Applications of Genomics Technologies to Enhance Rate of Genetic Progress for Yield of Maize within a Commercial Breeding Program

William S. Niebur, J. Antoni Rafalski, Oscar S. Smith, Mark Cooper

Pioneer Hi-Bred International Inc., 7300 N.W. 62nd Avenue, P.O. Box 1004, Johnston, Iowa 50131, USA

Abstract
The commercial grain yield of maize in the U.S. Corn Belt has undergone significant genetic improvement since the initiation of hybrid breeding in the first half of the 20th Century. Designing effective molecular enhanced breeding strategies that improve on the outcomes of the conventional pedigree breeding strategies is a challenge for commercial maize improvement programs today. Most of the traits of importance to the breeder are genetically complex and quantitative in nature. Genetically mapping these traits in elite breeding populations is difficult. Conventional mapping approaches in structured populations are often not relevant to the elite crosses of a maize breeder, are time consuming to create and even when successful usually only offer low levels of mapping resolution (ca. 10 to 20 cM). Therefore, considerable interest has emerged in the use of alternative methods for the mapping of quantitative traits. Genetic association mapping methods have been the subject of great interest in human genetics. More recently genetic association studies have been performed in crop plants, with some promising results. Much work still remains to develop association mapping methodology that is applicable to properties, such as pedigree relationships, among the elite populations of a maize breeder. We will discuss the utility of these approaches for mapping complex traits and issues for consideration in their application to breeding.

Media Summary
New methods of mapping offer a promising route to the identification of genetic markers for the genes of complex traits and provide a foundation for molecular enhanced breeding of maize.

Key Words
Genetic Diversity, SNP, Linkage Disequilibrium, Association Mapping, Pedigree, Forward Selection

Progress from conventional breeding: The foundation
The ultimate measure of success of a commercial maize breeding program is consistent widespread adoption by the target industry of the sequence of commercial hybrid products that it develops. Following the founding of the company in the 1920s, Pioneer has had a long history of developing successful maize hybrids (Duvick et al. 2004). The success of this long-term maize breeding program has relied on a continual commitment to research. Outcomes from the research have contributed to the development of a proprietary pool of elite germplasm and have created knowledge about the phenotypes and breeding value of the germplasm. The succession of Pioneer maize breeders involved in developing the elite germplasm of today have investigated broadly and adopted a range of breeding strategies that are effective at manipulating the germplasm using the available knowledge. Information management infrastructure has been developed that supports the scaling up of the breeding strategies and enables the breeder to access the germplasm knowledge from across the research community. Throughout the history of the breeding program, different research emphases have emerged and have been evaluated. The breeding strategy and research tools that remain in place today are a mixture of those that have been proven to be useful and the next generation of ideas that are currently under investigation. This fluid and active research community has been enabled by significant long-term support of the community of maize geneticists, breeders and researchers. Here we consider that today's foundation for molecular breeding is the reference population of elite genetics to which we intend to apply the molecular technologies. To understand the nature of the elite genetics reference population of a long-term breeding program, and how it differs from some of the more conventional reference populations used in genetics and trait mapping, it is important to understand the history of the breeding process that has given rise to the elite germplasm of today.

*Foundations: Historical view of changes at the level of the hybrid phenotype*
For the majority of the 20th Century, pedigree breeding, combined with an extensive multi-environment testing program that was designed to measure the phenotypic performance of new genotypes across a large sample of the target population of environments, was the core breeding strategy that underpinned

the successful genetic improvement of grain yield and other agronomic traits (Duvick et al. 2004). In addition to conducting applied breeding programs with a primary responsibility for developing new products, the Pioneer maize breeders have continually investigated opportunities to improve practical breeding methodology. Thus, the application of pedigree breeding to maize in Pioneer has undergone a continual process of evolution and refinement over the years, with any changes adopted after extensive evaluation by a combination of empirical research and breeder experience, the latter obtained from attempting to scale up the methods and put them into practice.
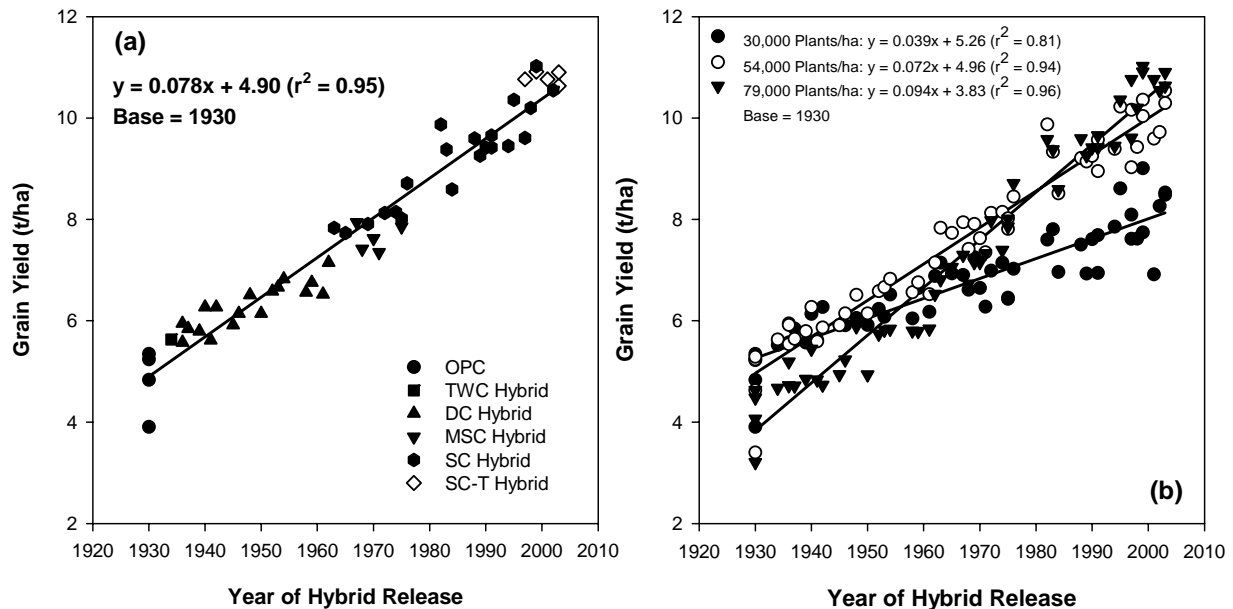
Figure 1 Changes in grain yield (Best Linear Unbiased Predictors, BLUPs) over time for a series of successful Pioneer maize hybrids, and four open-pollinated cultivars, evaluated in experiments conducted from 1990 to 2003. Entries were tested under different densities at each location-year combination: (a) yield at the density which gave the highest yield, (b) yield at low (~30,000 plants/ha), medium (~54,000 plants/ha) and high (~79,000 plants/ha).

Side-by-side comparisons of successful Pioneer maize hybrids commenced in 1972 and continue to this day (Figure 1; Duvick 2004). The objective of these studies was to monitor genetic progress for yield that was achieved by the breeding program and to examine associated changes in traits. Commencing with a sample of important open-pollinated cultivars (OPC) of the time, hybrids dating from the 1930s to present day are included in these so-called ERA experiments. The early hybrids were either double-cross (DC) or three parent crosses (TC). During the 1960s modified single-cross (MSC) and true single-cross (SC) hybrids appeared. Since this time SC hybrids have been the commercial hybrid product of choice for the US Corn Belt. From the mid-1990s successful single-cross hybrids with a transgene (SC-T) for specific traits, such as the Bt gene for resistance to European Corn Borer (*Ostrinia nubilalis* Hubner), were successfully commercialized. As justified on commercial success these transgenic hybrids were included in the ERA study. Over the course of this breeding history a number of crop management practices have changed. Plant populations (density) have increased and fertilizer inputs have increased. There is an important genotype-by-density interaction that has been actively exploited to realize the genetic gains for yield of maize. The early hybrids perform better at lower plant populations and the modern hybrids perform better at higher plant populations (Figure 1b). A number of reports of the results of these maize ERA studies have been published (Duvick 1977, 1992, Duvick et al. 2004). These studies have consistently demonstrated that genetic gain has been achieved for grain yield in the US Corn Belt and that these gains have resulted from the matching of superior genetics and management practices. To date these genetic gains for yield of maize can be described as linear with respect to time (e.g. Figure 1a).
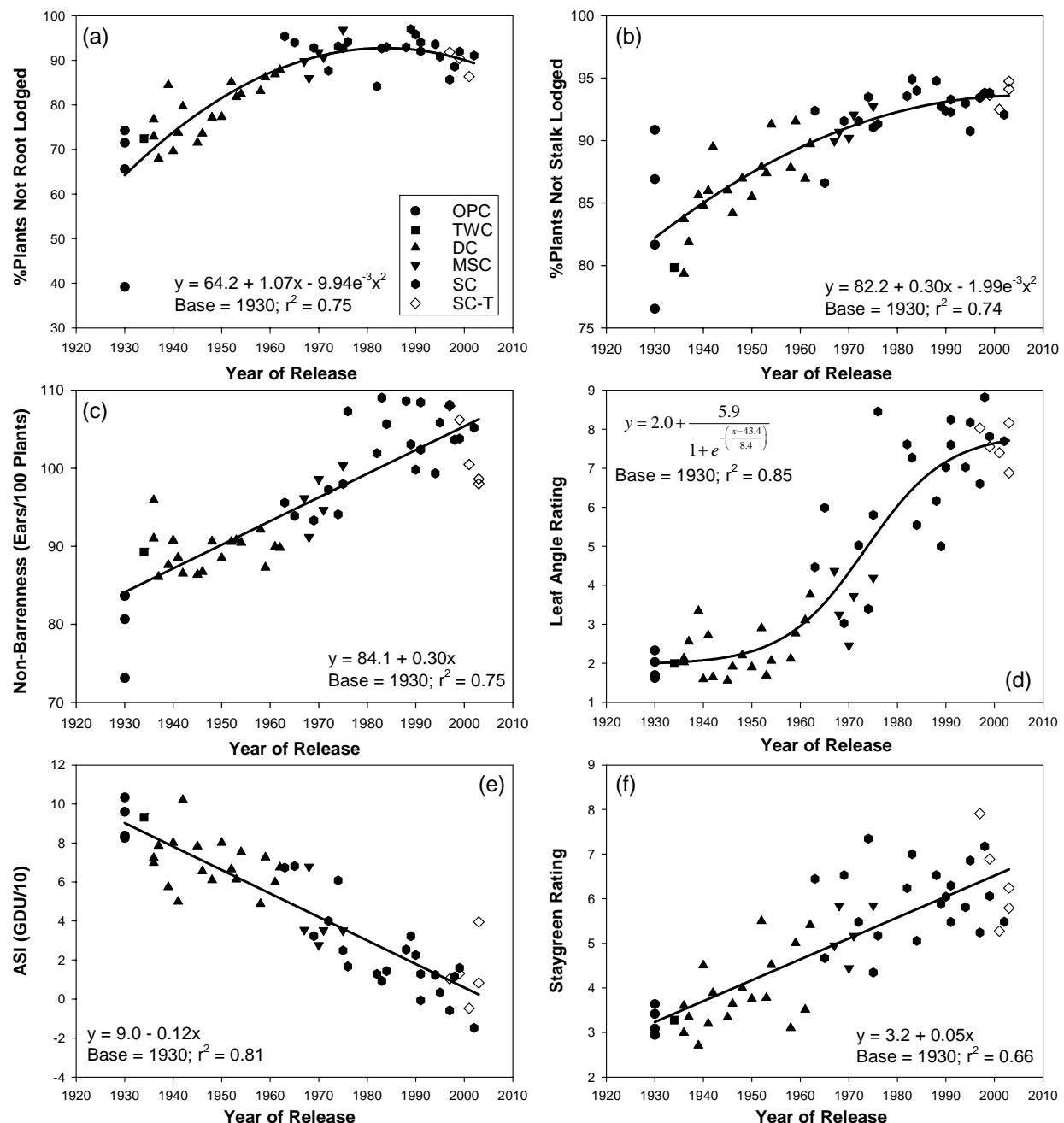
Figure 2 Changes in traits over time for a series of successful Pioneer maize hybrids and four open-pollinated cultivars, evaluated in experiments conducted from 1990 to 2003; (a) % of plants not root lodged, (b) % of plants not stalk lodged, (c) non-barrenness measured as the number of ears per 100 plants, (d) Leaf angle rating (1 = floppy leaves and 9 = erect leaves), (e) Anthesis to Silking Interval (ASI) measured as Growing Degree Days (GDU/10), (f) Staygreen rating (1 = not Staygreen and 9 = Staygreen). Entries were tested at three densities at each location-year combination: low (~30,000 plants/ha), medium (~54,000 plants/ha) and high (~79,000 plants/ha).

Analyses of the contributions of traits to the genetic gain for yield have shown that different traits were important at different times during the sequence of improvements (Figure 2). Early in the breeding program large gains were made in improving standability by reducing the susceptibility to both root and stalk lodging (Figure 2a,b) and by reducing barrenness (Figure 2c). At the time of the transition from double-cross to single-cross hybrids (1960s), and in association with increases in plant density, there was a rapid transition from floppy to erect leaf hybrids (Figure 2d). Throughout the history of the breeding program there has been a continual trend to reduce the anthesis to silking interval (ASI) (Figure 2e) and increase canopy staygreen (Figure 2f). For further details and recent reviews of the changes over the course of the breeding program see the publications by Duvick et al (2004). A synopsis or "strong

hypothesis" that can be drawn from the large body of analyses of genetic gain for yield that have been achieved by the Pioneer program is that higher yields of modern maize hybrids have been achieved by creating new combinations of genes for traits that together contribute to enhanced tolerance of the maize hybrid to the diverse range of stresses that are encountered within the target population of environments.

*Foundations: Historical view of changes at the level of the inbred genotype*
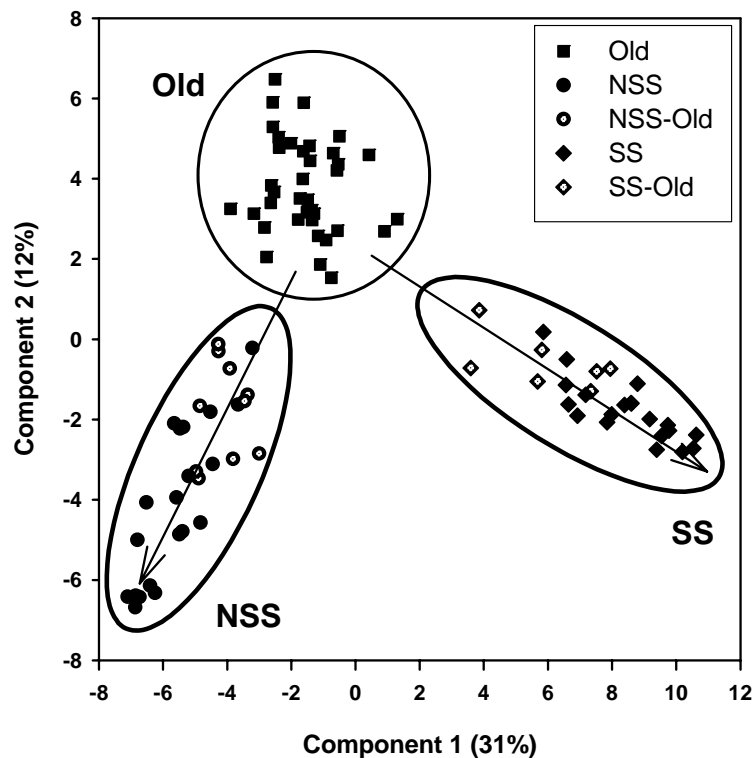
Figure 3 Inbred scores on the first two principal components from an analysis of SSR molecular marker profiles of the parents of the ERA hybrids. The large boundaries distinguish the three main groups of lines designated as Old, SS and NSS. SS = Stiff Stalk Synthetic, NSS = Non Stiff Stalk Synthetic, Old = the older inbred lines used prior to the formation of the SS and NSS heterotic groups. The arrows indicate the direction of the progression of inbred improvement within the SS and NSS heterotic groups.

Molecular markers can be used to examine some of the changes at the DNA sequence level that have been associated with the changes in yield and agronomic trait phenotypes. The inbred parents of the series of Pioneer hybrids examined in the ERA studies (Figures 1 and 2) were characterized by a set of simple sequence repeats (SSRs) that gave good coverage of all 10 chromosomes. The allele scores were used to measure the genetic distance among the inbred parents. A proximity matrix among the set of inbred lines was constructed and examined by both cluster and ordination analyses. A summary of important features of the genetic diversity among the parents was achieved by plotting the parent scores on the first two principal components (Figure 3). Three main inbred groups were observed based on the relative positions of the scores of the parents. There was a group of inbreds that represented the parents of the older double cross hybrids. The other two groups represented a distinction between newer inbreds and coincided with the two main heterotic groups of the single cross hybrids, one designated as a Stiff Stalk Synthetic (SSS) group and the other a Non-Stiff Stalk Synthetic (NSS) group. Within the SSS and NSS groups the older of these inbreds were generally located closer to the group of older inbred parents of the double cross hybrids. The separation of the SSS and NSS groups diverged with lower scores on component 2 and this was associated with greater divergence between the newer SSS and NSS inbreds. These descriptive analyses of patterns of SSR diversity among the parents of a sequence of successful hybrids developed by the breeding program are indicative that significant changes have occurred at the DNA sequence level within the maize genome over the history of the breeding program. The early Pioneer maize breeders created the heterotic groups of today and the modern maize breeders have further improved the inbreds within the heterotic groups to provide the commercial single-cross hybrids of today.

Foundations: Historical view of changes at the level of DNA sequence variation
While Figure 3 is indicative of some temporal patterns of change at the level of the whole genome over the course of the breeding program it does not provide information on the functional variation that is associated with the changes in trait phenotypes (Figures 1 and 2). It can be expected that many of the changes in the allelic composition of genes that have occurred over the course of the breeding program will be complex. Examples of the types of change that can be observed at an individual locus are shown for two SSR loci (Figure 4). To describe trends of change in allele frequency over time the hybrids were grouped by decade of release. The inbreds were then grouped on whether they were used as a female or a male in the hybrids. This classification was somewhat arbitrary during the double cross hybrid phase (1930s to 1950s). However, from the 1960s the distinction between female and male coincides with the SSS and NSS heterotic groups, respectively. Within each decade the frequency of each allele was determined for the female and male inbreds and also for the hybrids.
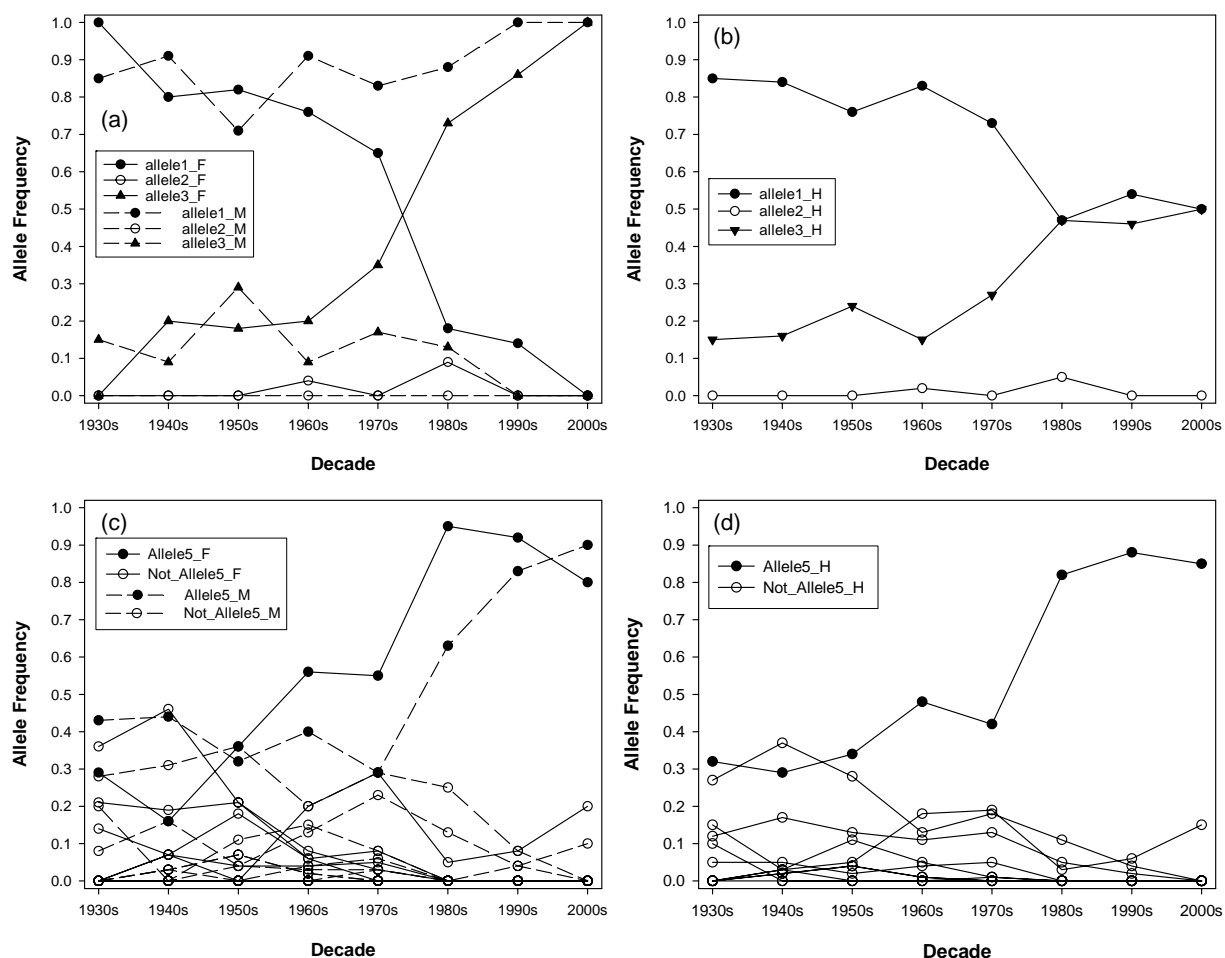
Figure 4 Frequencies of alleles for two unlinked SSR loci, observed by decade, in the female and male pools of inbred lines and the hybrids of the inbred lines: SSR locus 1 inbreds (a) and hybrids (b); SSR locus 2 inbreds (c) and hybrids (d).

For SSR locus 1 (Figure 4a and b) three alleles were identified. Two of the alleles (designated as alleles 1 and 3) were observed to have a high frequency during some time periods, and the other (designated as allele 2) was consistently rare (Figure 4a). Initially allele 1 dominated in the male and female inbreds of the early hybrids. Over time allele 1 was retained at a high frequency in the male pool of inbreds and it was swept out of the female pool. In the female pool allele 1 was eventually replaced by allele 3, which started at a low frequency. Consequently in the single cross hybrids both alleles were brought together consistently from the 1980s onward (Figure 4b). For SSR locus 2 (Figure 4c and d) there was a different pattern of change over time observed than that for SSR locus 1. At this locus there were 11 alleles observed. Initially, in the early decades none of these alleles clearly dominated. However, over the course of the breeding program the same allele (designated as allele 5) came to dominate both the female and

male pools of inbreds from the 1970s onward (Figure 4c) and therefore this allele was consistently observed at a high frequency in the single cross hybrids from the 1970s onward (Figure 4d).

Elite genetics as the foundation for molecular breeding
The key point we emphasize here is that foundation on which we are building molecular breeding strategies for maize is the advanced cycles of a mature and highly successful pedigree breeding program. Many genetic and associated phenotypic changes have already been made in the process of developing the elite germplasm available to the maize breeders of today. Significant long-term genetic progress has already been made for grain yield (Figure 1), thus the bar is set high for any methodology that is under investigation to enhance the success rate of this breeding program. It is likely that many of the trait combinations, and associated genetic variation, that can still be exploited by less mature breeding programs has already been extensively evaluated and if effective utilized in the course of achieving the historical genetic gains for yield within the Pioneer maize breeding program. Thus, many of the molecular breeding questions that are relevant to other breeding programs may be of less relevance to the Pioneer program. With this view in mind we discuss some applications of genomics technologies to enhance rate of gain for yield in maize breeding.

Genomics and Mapping Traits: Associative Approaches
With the availability of a continually expanding toolkit of genomic technologies the maize breeder now has a growing number of options for the conduct of a molecular enhanced breeding program. As for all previous modifications to maize breeding strategies, the same rules for adoption of molecular breeding technologies apply, an advantage over the incumbent breeding strategy has to be demonstrated. There are some logical steps to be considered in evaluating the use of genomics technologies to maize breeding: (1) demonstrating the feasibility of associating genetic variation at the DNA sequence level with phenotypic variation for important traits, (2) demonstrating that the knowledge of the gene-to-phenotype associations provides additional value to the current phenotypic evaluation process, (3) scaling up the use of the molecular technologies for high throughput application, (4) demonstrating the improvements in genetic gain and product development from the molecular enhanced breeding strategies. Here we consider aspects of these steps in relation to the implementation of marker-assisted selection (MAS) for maize in a large commercial breeding program.

Genetic mapping of quantitative traits
The goal of a plant geneticist is to associate (correlate) the alleles (or haplotypes) defined at the DNA sequence level, with the phenotypic variation for agronomic and quality traits. Several methods can and have be used to genetically map quantitative traits in maize. Some of the more common approaches are discussed below.

Quantitative trait locus (QTL) mapping is perhaps the most widely used today. In this approach two parents, differing for the trait of interest, are crossed to develop a reference segregating population for mapping the trait. Many generations can be derived from the initial cross to be used for mapping; e.g. F2 or F2-derived F3 families, BC generations, Recombinant Inbred Lines (RILs) developed by continual selfing of random individuals sampled from the F2 generation to an advanced level of inbreeding (e.g. F7), or generation of doubled haploids (DHs) from the F1 individual or some later generations following inbreeding by selfing. For quantitative traits it is desirable to create a reference population that can be maintained indefinitely. This allows individuals to be replicated within and across environments to improve precision of trait measurement and testing for QTL-by-Environment (Q×E) interactions. Therefore, RILs and DHs are often a preferred reference population for mapping. To map a trait the individuals sampled from the reference population are measured for the trait phenotype in appropriate samples of environmental conditions and genotyped using an appropriate marker system, such as SSRs, RFLPs, AFLPs or some other method of choice. For the chosen mapping reference population, a set of statistical expectations are defined for the linkage relationships between marker positions on the genetic map and for the likelihood of the presence of a QTL in positions relative to the marker positions on the map. Based on these sets of expectations genome scans can be conducted to test for the presence of a segregating QTL associated with the phenotypic variation for the trait being located at genomic positions defined by the genetic map. Many methods have been developed for conducting such genome wide testing for QTL. Today a widely used method is Composite Interval Mapping (CIM). Specialized software has been developed to implement CIM for a range of mapping reference populations. For many

reference populations the map resolution from this approach is usually modest, on the order of 10-40 cM. Further work will usually be necessary to achieve greater map resolution and identify markers that will be useful for breeding. Typically, QTL of interest can be investigated further by establishing populations where greater numbers of recombination events occur within the region defined to contain the QTL. One approach that can be used involves continual backcrossing of the QTL into a background where the particular QTL allele is absent in order to create a set of Near Isogenic Lines (NILs). This can be done in parallel for multiple QTL and continued to achieve a defined level of higher resolution. Localizing QTL to regions in the order of 1 to 5 cM provides a basis for developing markers for some breeding purposes and may serve as an entry point for isolation of a gene associated with the QTL, "map based cloning". Several QTL have been cloned, but this approach is laborious (Frary et al. 2000; Yano et al. 2000) and is not necessary for application of Marker Assisted Selection (MAS).

The results that have been achieved by applying QTL mapping approaches to biparental combinations of old and elite maize lines suggest that there are likely to be many small QTL for the traits of interest to the breeder in the elite reference populations. In one study, a large F4 mapping population of approximately 1000 lines was established and tested at multiple locations for two years (Openshaw and Frascaroli 1997). Early analyses of this data set indicated large numbers of small QTL were the basis of the genetic architecture for most traits. Subsequent work (Unpublished) has shown that for quantitative traits such as grain yield Q×E interactions are commonplace and make important contributions to the small QTL effects observed for mean performance across environments.

Segmental introgression library mapping involves using backcrossing with molecular markers to create a set of inbred lines, where each inbred contains a different chromosome segment from the genome of a donor line that possesses a trait phenotype of interest for mapping within the recurrent parent background. Multiple molecular markers are used to genotype individuals during the backcross process to identify the chromosome segments from the donor within the recurrent background. The target is to have a series of inbreds with single different segments where the complete set represents the whole genome of the donor parent. The target size of the donor segments depends on the initial mapping resolution that the geneticist seeks. The smaller the target segment size the greater the number of segmental introgression lines and effort required. A reasonable target segment size may be in the order of 1/4 to 1/10 of a chromosome. Once a segmental introgression set is developed, the contribution of each chromosome segment to the trait phenotype can be directly ascertained in appropriate test environments. Once individual segments have been evaluated, epistatic interactions between segments can be studied by making the appropriate crosses among the different introgression lines to create different segment combinations. The introgression lines can be used as a starting point for fine mapping a segment of interest by conducting further generations of backcrossing, using markers to identify recombination events within the segment. A relatively significant commitment of marker resources is necessary in the process of creation of the segmental introgression library set of lines.

Association mapping relies on a large number of historical recombination events, distributed throughout the genome and operating within a diverse population of individuals. Large numbers of randomly distributed recombination events progress physically separated genomic regions towards a state of linkage equilibrium. Two regions within the genome that are in linkage equilibrium will show no correlation of allele combinations between the two regions. The greater the number of recombination events the more widespread will be the extent of linkage equilibrium and the smaller will be the expected size of the physical region that remains in linkage disequilibrium. The approach of much of the genome towards a state of linkage equilibrium enables detection of linkage associations between polymorphic markers and the segregating genes that influence the phenotypic variation for traits among the individuals within the diverse population. The smaller the physical size of the regions that remain in linkage disequilibrium following the historical recombination events, the finer is the expected mapping resolution that is achievable. Concomitantly, with more widespread linkage equilibrium a higher density of marker coverage is required across the genome.
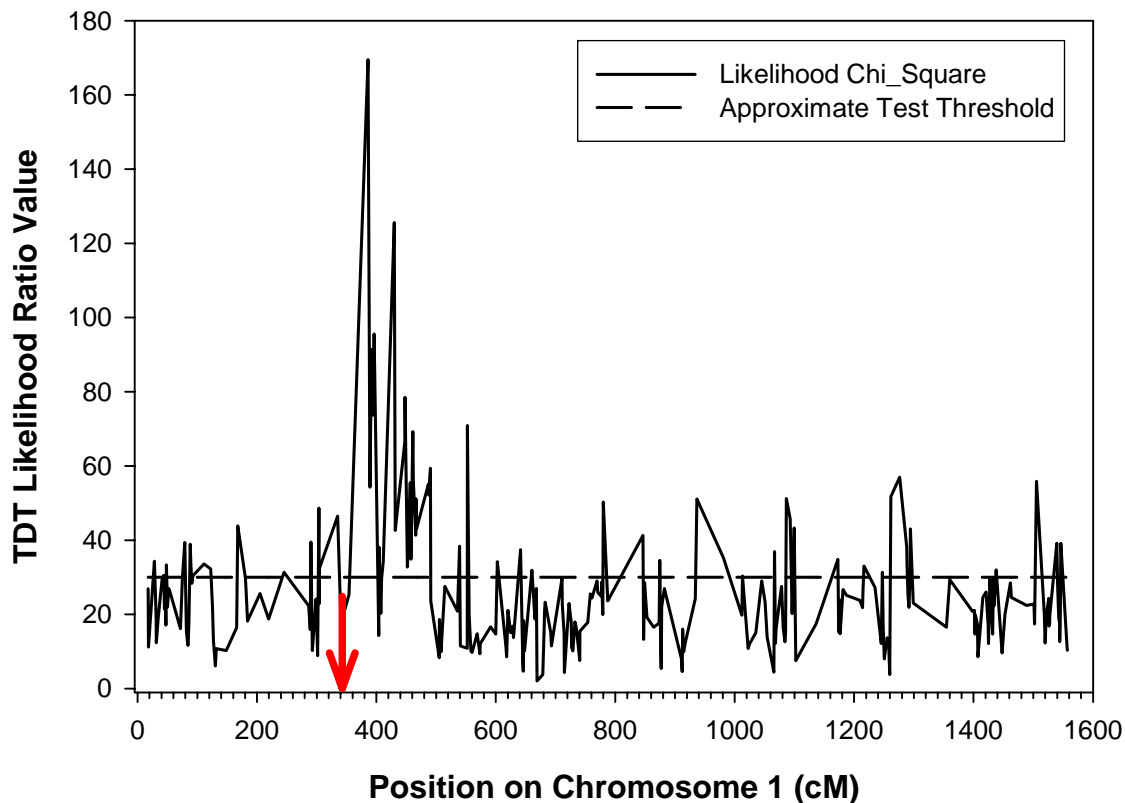
Figure 5 Significance of the results of the Transmission Disequilibrium Test (TDT) for an association between sequence haplotypes at positions on Chromosome 1 of maize and the red cob color phenotype examined in a set of elite maize inbred lines. The red arrow indicates the approximate location of the pericarp 1 (p1) gene.

A major advantage of the association mapping approach is that if a suitable reference population of lines is already available, e.g. a collection of random lines obtained from a germplasm bank, and there is a sufficiently dense genetic map, there is no need to spend the long periods of time required to develop the specialized mapping populations, that is needed for the other methods discussed above. However, while the association mapping concept is appealing there are a number of factors that can compromise the success of this approach. Essentially these are any population history or biological factors other than physical linkage that can give rise to a correlation between alleles between regions of the genome. These include, factors that influence the frequency and distribution of recombination events around the genome, population bottlenecks and associated random drift, non-random mating and resultant population structure effects, including the pedigree structures established over the history of a breeding program, selection for allele combinations from multiple non-contiguous genes, and the interactions of these factors with each other and the historical patterns of recombination events across the genome. Thus, we expect to see different levels and patterns of linkage disequilibrium across the genome in different reference populations. Given the history of maize breeding, discussed above, that has given rise to the elite genetic populations of today many of these factors must be taken into consideration in the application of association mapping approaches to maize breeding. While these complications will require careful application of association mapping to elite populations this approach will be feasible for a number of traits. For example, for the simple trait phenotype red cob color that varies among elite inbred lines it is possible to associate sequence variation in the pericarp 1 gene (p1) with variation in the trait phenotype in a genome scan association mapping study (Figure 5).

Heterogeneous intermated populations that are created by intermating multiple parents for multiple cycles have been suggested for trait mapping. This reference population structure may be viewed as intermediate between the specific biparental populations for classical QTL mapping and the large diverse populations used for association mapping. In this approach the populations are created deliberately by selecting a diverse set of parents (e.g. perhaps somewhere between 4 and 20) that are representative of the trait variation in the breeding germplasm. These parents are then intercrossed for a number of generations to allow recombination to reduce the extent of linkage disequilibrium across the genome and thus increase

genetic resolution (Mott et al., 2000). An advantage of this approach over the association mapping approach is that many of the complications introduced from unknown historical population structures can be avoided by managing the intercrossing of the selected parents. The advantage over the classical biparental QTL mapping study is that the reference population for the study is broadened to be more representative of the germplasm of the breeding program.

Complex pedigree mapping is a logical approach for any breeding program where the reference germplasm has been developed by a pedigree breeding process. In this approach the historical pedigree breeding structure created by the breeding program is explicitly defined and used in the testing for the presence of QTL. To implement such a method reliable pedigree records are required. Selected individuals from key positions throughout the pedigree are both genotyped and phenotyped for the traits of interest. For many breeding programs historical phenotypic data may already be available and can be analyzed for use by itself or in combination with specifically designed experiments that directly compare the individuals from different cycles of the pedigree breeding program. While this is a natural reference population for a breeding program, the resulting population structure that emerges from pedigree breeding is highly complex, which makes analyses of these data challenging. Methods for analyses of these pedigree structured data sets have been proposed (Bink et al. 2002; Yi and Xu 2001; Jansen et al. 2003).

Mapping: A path forward
Several examples of candidate gene based genetic association mapping in maize have been reported (Remington et al. 2001; Thornsberry et al. 2001). The next challenge is whole genome scan association mapping, the success of which is also likely to be dependent on the availability of candidate genes. Here it is necessary to identify relevant populations with sufficient linkage disequilibrium to allow association mapping, combined with a technically accessible number of genetic markers, perhaps in the order of 2,000 to 10,000. Heterogeneous stock populations may be suitable (Mott et al. 2000), but their development takes many years. It is likely that elite maize populations may contain sufficient linkage disequilibrium for whole genome scan mapping. However, we need to develop a better understanding of how to deal with population structure issues, which can confound the analysis. Also, a better understanding of the statistical power of different methodologies is needed. Once association mapping has identified candidate regions, validation of the associations can be undertaken simultaneously with higher resolution mapping by targeted genotyping of high-resolution more diverse populations.

In a number of trait mapping studies that have been reported in the literature, a range in the size of QTL effects has been observed, with a few major QTL and many minor QTL. The resulting distribution of these QTL effects can be been approximated by the exponential distribution. This observation has been used by some to argue against the classical infinitesimal polygene model used in quantitative genetics, in favor of alternative models that allow for the presence of major and minor genes. Mackay (2001) gives a recent review of the theoretical bases and merits of this argument, and some alternative sources of evidence that relate to this debate on the appropriate genetic model for quantitative traits. An important consideration is that the observation of a distribution of QTL effect sizes is not decisive in itself, since there are a number of factors, including experimental design issues, that can result in this observation at the level of the QTL estimates, even when the genes themselves do not have such a distribution of effects on the phenotype (Beavis 1998). The question of what is the appropriate genetic model for the standing variation of quantitative traits in the elite populations of a long-term breeding program, and whether this may differ from that of natural populations or diverse crosses, is an open question that is currently under investigation. See Cooper et al. (2004), elsewhere in these proceedings for further discussion of this topic.

Genetic diversity and linkage disequilibrium in elite maize germplasm
All trait mapping approaches, and the ultimate success of their application in MAS, rely on the extent of functional allelic variation for traits and the patterns of linkage disequilibrium in both the reference population of germplasm that is used to establish sequence to phenotype associations and importantly in the target germplasm within which forward selection is conducted. To understand the potential power of association mapping approaches in maize breeding we have examined genetic diversity and linkage disequilibrium for elite maize germplasm. One approach that can be used is to re-sequence a large number of segments of the genome for a representative set of germplasm. The 3'-UTR of maize genes can be used for this purpose (Ching et al. 2002). Short, EST 3'-end derived PCR amplicons can be used for Single Nucleotide Polymorphism (SNP) identification. Using this approach it is found that at the DNA sequence

level, maize is highly diverse and SNPs and insertion-deletion (INDEL) mutants are found at high frequency. In one study the SNP frequency was on average 1 in 60 bp in non-coding segments, and 1 in 120 bp in exons. These estimates are consistent with subsequent larger studies we have conducted. With this high SNP frequency many SNP haplotypes are possible within small sequence segments. However, in general in our elite breeding germplasm we have observed a smaller than expected number of SNP haplotypes, approximately ranging from 3 to 8 per 500-600 bp segment, at most genetic loci. This situation can be favorable for some applications of genome wide association mapping when it is combined with adequate levels of linkage disequilibrium among physically linked sequences.

Large differences in the extent of linkage disequilibrium have been observed in different maize reference populations (Rafalski 2002). In broadly based populations of maize, linkage disequilibrium declines rather rapidly, within two to several hundred bp, or, at most, several kb (Remington et al. 2001; Tenaillon et al. 2001). However, in elite populations, with ancestry that can be traced back in relatively recent history to a narrow set of parents, instances of long range linkage disequilibrium extending for more than 100 kb have been observed. For example, at the Adh1 locus on chromosome 1, linkage disequilibrium extends to significantly more than 100 kb in either direction from the Adh1 gene itself, in two different sets of maize elite inbred lines (Jung et al. 2004). The observation that elite maize populations can exhibit high long-range linkage disequilibrium has implications for association mapping strategies.

The choice between candidate gene-based association mapping and whole genome scanning is determined to a large extent by the amount of linkage disequilibrium in the reference germplasm. In the presence of considerable linkage disequilibrium a whole genome scanning approach is appropriate with practical levels of genome marker coverage. In contrast if linkage disequilibrium declines rapidly with physical distance, as has been reported for broad-based populations, then it may be impractical to achieve sufficient genome marker coverage. In the latter case, an alternative approach is the candidate gene approach to association mapping. Implementing a candidate gene approach requires some level of knowledge of the biology of the traits under consideration. Today this is practical for some traits where underlying biochemical and developmental pathways have been well studied. For example, the involvement of lignin in the silage quality of maize is well established (Cherney et al. 1991). Therefore, genes involved in lignin biosynthesis would be reasonable candidates for the analysis of associations with silage quality.

Molecular Enhanced Breeding: Marker Assisted Selection
The preferred implementation of MAS will depend on the genetic complexity of the target traits. For "simple traits", where phenotypic variation is accounted for by a few major QTL or candidate genes, mapping of these major QTL to a resolution of around 5 cM, may identify the target allele combination to select towards. This strategy appears to have been appropriate in some cases (Castro et al. 2003; Bouchez et al. 2002). For complex traits, such as grain yield and tolerance to many of the abiotic stresses, it is anticipated from the molecular evidence available today that many of the genes responsible for standing genetic variation in elite populations will likely be components of networks of genes determining trait variation and trait contributions to hybrid performance. Furthermore, this standing variation for complex traits is likely to be in part a consequence of sequence variation that influences the regulatory processes involved in determining gene expression and function as well as variation in the structural sequence that codes for the gene products. There has long been discussion within quantitative genetics and breeding of the importance and potential effects of epistasis, in the form of gene-by-gene interactions, and gene-by-environment interactions. While many of the questions that are part of this discussion are not resolved, today we are starting to get a glimpse of potential mechanistic bases of these components of gene action and interaction. In cases where the "context dependencies" that arise from these interactions are important in determining trait phenotypes it can be important for the short-term and long-term outcomes of the breeding program to appropriately accommodate for their effects in the breeding strategy (Podlich et al. 2004). In these situations we anticipate the need for an iterative process of QTL mapping that will involve multiple cycles of trait mapping that is conducted within large reference populations constructed to represent multiple elite crosses.

Challenges for effective molecular breeding

Today the challenge for the design of effective molecular breeding strategies is the same as the challenge that was faced by the previous Pioneer maize breeders when they were improving on the earlier implementations of pedigree breeding methodology, i.e. the need to develop a robust methodology that enhances the sustainable rate of genetic gain for the target traits in the long-term and creates multiple opportunities for successful new products in the short to medium term. We can be more specific on this point by recognizing that the breeding program is continually working with a continuum of traits, some are genetically simple, relatively speaking, while others are genetically more complex. There is a need to balance the breeding effort across these trait targets in order to ensure progress is made on all fronts. It is important to exploit the opportunities that exist today while at the same time tackling the more challenging traits with a long-term view to success.

Technical challenges likely to be a priority for molecular breeding research in breeding programs include:
1. Mapping important traits to appropriate levels of resolution and completeness in the reference breeding population in order to create predictive gene-to-phenotype models of traits that can be used to augment the gains that are currently achievable by phenotypic selection.
2. Achieving high throughput molecular and phenotypic characterization of elite breeding programs.
3. Designing experimental trait mapping strategies and developing appropriate quantitative genetic methods with sufficient power to analyze multiple traits in large data sets that consist of multiple forms and sources of molecular and phenotypic data.
4. Training the next generation of plant breeders to be effective at critically evaluating the application of genomic technologies to enhance the effectiveness of breeding.

In addition to the above, technical challenges that are also likely to be a priority for large commercial programs include:
1. Cost effective scaling up of molecular breeding methods that have proven to enhance the long-term and short-term product development outcomes.
2. Developing a high volume information management infrastructure that enables the breeder to access genetic and phenotypic knowledge of germplasm as needed.
3. Training of the next generation of plant breeders to be effective in molecular enhanced breeding strategies that are larger, more complex and that apply genetic research and breeding methods that are currently beyond the scale of experience of the institutions currently undertaking the formal training of plant breeders.

Now that genomic DNA sequences are available, or will soon be available, for several plant species, understanding the patterns of genetic diversity within each species, and, in the case of cultivated species, within relevant breeding germplasm, is the next challenge. Tools to develop such understanding exist, and their cost is decreasing. Obtaining high-quality phenotypic data remains a significant challenge. Most traits of interest to maize breeders are strongly affected by environment, usually necessitating complex and costly experimental designs. This is a strong argument for the identification of versatile populations, which may be used for association mapping of many traits, and for which high-density genotyping and precision phenotyping will be justified. The work towards understanding the capabilities of this approach is just beginning. In the process we will learn a lot about natural variation affecting quantitative traits, and about diversity and recombination within plant genomes. In turn we will learn a lot about how knowledge of sequence to phenotype associations can be used to enhance the success of maize product development.

References

Beavis, W.D. 1998. QTL analysis: Power, precision and accuracy. In 'Molecular analysis of complex traits'. (Ed. A.H. Paterson), pp. 145-161. (CRC Press, Boca Raton, FL).

Bink M, Uimari P, Sillanpaa J, Janss G, Jansen C (2002). Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theoretical and Applied Genetics 104, 751-762.

Bouchez A, Hospital F, Causse M, Gallais A, Charcosset A (2002). Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics 162, 1945-1959.

Castro AJ, Chen X, Corey A, Filichkina T, Hayes PM, Mundt C, Richardson K, Sandoval-Islas S, Vivar H (2003). Pyramiding and validation of quantitative trait locus (QTL) alleles determining resistance to barley stripe rust: Effects on adult plant resistance. Crop Science 43, 2234-2239.

Cherney JH, Cherney DJR, Akin DE, Axtell JD (1991). Potential of brown-midrib, low lignin mutants for improving forage quality. Advances in Agronomy 46.

Ching A, Caldwell K, Jung M, Smith O, Tingey S, Morgante M, Rafalski A (2002). SNP frequency and haplotype structure of 18 maize genes. BMC Genetics 3, 19.

Cooper M, Podlich DW, Smith OS (2004). Complex traits and gene to phenotype models. Proceedings of the 4th International Crop Science Congress, 26 Sept – 1 Oct 2004, Brisbane, Australia, Published on CDROM.

Duvick DN (1977). Genetic rates of gain in hybrid maize yields during the past 40 years. Maydica 22, 187-196.

Duvick DN (1992). Genetic contributions to advances in yield of U.S. maize. Maydica 37, 69-79.

Duvick DN, Smith JSC, Cooper M (2004). Long-term selection in a commercial hybrid maize breeding program. Plant Breeding Reviews 24(2), 109-151.

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000). Fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289, 85-88.

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

Jung M, Ching A, Bhattramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004). Linkage disequilibrium and sequence diversity in a 500 kbp region around the *adh1* locus in elite maize germplasm. Theoretical and Applied Genetics (In Press).

Mackay TFC (2001). The genetic architecture of quantitative traits. Annual Reviews in Genetics 35, 303-339.

Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. Proceedings of the National Academy of Sciences 97, 12649-12654.

Openshaw SJ and Frascaroli E (1997). QTL detection and marker-assisted selection for complex traits in maize. In 'Proceedings of the 52nd Annual Corn and Sorghum Research Conference'. pp. 44-53. (American Seed Trade Association, Washington D.C., USA).

Podlich DW, Winkler CR, Cooper M (2004). Mapping As You Go: An effective approach for marker-assisted selection of complex traits. Crop Science (In Press).

Rafalski, JA (2002). Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5, 94-100.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley, J, Kresovich S, Goodman MM, Buckler ES (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proceedings of the National Academy of Sciences 98, 11479-11484.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. Mays L.). Proceedings of the National Academy of Sciences 98, 9161-9166.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28, 286-289.

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

Yi N, Xu S (2001). Bayesian mapping of quantitative trait loci under complicated mating designs. Genetics 157, 1759-1771.