ELSEVIER

Genetic association mapping and genome organization of maize

Jianming Yu and Edward S Buckler

Association mapping, a high-resolution method for mapping quantitative trait loci based on linkage disequilibrium, holds great promise for the dissection of complex genetic traits. The recent assembly and characterization of maize association mapping panels, development of improved statistical methods, and successful association of candidate genes have begun to realize the power of candidate-gene association mapping. Although the complexity of the maize genome poses several significant challenges to the application of association mapping, the ongoing genome sequencing project will ultimately allow for a thorough genome-wide examination of nucleotide polymorphism-trait association.

Addresses
Institute for Genomic Diversity and United States Department of Agriculture – Agricultural Research Service, and Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, USA

Corresponding author: Buckler, Edward S (esb33@cornell.edu)

Introduction

Most traits of agricultural or evolutionary importance are controlled by multiple quantitative trait loci (i.e. complex traits). Genetic mapping and molecular characterization of these functional loci facilitates genome-aided breeding for crop improvements such as disease resistance, efficiency of fertilizer use, and drought tolerance. Two of the most commonly used tools for dissecting complex traits are linkage analysis and association mapping [1,2]. Linkage analysis exploits the shared inheritance of functional polymorphisms and adjacent markers within families or pedigrees of known ancestry. Linkage analysis in plants has been typically conducted with experimental populations that are derived from a bi-parental cross. Although based on the same fundamental principles of genetic recombination as linkage analysis, association mapping examines this shared inheritance for a collection of individuals often with unobserved ancestry. As the unobserved ancestry can extend thousands of generations, the shared inheritance will only persist for adjacent loci after these many generations of recombination. Essentially, association mapping exploits historical and evolutionary recombination at the population level [3,4].

By exploring deeper population genealogy rather than family pedigree, association mapping offers three advantages over linkage analysis: much higher mapping resolution; greater allele number and broader reference population; and less research time in establishing an association [5,6] (Figure 1).

Linkage analysis and association mapping, however, are complimentary to each other in terms of providing prior knowledge, cross-validation, and statistical power [7••]. Systematic comparisons of these two different approaches have been reviewed elsewhere both in general [8•] and more specifically in maize [7••]. Procedures for conducting an association mapping study in plants have also been well documented [7••,9]. Here, we will focus on recent advances in association mapping conducted in maize, and discuss maize genome structure and its implications for association mapping.

Linkage disequilibrium

The comparatively high-resolution provided by association mapping is dependent upon the structure of linkage disequilibrium (LD) across the genome. Linkage disequilibrium (LD) refers to the non-random association of alleles between genetic loci. Many genetic and non-genetic factors, including recombination, drift, selection, mating pattern, and admixture (i.e. a population of subgroups with different allele frequencies), affect the structure of LD [6,10]. The key to association mapping is the LD between functional loci and markers that are physically linked. The decay of LD over physical distance in a population determines the density of marker coverage needed to perform an association analysis. For example, if LD decays rapidly, then a higher marker density is required to capture markers located close enough to functional sites.

Studies have shown that LD levels vary both within and between species [6]. For example, LD extends less than 1000 bp [11] for maize landraces and roughly 2000 bp for diverse maize inbred lines [4], but can be as high as 100 kb for commercial elite inbred lines [12]. LD decay can also vary considerably from locus to locus. For example, significant LD was observed up to 4 kb for the *Y1* locus (encoding phytoene synthase), but was seen at only 1 kb for *PSY2* (a putative phytoene synthase) in the same maize population [13••]. A more recent study showed that LD extends over 800 kb around *Y1* [14•],

Figure 1
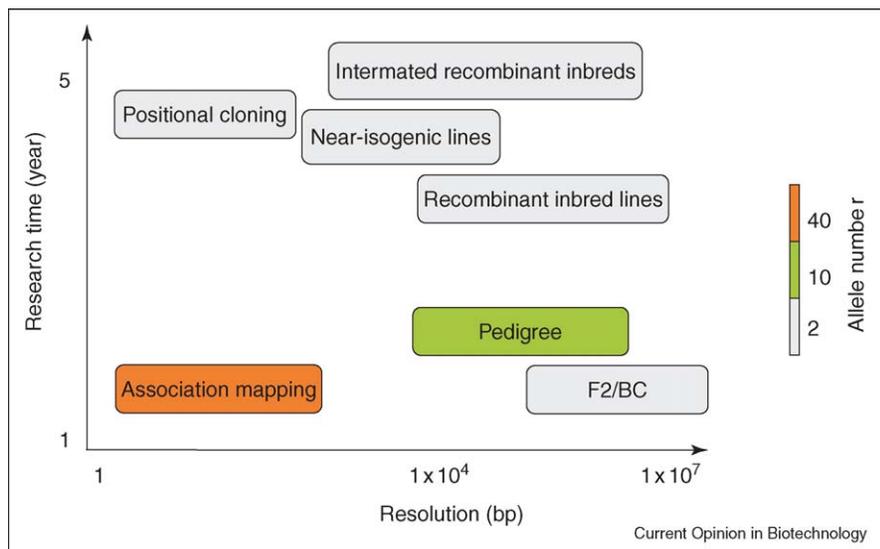
Schematic comparison of various methods for identifying nucleotide polymorphism trait association in terms of resolution, research time and allele number. BC, backcross.

Genome structure

A recent, large-scale sequence study revealed that the maize genome contains approximately 59 000 genes, accounting for 7.5% of the genome [23••]. Over half of the genome (58%) is composed of all types of repeat elements, mainly retroelements and DNA transposons. Unknown sequences occupy the space between these known repeat elements and identifiable coding regions, accounting for the remaining 34.5%. Although about one-third of maize genes are organized in tandem arrays, fewer than half are present in two orthologous copies, indicating a heavy loss of unlinked duplicated genes during the diploidization process following the hybridization of two progenitors [23••].

On the basis of single nucleotide polymorphism analysis, another study predicted that about 1200 maize genes were targets of selection during maize domestication or subsequent improvement by modern breeding [24]. Of these, several candidate genes with putative functions in plant growth were found to be clustered near quantitative trait loci (QTL) that contribute to phenotypic differences between maize and teosinte, the closest wild relative of maize. Association mapping offers a powerful opportunity to continue the work necessary to validate these co-localizations between candidate genes and QTL.

Two other recent studies on the maize genome revealed some potential difficulties for association mapping, owing to sequence non-homology among maize inbred lines [25,26••]. In both studies, researchers found that the clusters of retrotransposons differ markedly in make-up and location in different maize inbred lines. Gene movement by *Helitron* transposons has been offered as an explanation for this haplotype variability [27,28,29••, 30]. This sequence non-homology reduces recombination and preserves LD, thereby limiting the success of association mapping. If candidate genes are located within a long, non-colinear chromosome region, association analysis could result in the mapping of unrelated genes. Non-homologous sequences identified thus far, however, have often been found to be gene fragments rather than intact genes [28,29••]. The impact of these sequences on candidate genes, gene expression and phenotype will require further investigation.

Mapping populations

To date, a limited number of association mapping populations have been publicly reported in maize, perhaps owing to the direct economic value such results hold for private seed companies. The first public maize association mapping population consisted of 102 diverse inbred lines [4]. Newer versions have been characterized more recently [31,32•], the latest of which includes 302 maize inbred lines representing the diversity present in public sector breeding programs around the world. A public maize association mapping population with diverse germplasm bases has also been assembled at the Institut National de la Recherche Agronomique (INRA) [33], and an additional population for mapping endosperm

color has been assembled with 75 public and private maize inbred lines [13[••]].

A large-scale maize QTL mapping population (Nested Association Mapping, NAM), comprising 5000 recombinant inbred lines (RIL) derived from crossing each of 25 diverse maize inbred lines to B73, is currently under development (Molecular and Functional Diversity of the Maize Genome Project; http://www.panzea.org). These lines were chosen to maximize the genetic diversity in maize. As both LD and linkage information can be simultaneously exploited, this population will provide the maize research community with a unified mapping resource that bridges linkage analysis and association mapping.

Population structure

Samples used in association mapping studies can be grouped by the level of population structure and within-group familial relatedness [34[••]] (Figure 2). The concern about population structure is that LD can be caused by admixture of subpopulation, which leads to false-positive results if not correctly controlled in statistical analysis. Such false-positives arise when testing random genetic markers with different frequencies in subpopulations for a trait with parallel phenotypic differences. The complex evolutionary and breeding history in maize [31,32[•]] and other species [22,35] has undoubtedly created both population structure and complex familial relationships. To reduce this risk, estimates of population structure must be included in association analysis.

If, however, the distribution of functional alleles is highly correlated with population structure, statistically controlling for population structure can result in false-negatives, particularly for small sample sizes. Flowering time in maize appears to be one trait for which this phenomenon is common [32[•]], and other traits under local adaptation or diversifying selection in different subpopulations may be effected as well. Association studies, therefore, are best carried out in independent populations with a large sample size.

Two recent studies in maize serve to illustrate the above scenario. In an attempt to validate the function of the *Dwarf8* (*D8*) locus, 71 elite European inbred lines were genotyped for *D8* polymorphism and phenotyped for flowering time [36]. Although significant association was detected without controlling for population structure, no association resulted when the population structure was controlled. By contrast, the association of *D8* polymorphism with flowering time has been validated in a large association mapping population of 375 maize inbred lines [33].

Statistical approaches

Different statistical approaches have been designed to deal with the population structure issue for different

Figure 2
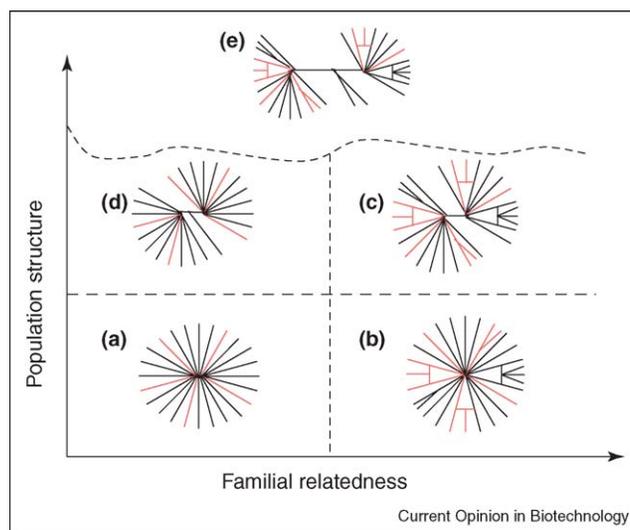
Schematic diagram of the different types of population encountered in association mapping studies. Examples and relevant statistical methods for the analysis of the different population types are described. **(a)** Ideal sample with subtle population structure and familial relatedness (e.g. Centre d'Etude du Polyphorphisme Humain [CEPH] grandparents), regression and genomic control (GC). **(b)** Family-based sample (e.g. CEPH, Utah family), transmission disequilibrium test (TDT), quantitative transmission disequilibrium test (QTDT), GC, and mixed model (pedigree-based coancestry matrix and relative kinship matrix). **(c)** Sample with population structure (e.g. human admixture), structured association (SA) and GC. **(d)** Sample with both population structure and familial relationships (e.g. maize association panel), SA, GC, mixed model (population structure [Q] plus relative kinship matrix [K]). **(e)** Sample with severe population structure and familial relationships (e.g. rice or *Arabidopsis* association mapping panel), methods unknown. The red and black color scheme indicates the polymorphism and diversity.

association samples [34[••]] (Figure 2). For family-based samples, the transmission disequilibrium test (TDT) has long been used to study the genetic basis for human disease, whereas the quantitative TDT (QTDT) has been employed in the dissection of quantitative traits. To address the issue of population structure in population-based samples, genomic control (GC) and structured association (SA) are the two most common methods utilized in both human and plant studies. With GC, a set of random markers is used to estimate the degree of inflation of the test statistics generated by population structure, assuming such structure has a similar effect on all loci [37]. By contrast, SA analysis first uses a set of random markers to estimate population structure (Q), and then incorporates this estimate into further statistical analysis [38–40]. Modification of SA with logistic regression has been used in previous association studies [3,7[••]], and a general linear model version is available in TASSEL (http://www.maizegenetics.net).

A unified mixed-model approach for association mapping that accounts for multiple levels of relatedness has

recently been developed [34[••]]. In this method, random markers are used to estimate Q and a relative kinship matrix (K), which are then fit into a mixed-model framework to test for marker-trait association. Application of this new method to maize quantitative traits and human gene expression data resulted in improved control of both type I and type II error rates when compared with other methods. As this mixed-model approach crosses the boundary between family-based and population-based samples, it provides a powerful complement to currently available methods for association mapping [34[••]].

Examples of association mapping studies

In the first candidate-gene association mapping study in plants, DNA sequence polymorphisms within the *D8* locus were associated with flowering time [3]. This research marked the first empirical association study in any organism for which background molecular markers were used to control for population structure [41]. Later studies of the same population associated the candidate gene *su1* with sweetness taste [42], *bt2*, *sh1* and *sh2* with kernel composition, and *ae1* and *sh2* with starch pasting properties [7[••]]. In this latter study, principle component analysis was used to cluster phenotypic traits into three major groupings before association analyses, which served to reduce multiple testing and also facilitated the interpretation of the results for many correlated traits. In a separate study, candidate genes *a1* and *whp1* were associated with maysin synthesis after controlling for a previously determined epistatic *p* locus [43], illustrating the importance of incorporating known candidate genes in ensuing analyses.

Association mapping has also been used to successfully associate candidate gene *Y1* with maize endosperm color [13[••]], a result later substantiated by linkage analysis [44]. A follow-up study on sequence diversity and LD around the *Y1* region revealed a significant reduction in nucleotide diversity. This selective sweep extends further upstream of *Y1* [14[•]]. The extensive LD around the *Y1* region is purportedly caused by the qualitative nature of the trait, recent timing of selection, and partial genetic isolation of yellow germplasm after selection [13[••]].

Progress continues to be made in deciphering the number of QTL underlying complex traits in maize. A comprehensive linkage analysis study detected approximately 50 QTL underlying oil concentration in the maize kernel [45], while QTL meta-analysis found 62 consensus QTL for flowering time [46]. *In silico* mapping of QTL using a mixed-model approach has been developed to exploit the available genotypic, phenotypic, and pedigree data in maize breeding programs [47]. Recent studies have also shown that gene discovery can be initiated by analyzing existing data for pedigreed maize inbred lines or hybrids [47–49]. Simulation work, however, revealed that addi-

tional effort is needed to reduce the false discovery rate [49], possibly owing to the extensive LD found within pedigreed material [12,50].

As the above examples illustrate, association mapping is especially useful for dissecting candidate genes underlying Mendelian traits (e.g. *Y1* for endosperm color and *su1* for sweetness taste), owing to their relatively simple genetics (few loci and accurate phenotypic measurement) and strong imposed selection. For more complex traits, candidate genes with relatively large effects on traits with relatively high heritability (e.g. *D8* for flowering time) will associate first. Association has, however, been successfully established for traits with only moderate heritability, such as starch concentration in maize.

Conclusions

Although the maize genome presents many technical challenges [51], the first genome sequencing project was funded by the National Science Foundation of USA (NSF), the United States Department of Agriculture (USDA), and the Department of Energy (DOE) of USA in November 2005. With the genome sequence in place, comprehensive gene discovery can be initiated, providing enormous opportunity for candidate-gene association mapping studies. Moreover, as draft sequencing of diverse inbred lines becomes increasingly practical, the feasibility of genome-wide association mapping can be further investigated.

Mutational studies, molecular and biochemical analyses, linkage analysis, comparative genetics, and transgenic studies remain essential building blocks in the further advancement of association mapping, providing candidate genes and validating newly reported associations. The availability of the Nested Association Mapping (NAM) population in the next couple of years will permit a high-resolution genome scan. Although genotyping continues to decrease in cost, the expense and precision of phenotyping pose lingering challenges that must be addressed [32[•]]. Additional challenges include non-additive genetic effects and genotype and environment interactions (i.e. genotypes differ in their relative performance across environments) that are commonplace with the evaluation of diverse germplasm in diverse environments [7[••]]. However, with a better understanding of simple scenarios, we will ultimately move towards more complex issues such as dominance, epistasis, genotype and environment interaction, and heterosis.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Risch N, Merikangas K: **The future of genetic studies of complex human diseases**. *Science* 1996, **273**:1516-1517.

2. Mackay TF: **The genetic architecture of quantitative traits**. *Annu Rev Genet* 2001, **35**:303-339.

3. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: **Dwarf8 polymorphisms associate with variation in flowering time**. *Nat Genet* 2001, **28**:286-289.

4. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES: **Structure of linkage disequilibrium and phenotypic associations in the maize genome**. *Proc Natl Acad Sci USA* 2001, **98**:11479-11484.

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

6. Flint-Garcia SA, Thornsberry JM, Buckler ES: **Structure of linkage disequilibrium in plants**. *Annu Rev Plant Biol* 2003, **54**:357-374.

7. Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM,
•• Buckler ES: **Dissection of maize kernel composition and starch production by candidate gene association**. *Plant Cell* 2004, **16**:2719-2733.
An association study of six candidate genes involved in the starch biosynthetic pathway with diverse maize inbred lines. Principle component analysis was applied to summarize the information on the correlated phenotypic traits for kernel composition and starch pasting properties. Good discussion on association mapping versus linkage analysis.

8. Hirschhorn JN, Daly MJ: **Genome-wide association studies for
• common diseases and complex traits**. *Nat Rev Genet* 2005, **6**:95-108.
A comprehensive review of genome-wide association studies in humans, including rationale, power, efficiency, approaches, analysis and interpretation.

9. Whitt SR, Buckler ES: **Using natural allelic diversity to evaluate gene function**. *Methods Mol Biol* 2003, **236**:123-140.

10. Gaut BS, Long AD: **The lowdown on linkage disequilibrium**. *Plant Cell* 2003, **15**:1502-1506.

11. Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS: **Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.)**. *Proc Natl Acad Sci USA* 2001, **98**:9161-9166.

12. Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ: **SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines**. *BMC Genet* 2002, **3**:19.

13. Palaisa KA, Morgante M, Williams M, Rafalski A: **Contrasting
•• effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci**. *Plant Cell* 2003, **15**:1795-1806.
Maize phytoene synthase *Y1* was associated with endosperm color. Nucleotide diversity was found to be equivalent at *PSY2*, another putative phytoene synthase, but was 19-fold less in yellow maize inbred lines than in white.

14. Palaisa K, Morgante M, Tingey S, Rafalski A: **Long-range
• patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep**. *Proc Natl Acad Sci USA* 2004, **101**:9885-9890.
A follow-up study on the surrounding region of the *Y1* gene, which was previously associated with determining maize endosperm color. This is the largest extension of LD (200 kb upstream and 600 kd downstream of *Y1*) reported in a maize candidate gene so far, possibly owing to the qualitative nature of the trait, strong selection in short time, and partial isolation.

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

16. Ingvarsson PK: **Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (Populus tremula L., Salicaceae)**. *Genetics* 2005, **169**:945-953.

17. Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB: **Nucleotide diversity and linkage disequilibrium in loblolly pine**. *Proc Natl Acad Sci USA* 2004, **101**:15255-15260.

18. Morrell PL, Toleno DM, Lundy KE, Clegg MT: **Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp. spontaneum) despite high rates of self-fertilization**. *Proc Natl Acad Sci USA* 2005, **102**:2442-2447.

19. Garris AJ, McCouch SR, Kresovich S: **Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (Oryza sativa L.)**. *Genetics* 2003, **165**:759-769.

20. Hamblin MT, Mitchell SE, White GM, Gallego J, Kukatla R, Wing RA, Paterson AH, Kresovich S: **Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of sorghum bicolor**. *Genetics* 2004, **167**:471-483.

21. Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ *et al.*: **The extent of linkage disequilibrium in Arabidopsis thaliana**. *Nat Genet* 2002, **30**:190-193.

22. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R *et al.*: **The pattern of polymorphism in Arabidopsis thaliana**. *PLoS Biol* 2005, **3**:e196.

23. Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y,
•• Wei F, Fuks G, Soderlund CA, Mayer KF *et al.*: **Sequence composition and genome organization of maize**. *Proc Natl Acad Sci USA* 2004, **101**:14349-14354.
A comprehensive sequence analysis of the maize genome. The maize genome was found to contain 59 000 genes, accounting for 7.5% of the genome. All types of repeat elements make up 58% of the genome, and unknown sequences account for the remaining 34.5%. Although one-third of the maize gene is organized in tandem arrays, fewer than half of the genes are present in two orthologous copies.

24. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS: **The effects of artificial selection on the maize genome**. *Science* 2005, **308**:1310-1314.

25. Fu H, Dooner HK: **Intraspecific violation of genetic colinearity and its implications in maize**. *Proc Natl Acad Sci USA* 2002, **99**:9573-9578.

26. Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A:
•• **Evolution of DNA sequence nonhomologies among maize inbreds**. *Plant Cell* 2005, **17**:343-360.
A large-scale examination of sequence nonhomology among maize inbred lines. From 2.8 Mb of sequence studied, more than half was found to be noncollinear between two maize inbred lines, B73 and Mo17, owing to the insertion of large numbers of long-terminal repeat (LTR) retrotransposons. The ratio between shared and nonshared LTR retroelements was 27:62 and that of genes was 45:27.

27. Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, Llaca V, Tanyolac B, Park YJ, Jeong OY, Bennetzen JL *et al.*: **Gene loss and movement in the maize genome**. *Genome Res* 2004, **14**:1924-1931.

28. Lai J, Li Y, Messing J, Dooner HK: **Gene movement by Helitron transposons contributes to the haplotype variability of maize**. *Proc Natl Acad Sci USA* 2005, **102**:9068-9073.

29. Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A:
•• **Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize**. *Nat Genet* 2005, **37**:997-1002.
A genome-wide comparison of gene content in two maize inbred lines, B73 and Mo17. Among 20 656 genes and gene fragments examined in allelic bacterial artificial chromosome (BAC) contigs, 20.6% were not shared between the two lines. Eight of nine genic insertion sites examined appeared to be mediated by helitron with an unknown mechanism.

30. Brunner S, Pea G, Rafalski A: **Origins, genetic organization and transcription of a family of non-autonomous helitron elements in maize**. *Plant J* 2005, **43**:799-810.

31. Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J:
**Genetic structure and diversity among maize inbred lines
as inferred from DNA microsatellites**. *Genetics* 2003,
165:2117-2128.

32. Flint-Garcia SA, Thuillet A, Yu J, Pressoir G, Romero SM,
•   Mitchell SE, Doebley JF, Kresovich S, Goodman MM, Buckler ES:
**Maize association population: A high resolution platform for
QTL dissection**. *Plant J* 2005, **44**:1054-1064.
A systematic description of a maize association mapping panel of 302
diverse inbred lines. Population structure estimates of the whole set, and
heritability estimates for 60 traits based on replicated measurement from
a subset, are presented.

33. Camus-Kulandaivelu L, Veyrieras J-B, Madur D, Combes V,
Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D,
Charcosset A: **Maize adaptation to temperate climate:
relationship with population structure and polymorphism in
the *Dwarf8* gene.** *Genetics* 2006; published ahead of print doi:
10.1534/genetics.105.048603.

34. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M,
•• Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al.*:
**A unified mixed-model method for association mapping
that accounts for multiple levels of relatedness**. *Nat Genet*
2006, **38**:203-208.
Reports new statistical methods for association mapping that simulta-
neously account for both population structure and familial relatedness
within a sample. Marker-based relative kinship and population structure
were simultaneously introduced into the traditional mix-model frame-
work. The method crosses the boundary between family-based and
mixed association samples and is demonstrated with maize quantitative
traits and human gene expression examples.

35. Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S:
Genetic structure and diversity in *Oryza sativa* L.
Genetics 2005, **169**:1631-1638.

36. Andersen JR, Schrag T, Melchinger AE, Zein I, Lubberstedt T:
**Validation of Dwarf8 polymorphisms associated with
flowering time in elite European inbred lines of maize (*Zea
mays* L.)**. *Theor Appl Genet* 2005, **111**:206-217.

37. Devlin B, Roeder K: **Genomic control for association studies**.
Biometrics 1999, **55**:997-1004.

38. Pritchard JK, Rosenberg NA: **Use of unlinked genetic markers
to detect population stratification in association studies**.
Am J Hum Genet 1999, **65**:220-228.

39. Pritchard JK, Stephens M, Donnelly P: **Inference of population
structure using multilocus genotype data**. *Genetics* 2000,
155:945-959.

40. Falush D, Stephens M, Pritchard JK: **Inference of population
structure using multilocus genotype data: linked loci and
correlated allele frequencies**. *Genetics* 2003, **164**:1567-1587.

41. Pritchard JK: **Deconstructing maize population structure**.
Nat Genet 2001, **28**:203-204.

42. Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES:
Genetic diversity and selection in the maize starch pathway.
Proc Natl Acad Sci USA 2002, **99**:12959-12962.

43. Szalma SJ, Buckler ES, Snook ME, McMullen MD: **Association
analysis of candidate genes for maysin and chlorogenic
acid accumulation in maize silks**. *Theor Appl Genet* 2005,
110:1324-1333.

44. Wong JC, Lambert RJ, Wurtzel ET, Rocheford TR: **QTL and
candidate genes phytoene synthase and zeta-carotene
desaturase associated with the accumulation of carotenoids
in maize**. *Theor Appl Genet* 2004, **108**:349-359.

45. Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D,
Hauge B, Lai C, Clark D, Rocheford TR, Dudley JW: **The genetic
architecture of response to long-term artificial selection for
oil concentration in the maize kernel**. *Genetics* 2004,
168:2141-2155.

46. Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L,
Murigneux A, Charcosset A: **Genetic architecture of flowering
time in maize as inferred from quantitative trait loci meta-
analysis and synteny conservation with the rice genome**.
Genetics 2004, **168**:2169-2185.

47. Parisseaux B, Bernardo R: **In silico mapping of quantitative trait
loci in maize**. *Theor Appl Genet* 2004, **109**:508-514.

48. Zhang YM, Mao Y, Xie C, Smith H, Luo L, Xu S: **Mapping
quantitative trait loci using naturally occurring genetic
variance among commercial inbred lines of maize (*Zea mays*
L.)**. *Genetics* 2005, **169**:2267-2275.

49. Yu J, Arbelbide M, Bernardo R: **Power of *in silico* QTL mapping
from phenotypic, pedigree, and marker data in a hybrid
breeding program**. *Theor Appl Genet* 2005, **110**:1061-1067.

50. Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M,
Reif JC: **Linkage disequilibrium in European elite maize
germplasm investigated with SSRs**. *Theor Appl Genet* 2005,
111:723-730.

51. Martienssen RA, Rabinowicz PD, O'Shaughnessy A, McCombie
WR: **Sequencing the maize genome**. *Curr Opin Plant Biol* 2004,
7:102-107.