oligonucleotide array to develop a prediction model for the value of parental maize lines in relation to the grain yield performance of their hybrid progeny. This study found that predictions based on 50 well chosen genes were as accurate as predictions based on 5000 random genes. Similarly, when comparing the predictive value of random markers versus selected markers based on association analysis in a panel of related breeding lines, we have found that 250 of the best markers provided the same level of prediction as 1500 random markers (W. Wen et al., unpublished data, 2010).

Prospects and Priorities for Future Applications of Association Mapping in Maize Improvement

As discussed previously, in outcrossing species, very large populations are needed for gene discovery of quantitative traits for which each gene contributes very little to the phenotypic variation. In plant studies, particularly for traits that must be evaluated in replicated multilocational field trials, it has not been possible for many public programs to work with populations beyond about 1000 individuals. Because the NAM population contains 5000 individuals across 25 interrelated populations, it has required the collaboration of many research groups to handle the necessary phenotyping and data processing work. This type of study represents a shift to "big science" in crop genomics research that was first seen in human genomics research in the 1990s and that fueled a paradigm shift in the type of challenges addressed by public research and in turn the pharmaceutical industry (Psaty et al., 2007). Genome-wide association studies may provide the possibility for many small public research projects to participate in "big science" research, by enabling them to combine independent populations and studies. This type of aggregation of data sets into one GWA study has been successfully applied in wheat, a self-pollinated species (Crossa et al., 2007), but it has not yet been validated in maize. However, since many independent maize association mapping panels of different sizes and with different genetic backgrounds have been developed and phenotyped for the same or similar traits (see Table 1), it would be an excellent use of existing resources to combine these into a single GWA study that should have a substantially higher power to detect QTL of small effect.

Meanwhile, the potential throughput of genotyping systems is rapidly increasing and unit costs are consistently falling. This trend seems set to continue with the development of next generation sequencing techniques. Large-scale GWA studies of thousands of individuals using common and high-density markers is already a reality and screening of tens of thousands of individuals through resequencing will soon also become a realistic option in several crop species including maize. These approaches will greatly help to increase our understanding of the genetic architecture of maize complex traits and to begin the long journey toward true design-led molecular breeding. Meanwhile, all newly identified trait-targeted markers have the potential to improve the efficiency of MAS, MARS, and GWS.

Unfortunately the facilitating technology for modern molecular breeding is still highly expensive and the operational reagents and services are still highly inaccessible for many developing countries (Tester and Langridge, 2010). Fortunately, some organizations are trying to fill this gap, for example, the Generation Challenge Programme (GCP) is establishing a Molecular Breeding Platform (MBP) (http://wiki.cimmyt.org/confluence/display/MBP/Home [verified 8 Dec. 2010]) for breeders across the world especially those in developing countries to facilitate their access to the best, most cost-effective molecular marker technologies. With the rapid development of next generation sequencing technologies it will soon be possible to genotype very large collections of germplasm by sequencing (so-called "genotyping by sequencing") (Huang et al., 2009). Access to genotyping by sequencing services will allow maize breeders in developing countries to bridge this technology transition much more readily than previous transitions.

Many research institutes and university departments in high-income countries and emerging economies (China, India, Mexico, Brazil, South Africa, etc.) have large, well-funded agricultural research programs that are easily able to participate in large-scale gene identification projects in maize. However, programs in most low-income countries (including the Consultative Group on International Agricultural Research [CGIAR] centers) may be too small to establish their own gene identification programs, but they can apply the knowledge gained from studies in advanced research organizations to their marker-based breeding programs. Sadly there is still an information gap for many developing country scientists who are unable to access subscription-based journals or gain travel funds to attend international conferences. Collaborative programs such as those coordinated by the International Agricultural Research Institutes (IARIs) of CGIAR are working to help developing country partners to access up-to-date information and tailored technical support throughout all stages of establishing and implementing molecular breeding programs.

It is clear that the initiatives described above are just a small beginning and we echo the appeal for a long-term strategic plan for global coordination of maize research that has recently been elaborated for rice (Zhang et al., 2008). This would include collating all the existing phenotypic data in a single publicly accessible database (such as the Maize Genetics and Genomics Database [http://www.maizegdb.org {verified 6 Dec. 2010}] and Gramene [http://www.gramene.org {verified 6 Dec. 2010}]) and coordinating future phentoyping efforts on a global level as well as developing cooperative genotyping programs to increase efficiency and maximize opportunities for integration. All results should be made publicly available in the shortest possible time, and training of potential users, particularly those from developing countries, should be performed within ongoing molecular breeding programs as an essential component of all projects.

CONCLUSIONS

Association mapping offers great potential to enhance maize genetic improvement. This will certainly be strengthened by the use of high throughput and cost effective next generation sequencing techniques that will enable GWA studies to become a popular and routine approach in maize. However, association mapping remains complementary to rather than a replacement for linkage mapping and other gene identification and validation techniques. Moreover, the contrast between the large number of variants with small effects identified by GWA studies versus the small number of genomic regions with large effects identified by linkage mapping remains a challenge to our current understanding of the genetic architecture of complex traits. Although, for practical applications, the integration of linkage mapping and association mapping approaches offers substantial opportunity to resolve the individual constraints of each approach while synergizing their respective strengths. Nevertheless, population structure remains a big limitation for association studies that requires careful choice of germplasm and the development of advanced statistical approaches. In addition, as the size of populations and the density of marker screening rapidly increase, so does the probability of detecting nonlinked (false) associations. These issues reinforce the need to independently validate candidate genes and/or markers in diverse genetic backgrounds (independent populations) to eliminate false positives. Inevitably, this brings us back to the need for large-scale cost-effective precision phenotyping, which remains a major logistical challenge and bottleneck to the development of molecular breeding programs. Nevertheless, significant progress is being made in facilitating technologies for such phenotyping. Finally, there is undoubtedly an urgent need to bridge the gap between genomics researchers and molecular breeders in developed and developing countries, and particularly to share new knowledge faster and to enable genetic improvement gains (especially in Africa) to catch up with those in the leading producer countries. Emerging economies such as China and India, Brazil and Mexico, and South Africa have a major role in bridging this technological divide in maize breeding. If successful, millions of dollars of genomics research investment may finally benefit the poorest people in the world.

Acknowledgments

Helpful comments were provided by Drs. Peter Balint-Kurti (USDA-ARS, Raleigh, North Carolina), Rowena Y. Kelley (Mississippi State University), and Ian Mackay (NIAB) during the development of this paper. We also thank Ms. Qing Li for preparation of the Table 1. The development of this paper was supported in part by the Drought-Tolerant Maize for Africa project, funded by the Bill and Melinda Gates Foundation and the National Hi-Tech Research and Development Program of China.

References

Alexandrov, N.N., V.V. Brover, S. Freidin, M.E. Troukhan, T.V. Tatarinova, H. Zhang, T.J. Swaller, Y.P. Lu, J. Bouck, R.B. Flavell, and K.A. Feldmann. 2009. Insights into corn genes derived from large-scale cDNA sequencing. Plant Mol. Biol. 69:179–194.

Alonso-Blanco, C., B. Mendez-Vigo, and M. Koornneef. 2005. From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. Int. J. Dev. Biol. 49:717–732.

Altshuler, D., M.J. Daly, and E.S. Lander. 2008. Genetic mapping in human disease. Science 322(5903):881–888.

Andersen, J.R., T. Schrag, A.E. Melchinger, I. Zein, and T. Lubberstedt. 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, J.R., I. Zein, G. Wenzel, B. Krützfeldt, J. Eder, M. Ouzunova, and T. Lübberstedt. 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, J.R., I. Zein, G. Wenzel, B. Darnhofer, J. Eder, M. Ouzunova, and T. Lübberstedt. 2008. Characterization of phenylpropanoid pathway genes within European maize (*Zea mays* L.) inbreds. BMC Plant Biol. 8:2.

Atwell, S., Y.S. Huang, B.J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, D. Meng, A. Platt, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Muliyati, X. Zhang, M.A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. de Meaux, J.R. Ecker, N. Faure, J.M. Kniskern, J.D. Jones, T. Michael, A. Nemri, F. Roux, D.E. Salt, C. Tang, M. Todesco, M.B. Traw, D. Weigel, P. Marjoram, J.O. Borevitz, J. Bergelson, and M. Nordborg. 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature doi:10.1038/nature08800.

Beló, A., P. Zheng, S. Luck, B. Shen, D.J. Meyer, B. Li, S. Tingey, and A. Rafalski. 2008. Whole genome scan detects an allelic variant with increased oleic acid levels in maize. Mol. Genet. Genomics 279:1–10.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082–1090.

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J.C. Glaubitz, M.M. Goodman, C. Harjes, K. Guill, D.E. Kroon, S. Larsson, N.K. Lepak, H. Li, S.E. Mitchell, G. Pressoir, J.A. Peiffer, M.O. Rosas, T.R. Rocheford, M.C. Romay, S. Romero, S. Salvo, H.S. Villeda, H.S. da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M.D. McMullen. 2009. The genetic architecture of maize flowering time. Science 325:714–718.

Buckler, E.S., and N.M. Stevens. 2005 Maize origins, domestication, and selection. p. 67–90. *In* T.J. Motley, N. Zerega, and H. Cross (ed.) Darwin's harvest. Columbia Univ. Press, New York, NY.

Butrón, A., Y.C. Chen, G.E. Rottinghaus, and M.D. McMullen. 2010. Genetic variation at *bx1* controls DIMBOA content in maize. Theor. Appl. Genet. 120:721–734.

Camus-Kulandaivelu, L., J.B. Veyrieras, D. Madur, V. Combes, M. Fourmann, S. Barraud, P. Dubreuil, B. Gouesnard, D. Manicacci, and A. Charcosset. 2006. Maize adaptation to temperate climate: Relationship between population structure and polymorphism in the *Dwarf8* gene. Genetics 172:2449–2463.

Chan, E.K., H.C. Rowe, and D.J. Kliebenstein. 2010. Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. Genetics 85:991–1000.

Ching, A., K.S. Caldwell, M. Jung, M. Dolan, O.S. Smith, S. Tingey, M. Morgante, and A.J. Rafalski. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 3:19.

Chintamanani, S., S.H. Hulbert, G.S. Johal, and P.J. Balint-Kurti. 2010. Identification of a maize locus that modulates the hypersensitive defense response, using mutant-assisted gene identification and characterization. Genetics 184:813–825.

Clark, R.M., T.N. Wagler, P. Quijada, and J. Doebley. 2006. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. Nat. Genet. 38:594–597.

Cockram, J., H. Jones, F.J. Leigh, D. O'Sullivan, W. Powell, D.A. Laurie, and A.J. Greenland. 2007. Control of flowering time in temperate cereals: Genes, domestication, and sustainable productivity. J. Exp. Bot. 58:1231–1244.

Collier, P., G. Conway, and T. Venables. 2008. Climate change and Africa. Oxf. Rev. Econ. Policy 24:337–353.

Crossa, J., J. Burgueño, S. Dreisigacker, M. Vargas, S.A. Herrera-Foessel, M. Lillemo, R.P. Singh, R. Trethowan, M. Warburton, J. Franco, M. Reynolds, J.H. Crouch, and R. Ortiz. 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889–1913.

Devlin, B., and K. Roeder. 1999. Genomic control for association studies. Biometrics 55:997–1004.

Dooner, H.K., and L. He. 2008. Maize genome structure variation: Interplay between retrotransposon polymorphisms and genic recombination. Plant Cell 20:249–258.

Ducrocq, S., D. Madur, J.B. Veyrieras, L. Camus-Kulandaivelu, M. Kloiber-Maitz, T. Presterl, M. Ouzunova, D. Manicacci, and A. Charcosset. 2008. Key impact of *Vgt*1 on flowering time adaptation in maize: Evidence from association mapping and ecogeographical information. Genetics 178:2433–2437.

Ejeta, G. 2010. African Green Revolution needn't be a mirage. Science 327:831–832.

Ersoz, E.S., J. Yu, and E.S. Buckler, IV. 2009. Applications of linkage disequilibrium and association mapping in maize. p. 173–195. *In* A.L. Kriz and B.A. Larkins (ed.) Molecular genetic approaches to maize improvement. Springer-Verlag, New York, NY.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Fan, C.C., Y.Z. Xing, H.L. Mao, T.T. Lu, B. Han, C.G. Xu, X.H. Li, and Q.F. Zhang. 2006. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor. Appl. Genet. 112:1164–1171.

Fan, C.C., S.B. Yu, C.R. Wang, and Y.Z. Xing. 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.

FAO. 2009a.Food security statistics. Available at http://www.fao.org/economic/ess/food-security-statistics/en (verified 6 Dec. 2010). Food and Agriculture Organization of the United Nations, Rome.

FAO. 2009b. Global agriculture towards 2050. Briefing paper for FAO high-level expert forum on "How to feed the world 2050," Rome. 21–13 Oct. 2009. Available at http://www.fao.org/wsfs/world-summit/en (verified 6 Dec. 2010). Food and Agriculture Organization of the United Nations, Rome.

Fernie, A.R., and N. Schauer. 2009. Metabolomics-assisted breeding: A viable option for crop improvement? Trends Genet. 25:39–48.

Finkel, E. 2009. Imaging. With 'phenomics,' plant scientists hope to shift breeding into overdrive. Science 325:380–381.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357–374.

Flint-Garcia, S.A., A.C. Thuillet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant J. 44:1054–1064.

Frisch, M., A. Thiemann, J. Fu, T.A. Schrag, S. Scholten, and A.E. Melchinger. 2010. Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor. Appl. Genet. 120:441–450.

Fu, Z., J. Yan, Y. Zheng, M.L. Warburton, J.H. Crouch, and J.S. Li. 2010. Nucleotide diversity and molecular evolution of the *PSY1* gene in Zea mays compared to some other grass species. Theor. Appl. Genet. 120:709–720.

Gaut, B.S., and A.D. Long. 2003. The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506.

Goff, S.A., D. Ricke, T.H. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma, D. Hadley, D. Hutchison, C. Martin, F. Katagiri, B.M. Lange, T. Moughamer, Y. Xia, P. Budworth, J. Zhong, T. Miguel, U. Paszkowski, S. Zhang, M. Colbert, W.L. Sun, L. Chen, B. Cooper, S. Park, T.C. Wood, L. Mao, P. Quail, R. Wing, R. Dean, Y. Yu, A. Zharkikh, R. Shen, S. Sahasrabudhe, A. Thomas, R. Cannings, A. Gutin, D. Pruss, J. Reid, S. Tavtigian, J. Mitchell, G. Eldredge, T. Scholl, R.M. Miller, S. Bhatnagar, N. Adey, T. Rubano, N. Tusneem, R. Robinson, J. Feldhaus, T. Macalma, A. Oliphant, and S. Briggs. 2002. Draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100.

Gordon, D., and S.J. Finch. 2005. Factors affecting statistical power in the detection of genetic association. J. Clin. Invest. 115:1408–1418.

Gore, M.A., J.M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, B.L. Hurwitz, J.A. Peiffer, M.D. McMullen, G.S. Grills, J. Ross-Ibarra, D.H. Ware, and E.S. Buckler. 2009. A first-generation haplotype map of maize. Science 326:1115–1117.

Guillet-Claude, C., C. Birolleau-Touchard, D. Manicacci, M. Fourmann, S. Barraud, V. Carret, J.P. Martinant, and Y. Barrière. 2004a. Genetic diversity associated with variation in silage corn digestibility for three *O*-methyltransferase genes involved in lignin biosynthesis. Theor. Appl. Genet. 110:126–135.

Guillet-Claude, C., C. Birolleau-Touchard, D. Manicacci, P.M. Rogowsky, J. Rigau, A. Murigneux, J.P. Martinant, and Y. Barrière. 2004b. Nucleotide diversity of the *ZmPox3* maize peroxidase gene: Relationships between a MITE insertion in exon 2 and variation in forage maize digestibility. BMC Genet. 5:19.

Gupta, P.K., S. Rustgi, and R.R. Mir. 2008. Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18.

Harjes, C.E., T.R. Rocheford, L. Bai, T.P. Brutnell, C.B. Kandianis, S.G. Sowinski, A.E. Stapleton, R. Vallabhaneni, M. Williams, E.T. Wurtzel, J. Yan, and E.S. Buckler. 2008. Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. Science 319:330–333.

Hedrick, P.W. 1987. Gametic disequilibrium measures: Proceed with caution. Genetics 117:331–341.

Huang, X., Q. Feng, Q. Qian, Q. Zhao, L. Wang, A. Wang, J. Guan, D. Fan, Q. Weng, T. Huang, G. Dong, T. Sang, and B. Han. 2009. High-throughput genotyping by whole-genome resequencing. Genome Res. 19:1068–1076.

Hunter, K.W., and N.P. Crawford. 2008. The future of mouse QTL mapping to diagnose disease in mice in the age of whole-genome association studies. Annu. Rev. Genet. 42:131–141.

Johnson, R. 2004. Marker-assisted selection. Plant Breed. Rev. 24:293–309.

Jorgenson, E., and J.S. Witte. 2006. A gene-centric approach to genome-wide association studies. Nat. Rev. Genet. 7:885–891.

Jung, M., A. Ching, D. Bhattramakki, M. Dolan, S. Tingey, M. Morgante, and A. Rafalski. 2004. Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. Theor. Appl. Genet. 109:681–689.

Khush, G.S. 2001. Green revolution: The way forward. Nat. Rev. Genet. 2:815–822.

Khush, G. 2005. What it will take to feed 5.0 billion rice consumers in 2030. Plant Mol. Biol. 59:1–6.

Kim, S., V. Plagnol, T.T. Hu, C. Toomajian, R.M. Clark, S. Ossowski, J.R. Ecker, D. Weigel, and M. Nordborg. 2007. Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. 39:1151–1155.

Koebner, R. 2003. MAS in cereals: Green for maize, amber for rice, still red for wheat and barley. *In* Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding? Turin, Italy. 17–18 Oct. 2003. Available at http://www.fao.org/biotech/docs/Koebner.pdf (verified 6 Dec. 2010). Food and Agriculture Organization of the United Nations, Rome.

Kojima, S., Y. Takahashi, Y. Kobayashi, L. Monna, T. Sasaki, T. Araki, and M. Yano. 2002. *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol. 43:1096–1105.

Kover, P.X., W. Valdar, J. Trakalo, N. Scarcelli, I.M. Ehrenreich, M.D. Purugganan, C. Durrant, and R.A. Mott. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. 5:e1000551.

Krill, A.M., M. Kirst, L.V. Kochian, E.S. Buckler, and O.A. Hoekenga. 2010. Association and linkage analysis of aluminum tolerance genes in maize. PLoS One 5:e9958.

Laurie, C.C., S.D. Chasalow, J.R. LeDeaux, R. McCarroll, D. Bush, B. Hauge, C. Lai, D. Clark, T.R. Rocheford, and J.W. Dudley. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141–2155.

Li, J., and T. Jiang. 2005. Haplotype-based linkage disequilibrium mapping via direct data mining. Bioinformatics 21:4384–4393.

Li, Q., L. Li, X. Yang, G. Bai, M.L. Warburton, J. Dai, J. Li, and J. Yan. 2010b. Function, relationship, and evolutionary fate of two maize genes orthologous to rice GW2 associated with kernel size and weight. BMC Plant Biol. 10:143.

Li, Q., X. Yang, G. Bai, M.L. Warburton, G. Mahuku, M. Gore, J. Dai, J. Li, and J. Yan. 2010a. Cloning and characterization of a putative *GS3* ortholog involved in maize kernel development. Theor. Appl. Genet. 120:753–763.

Long, A.D., and C.H. Langley. 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res. 9:720–731.

Lübberstedt, T., I. Zein, J.R. Andersen, G. Wenzel, B. Krützfeldt, J. Eder, M. Ouzunova, and S. Chun. 2005. Development and application of functional markers in maize. Euphytica 146:101–108.

Lupski, J.R., J.G. Reid, C. Gonzaga-Jauregui, D. Rio Deiros, D.C. Chen, L. Nazareth, M. Bainbridge, H. Dinh, C. Jing, D.A. Wheeler, A.L. McGuire, F. Zhang, P. Stankiewicz, J.J. Halperin, C. Yang, C. Gehman, D. Guo, R.K. Irikat, W. Tom, N.J. Fantin, D.M. Muzny, and R.A. Gibbs. 2010. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N. Engl. J. Med. 362:1181–1191.

Lusis, A.J., and P. Pajukanta. 2008. A treasure trove for lipoprotein biology. Nat. Genet. 40:129–130.

Mackay, I., and W. Powell. 2007. Methods for linkage disequilibrium mapping in crops. Trends Plant Sci. 12:57–63.

Mackay, T.F. 2009. A-maize-ing diversity. Science 325:688–689.

Manenti, G., A. Galvan, A. Pettinicchio, G. Trincucci, E. Spada, A. Zolin, S. Milani, A. Gonzalez-Neira, and T.A. Dragani. 2009. Mouse genome-wide association mapping needs linkage analysis to avoid false-positive loci. PLoS Genet. 5(1):e1000331.

Mangelsdorf, P.C., and G.S. Fraps. 1931. A direct quantitative relationship between vitamin A in corn and the number of genes for yellow pigmentation. Science 73:241–242.

Manicacci, D., L. Camus-Kulandaivelu, M. Fourmann, C. Arar, S. Barrault, A. Rousselet, N. Feminias, L. Consoli, L. Francès, V. Méchin, A. Murigneux, J.L. Prioul, A. Charcosset, and C. Damerval. 2009. Epistatic interactions between *opaque2* transcriptional activator and its target gene *CyPPDK1* control kernel trait variation in maize. Plant Physiol. 150:506–520.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F. Mackay, S.A. McCarroll, and P.M. Visscher. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753.

Martin, G. 2010. The maize infinium 60K array – Design, first data and potential applications. *In* Abstr. Plant Animal Genome Conf., XVIII, San Diego, CA. 9–13 Jan. 2010. Available at http://www.intl-pag.org/18/18i-illumina.html (verified 8 Dec. 2010).

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, P. Brown, C. Browne, M. Eller, K. Guill, C. Harjes, D. Kroon, N. Lepak, S.E. Mitchell, B. Peterson, G. Pressoir, S. Romero, M. Oropeza Rosas, S. Salvo, H. Yates, M. Hanson, E. Jones, S. Smith, J.C. Glaubitz, M. Goodman, D. Ware, J.B. Holland, and E.S. Buckler. 2009. Genetic properties of the maize nested association mapping population. Science 325:737–740.

Messmer, R., Y. Fracheboud, M. Bänziger, M. Vargas, P. Stamp, and J.M. Ribaut. 2009. Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. Theor. Appl. Genet. 119:913–930.

Metzker, M.L. 2009. Sequencing technologies – The next generation. Nat. Rev. Genet. 11:31–46.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Montes, J.M., A.E. Melchinger, and J.C. Reif. 2007. Novel throughput phenotyping platforms in plant genetic studies. Trends Plant Sci. 12:433–436.

Moose, S.P., and R.H. Mumm. 2008. Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol. 147:969–977.

Mott, R., C.J. Talbot, M.G. Turri, A.C. Collins, and J. Flint. 2000. A method for fine mapping quantitative trait loci in outbred animal stocks. Proc. Natl. Acad. Sci. USA 97:12649–12654.

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, and E.S. Buckler. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202.

Ng, S.B., K.J. Buckingham, C. Lee, A.W. Bigham, H.K. Tabor, K.M. Dent, C.D. Huff, P.T. Shannon, E.W. Jabs, D.A. Nickerson, J. Shendure, and M.J. Bamshad. 2010. Exome sequencing identifies the cause of a Mendelian disorder. Nat. Genet. 42:30–35.

Palaisa, K.A., M. Morgante, M. Williams, and A. Rafalski. 2003. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795–1806.

Paterson, A.H., Y.R. Lin, Z.K. Li, K.F. Schertz, J.F. Doebley, S.R.M. Pinson, S.C. Liu, J.W. Stansel, and J.E. Irvine. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718.

Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:e190.

Peleman, J., and J. van der Voort. 2003. Breeding by design. Trends Plant Sci. 8:330–334.

Pressoir, G., P.J. Brown, W. Zhu, N. Upadyayula, T. Rocheford, E.S. Buckler, and S. Kresovich. 2009. Natural variation in maize architecture is mediated by allelic differences at the PINOID co-ortholog *barren inflorescence2*. Plant J. 58:618–628.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909.

Pritchard, J.K., M. Stephens, N.A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. Am. J. Hum. Genet. 67:170–181.

Psaty, B.M., D. Arnett, and G. Burke. 2007. A new era of cardiovascular disease epidemiology. JAMA 298:2060–2062.

Purcell, S., S.S. Cherny, and P. C. Sham. 2003. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19: 149–150.

Rafalski, J.A. 2010. Association genetics in crop improvement. Curr. Opin. Plant Biol. 13:174–180.

Robertson, G.P., and P.M. Vitousek. 2009. Nitrogen in agriculture: Balancing the cost of an essential resource. Annu. Rev. Environ. Resour. 34:97–125.

Rostoks, N., L. Ramsay, K. MacKenzie, L. Cardle, P.R. Bhat, M.L. Roose, J.T. Svensson, N. Stein, R.K. Varshney, D.F. Marshall, A. Graner, T.J. Close, and R. Waugh. 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc. Natl. Acad. Sci. USA 103:18656–18661.

Salse, J., B. Piégu, R. Cooke, and M. Delseny. 2004. New in silico insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. Plant J. 38:396–409.

Salvi, S., G. Sponza, M. Morgante, D. Tomes, X. Niu, K.A. Fengler, R. Meeley, E.V. Ananiev, S. Svitashev, E. Bruggemann, B. Li, C.F. Hainey, S. Radovic, G. Zaina, J.A. Rafalski, S.V. Tingey, G.-H. Miao, R.L. Phillips, and R. Tuberosa. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc. Natl. Acad. Sci. USA 104:11376–11381.

Salvi, S., and R. Tuberosa. 2005. To clone or not to clone plant QTLs: Present and future challenges. Trends Plant Sci. 10:297–304.

Schaid, D.J. 2004. Evaluating associations of haplotypes with traits. Genet. Epidemiol. 27:348–364.

Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, P. Minx, A.D. Reily, L. Courtney, S.S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S.M. Rock, E. Belter, F. Du, K. Kim, R.M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S.M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L.

Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M.J. Levy, L. McMahan, V.P. Buren, M.W. Vaughn, K. Ying, C.T. Yeh, S.J. Emrich, Y. Jia, A. Kalyanaraman, A.P. Hsia, W.B. Barbazuk, R.S. Baucom, T.P. Brutnell, N.C. Carpita, C. Chaparro, J.M. Chia, J.M. Deragon, J.C. Estill, Y. Fu, J.A. Jeddeloh, Y. Han, H. Lee, P. Li, D.R. Lisch, S. Liu, Z. Liu, D.H. Nagel, M.C. McCann, P. SanMiguel, A.M. Myers, D. Nettleton, J. Nguyen, B.W. Penning, L. Ponnala, K.L. Schneider, D.C. Schwartz, A. Sharma, C. Soderlund, N.M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T.K. Wolfgruber, L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J.L. Bennetzen, R.K. Dawe, J. Jiang, N. Jiang, G.G. Presting, S.R. Wessler, S. Aluru, R.A. Martienssen, S.W. Clifton, W.R. McCombie, R.A. Wing, and R.K. Wilson. 2009. The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112–1115.

Soderlund, C., A. Descour, D. Kudrna, M. Bomhoff, L. Boyd, J. Currie, A. Angelova, K. Collura, M. Wissotski, E. Ashley, D. Morrow, J. Fernandes, V. Walbot, and Y. Yu. 2009. Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. PLoS Genet. 5:e1000740.

Sprague, G.F., and J.W. Dudley (ed.). 1988. Corn and corn improvement, 3rd ed. Agron. Monogr. 18. ASA, CSSA, SSSA, Madison, WI.

Strable, J., and M.J. Scanlon. 2009. Maize (*Zea mays*): A model organism for basic and applied research in plant biology. CSH Protoc. 2009(10):pdb.emo132.

Szalma, S.J., E.S. Buckler, M.E. Snook, and M.D. McMullen. 2005. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor. Appl. Genet. 110:1324–1333.

Takeda, S., and M. Matsuoka. 2008. Genetic approaches to crop improvement: Responding to environmental and population changes. Nat. Rev. Genet. 9:444–457.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley, and B.S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA 98:9161–9166.

Tester, M., and P. Langridge. 2010. Breeding technologies to increase crop production in a changing world. Science 327:818–822.

Thornsberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E.S. Buckler. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. 28:286–289.

Tian, F., N.M. Stevens, and E.S. Buckler. 2009. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proc. Natl. Acad. Sci. USA 106(Supplement 1):9979–9986.

Tracy, W.F., S.R. Whitt, and E.S. Buckler. 2006. Recurrent mutation and genome evolution: Example of *sugary1* and the origin of sweet maize. Plant Gen. 1:S49–S54.

Vallabhaneni, R., and E.T. Wurtzel. 2009. Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. Plant Physiol. 150:562–572.

Visscher, P.M. 2008. Sizing up human height variation. Nat. Genet. 40:489–490.

Walbot, V. 2009. 10 reasons to be tantalized by the B73 maize genome. PLoS Genet. 5:e1000723.

Waugh, R., J.L. Jannink, G.J. Muehlbauer, and L. Ramsay. 2009. The emergence of whole genome association scans in barley. Curr. Opin. Plant Biol. 12:218–222.

Wilson, L.M., S.R. Whitt, A.M. Ibanez, T.R. Rocheford, M.M. Goodman, and E.S. Buckler. 2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733.

Xing, Y.Z., and Q.F. Zhang. 2010. Genetic and molecular basis of rice yield. Annu. Rev. Plant Biol. 61:11.1–11.22.

Yan, J.B., C.B. Kandianis, C.E. Harjes, L. Bai, E.H. Kim, X.H. Yang, D. Skinner, Z.Y. Fu, S. Mitchell, Q. Li, M.G.S. Fernandez, M. Zaharieva, R. Babu, Y. Fu, N. Palacios, J.S. Li, D. DellaPenna, T.P. Brutnell, E.S. Buckler, M.L. Warburton, and T. Rocheford. 2010b. Rare genetic variation at Zea mays *crtRB1* increases β-carotene in maize grain. Nat. Genet. 42:322–327.

Yan, J.B., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen, and J. Crouch. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4:e8451.

Yan, J.B., H. Tang, Y.Q. Huang, Y.L. Zheng, and J.S. Li. 2004. Comparative analysis of QTL for important agronomic traits between maize and rice. Acta Genet. Sin. 31:1401–1407.

Yan, J.B., X.H. Yang, T. Shah, H. Sánchez, J.S. Li, M.L. Warburton, Y. Zhou, J. Crouch, and Y.B. Xu. 2010a. High-throughput SNP genotyping with the GoldenGate assay in maize. Mol. Breed. 25:441–451.

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569.

Yano, M., Y. Katayose, M. Ashikari, U. Yamanouchi, L. Monna, T. Fuse, T. Baba, K. Yamamoto, Y. Umehara, Y. Nagamura, and T. Sasaki. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell 12:2473–2484.

Yu, J., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17:155–160.

Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203–208.

Zhang, N., A. Gur, Y. Gibon, R. Sulpice, S. Flint-Garcia, M.D. McMullen, M. Stitt, and E.S. Buckler. 2010a. Genetic analysis of central carbon metabolism unveils an amino acid substitution that alters maize NAD-dependent isocitrate dehydrogenase activity. PLoS One 5:e9991.

Zhang, Q., J. Li, Y. Xue, B. Han, and X.W. Deng. 2008. Rice 2020: A call for an international coordinated effort in rice functional genomics. Mol. Plant 1:715–719.

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010b. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355–360.

Zheng, P., W.B. Allen, K. Roesler, M.E. Williams, S. Zhang, J. Li, K. Glassman, J. Ranch, D. Nubel, W. Solawetz, D. Bhattramakki, V. Llaca, S. Deschamps, G.Y. Zhong, M.C. Tarczynski, and B. Shen. 2008. A phenylalanine in *DGAT* is a key determinant of oil content and composition in maize. Nat. Genet. 40:367–372.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Gen. 1:5–20.

Zhu, C., and J. Yu. 2009. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics 182:875–888.