Chapter 25
Mining Natural Variation for Maize Improvement: Selection on Phenotypes and Genes

Shilpa Sood, Sherry Flint-Garcia, Martha C. Willcox and James B. Holland

Contents

Abstract Maize is highly genetically and phenotypically diverse. Tropical maize and teosinte are important genetic resources that harbor unique alleles not found in temperate maize hybrids. To access these resources, breeders must be able to extract favorable unique alleles from tropical maize and teosinte from their population

J. B. Holland (✉) · S. Sood
Department of Crop Science, North Carolina State University,
Raleigh, NC 27695, USA
e-mail: james_holland@ncsu.edu

J. B. Holland
USDA-ARS Plant Science Research Unit, Raleigh, NC 27695, USA

S. Sood
e-mail: shilpa_sood@ncsu.edu

S. Flint-Garcia
USDA-ARS Plant Genetics Research Unit, Columbia, MO 65211, USA

Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

M. C. Willcox
Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT),
Texcoco, Edo. de México, México

R. Tuberosa et al. (eds.), *Genomics of Plant Genetic Resources,*
DOI 10.1007/978-94-007-7572-5_25,

genomic context, where they are linked with many undesired alleles that confer adaptation to tropical environments, ancient farming methods, or wild growth habit (in the case of teosinte). Long-term traditional breeding efforts have demonstrated the value of diverse germplasm to improve maize productivity, while also enhancing the genetic base of cultivated varieties. Genomics provides new opportunities to identify the genes affecting important agronomic traits and to estimate the wide range of allelic effects at such genes. New approaches to complex trait analysis, including joint multiple population analysis, genome-wide association analysis, and genomic selection, can leverage high throughput sequencing and genotyping technologies to improve our understanding of the genome-wide distribution of allele effects across the wide genetic variation in the primary gene pool of maize. Implementing this information for practical maize improvement remains a challenge.

Keywords Maize · Teosinte · Allelic effect · Genome sequencing · Genome-wide association analysis · Linkage drag · Genomic selection · SNP · Candidate gene · Haplotype · Adaptation · Productivity · Zea mays

Maize (*Zea mays* L. subsp. *mays*) is an extremely genetically variable crop, adapted to a wide range of habitats, from latitude 40° S to 58° N and including the tropics (Mangelsdorf 1974). In México alone, maize is adapted to environments from 0 to 2900 masl and with 426–4245 mm annual rainfall (Ruiz et al. 2008). The wide genetic variation and adaptation of maize is reflected in its amazing phenotypic diversity for many morphological, developmental, agronomic, and reproductive traits (Kuleshov 1933).

Maize was presumably originally domesticated 5–10,000 years ago in or near Southern México from a progenitor similar to the extant wild teosinte, *Z. mays* subsp. *parviglumis*, hereafter *parviglumis* (Matsuoka et al. 2002). Stringent selection for rare combinations of mutations in a relatively small number of key domestication loci in the earliest phase of the domestication process, followed by thousands of generations of artificial selection for increased ear size and kernel production per plant subjected maize to a population bottleneck, reducing its genetic diversity relative to teosinte (Doebley 2004; Wright et al. 2005). Nevertheless, modern maize retains higher sequence diversity than humans or Drosophila (Tenaillon et al. 2001). The predominantly outcrossing mating system of maize, its exposure to selection for adaptation in very diverse environments and for distinct purposes, and the potential for gene flow between maize and its sympatric wild relatives near its center of origin in México all contributed to the relatively high genetic diversity within maize compared to other crops.

Breeders would like to exploit this substantial genetic variation for the purpose of improving elite maize hybrids for important agronomic traits including grain yield and quality, disease and insect resistance, and abiotic stress resistance. Historically, breeders attempted to measure, classify, and exploit maize genetic variation based on observable phenotypic variation. Maize geneticists pioneered the development of molecular marker systems in plants, which provide a means to directly assay genetic variation. In recent years, the ability to characterize genetic variation in maize at the

sequence level has improved dramatically, permitting unprecedented opportunities to identify specific genes (or non-coding sequences) controlling phenotypic variation, and to expose the underlying allelic to direct genic selection.

This review of the use of natural genetic variation in maize will follow the historical development of methodological approaches, from strictly phenotype-based evaluation, classification, and selection, the successes and failures of which are well documented, to current approaches based on gene identification and allele mining, which are just beginning to be tested. Rather than ignore decades of work on phenotype-based breeding with diverse maize, we believe that lessons learned from this research provide a useful framework for considering the likely advantages and disadvantages of modern gene-based selection.

25.1 Maize History and Classification

Maize spread through the Americas following its domestication in Southern México approximately 5,000–10,000 years ago (Matsuoka et al. 2002; Piperno et al. 2009; Van Heerwaarden et al. 2011), resulting in a distribution ranging from the Gaspe Peninsula in modern day Canada (> 40° N) to Chile and Argentina (nearly 40° S) before the arrival of Columbus (Weatherwax 1954). The spread of maize northward from its center of origin has been studied in some detail, revealing that maize was grown east of the Mississippi River by about 2000 years BP ago (Crawford et al. 2006), but that it remained a minor component of the early agricultural system in this area until about 1200 years BP (Smith 1989). A dramatic shift to a maize-based agriculture in North America occurred between 1,200 and 900 years BP; the evolution of the early maturing Northern Flint type was likely an important component of this transition, but the biological changes of maize occurred within dramatic cultural changes that happened during this time. The subsequent Colombian exchange (Crosby 1972) resulted in the relatively rapid dissemination of maize to Europe, followed by Asia and Africa. The natural selection for adaptation to widely diverse ecological habitats combined with artificial selection for human food and ceremonial uses (Weatherwax 1954; Hernández 1985) was the basis for the fantastic display of phenotypic diversity among maize varieties from around the world (Fig. 25.1).

The tremendous variability within maize was recognized early on, and initial attempts to classify the different types of maize were rather artificial, focusing on endosperm type (Sturtevant 1899). Anderson and Cutler (1942) introduced the concept of maize races as a way to delineate groups in which individuals have "a significant number of genes in common," and suggested some characteristics of the reproductive organs (tassel and ear) that would be useful for grouping maize into races. Such methods, along with geographic origins, informed the large-scale efforts to collect and classify the landraces grown in Latin America in the 1940s and 1950s (Goodman and Brown 1988). This classification effort was performed more or less independently for each country or region, such that relationships among maize populations from different areas were not formally considered, although some race names

Fig. 25.1 A small sample of the phenotypic variability among Latin American races of maize for ear and kernel morphology. Each ear represents a different race, grown under common conditions in a winter nursery in Homestead, FL by Dr. M.M. Goodman. (Photographs by Dr. Jesús Sánchez-Gonzalez)

were used for maize found in different regions. Furthermore, the classifications were performed on a somewhat *ad hoc* basis, without formally defining what "significant number of genes" or character similarity was sufficient to define a race; the authors intended these racial groupings as only preliminary steps in classification of maize (Holland and Nelson 2010). About 250 races have been named for maize of the Americas (M.M. Goodman, pers. comm.), and collection and classification efforts continue to this day in regions where traditional landraces are still grown in México (Ron Parra et al. 2006; Rincón et al. 2010).

Goodman and colleagues formalized the classification of maize and studied relationships among races from different countries using numerical taxonomy (reviewed in Goodman and Brown 1988; Holland and Nelson 2010). The development of isozymes as a genetic marker system in maize provided geneticists with a method to measure relationships among groups without the confounding influence of environment on phenotypic characters. A series of studies by Goodman and colleagues (Goodman and Stuber 1983; Doebley et al. 1984; Doebley et al. 1985; Bretting et al. 1987; Doebley et al. 1988; Bretting et al. 1990; Sanchez and Goodman 1992a, b; Sanchez et al. 2000a, b, 2006, 2007) measured the genetic variation at neutral isozyme loci among and within landraces from the Americas. A key finding of these studies was the very high level of genetic variation within accessions (generally

representing a sample of ears from a single field or village) and races. Genetic differentiation among accessions within a race or among races tends to be low (typically less than 20 %), indicating that races and accession groupings account for only a limited amount of genetic variation; the remaining bulk of genetic variation can be found within collections (Sanchez et al. 2000a, b). Where genetic variation follows racial groupings, it is often strongly associated with geography and ecology, altitude in particular (Bretting et al. 1990; Sanchez et al. 2000a). Further, races differ for the amount of variation they contain, with widespread races, particularly those from Mesoamerica, possessing more alleles per locus than races used as specialty varieties and with restricted geographic ranges (Sanchez et al. 2000a). Finally, rare alleles are exceedingly common: 65 % of alleles had frequency of 1 % or less in the Mexican races analyzed by Sanchez et al. (2000a, b). Reif et al. (2006) largely confirmed these findings with SSR analyses of Mexican landraces. Pressoir and Berthaud (2004a, b) measured both SSR and trait variation within and among landrace samples collected from a small region of México, finding strong differentiation among populations from different villages for certain ear traits, but almost no differentiation for random SSR markers. They interpreted these apparently contradictory results as evidence that gene flow is very common among villages (facilitated by regular seed exchanges), reducing differentiation for most of the genome, but that strong divergent local selections for specific traits result in differentiation at those loci controlling the targeted traits.

SSR evaluations of maize landraces from throughout the Americas indicate that at the broadest scale, American landraces can be grouped into four geographically-based clusters: highland Mexican, Northern United States, lowland tropical, and highland Andean (Vigouroux et al. 2008). Landraces from some geographic areas represent mixtures of these mega-groupings: e.g., Southeastern USA landraces appear to have originated from a mixture of Northern USA and tropical lowland types, whereas lowland Brazilian maize appears to have arisen from admixture between Andean and tropical lowland groups. Variation among landraces within these mega groups is highest for Mexican and lowest for the Andean and Northern US landraces, which represent the extremes of geographic spread from the center of origin. SSR studies also clarified the relationships between landraces of Europe and the Americas, suggesting two distinct introductions of maize to Europe: first by Caribbean maize and later by Northern Flint types (Rebourg et al. 2003; Dubreuil et al. 2006).

In contrast to the high levels of molecular variation observed in landraces and tropical germplasm in general, modern temperate hybrids exhibit high degrees of relatedness arising from the use of a limited set of founder lines (Smith et al. 1992; Duvick et al. 2004). A 23 % reduction in sequence variation was observed between landraces and public USA inbreds (Tenaillon et al. 2001), and a further reduction between public inbreds and private industry hybrids might be expected. Indeed, comparison of public and private industry inbreds (expired Plant Variety Protection) demonstrates limited genetic variation among many private inbreds, but also reveals some unique germplasm groups developed by private industry that were not represented in publicly developed lines (Nelson et al. 2008).

25.2 Breeding to Enhance Genetic Diversity in Elite Materials

The narrow genetic base of maize hybrids in the United States (all derived from only one of the 250 or so named races, the Corn Belt Dents) relative to the global diversity of the crop was recognized early on (Anderson 1944). Brown (1953) recommended the use of exotic germplasm to ameliorate the narrow genetic base of US maize and increase long-term potential for yield grain. Tropical maize, in particular, was identified as harboring the most genetic variation for observable characters, and as such the source of exotic germplasm most likely to have unique (and hopefully favorable) alleles absent from the Corn Belt Dents (Gerrish 1983; Goodman 1985; Tallury and Goodman 1999). In principal, the potential utility of broadening the genetic base of temperate maize is widely accepted, however the difficulties encountered by breeders in overcoming poor adaptation of tropical maize to temperate regions have hindered all efforts to broaden the genetic base of hybrid varieties grown in the USA (Hallauer 1978). Using tropical germplasm for breeding in temperate regions is hampered by: a lack of information needed to rationally choose exotic materials from among the tens of thousands of available sources; photoperiod sensitivity; a significant gap in agronomic quality between elite U.S. materials and exotic races; severe inbreeding depression; and undesirable agronomic characters such as weak roots and stalks, excessive plant and ear height, susceptibility to smut, and high grain moisture (Hallauer 1978; Goodman 1985, 1992).

Furthermore, the Corn Belt Dents have been generally recognized as one of the most, if not the most, inherently productive races of maize. It has been argued that their dominance in temperate regions is not by chance, but rather because they represent a hybrid race (admixture of Northern Flints and Southern Dents) and have the longest history of selection in the Corn Belt region of the USA (Troyer 1999). Recall that the Native American and Mesoamerican peoples that have the longest history as maize breeders did not grow maize in the grasslands that are now the Corn Belt region (Weatherwax 1954); rather, maize was not selected for these regions until the relatively recent transformation of the Midwestern prairies into farmland. Nevertheless, it is unlikely that the Corn Belt Dent race has a monopoly on all of the useful genes available in maize, and therefore, exotic germplasm may be useful for the improvement of U.S. maize (Brown 1975; Geadelmann 1984). Furthermore, as the dominant corn growing environments change due to global climate change and climatic events such as drought become more frequent, the utility of maize types selected by Native Americans for harsher environments may become essential.

Sources of exotic germplasm available to breeders include landrace accessions, composite populations, inbred lines, and hybrids. The breeding experience of Goodman et al. (2004) with tropical maize germplasm in the temperate USA has been a clear demonstration of the substantially greater utility of tropical hybrids and inbreds as breeding parents, as compared to recurrent selection populations, or, worse, landrace collections *per se*. The previous efforts of breeders in tropical regions in purging deleterious alleles during inbreeding to develop lines should be taken advantage of if at all possible. Starting with tropical hybrids, inbred lines with purely

tropical backgrounds but adapted to temperate regions have been developed by traditional breeding methods; although these lines have relatively poor performance *per se*, they produce hybrids with very high yield potential in some cases (Holley and Goodman 1988; Uhr and Goodman 1995a, b; Tallury and Goodman 1999; Goodman et al. 2000; Goodman 2004).

Breeding with landraces *per se* is more challenging, and the first difficulty is deciding which landrace accession to choose to use in a breeding program. Approximately 20,000 unique accessions of Latin American maize are stored in germplasm banks worldwide (Goodman 1983), and besides race name, there is often no information available to guide selection of starting materials. Evaluation of accessions for yield and agronomic performance *per se* in environments to which they are adapted is probably the most efficient and useful criterion for selection of breeding material. Castillo-Gonzalez and Goodman (1989) evaluated about 1,300 Latin American accessions in short daylength nurseries, and used their yield levels from this experiment as a culling criterion. The best accessions selected from this evaluation were crossed to a temperate line and selection within these breeding crosses resulted in the development of families and inbreds with acceptable adaptation and superior combining ability. (Holland and Goodman 1995; Tarter et al. 2003). Following this model, The Latin American Maize Project was undertaken to evaluate as many possible accessions in their home environments. Landrace collections were evaluated in their countries of origin (Salhuana et al. 1998) and the best collections were advanced to the Germplasm Enhancement of Maize (GEM) program (Pollak 2003). In the traditional GEM breeding protocol, superior landraces are crossed to elite proprietary inbreds, and the segregating populations are made available to GEM cooperators. Early generation selection is followed by extensive evaluations, and numerous lines with superior agronomic performance have been released for public use (Balint-Kurti et al. 2006). These programs have demonstrated the excellent potential of tropical maize germplasm for improving temperate material.

25.3 QTL Analysis and its Discontents

Many agriculturally and evolutionarily important traits in plants are quantitative in nature. Phenotypic variation for these traits is caused by a combination of segregation at multiple quantitative trait loci (QTL), the environment, and the interaction between genes and the environment (Mackay 2001). Two of the most commonly used approaches to dissect genes underlying complex quantitative traits are linkage analysis and association mapping (Mackay 2001; Risch and Merikangas 1996). Linkage analysis utilizes the shared inheritance of functional polymorphisms and adjacent markers within families or pedigrees of known ancestry. In plants, linkage analysis has been traditionally conducted with experimental populations derived from a biparental cross, such as F_2, backcross or recombinant inbred lines. Following the initial successes of identifying QTL in plants (Edwards et al. 1987; Paterson et al. 1988), methods to use QTL information to enhance selection of quantitative

traits were developed (Stuber and Edwards 1986; Lande and Thompson 1990). In particular, Tanksley and colleagues recognized the potential utility of linkage mapping approaches to aid the identification of unique favorable alleles in wild relatives and germplasm collections, and their subsequent incorporation into elite breeding populations (Tanksley and Nelson 1996; Tanksley and McCouch 1997). Indeed, the major practical success of QTL mapping has been the identification and marker-aided selection of QTL with moderate to large effects on biotic and abiotic stress resistances in several self-pollinating crops (Young 1999; Frary et al. 2000; Monforte and Tanksley 2000; Holland 2004; Pumphrey et al. 2007; Venuprasad et al. 2011).

Tuberosa and Salvi (2009) reviewed progress in QTL mapping in maize, citing several cases where QTL with moderate effects on complex traits such as abiotic stress resistance were identified, providing potential marker-assisted selection targets. In general this has been aided by the physiological dissection of complex traits into component traits (e.g., traits such as root architecture, leaf morphology, or anthesis-silk interval that influence grain yield), which have simpler genetic control when evaluated under controlled environmental conditions (Tuberosa and Salvi 2009). In general, however, most quantitative traits in maize appear to be under more complex genetic control relative to self-pollinating species. Thus, the genetic control is distributed across many loci, resulting in numerous loci with small effects, a situation in which QTL mapping has limited power and poor accuracy in typical mapping population sizes of a few hundred progeny lines (Beavis 1998; Melchinger et al. 1998). In such cases, very large population sizes are required to obtain accurate estimates of QTL positions and effects (Laurie et al. 2004; Schön et al. 2004; Holland 2007). Furthermore, the very high genetic diversity and low levels of linkage disequilibrium in diverse maize populations hinders the translation of QTL effect estimates from mapping populations to breeding populations representative of elite breeding programs (Holland 2004; Holland 2007; Bernardo 2008).

Recognizing the limited inferences that can be drawn from traditional biparental mapping populations and the difficulty in applying QTL mapping information to general breeding populations, maize geneticists pioneered methods to increase mapping resolution with advanced intercross line (AIL) populations and to broaden the inference space of QTL analyses by combining QTL mapping information across populations and pedigrees. The intermated B73 $\times$ Mo17 AIL population was derived by selfing lines to high levels of homozygosity following four generations of random mating, creating four times as many recombination events within small intervals compared to the initial F_2 generation (Lee et al. 2002; Sharopova et al. 2002; Winkler et al. 2003). This population serves as the community standard high resolution mapping population used to connect the B73 genome sequence and genetic map (Fu et al. 2006; Schnable et al. 2009), and also has been used for high resolution QTL mapping (Balint-Kurti et al. 2007; Lauter et al. 2008; Rodriguez et al. 2008; Zhang et al. 2010a). Other maize AILs have been used for genetic and QTL mapping (Falque et al. 2005, Falke et al. 2006; Huang et al. 2010b).

Meta-analysis of multiple independent QTL studies has been used to synthesize results with respect to a common consensus genetic map, highlighting genome regions that are consistently associated with variation for a trait across populations and

environments (Chardon et al. 2004) or improving the precision of QTL localization (Kump et al. 2010). A more direct approach to integrate QTL information across populations is joint population QTL mapping (Rebai et al. 1997; Blanc et al. 2006; Buckler et al. 2009; Coles et al. 2010). Joint linkage analysis increases power and resolution of QTL mapping, permits tests of QTL effect interactions with genetic backgrounds, and permits direct comparison of multiple allele effects, enhancing understanding of genetic heterogeneity (Holland 2007). Methods to combine information across more complex pedigrees have also been developed in the context of maize breeding programs (Zhang et al. 2005).

25.4 Association Analysis

Although linkage-based QTL mapping has been useful in identifying a number of genes affecting qualitative and quantitative traits, and despite substantial methodological advances pioneered in maize, several factors have hindered the translation of QTL mapping studies into breeding tools: the limitations of QTL mapping resolution (typically 10–20 cM, Holland 2007), accuracy of effect estimation, and sampling of allelic variation (typically only two alleles per locus, Holland 2004; Bernardo 2008). An alternative to linkage-based QTL mapping is association analysis, also known as association mapping or linkage disequilibrium mapping. Association analysis is based on gametic phase disequilibrium (commonly, although inaccurately, referred to as linkage disequilibrium, LD) to study the relationship between phenotypic variation and genetic polymorphisms. By focusing on diverse germplasm of unrelated ancestry, association analysis aims to sample genomes that have undergone thousands of generations of recombination since their descent from a common ancestor. As such, association mapping makes use of ancient as well as evolutionary recombination at the population level (Risch and Merikangas 1996; Remington et al. 2001; Thornsberry et al. 2001; Yu and Buckler 2006). The reduced correlations between even very closely linked loci potentially enables very high resolution marker-phenotype associations (Buckler and Thornsberry 2002; Flint-Garcia et al. 2005). Originally developed to identify genes involved in human diseases (Kerem et al. 1989; Corder et al. 1994), association mapping has become increasingly popular in plants in the last decade (Hauser et al. 2001; Thornsberry et al. 2001; Wilson et al. 2004; Szalma et al. 2005; Breseghello and Sorrells 2006a; Ehrenreich et al. 2009) because of advances in high throughput genomic technologies that provide dense coverage of the genome, the interest among breeders to identify novel and superior alleles, and improvements in statistical analysis methods. Advantages of association mapping over linkage mapping include the potential to survey effects of many alleles per locus, reduced cost and time to assemble an association mapping panel compared to creating structured populations for linkage analysis, and higher mapping resolution (Breseghello and Sorrells 2006a; Yu and Buckler 2006).

The resolution of association mapping is dependent upon the extent of linkage disequilibrium (LD), which, in turn, depends on recombination, genetic drift, selection, mating pattern and population admixtures; these factors vary both within species and between species (Flint-Garcia et al. 2003; Gaut and Long 2003). In maize, significant levels of LD extend less than 1 kb for landraces (Tenaillon et al. 2001) and almost 2 kb for diverse inbred lines (Remington et al. 2001), but much farther in collections of elite commercial inbred lines (Ching et al. 2002; Rafalski 2002). Thus, in diverse maize samples, the rapid breakdown of LD is sufficient to permit gene-level mapping resolution, and is an ideal method to test the phenotypic effects of candidate genes, as has been done in maize for a handful of genes known or hypothesized to act as regulators or structural components of biochemical or developmental pathways. These include genes for flowering time, kernel composition, and secondary metabolite concentrations (Thornsberry et al. 2001; Whitt et al. 2002; Palaisa et al. 2003; Wilson et al. 2004; Andersen et al. 2005; Szalma et al. 2005; Camus-Kulandaivelu et al. 2006; Harjes et al. 2008; Yan et al. 2010). Unfortunately, our understanding of genetic regulation is insufficient to reliably identify candidate genes for the vast majority of agriculturally important traits.

In the absence of a candidate gene list likely to contain causal loci, researchers must rely on random markers to sample the entire genome, in so-called genome-wide association studies (GWAS). The rate of decay of LD over physical distance determines the density of marker coverage needed to perform whole genome association analysis. In some self-pollinated crops or highly related maize populations with very extensive LD, one marker placed every cM or so can be sufficient to tag all segregating sites, but the resulting mapping resolution will be low (Breseghello and Sorrells 2006b; Rostoks et al. 2006; Hyten et al. 2007). In diverse maize, where LD decays rapidly, very high marker density is required to ensure a high probability that at least one marker is in high LD with causal loci (Yu and Buckler 2006). Gore et al. (2009) estimated that more than 10 million SNPs will be required to adequately conduct genome-wide association analysis in maize. The use of new high-throughput techniques, which allow genotyping hundreds of thousands of SNPs in a single assay and the creation of high density SNP haplotype maps in different plant species (Clark et al. 2007; Gore et al. 2009), has significantly boosted the application of association analysis in genome-wide scans for complex traits (Atwell et al. 2010; Brachi et al. 2010; Huang et al. 2010a; Kump et al. 2011; Poland et al. 2011; Ramsay et al. 2011; Tian et al. 2011). Association mapping undoubtedly has tremendous potential in dissecting the complex traits in plants and especially maize given its extensive phenotypic and molecular diversity. Several association mapping populations have been assembled in maize for various objectives (Andersen et al. 2005; Flint-Garcia et al. 2005; Camus-Kulandaivelu et al. 2006; Yan et al. 2009; Yang et al. 2010; Hansey et al. 2011).

Gene-phenotype associations may arise due to: causality (these are the associations we are most interested in), LD arising from physical proximity between marker and causal site (these can be useful in marker assisted selection), and LD arising from population structure. Population structure can cause highly significant associations between a marker and a phenotype, even when the marker is not physically linked to

any causative loci (Pritchard 2001; Thornsberry et al. 2001). Therefore, it is important to include estimates of population structure in the association analysis (Flint-Garcia et al. 2005). Various statistical approaches have been designed to control for population structure in different association samples such as the general linear model based approaches: genomic control (Devlin and Roeder 1999), and structured association (Pritchard et al. 2000) for population-based samples and transmission disequilibrium test (Abecasis et al. 2000) for family-based samples. Unified mixed linear model (MLM), and modified "compressed MLM" approaches appear to be superior to previously developed methods for association analysis in maize and other species (Yu et al. 2006; Zhang et al. 2010b). These methods can be used in the context of candidate gene association tests or GWAS (Zhang et al. 2010b). GWAS introduces computational challenges associated with conducting very large numbers of statistical tests, although increases in computer processing speed and improvements in algorithmic efficiency have permitted their application even with huge numbers of SNP tests in GWAS (Kang et al. 2008; Zhang et al. 2010b; Lippert et al. 2011). In addition, GWAS is confronted with the difficulty of determining significance thresholds for thousands or millions of statistical tests, although False Discovery Rate methods are very helpful in this regard as they are computationally tractable even with large numbers of tests (Benjamini and Yekutieli 2005).

25.5 Linkage and Association Analysis in Nested Association Mapping Populations

Linkage mapping and association analysis are complementary in many ways: linkage mapping has high power but low resolution, while association analysis has low power and high resolution. Linkage mapping uses structured populations to its advantage, while association analysis is hindered by population structure. To integrate the advantages of linkage analysis and association mapping into a single strategy, a large-scale set of inter-related maize mapping families was created to facilitate dissection of complex genetic variation underlying quantitative traits. The maize Nested Association Mapping (NAM) population was created to capture significant genetic variation and low linkage disequilibrium in a sample of lines that were used as founders to create multiple biparental linkage mapping populations (Yu et al. 2008; McMullen et al. 2009). The NAM population was created by crossing the inbred reference line B73 to 25 inbred lines that are representative of maize diversity (Flint-Garcia et al. 2005). From each cross, 200 recombinant inbred lines were derived by self-fertilization, resulting in a total of 5,000 RILs (McMullen et al. 2009). B73 was chosen as a reference line because of its role in the physical map and whole-genome sequence (Schnable et al. 2009). The other 25 parents maximize the diversity among the RIL families. More than half of these diverse lines are tropical in origin, nine are temperate lines, two are sweet corn lines and one is a popcorn line (McMullen et al. 2009). Thus, NAM is a specific case of inter-related mapping population mating designs referred to as a reference design. Although there are theoretically better mating designs for joint population QTL mapping (Verhoeven et al.

2006), the reference design was selected for practical reasons. Creating RILs with 50 % pedigree contribution from a broadly-adapted temperate line ensured that even the effects of alleles from tropical inbred founders would not be highly confounded with poor adaptation to temperate environments. Each NAM line was genotyped with a common panel of 1,106 SNPs selected to cover the genome and to be polymorphic within most families, and the use of the common map permits investigation of recombination frequency differences among families and simplifies implementation of joint linage QTL analysis (Buckler et al. 2009; McMullen et al. 2009).

The power of NAM for QTL analysis has been demonstrated by dissecting the genetic architecture of flowering time in maize. Joint linkage QTL mapping revealed 36–39 QTL affecting time to anthesis or silking (Buckler et al. 2009). All the identified QTL exerted small effect on the phenotype in an additive manner. In fact, the largest effect QTL for days to silking (DS) only had an additive effect of 1.7 days relative to B73. The complexity of flowering time in maize, in contrast to the identification of genes with very large effects on flowering time in self-pollinating species such as wheat and rice (Cockram et al. 2007; Izawa 2007), is likely due in part to the predominantly outcrossing mating system of maize. Rare mutations with large effects on flowering time would be associated with reproductive isolation and self-fertilization, and a consequent decrease in fitness in the progeny carrying the mutation. In addition, little evidence of epistasis or genotype-environment interactions (GEIs) was revealed, although the testing environments all had long daylengths; greater GEI would be expected when comparing across environments of distinct photoperiods (Buckler et al. 2009).

The most powerful application of NAM is the ability to efficiently conduct GWAS. NAM provides several substantial advantages for GWAS relative to diverse line association panels. First, the framework linkage map of 1,106 SNPs permits efficient imputation or "projection" of founder line SNP variation onto the entire RIL panel (Yu et al. 2008). For example, the maize haplotype map (HapMap version 1) consists of 1.6 million SNPs identified among the founders of the NAM population (Gore et al. 2009). Since the physical positions of those SNPs in the B73 reference sequence and their allelic composition among the founders are known, their probable allelic status in the NAM RILs can be imputed easily based on the flanking markers of the linkage map. Thus, the 1.6 M HapMap SNPs could be accurately imputed onto 5,000 mapping lines by sequencing only the 26 founders.

A second major advantage of GWAS in NAM is its known population structure. Whereas population structure in diverse line panels must be estimated with random markers, the structure of NAM is known: there are 25 families and there is no structure within families because the RILs within families were derived randomly. Thus, population structure is accounted for completely in the analysis simply by fitting the family main effect. Furthermore, the ability to conduct joint linkage QTL analysis and GWAS in the same population provides an additional advantage: the joint linkage QTL model can be used to account for genetic variation outside of the region being tested in GWAS, thus increasing the power for GWAS. The random derivation of RILs within each family also eliminates any unlinked LD that existed among the

founder lines, and dissipates LD among linked SNPs to an extent determined by the strength of linkage (Kump et al. 2011).

NAM-GWAS was conducted for leaf architecture traits (Tian et al. 2011) and two foliar diseases of maize (Kump et al. 2011; Poland et al. 2011). Similar to flowering time, the joint linkage QTL analysis detected between 29 to 36 loci for different traits (Poland et al. 2011; Tian et al. 2011). Again, most of the QTL effects were small, but together they explained more than 77–83 % of the genetic variance among RILs.

In these three studies, GWAS using 1.6 M HapMap SNPs identified between 203 and 295 SNPs with strong association with a trait. Among the associated SNPs, only 30–50 % of the SNPs for all three traits were in the QTL regions identified through linkage analysis. To some extent the incomplete overlap of QTL and associated SNP positions can be explained by low SNP coverage and differences in power of the two analyses. Some of the causative SNPs might have been missed in GWAS because the power to detect small effects that are segregating in only one or few crosses is limited (Holland 2007; Haley 2011). Also, complete marker saturation of the maize genome has been estimated to require ten times more SNPs than the current 1.6 M HapMap SNPs (Gore et al. 2009). Therefore it can be assumed that many of the causative SNPs were missed in the GWAS analyses due to a lack of linkage disequilibrium with the tested SNPs. Finally, only SNPs and small insertion/deletion polymorphisms were considered in these studies, whereas structural variation such as copy number variation and presence-absence variation among maize inbreds may also play a role in complex phenotypes (Lai et al. 2010; Swanson-Wagner et al. 2010; Eichten et al. 2011).

In spite of these limitations, each NAM-GWAS study (Kump et al. 2011; Poland et al. 2011; Tian et al. 2011) was able to successfully identify causal variation in or around several genes whose predicted functions are consistent with their association with the phenotype. The GWAS results provide candidate genes that can be tested further by fine-mapping and isolating individual loci affecting these important agronomic traits. However, one cannot be certain that SNPs identified as associated with a phenotypic trait in NAM-GWAS are in fact causal. In some cases, longer-range LD was observed between SNPs on the same chromosome because of the limited sampling of founders (Kump et al. 2011). It is likely that some proportion of SNPs associated with a trait in NAM-GWAS will turn out to be in LD with causal variants, perhaps at linked loci. The possibility that causal variants exist in non-coding regions (Clark et al. 2006; Salvi et al. 2007) and the lack of annotation for many predicted maize genes also hinders the interpretation of GWAS results in terms of biology.

Recently, the maize haplotype map has been expanded in terms of density of SNPs scored, types of variants scored (read depth variants, a proxy for copy number variation), and germplasm (now including more maize lines and some teosinte inbreds; Chia et al. 2012). This second generation maize haplotype map (HapMap II) provides more than 27 million SNPs and a thousand read depth variants scored on the NAM founders (www.panzea.org). This has provided more power for the most recent NAM-GWAS studies, but also complicates their interpretation, as separating false from true positive signals and causative variants from variants associated by LD becomes more difficult (Hung et al. 2012).

25.6 QTL Fine-Mapping

One way to proceed with identifying the causal variants associated with a QTL or a SNP associated with a phenotype is to conduct high-resolution fine mapping to resolve the variant to a single gene or single non-coding region. While QTL "cloning" is still a major challenge for most quantitative traits, it has been accomplished in a number of cases (Frary et al. 2000; Fridman et al. 2004; Salvi and Tuberosa 2005; Salvi et al. 2007). Selecting appropriate plant material before initiating a QTL fine-mapping and cloning experiment is perhaps the most important aspect of QTL characterization. Near-isogenic lines (NILs) and introgression lines (ILs) are often ideal material with which to initiate high-resolution mapping or positional cloning efforts (Fridman et al. 2004; Eichten et al. 2011). NILs generally refer to sets of lines differing from some common recurrent parent inbred by a small proportion of donor genome (in this case including the target QTL). Introgression lines (ILs) are backcross-derived lines containing segments from wild relatives (or exotic germplasm) such that an IL library would contain the entire genome of wild donor or exotic parent in the recurrent parent background (Zamir 2001; Salvi and Tuberosa 2005; Salvi et al. 2007).

Fine-mapping follows from QTL identification and development of a NIL pair differing only at the target QTL region. Markers defining the QTL region are used to select rare progeny with recombinant chromosomes from the segregating population derived from crossing the NIL pair (Fig. 25.2). Analysis of cosegregation between the phenotype and high density markers within the target region obtained from large insert genomic libraries (bacterial or yeast artificial chromosomes), a reference genome sequence where available, or next-generation sequencing technologies (Elshire et al. 2011) can resolve the QTL position to less than one cM genetic and less than Mb physical distances if sufficient recombinant progeny are obtained and marker density is high enough. With well-annotated sequence information on the narrowed QTL interval, candidate genes can be identified and their allelic sequence variants determined (Paran and Zamir 2003). Once candidate genes are identified, the final step is usually validating the phenotypic effect of critical sequence variants by genetic complementation tests, genetic engineering approaches such as RNA interference (RNAi), gene expression analyses, or by reverse genetic strategies like TILLING, T-DNA or transposon tagging. Several QTLs have been isolated in maize using these basic guidelines, e.g. *tb1* (Doebley et al. 1997; Doebley 2004), *tga1* (Wang et al. 2005), *DGAT1-2* (Zheng et al. 2008), *Rcg1* (Frey 2006; Frey et al. 2011), and *Vgt1* (Salvi et al. 2007).

Once an individual gene affecting a quantitative trait is isolated, the next step is to assess the molecular basis of allelic variation for that trait. In maize, variation in QTL alleles has been identified in both coding and regulatory regions of single genes. The cloned domestication QTL *tb1* (*teosinte branched1*) controls the difference in apical dominance between maize and its progenitor teosinte (Doebley et al. 1997), but no causal nucleotide differences have been observed in the *tb1* coding region between maize and teosinte alleles. Instead, the functional variation seems to lie in the regulatory region upstream of the *tb1* gene (Doebley et al. 1997; Wang et al. 1999; Clark

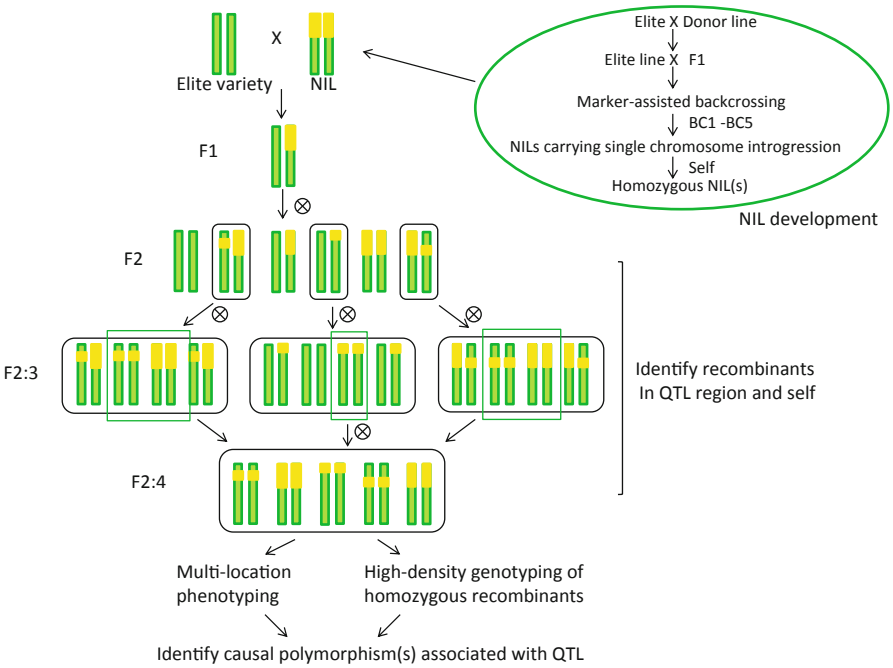

Fig. 25.2 Flow chart for high-resolution genetic mapping. Following the localization of a QTL to a ~ 10–30 cM region of the genetic map, the causal gene(s) may be identified following homogenization of the genetic background and screening large segregating progenies for recombination events within the QTL interval. Multiple rounds of mapping and screening may be required to achieve gene-level resolution

et al. 2006). A transposon insertion in the regulatory region of *tb1* was shown to partially explain the increased apical dominance in maize compared to teosinte (Studer et al. 2011). Similarly, the flowering time QTL *Vgt1* (*vegetative to generative1*) is due to allelic variation in a noncoding region about 70 kb upstream of an *Ap2*-like transcription factor (Salvi et al. 2007). *Vgt1* acts as a cis-regulatory element that controls the expression of downstream genes (Salvi et al. 2007). In contrast, functional variation at the *tga1* (*teosinte glume architecture1*) locus which controls the differences in fruitcase/ear structure between maize and teosinte is due to the single amino acid substitution in the tga1 protein (Wang et al. 2005). Similarly, *DGAT1-2*, a major QTL for high oil content, is caused by an amino acid insertion in the DGAT1-2 protein in the ancestral allele that causes low oil content (Zheng et al. 2008). In addition to sequence variation leading to amino acid changes in proteins and altered regulatory activity, copy number variation (CNV) or presence absence variation (PAV) may underlie QTL. For example, *Rcg1* is a major QTL conferring resistance to anthracnose stalk rot, at which the resistant and susceptible alleles differ by the presence of an entire gene (Frey 2006; Frey et al. 2011). Only 5 % of the US germplasm carries the resistant allele, but the allele is at higher frequency in tropical germplasm

(Frey 2006). The number of QTLs cloned so far is very small (Salvi and Tuberosa 2005; Doebley et al. 2006) and thus our knowledge about the genes and sequences causing such vast phenotypic variation is also very limited. Therefore, although difficult and costly, fine-mapping and cloning experiments provide unique information about the molecular basis of QTL.

Precise estimates of QTL positions are now available from joint linkage and GWAS analysis implemented in NAM (Buckler et al. 2009; Kump et al. 2011; Poland et al. 2011; Tian et al. 2011). We and others are attempting to identify the sequence variants (quantitative trait nucleotides, QTN) underlying some of these QTL by fine-mapping them in an effort to better understand the genetic control of complex traits and to validate and refine the statistical methods used for QTL mapping in NAM. In order to achieve detailed genetic characterization of QTLs underlying agronomic traits studied in NAM population, we have utilized a small sample of a series of near isogenic lines (NILs) carrying introgressions from the NAM founders in B73 genetic background developed by Syngenta AG (Pennisi 2008). We selected several NILs to target alleles with predicted significant effects on plant height, flowering time, and kernel composition as starting materials for fine-mapping. These NILs were crossed to B73 and segregating F_2 populations were screened with markers defining the introgression regions to identity recombinant progeny. Marker selection for homozygous selfed progenies of these F2s was then used to obtain homozygous stocks carrying recombined-chromosomes in the target region (Fig. 25.2). With the availability of several million HapMap SNPs (www.panzea.org) and cost-effective genotyping platforms (including genotyping by sequencing, Elshire et al. 2011), it has become relatively efficient to densely genotype large number of genotypes in short time. However, a major challenge rests in accurately phenotyping the recombinant lines in the fine-mapping experiments, where the QTL has only a small phenotypic effect (Price 2006). Nonetheless, by replicating the phenotyping experiments within and across locations and by combining multiple data sets so as to increase the heritability of the trait (for low heritability traits), it might be possible to detect small QTL effects in fine-mapping populations. For now, it remains to be seen what sorts of genes are responsible for small effect variation in quantitative traits (or if the variants are coding regions at all), and the extent to which QTL identified in NAM were mapped precisely or how often QTL often represent statistical fusions of multiple linked genes that separate into distinct and possibly undetectable effects by high resolution mapping (Studer and Doebley 2011).

25.7 Marker-Based Selection for Complex Traits in Maize

Whereas marker-based selection has become routine for genes or QTL with moderate to large effects on agronomic traits in several self-pollinating species (Cahill and Schmidt 2004; Dubcovsky 2004; Collard and Mackill 2008; Jena and Mackill 2008; Yan et al. 2010), this has not generally been the case in maize to date, although there are a handful of specific traits where the substitution of markers for

phenotypic selection could become routine, e.g., kernel β-carotene content (Harjes et al. 2008;Yan et al. 2010) and anthracnose stalk rot resistance (Frey et al. 2011). Markers have been used to enhance phenotypic selection for quantitative traits like yield in maize (Crosbie et al. 2006; Eathington et al. 2007), but the implementation of marker-based QTL selection in maize has been hindered by the high diversity and low linkage disequilibrium in maize, both of which result in population-dependent marker-trait associations. In other words, QTL mapped in one biparental population may have little or no relation to the QTL segregating for the same trait in other breeding populations (Holland 2004). Efficient use of markers to enhance selection response for polygenic traits in maize requires identification of causal nucleotides (QTN) to use as reliable selection targets or breeding methods that more reliably relate genomic information to breeding values (Holland 2004; Bernardo 2008).

Although we are far from having lists of favorable agronomic trait QTN alleles from exotic germplasm, the very long-term goal of GWAS is to create such lists. Identifying QTN is most likely to occur first for component traits of very complex traits, because the components are more likely to be under simpler genetic control than the complex traits, as already demonstrated for QTL for yield components (Tuberosa and Salvi 2009). Once QTN are reliably known, diverse germplasm collections can be more effectively mined for unique allelic variants that may prove beneficial but are absent from elite breeding populations. Optimally, introgression libraries targeting a wide range of diversity at target regions could be used to estimate the allelic effects of the QTN. Developing nearly-isogenic stocks for each target gene from a wide range of germplasm is likely to be cost-prohibitive, however. If sequence information is available from very diverse germplasm collections, however, the sequence information can be used to selectively choose a small number of donors carrying the different variants at a gene. Unfortunately, given the very high level of sequence diversity in maize, it may generally be unclear which of the many sequence variants within or around a gene should be targeted, and there may be many haplotypes to test. An alternative strategy would be to evaluate earlier backcross generations, which will not as effectively isolate the QTN effect from the donor background, but are much easier and faster to develop. Coles et al. (2011) used this approach to validate several photoperiod response QTL and also to investigate the effects of QTL alleles derived from distinct backgrounds, revealing a surprising degree of variation in photoperiod response effect among tropical donor lines.

Another approach to characterize the effects of QTN across diverse germplasm sources would be to create synthetic populations that include contributions from very diverse germplasm sources, but which have been random-mated for a large number of generations to reduce linkage disequilibrium around most target genes, allowing high resolution association analysis of diverse allele effects to be conducted without the impediment of distinct alleles being nested within population subgroups, as can happen with association mapping in existing diverse germplasm panels. Finally, heterogeneous inbred families (HIFs) (Tuinstra et al. 1997) segregating at target regions in NAM or other mapping panels containing a diverse range of parents could be mined for near-isogenic pairs differing for alleles at that target region. HIFs are generally only available after mapping line development, however, and differences

in allelic contrasts among HIF pairs representing different parental combinations may be due to QTL—by—background interactions as well as allelic main effect differences.

If favorable QTN alleles can be identified, they can be selected for with diagnostic DNA marker assays in a straightforward manner, enabling breeders to more easily move them from exotic backgrounds to distantly related elite material. In the absence of knowledge of the sequence variants responsible for favorable QTL effects, however, selection for QTL alleles across unrelated populations is not expected to be effective because of genetic heterogeneity for complex traits (Holland 2007). To overcome the difficulty of genetic heterogeneity for QTL across breeding populations, industrial-scale breeding programs have implemented QTL mapping and selection within individual families. A typical breeding scheme might be as follows: (1) topcross doubled haploid (DH) lines from a breeding cross, (2) phenotype topcrosses in replicated trials in target production environment, (3) genotype the DH lines, (4) intermate selected DHs in year-round nursery, (5) repeat intermatings among individual progeny plants in year-round nurseries following seed or seedling selection for a desired marker profile, (6) create lines and topcrosses from final recurrent selection step, and (7) return lines and topcrosses to target production environment for re-evaluation. By mapping QTL for each family to be targeted for marker-based selection, the breeder is able to use QTL information directly relevant to the breeding family (Podlich et al. 2004). For example, Monsanto Co. has implemented marker-assisted recurrent selection in many breeding families, in which recurrent selection is conducted within each biparental family to increase the frequency of favorable QTL alleles mapped independently in each family (Crosbie et al. 2006; Eathington et al. 2007). This approach appears to be successful at enhancing genetic gains for yield above phenotypic selection, but the investment required in marker, breeding, winter nursery, statistical, and management infrastructure to use this form of MAS is very costly.

An alternative approach to implementing MAS within breeding families follows a similar breeding scheme, but uses genomic selection (GS) methods to create the marker-based selection index instead of QTL mapping to create the marker selection index, one can instead. Originally developed for animal breeding (Meuwissen et al. 2001), GS was introduced in the plant breeding literature in a maize breeding context similar to the one outlined above by Bernardo and Yu (2007). GS avoids the problem of distinguishing between false and true positive QTL, which underlies many of the statistical problems of QTL effect estimation and use for breeding (Beavis 1998; Schön et al. 2004), and instead fits all markers into a phenotype prediction model based on observed data. Obviously, this results in highly over-parameterized models, whose solution requires specialized statistical techniques which depend on assumptions made about the distribution of QTL effects (Bernardo and Yu 2007; Heffner et al. 2009; de los Campos et al. 2010; Lorenz et al. 2011). The distinguishing feature of all GS methods is that accurate estimation of *individual* marker effects is simply not a goal; instead, the objective is to obtain a model in which the *combined* effects of all marker loci provide accurate predictions of breeding values. GS generally provides greater response to selection than QTL-based marker-assisted recurrent selection when implemented within families (Bernardo and Yu 2007). Again, however,

the practical application of GS on a wide scale will require massive infrastructure to combine genotyping, off-season nursery management, statistical analysis, and very accurate seed and plant tracking.

Bernardo (2009) suggested that GS would be effective in exotic germplasm breeding programs in maize. He simulated GS in adapted—by—exotic populations with varying proportions of loci at which the exotic parent had the favorable allele. Phenotypic response to GS was predicted to be better in F_2-derived populations than in backcross populations because of the higher probability of recovering favorable exotic alleles. Although GS appeared to reliably increase the frequency of favorable alleles from the adapted parent, the frequency of favorable alleles from the donor parent could decrease when the donor parent had a lower frequency of favorable alleles and when favorable exotic alleles were linked to unfavorable alleles (Bernardo 2009). Unfortunately, the reality is that we expect unfavorable linkage disequilibrium and relatively low frequencies of favorable exotic alleles to be the rule rather than the exception in germplasm incorporation programs.

An alternative approach to implementing GS is to attempt to build GS prediction models based on information from diverse breeding lines, such as those that might represent an entire breeding program (Heffner et al. 2009; Albrecht et al. 2011; Crossa et al. 2010; Lorenz et al. 2011) or even global maize diversity. Initial empirical tests of the predictive accuracy of GS models suggest that they should be effective for improving selection response among and within breeding crosses of elite lines (Albrecht et al. 2011; Crossa et al. 2010; Riedelsheimer et al. 2012). Prediction accuracy of breeding values for lines not closely related to the training populations is likely to be poor (Windhausen et al. 2012), however further research is needed to clarify the potential for GS in assisting breeding progress in adapted—by—exotic cross populations (Hamblin et al. 2011). A major difficulty that will continue to impede breeding by any method in such populations is the high proportion of linkages between unique favorable alleles from the exotic parent and alleles at nearby loci that cause problems in adaptation or poor agronomic performance. The crux of the problem is that GS methods are designed to select plants with highest predicted breeding value across their genome, but unadapted germplasm will almost always tend to have poor whole-genome breeding values even when it carries unique favorable alleles at a subset of loci. Thus, for GS to be effective at both increasing the mean genetic value of a breeding population and increasing the frequency of favorable alleles derived from exotic parents, it may need to be implemented in breeding populations that have undergone several to many generations of random-mating to break up linkages between favorable and unfavorable alleles derived from the exotic parent. Developing GS models in early generations of adapted—by—exotic cross populations may simply result in selection of progenies with higher proportions of adapted alleles.

In the following sections we outline three ongoing projects of which we are aware that are attempting to incorporate advanced genomics tools and strategies to increase the ability of maize breeders to identify favorable alleles and sources of germplasm for breeding.

25.8 GEM Allelic Diversity Project

The traditional GEM Project protocol involves selection among large numbers of early generation segregants from a limited number of exotic—by—adapted breeding families each year. Progress has been made in identifying lines with superior breeding values in these crosses, but this method restricts the number of breeding families that can be tested. In essence, the GEM project has been able to sample only a limited proportion of the favorable landrace accessions (representing a total of 24 races) identified by the LAMP project, and has not explored many other landraces deemed unacceptable by LAMP (Krakowsky et al. 2008). Furthermore, since each breeding cross involves crossing a landrace with one or two proprietary inbred lines, and the proprietary inbreds differ among crosses, GEM breeding crosses are not easily amenable to genomic analysis. Finally, lines released from the GEM program must meet minimal culling criteria for topcross yield potential and agronomic performance. For breeders interested in using exotic germplasm for specialty traits unrelated to yield, the potential elimination of lines carrying unique characteristics because of poor yield and agronomic performance may be undesirable. Therefore, in addition to the traditional breeding objective of the GEM project outlined above, a more recent effort organized by the GEM project has been to sample all of the Latin American races of maize via the "GEM Allelic Diversity Project".

The Allelic Diversity project protocol involves crossing a sample of each race of the Latin American maize (about 250–300 collections in all) to each of two Pioneer Hybrid Corn Belt Dent inbred lines with expired Plant Variety Protection certificates. A small sample (3–5) of DH or selfed RILs from each cross will be created and propagated by self-fertilization if possible. The only selection criterion will be the line's capacity to reproduce itself; otherwise, these will represent random, unselected lines from each cross. The resulting set of $\sim$ 1,500 lines should represent the widest sample yet available of maize allelic diversity in common adapted genetic backgrounds and will serve as an excellent platform for allele mining (Krakowsky et al. 2008).

25.9 Seeds of Discovery—Large-scale Genotyping and Phenotyping of CIMMYT Germplasm

An initiative known as Seeds of Discovery funded by the Mexican Secretaria de Agricultura, Ganaderia, Desarrollo Rural, Pesca, y Alimentacion (SAGARPA) is being conducted collaboratively through CIMMYT with a number of Mexican Institutions. The objective of the Seeds of Discovery Initiative is to provide information on maize genetic resources that will facilitate their use by maize breeders in the developing world. A number of components are involved in this process: (1) genotyping all maize collections within Mexican Germplasm banks, (2) phenotypic characterization of Mexican germplasm bank accessions, (3) identification of haplotypes or alleles with effects on specific characteristics, (4) estimation of haplotype or allele

frequencies among and between representative accessions of races, (5) the creation of elite bridge lines carrying specific alleles or haplotype regions, and (6) formation of a web portal to deliver this information to breeders worldwide.

Priorities for phenotyping are related to climate change and delivering new alleles to breeders that will promote food security worldwide. These priorities are: drought and heat tolerance, resistance to diseases with expanding ranges due to climatic changes, and nitrogen and phosphorus use efficiency (Collins et al. 2008; Tuberosa 2012). Quality parameters for human consumption will also be included in the phenotypic characterization, in particular those characteristics most important for human consumption in México. Within México there is a strong desire to provide self-sufficiency in maize production. Half of the land devoted to maize production in México is planted to native landraces. To increase production in México, yields must increase both in areas devoted to improved maize cultivars (mostly hybrids) as well as those devoted to landraces. This can be accomplished by either displacing landraces with hybrids or by increasing yields of landraces, but the latter approach is preferred to maintain the genetic and cultural diversity of maize in México. A major objective of this project is to evaluate of the ability of landraces to produce a crop under suboptimal conditions while recognizing the specific culinary properties of diverse maize landraces. As the center of origin of maize, the continued use of landraces in México is important for the preservation of maize diversity as a world resource.

Phenotyping of a large GWAS experiment is currently underway. This initial experiment aims to identify favorable alleles for complex traits harbored in the 4,000 accessions of CIMMYT's breeders' core collection. One plant from each of the 4,000 accessions was used to pollinate a CIMMYT hybrid tester, and DNA was isolated from each landrace plant sampled. Genotyping by sequencing (Elshire et al. 2011) will be performed on the one plant per accession used as the male parent of each topcross. Topcross entries were assigned to sets of about 600 entries each according to the origin of the accession (lowland, subtropical or highland) to target entries to appropriate environments and to accommodate limitations of phenotyping capacity of collaborators. Although the design is not balanced, partial balance was achieved by including multiple sets (up to 2,200 topcross entries) in several environments, by including repeated check cultivars within and across environments, and by ensuring that at least 10 % of the accession topcross entries were planted at multiple locations. Testing locations used in the first season were a combination of Instituto National de Investigaciones Forestales, Agricolas Y Pecuaria (INIFAP) stations, INIFAP managed farmer's fields, Mexican University Field Stations, CIMMYT experiment stations, and one farmer's field managed by Pioneer Hybrid. The first season's evaluation will emphasize disease reaction, low nitrogen tolerance and agronomic traits. A follow-up evaluation in the second season will emphasize drought resistance, and quality for human consumption. This GWAS experiment will allow estimation of haplotype effects in topcrosses in target environments in México, serve as a training population for GS, and provide an estimate of how haplotype frequencies are distributed across germplasm groups.

Another component of the Seeds of Discovery project is to provide funding and coordinate projects proposed by Mexican institutions. In parallel with the topcross GWAS experiment, INIFAP is leading a large project to genotype and evaluate *per se* performance phenotypes of their most recent germplasm collection. These 6,000 were collected between 2008 and 2010 in México and are associated with accurate passport data, including GPS coordinates of collection sites. The goal is to produce *per se* phenotypic data that can be correlated with genotypic data at an allele frequency level. A first step is to determine through GIS data the most representative 1,200 accessions (20 %) to represent the agroclimatic diversity of México (Ruiz et al. 2008). The project is currently producing full-sib families from 30 plants per accession which have been individually sampled for DNA extraction. The hope is to estimate allele frequencies within 12 full sib families representing 24 individuals per accession. The 12 full sib families per accession will be phenotyped at multiple sites within their area of adaptation (e.g., tropical, subtropical, or highland environments). The phenotyping will be based on priorities within the adaptation zone. For example, drought tolerance and ear rot resistances are a priority for all target environments, but heat tolerance is a priority specifically for subtropical and tropical environments.

The collection will also be used to evaluate germplasm for specific culinary uses important in Mexican culture and food markets. Those uses which have the greatest added value for small farmers are a priority, such as pozole (hominy), elotes (sweet corn or green ears), and totomoxtle (husks for wrapping tamales). In addition, specific kernel quality characteristics for tortilla production will be evaluated.

Diversity of agroclimates and landraces make logistical considerations for phenotypic evaluation a challenge for this project. The GWAS experiment consists of maize accessions adapted to tropical, subtropical, and highland environments as well as temperate materials from South America. Phenotypic evaluation of these divergent materials cannot be conducted in a single common environment, and, further, the number of entries to be tested exceeds the phenotyping capacity of most cooperators. Therefore, the entries were partitioned into sets of materials based on agroclimate and maturity with 10 % overlapping entries in order to accommodate the logistical constraints of collaborators. The use of repeated commercial checks which have wide adaptation within México as well as use of repeated entries overlapping sites will permit combined statistical analysis of all environments. For specific disease hotspots we are using farmers' fields with high incidence of disease pressure. These fields are rainfed which presents specific constraints; rainfed conditions are particularly difficult to manage because planting occurs after the raining season starts, and the initiation of the rainy season in México has become unpredictable in recent years in México with rains starting a month or more later than historically expected. After the rains initiate, land must be prepared in a window of a few days when the fields are dry enough to enter with a tractor, but this is also unpredictable.

Additional logistical constraints in these very large-scale experiments, in addition to appropriate site selection and management, include sample tracking and environmental characterizations. Handheld data collection computers enabled with bar code readers will facilitate accurate data collection. Weather stations at each site and soil characterization will also be used to characterize the environmental factors related

to genotype-by-environment interactions. Investments for improving infrastructure at collaborator sites within México are underway to provide multiple sites with capability to conduct managed drought and heat screenings, as well as to improve seed storage capabilities and seed tracking management systems that will be necessary for the quantity of material to be evaluated during this initiative. Joint training of personnel between CIMMYT and Mexican Institutions is also an important part of this project, particularly the data management and bioinformatics portion of the project.

25.10 Bridging the Domestication Bottleneck with Teosinte Introgression Libraries

While maize inbreds and landraces contain an incredible amount of genetic diversity relative to other crop species, teosinte contains even more diversity than landraces and inbreds. Various population genetics studies indicate that maize inbreds retain approximately 60 % of the variation present in teosinte (Wright et al. 2005), and approximately 80 % of the variation present in landraces (Tenaillon et al. 2001). All genes across the genome experienced the domestication and/or breeding bottlenecks, resulting in moderate reductions in variation in maize relative to teosinte (Tenaillon et al. 2004). However, genes targeted by artificial selection during domestication and/or improvement have greatly reduced variation, as the combined effect of the bottleneck and selection is much more severe (Innan and Kim 2004). Thus teosinte should harbor more diversity for all genes compared to maize, and much more diversity for those genes that were targets of selection during domestication and/or plant breeding.

A population genetics study involving large-scale resequencing in maize revealed that 2–4 % of maize genes were targets of artificial selection during domestication and/or plant breeding (Wright et al. 2005). It is currently unknown what proportion of these selected genes were targets of selection during domestication (diversity lost between teosinte and landraces) versus improvement (diversity lost between landraces and inbred lines), or both. However, the implication of this study is striking. When considering the conservatively estimated filtered gene set of 32,690 genes (Schnable et al. 2009) or the more liberal estimate of 59,000 genes in maize (Messing et al. 2004), this implies that between 650 to 1,200 maize genes have experienced artificial selection, and have little or no sequence diversity in modern diverse inbreds, although they do in teosinte. Whereas these $\sim$1,000 genes appear to have been under strong selection during the domestication and breeding of maize, this does not necessarily imply that the allele fixed in maize is the optimal allele for all modern environments and production systems. Furthermore, it is possible that suboptimal alleles were fixed in maize due to hitchhiking by tight linkage with a favorable allele at a nearby locus under selection (Tenaillon et al. 2002).

A large number of teosinte accessions can be obtained from either the USDA Plant Introduction Station in Ames, Iowa or the CIMMYT germplasm bank in México. However, direct comparison of maize to teosinte *per se* for any given trait is not

appropriate, as many of the undesirable teosinte traits (photoperiod sensitivity, incongruous plant architecture, lack of a true ear, the hard seed coat around the seed) mask potentially useful traits. Hence, teosinte must be crossed with maize to create germplasm the can be compared more equitably to maize. To this end, a set of introgression lines (ILs) is being developed from 10 *parviglumis* accessions in the B73 background.

Maize and *parviglumis* readily hybridize, both in the wild (Ellstrand et al. 2007) and in the nursery, given the proper conditions. As a short day plant, teosinte flowering is delayed under the long photoperiods of temperate US locations, and the first frost usually occurs prior to teosinte flowering. However, most teosintes can be induced to flower under short day conditions (Emerson 1924), and tassels can be observed in *parviglumis* within six weeks when grown in a day-neutral winter nursery site or growth chamber under short day conditions. When the objective is to a make large numbers of crosses, it is easiest to conduct initial crosses involving teosinte in a winter nursery setting.

Using teosinte as the pollen parent in controlled pollinations is significantly easier than as the female parent for several reasons. Shoot-bagging teosinte is very difficult as silks often emerge from the axil prior to ear shoot appearance. There is also a potential for gametophyte factors to discriminate against or exclude pollen not carrying the same allele, although this is mostly a problem with the sister subspecies *Zea mays* ssp. *mexicana* (Kermicle and Allen 1990; Nelson 1994). Finally, a successful pollination using teosinte as the female would result in an ear with only 5–12 seeds, thus requiring significantly more work to generate large numbers of progeny.

The F_1 hybrids are sometimes still photoperiod sensitive, flowering around September 15 in Missouri and resulting in highly tillered plants with long lateral branches (although see Rogers 1950). Again, these symptoms appear to be alleviated in a short day environment. However, beginning with BC_1 plants, the process of backcrossing in a temperate environment becomes much easier (Fig. 25.3). The ultimate goal of the project is to produce BC4 derived ILs, with the expected amount of teosinte being 3–5 % per line.

There are several ways that these introgression libraries will be used, and the applications described herein are interrelated. A very basic application is to explore empirical questions related to the processes of domestication and artificial selection. As described above, approximately 1,000 genes were targets of selection. Which genes are they and what are their functions? What traits were targeted by artificial selection during domestication/breeding? Are these selected genes relevant to agriculture today? An excellent example concerns a selected gene (AY104948) that has homology to the Arabidopsis *Auxin response factor1* (*ARF1*), a transcription factor with a putative function in plant growth. *ARF1* has very high levels of sequence diversity in teosinte, but almost no sequence diversity in maize inbreds (Wright et al. 2005). Auxin is clearly involved in apical dominance in plants, so it is possible that *ARF1* acts in a manner similar to *teosinte branched1* (Doebley et al. 1995). If so, we can postulate a corresponding phenotypic effect of the teosinte allele of *ARF1* in a maize background, such as increased tillering, increased lateral branching, and/or increased number of ears. However, preliminary studies of the teosinte introgression

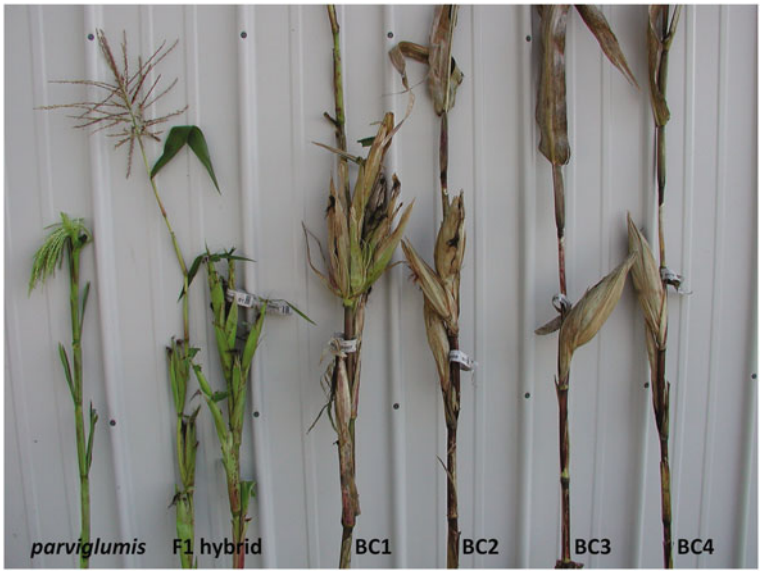

Fig. 25.3 Regression of progeny phenotypes to the B73 recurrent parent under repeated backcrossing of a maize-teosinte hybrid to the maize parent. Note, the ear heights shown not do not represent the ear height on the plant. (Photographs by Sherry Flint-Garcia)

lines do not show an effect of *ARF1* on any of these traits. A comprehensive analysis of each of the ~ 1,000 selected genes is needed, and the teosinte introgression lines will play a vital role in testing hypotheses.

A second application is to evaluate and compare the range of allelic effects of teosinte to those of maize. Recent studies of the NAM population reveal that allele effect series are prevalent; in many cases a QTL segregates in multiple NAM families, but the direction and/or magnitude of the allelic effect varies across the NAM founders (Buckler et al. 2009; Kump et al. 2011; Poland et al. 2011; Tian et al. 2011). For example, for flowering time, variants at different sequence positions in *vgt1* result in opposite effects: a MITE insertion in *vgt1* is responsible for the early flowering Northern Flint allele (Salvi et al. 2007) and SNPs in its target gene, *rap2.7*, are likely responsible for a late flowering tropical allele (Buckler et al. 2009). Because *parviglumis* harbors many unselected, often deleterious, alleles that have not been purged by domestication and improvement, it will likely contain alleles with opposing effects as compared to maize. Furthermore, we postulate that a loss of genetic variation across the genome during domestication and/or breeding results in a loss of phenotypic variation, and therefore reintroduction of variation from teosinte will result in greater phenotypic variation. Following this logic, we hypothesize that teosinte harbors stronger alleles for any given QTL than maize. These stronger-effect teosinte alleles may be useful for genetic studies, such as in QTL fine mapping experiments as discussed above, or in physiological studies, where the objective is not necessarily to improve maize but rather to understand the genetic and/or physiological basis of complex traits.

A third application is the use of teosinte allelic variation for trait improvement. A more directed approach is to identify pathways controlling the trait of interest, and reintroduce variation from teosinte for genes involved in the pathway. For example, three genes in the starch pathway show signatures of past selection: the small subunit of ADP-glucose pyrophosphorylase encoded by *brittle2*, the starch branching enzyme encoded by *amylose extender1*, and the debranching enzyme encoded by *sugary1* (Whitt et al. 2002). Restoration of allelic variation from teosinte for these three selected genes could result in increased kernel starch content or alternate forms of starch that may be useful as specialty industrial starches and healthy, resistant (slow-degrading) starches. A second approach is more trait-focused, where the genes controlling the trait are perhaps unknown, but where teosinte shows greater trait variation than maize. For example, teosinte seeds contain twice the kernel protein content and novel zein proteins as compared to maize (Flint-Garcia et al. 2009a), as well as altered amino acid content (Flint-Garcia et al. 2009b). We hypothesize that variation from teosinte can be used to increase protein content and improve protein quality of maize. Indeed, an independent group of researchers has demonstrated that alien introgression lines of *Zea mays* ssp. *mexicana* have increased yield, protein content, and essential amino acid content compared to control lines (Wang et al. 2008a; Wang et al. 2008b).

Some have made the argument that the "best" alleles were already selected during domestication, and that reintroducing variation from teosinte would reverse human efforts over the last 9,000 years. In a few select cases this is true. Understandably, we do not want to reintroduce *tga1* alleles that confer the stony fruit case that surrounds the teosinte seed (Dorweiler et al. 1993). However, domestication occurred in a very different environment and under very different cultural practices than the USA Corn Belt. If maize were domesticated from teosinte in a temperate environment under modern agricultural practices then alternate alleles may well have been selected for many traits. We can capitalize on the incredible amount of diversity in teosinte to search for valuable alleles to aid in scientific discovery and continued corn improvement.

Acknowledgments Research by SS, SF-G, and JBH is supported by US National Science Foundation (DBI-0321467 and IOS-0820619). We thank Drs. Jesús Sánchez-Gonzalez (University of Guadalajara) and Major M. Goodman (North Carolina State University) for the ears and photographs used in Fig. 25.1.

References

Abecasis G, Cardon L, Cookson W (2000) A general test of association for quantitative traits in nuclear families. Am J Human Genet 66:279–292

Albrecht T, Wimmer V, Auinger H-J et al (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Andersen JR, Schrag T, Melchinger AE et al (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

Anderson E (1944) The sources of effective germplasm in hybrid maize. Ann MO Bot Gard 31:355–361

Anderson E, Cutler H (1942) Races of *Zea mays*: I. Their recognition and classification. Ann MO Bot Gard 29:69–88

Atwell S, Huang YS, Vilhjalmsson BJ et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627–631

Balint-Kurti PJ, Blanco M, Millard M et al (2006) Registration of 20 GEM maize breeding germplasm lines adapted to the southern USA. Crop Sci 46:996–998

Balint-Kurti PJ, Zwonitzer JC, Wisser RJ et al (2007) Precise mapping of quantitative trait loci for resistance to southern leaf blight, caused by *Cochliobolus heterostrophus* race O, and flowering time using advanced intercross maize lines. Genetics 176:645–657

Beavis WD (1998) QTL analyses: Power, precision, and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, pp 145–162

Benjamini Y, Yekutieli D (2005) Quantitative trait loci analysis using the false discovery rate. Genetics 171:783–789

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

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49:419–425

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

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

Brachi B, Faure N, Horton M et al (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet 6:e1000940

Breseghello F, Sorrells ME (2006a) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci 46:1323–1330

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

Bretting PK, Goodman MM, Stuber CW (1987) Karyological and isozyme variation in West Indian and allied American mainland races of maize. Am J Bot 74:1601–1613

Bretting PK, Goodman MM, Stuber CW (1990) Isozymatic variation in Guatemalan races of maize. Am J Bot 77:211–225

Brown W (1953) Sources of germ plasm for hybrid corn. 8th Hybrid Corn Industry—Research Conference, pp 11–16

Brown WL (1975) A broader germplasm base in corn and Sorghum. 30th Annual Corn and Sorghum Research Conference, pp 81–89

Buckler ES, Holland JB, McMullen MM et al (2009) The genetic architecture of maize flowering time. Science 325:714

Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107–111

Cahill DJ, Schmidt DH (2004) Use of marker assisted selection in a product development breeding program. In: Fischer T, Turner N, Angus J, McIntyre L, Robertson M, Borrell A, Lloyd D (eds) New directions for a diverse planet: Proc 4th Int Crop Sci Congress, Brisbane, Australia

Camus-Kulandaivelu L, Veyrieras JB, Madur D et al (2006) Maize adaptation to temperate climate: Relationship between population structure and polymorphism in the *Dwarf8* gene. Genetics 172:2449–2463

Castillo-Gonzalez F, Goodman MM (1989) Agronomic evaluation of Latin American maize accessions. Crop Sci 29:853–861

Chardon F, Virlon B, Moreau L et al (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168:2169–2185

Chia JM, Song C, Bradbury PJ et al (2012) Maize hapmap 2 identifies extant variation from a genome in flux. Nat Genet 44:803–807
Ching A, Caldwell KS, Jung M et al (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3
Clark RM, Schweikert G, Toomajian C et al (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science 317:338–342
Clark RM, Wagler TN, Quijada P, Doebley J (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, Jones H, Leigh FJ et al (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. J Exp Bot 58:1231–1244
Coles ND, McMullen MD, Balint-Kurti PJ et al (2010) Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. Genetics 184:799–812
Coles ND, Zila CT, Holland JB (2011) Allelic effect variation at key photoperiod response QTL in maize. Crop Sci 51:1036–1049
Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos T Roy Soc B 363:557–572
Collins NC, Tardieu F, Tuberosa R (2008) QTL approaches for improving crop performance under abiotic stress conditions: where do we stand? Plant Physiol 147:469–486
Corder EH, Saunders AM, Risch NJ et al (1994) Protective effect of apolipoprotein-E type-2 allele for late-onset Alzheimer disease. Nat Genet 7:180–184
Crawford GW, Saunders D, Smith DG (2006) Pre-contact maize from Ontario, Canada: Context, chronology, variation, and plant association. In: Staller J, Tykot R, Benz B (eds) Histories of maize: multidisciplinary approaches to the prehistory, linguistics, biogeography, domestication, and evolution of maize. Academic Press, Burlington, pp 549–559
Crosbie TM, Eathington SR, Johnson GR et al (2006) Plant breeding: past, present, and future. In: Lamkey KR, Lee M (eds) Plant breeding: The Arnel R Hallauer International Symposium. Blackwell, Ames, pp 3–50
Crosby A (1972) The Columbian exchange: biological and cultural consequences of 1492. Greenwood, Westport, CT
Crossa J, de los Campos G, Perez P et al (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–U406
de los Campos G, Gianola D, Rosa GJM et al (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genet Res 92:295–308
Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 5:997–1004
Doebley J (2004) The genetics of maize evolution. Ann Rev Genet 38:37–59
Doebley J, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321
Doebley J, Stec A, Gustus C (1995) *tesosinte branched* and the origin of maize: Evidence for epistasis and the evolution of dominance. Genetics 141:333–346
Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488
Doebley J, Wendel JD, Smith JSC et al (1988) The origin of Cornbelt maize: the isozyme evidence. Econ Bot 42:120–131
Doebley JF, Goodman MM, Stuber CW (1984) Isoenzymatic variation in *Zea* (*gramineae)*. Syst Bot 9:204–218
Doebley JF, Goodman MM, Stuber CW (1985) Isozyme variation in the races of maize from Mexico. Am J Bot 72:629–639
Dorweiler J, Stec A, Kermicle J, Doebley J (1993) *Teosinte-Glume-Architecture-1*– a genetic locus controlling a key step in maize evolution. Science 262:233–235
Dubcovsky J (2004) Marker-assisted selection in public breeding programs: The wheat experience. Crop Sci 44:1895–1898

Dubreuil P, Warburton M, Chastanet M et al (2006) More on the introduction of temperate maize into Europe: Large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291
Duvick DN, Smith JSC, Cooper M (2004) Changes in performance, parentage, and genetic diversity of successful corn hybrids, 1930–2000. In: Smith CW, Betran FJ, Runge ECA (eds) Corn: origin, history, technology, and production. Wiley, New York, pp 65–97
Eathington SR, Crosbie TM, Edwards MD et al (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S-154–163
Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution, and types of gene action. Genetics 116:113–125
Ehrenreich IM, Hanzawa Y, Chou L et al (2009) Candidate gene association mapping of Arabidopsis flowering time. Genetics 183:325–335
Eichten SR, Foerster JM, de Leon N et al (2011) B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. Plant Physiol 156:1679–1690
Ellstrand NC, Garner LC, Hegde S et al (2007) Spontaneous hybridization between maize and teosinte. J Hered 98:183–187
Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. PLoS One 6:e19379
Emerson RA (1924) Control of flowering in teosinte. Short-day treatment brings early flowers. J. Hered. 15:41–48
Falke KC, Melchinger AE, Flachenecker C et al (2006) Comparison of linkage maps from F2 and three times intermated generations in two populations of European flint maize (*Zea mays* L.). Theor Appl Genet 113:857–866
Falque M, Decousset L, Dervins D et al (2005) Linkage mapping of 1454 new maize candidate gene loci. Genetics 170:1957–1966
Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrum in plants. Annu Rev Plant Biol 54:357–374
Flint-Garcia SA, Thuillet AC, Yu J et al (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant Journal 44:1054–1064
Flint-Garcia SA, Bodnar AL, Scott MP (2009a) Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. Theor Appl Genet 119:1129–1142
Flint-Garcia SA, Guill KE, Sanchez-Villeda H et al (2009b) Maize amino acid pathways maintain high levels of genetic diversity. Maydica 54:375–386
Frary A, Nesbitt TC, Frary A et al (2000) *fw*2.2: A quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–87
Frey TJ (2006) Fine mapping, cloning, verification, and fitness evaluation of a QTL, Rcg1, which confers resistance to *Colletotrichum graminicola* in maize. Ph.D. Thesis. Dep Plant and Soil Sciences. Univ. Delaware, Newark, DE
Frey TJ, Weldekidan T, Colbert T et al (2011) Fitness evaluation of *Rcg1*, a locus that confers resistance to *Colletotrichum graminicola* (Ces.) GW Wils. using near-isogenic maize hybrids. Crop Sci 51:1551–1563
Fridman E, Carrari F, Liu Y-S et al (2004) Zooming in on a quantitative trait for the tomato yield using interspecific introgressions. Science 305:1786–1789
Fu Y, Wen TJ, Ronin YI et al (2006) Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize. Genetics 174:1671–1683
Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506
Geadelmann JL (1984) Using exotic germplasm to improve northern corn. 39th Annual Corn & Sorghum Research Conference, pp 98–110
Gerrish EE (1983) Indications from a diallel study for interracial maize hybridization in the Corn Belt [Central USA]. Crop Sci 23:1082–1084

Goodman MM (1983) Racial diversity in maize. In: Williams LE, Gordon DT, Nault LR (eds) International Maize Virus Disease Colloquium and Workshop. Ohio Agricultural Research and Development Center, Wooster, pp 29–40
Goodman MM (1985) Exotic maize germplasm: Status, prospects, and remedies. Iowa State J Res 59:497–527
Goodman MM (1992) Choosing and using tropical corn germplasm. 47th Annual Corn & Sorghum Research Conference. Am. Seed Trade Assoc., Washington, DC, pp 47–64
Goodman MM (2004) Developing temperate inbreds using tropical maize germplasm: Rationale, results, conclusions. Maydica 49:209–219
Goodman MM, Brown WL (1988) Races of corn. In: Sprague GF, Dudley JW (eds) Corn and corn improvement. Am Soc Agron, Madison, pp 33–79
Goodman MM, Moreno J, Castillo F et al (2000) Using tropical maize germplasm for temperate breeding. Maydica 45:221–234
Goodman MM, Stuber CW (1983) Races of maize. VI. Isozyme variation among races of maize in Bolivia [*Zea mays*, corn]. Maydica 28:169–187
Gore MA, Chia JM, Elshire RJ et al (2009) A first-generation haplotype map of maize. Science 326:1115–1117
Haley C (2011) A cornucopia of maize genes. Nat Genet 43:87–88
Hallauer AR (1978) Potential of exotic germplasm for maize improvement. In: Walden DB (ed) Maize breeding and genetics. Wiley, New York, pp 229–247
Hamblin MT, Buckler ES, Jannink J-L (2011) Population genetics of genomics-based crop improvement methods. Trends Genet 27:98–106
Hansey CN, Johnson JM, Sekhon RS et al (2011) Genetic diversity of a maize association population with restricted phenology. Crop Sci 51:704–715
Harjes CE, Rocheford TR, Bai L et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319:330–333
Hauser MT, Harr B, Schlotterer C (2001) Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: Molecular analysis of the candidate gene *GLABROUS1*. Mol Biol Evol 18:1754–1763
Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12
Hernández E (1985) Maize and man in the Greater Southwest. Econ Bot 39:416–430
Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities. In: Fischer T, Turner N, Angus J, McIntyre L, Robertson M, Borrell A, Lloyd D (eds) New directions for a diverse planet: Proc 4th Int Crop Sci Congress, Brisbane, Australia
Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161
Holland JB, Goodman MM (1995) Combining ability of tropical maize accessions with U.S. germplasm. Crop Sci 35:767–773
Holland JB, Nelson PT (2010) Dedication: Major M. Goodman: Maize Geneticist and Breeder. Plant Breed Rev. Wiley, pp 1–29
Holley RN, Goodman MM (1988) Yield potential of tropical hybrid maize derivatives. Crop Sci 28:213–218
Huang X, Wei X, Sang T et al (2010a) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967
Huang Y-F, Madur D, Combes V et al (2010b) The genetic architecture of grain yield and related traits in *Zea maize* L. revealed by comparing intermated and conventional populations. Genetics 186:395–404
Hung HY, Shannon LM, Tian F et al (2012) ZmCCT and the genetic basis of day-length adaptation underlying the post-domestication spread of maize. Proc Natl Acad U S A 109:E1913–1921
Hyten DL, Choi IY, Song QJ et al (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937–1944

Innan H, Kim Y (2004) Pattern of polymorphism after strong artificial selection in a domestication event. Proc Nat Acad U S A 101:10667–10672
Izawa T (2007) Adaptation of flowering-time by natural and artificial selection in Arabidopsis and rice. J Exp Bot 58:3091–3097
Jena KK, Mackill DJ (2008) Molecular markers and their use in marker-assisted selection in rice. Crop Sci 48:1266–1276
Kang HM, Zaitlen NA, Wade CM et al (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723
Kerem BS, Rommens JM, Buchanan JA et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080
Kermicle JL, Allen JO (1990) Cross-incompatibility between maize and teosinte. Maydica 35:399–408
Krakowsky MD, Holley R, Deutsch JA et al (2008) Maize allelic diversity project. 50th Maize Genetics Conference, Washington, DC
Kuleshov NN (1933) World diversity of phenotypes of maize. J Amer Soc Agron 25:688–700
Kump KL, Bradbury PJ, Buckler ES et al (2011) Genome-wide association study of quantitative resistance to Southern leaf blight in the maize nested association mapping population. Nat Genet 43:163–168
Kump KL, Holland JB, Jung MT et al (2010) Joint analysis of near-isogenic and recombinant inbred line populations yields precise positional estimates for quantitative trait loci. Plant Genome 3:142–153
Lai JS, Li RQ, Xu X et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42:1027–U1158
Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756
Laurie CC, Chasalow SD, Ledeaux JR et al (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. Genetics 168:2141–2155
Lauter N, Moscou MJ, Habiger J, Moose SP (2008) Quantitative genetic dissection of shoot architecture traits in maize: towards a functional genomics approach. Plant Genome 1:99–110
Lee M, Sharopova N, Beavis WD et al (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (*IBM*) population. Plant Mol Biol 48:453–461
Lippert C, Listgarten J, Liu Y et al (2011) FaST linear mixed models for genome-wide association studies. Nat Meth 8:833–835
Lorenz AJ, Chao SM, Asoro FG et al (2011) Genomic selection in plant breeding: Knowledge and prospects. Adv Agron 110:77–123
Mackay TFC (2001) The genetic architecture of quantitative traits. Ann Rev Genet 35:303–309
Mangelsdorf PC (1974) Corn: its origin, evolution, and improvement. Belknap Press of Harvard University Press, Cambridge
Matsuoka Y, Vigouroux Y, Goodman MM et al (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Nat Acad U S A 99:6080–6084
McMullen MD, Kresovich S, Sanchez Villeda H et al (2009) Genetic properties of the maize Nested Association Mapping population. Science 325:737–740
Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveal low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403
Messing J, Bharti AK, Karlowski WM et al (2004) Sequence composition and genome organization of maize. Proc Nat Acad U S A 101:14349–14354
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
Monforte AJ, Tanksley SD (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristis and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor Appl Genet 100:471–479

Nelson OE (1994) The gametophyte factors of maize. In: Freeling M, Walbot V (eds) The maize handbook. Springer-Verlag, New York, pp 496–503
Nelson PT, Coles ND, Holland JB et al (2008) Molecular characterization of maize inbreds with expired U.S. plant variety protection. Crop Sci 48:1673–1685
Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795–1806
Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. Trends Genet 19:303–306
Paterson AH, Lander ES, Hewitt JD et al (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726
Pennisi E (2008) Plant sciences—Corn genomics pops wide open. Science 319:1333–1333
Piperno DR, Ranere AJ, Holst I et al (2009) Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, México. Proc Natl Acad Sci U S A 106:5019–5024
Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go. An effective approach for marker-assisted selection of complex traits. Crop Sci 44:1560–1571
Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Nat Acad U S A 108:6893–6898
Pollak LM (2003) The history and success of the public-private project on germplasm enhancement of maize (GEM). Adv Agron 78:45–87
Pressoir G, Berthaud J (2004a) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92:88–94
Pressoir G, Berthaud J (2004b) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. Heredity 92:95–101
Price AH (2006) Believe it or not, QTLs are accurate!. Trends Plant Sci 11:213–216
Pritchard JK (2001) Deconstructing maize population structure. Nat Genet 28:203–204
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Pumphrey MO, Bernardo R, Anderson JA (2007) Validating the QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Sci 47:200–206
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100
Ramsay L, Comadran J, Druka A et al (2011) *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene T*EOSINTE BRANCHED 1*. Nat Genet 43:169–172
Rebai A, Blanchard P, Perret D, Vincourt P (1997) Mapping quantitative trait loci controlling silking date in a diallel cross among four lines of maize. Theor Appl Genet 95:451–459
Rebourg C, Chastanet M, Gouesnard B et al (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903
Reif J, Warburton M, Xia X et al (2006) Grouping of accessions of Mexican races of maize revisited with SSR markers. Theor Appl Genet 113:177–185
Remington DL, Thornsberry JM, Matsuoka Y et al (2001) Structure of linkage disequilibrium and phenotypic assosiations in the maize genome. Proc Nat Acad U S A 98:11479–11484
Riedelsheimer C, Czedik-Eysenberg A, Grieder C et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220
Rincón SF, Castillo GF, Ruiz T NA (2010) Diversidad y distibución de los maíces nativos en Coahuila, México. SOMEFI, Chapingo, México
Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517
Rodriguez VM, Butron A, Malvar RA et al (2008) Quantitative trait loci for cold tolerance in the maize IBM population. Int J Plant Sci 169:551–556

Rogers JS (1950) The inheritance of photoperiodic response and tillering in maize-teosinte hybrids. Genetics 35:513–540
Ron Parra J, Sánchez-González JJ, Jiménez-Cordero AA et al (2006) Maíces nativos del Occidente de México I. Colectas 2004. Scientia-CUCBA 8:1–139
Rostoks N, Ramsay L, MacKenzie K et al (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Nat Acad U S A 103:18656–18661
Ruiz C JA, Puga ND, Sánchez G JJ et al (2008) Climatic adaptation and ecological descriptors of 42 Mexican maize races. Crop Sci 48:1502–1512
Salhuana W, Pollak LM, Ferrer M et al (1998) Breeding potential of maize accessions from Argentina, Chile, USA, and Uruguay. Crop Sci 38:866–872
Salvi S, Sponza G, Morgante M et al (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc Nat Acad U S A 104:11376–11381
Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304
Sanchez G JJ, Goodman MM (1992a) Relationships among Mexican and some North American and South American races of maize. Maydica 37:41–51
Sanchez G JJ, Goodman MM (1992b) Relationships among the Mexican races of maize. Econ Bot 46:72–85
Sanchez G JJ, Goodman MM, Stuber CW (2000a) Isozymatic and morphological diversity in the races of maize of Mexico. Econ Bot 54:43–59
Sanchez G JJ, Stuber CW, Goodman MM (2000b) Isozymatic diversity in the races of maize of the Americas. Maydica 45:185–203
Sanchez G JJ, Goodman MM, Bird RMK, Stuber CW (2006) Isozyme and morphological variation in maize of five Andean countries. Maydica 51:25–42
Sanchez G JJ, Goodman MM, Stuber CW (2007) Racial diversity of maize in Brazil and adjacent areas. Maydica 52:13–30
Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112–1115
Schön CC, Utz HF, Groh S et al (2004) Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485–498
Sharopova N, McMullen MD, Schultz L et al (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481
Smith BD (1989) Origins of agriculture in eastern North America. Science 246:1566–1571
Smith JSC, Smith OS, Wright S et al (1992) Diversity of U.S. hybrid maize germplasm as revealed by restriction fragment length polymorphisms. Crop Sci 32:598–604
Stuber CW, Edwards MD (1986) Genotypic selection for improvement of quantitative traits in corn using molecular marker loci. Proc 41st Ann Corn & Sorghum Res Conf. American Seed Trade Association, Chicago, IL, pp 70–83
Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat Genet 43:1160–1163
Studer AJ, Doebley JF (2011) Do large effect QTL fractionate? A case study at the maize domestication QTL *teosinte branched1*. Genetics 188:673–681
Sturtevant E (1899) Varieties of corn. US Dep Agr Off Exp Sta Bul 57
Swanson-Wagner RA, Eichten SR, Kumari S et al (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20:1689–1699
Szalma S, Buckler E, Snook M, McMullen M (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor Appl Genet 110:1324–1333
Tallury SP, Goodman MM (1999) Experimental evaluation of the potential of tropical germplasm for temperate maize improvement. Theor Appl Genet 98:54–61
Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063–1066

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tarter JA, Goodman MM, Holland JB (2003) Testcross performance of semiexotic inbred lines derived from Latin American maize accessions. Crop Sci 43:2272–2278

Tenaillon MI, Sawkins MC, Long AD et al (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Nat Acad U S A 98:9161–9166

Tenaillon MI, Sawkins MC, Anderson LK et al (2002) Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Genetics 162:1401–1413

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 21:1214–1225

Thornsberry JM, Goodman MM, Doebley J et al (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Tian F, Bradbury PJ, Brown PJ et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

Troyer AF (1999) Background of U.S. hybrid corn. Crop Sci 39:601–626

Tuberosa R (2012) Phenotyping for drought tolerance of crops in the genomics era. Frontiers in. Plant Physiol 3(347):1–25

Tuberosa R, Salvi S (2009) QTL for agronomic traits in maize production. In: Bennetzen JL, Hake SC (eds) Handbook of maize: its biology. Springer, New York, pp 501–541

Tuinstra MR, Ejeta G, Goldsbrough PB (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95:1005–1011

Uhr DV, Goodman MM (1995a) Temperate maize inbreds derived from tropical germplasm: I. Testcross yield trials. Crop Sci 35:779–784

Uhr DV, Goodman MM (1995b) Temperate maize inbreds derived from tropical germplasm: II. Inbred yield trials. Crop Sci 35:785–790

Van Heerwaarden J, Doebley J, Briggs WH et al (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci USA 108:1088–1092

Venuprasad R, Bool M, Quiatchon L, Atlin G (2011) A QTL for rice grain yield in aerobic environments with large effects in three genetic backgrounds. Theor Appl Genet 1–10

Verhoeven KJF, Jannink JL, McIntyre LM (2006) Using mating designs to uncover QTL and the genetic architecture of complex traits. Heredity 96:139–149

Vigouroux Y, Glaubitz JC, Matsuoka Y et al (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95:1240–1253

Wang H, Nussbaum-Wagler T, Li BL et al (2005) The origin of the naked grains of maize. Nature 436:714–719

Wang LZ, Xu CZ, Qu ML, Zhang JR (2008a) Kernel amino acid composition and protein content of introgression lines from *Zea mays* ssp *mexicana* into cultivated maize. J Cereal Sci 48:387–393

Wang LZ, Yang AF, He CM et al (2008b) Creation of new maize germplasm using alien introgression from *Zea mays* ssp *mexicana*. Euphytica 164:789–801

Wang R-L, Stec A, Hey J et al (1999) The limits of selection during maize domestication. Nature 398:236–239

Weatherwax P (1954) Indian corn in old America. McMillan, New York

Whitt SR, Wilson LM, Tenaillon MI et al (2002) Genetic diversity and selection in the maize starch pathway. Proc Nat Acad U S A 99:12959–12962

Wilson LM, Whitt SR, Ibañez AM et al (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Windhausen VS, Atlin GN, Hickey JM et al (2012) Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. G3 2:1427–1436

Winkler CR, Jensen NM, Cooper M et al (2003) On the determination of recombination rates in intermated recombinant inbred populations. Genetics 164:741–745

Wright SI, Bi IV, Schroeder SG et al (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Yan J, Shah T, Warburton ML et al (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:e8451

Yan JB, Kandianis CB, Harjes CE et al (2010) Rare genetic variation at *Zea mays crtRB1* increases beta-carotene in maize grain. Nat Genet 42:322–327

Yang XH, Yan JB, Shah T et al (2010) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet 121:417–431

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. Molec Breed 5:505–510

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

Yu J, Holland JB, McMullen M, Buckler ES (2008) Genetic design and statistical power of Nested Association Mapping in maize. Genetics 178:539–551

Yu JM, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zhang M, Montooth KL, Wells MT et al (2005) Mapping multiple quantitative trait loci by Bayesian classification. Genetics 169:2305–2318

Zhang NY, Gibon Y, Gur A et al (2010a) Fine quantitative trait loci mapping of carbon and nitrogen metabolism enzyme activities and seedling biomass in the maize IBM mapping population. Plant Physiol 154:1753–1765

Zhang ZW, Ersoz E, Lai CQ et al (2010b) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–362

Zheng P, Allen WB, Roesler K et al (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat Genet 40:367–372