Developing Marker-Assisted Selection Strategies for Breeding Hybrid Rice

Yunbi Xu*
RiceTec, Inc., P.O. Box 1305, Alvin, Texas 77512.
Present address: Department of Plant Breeding, Cornell University
Ithaca, New York 14853-1901

I. INTRODUCTION
II. FEATURES OF HYBRID BREEDING
   A. Hybrid Prediction
   B. Selection for Hybrid Performance
   C. Seed Production and Commercialization
   D. Grain Production
III. COMPONENTS OF MARKER-ASSISTED SELECTION
   A. Genetic Markers and Maps
   B. Marker Characterization
   C. Marker-Trait Associations
   D. Genotyping and High-throughput Genotyping Systems
   E. Data Management and Delivery
IV. GERmplasm EVALUATION
   A. Assessing Collection Redundancies and Gaps
   B. Monitoring Genetic Shifts
   C. Identifying Unique Germplasm
   D. Construction of Core Collection
   E. Germplasm Genotyping Database
V. TRAITS REQUIRING TESTCROSSING OR PROGENY TESTING
   A. Fertility Restoration
   B. Outcrossing
   C. Wide Compatibility

*The author gratefully acknowledges Drs. Susan R. McCouch, Barry Tillman, Mark Walton, Jules Janick, and three anonymous reviewers for their critical reading and invaluable comments and suggestions on the manuscript, and Dr. Junjian Ni for his help in collecting references. This article is dedicated to Emeritus Professor Zongtan Shen at Zhejiang University, Hangzhou, China, for his contributions to agriculture education, rice genetics, and plant breeding.
VI. ENVIRONMENT-DEPENDENT TRAITS
   A. Photoperiod/Temperature Sensitivity
   B. Environment-induced Genic Male Sterility
   C. Biotic and Abiotic Stresses

VII. QUALITY TRAITS
   A. Seed Traits
   B. Hybrid Seed Traits
   C. Selection of Quality Traits

VIII. GENE INTROGRESSION AND WHOLE GENOME SELECTION
   A. Gene Introgression
   B. Whole Genome Selection
   C. Selection for Multiple Genes/Traits
   D. Integrated Genetic Mapping and MAS
   E. Response to Selection

IX. PREDICTION OF HYBRID PERFORMANCE AND HETEROSIS
   A. Combining Ability and Heterosis
   B. Genetic Basis of Heterosis
   C. Construction of Heterotic Groups
   D. Hybrid Prediction

X. SEED QUALITY ASSURANCE
   A. Off-type
   B. False Hybrids

XI. GENERAL DISCUSSIONS
   A. Economic Consideration
   B. Bioinformatics and Breeding Database
   C. Opportunities and Challenges

LITERATURE CITED

LIST OF ABBREVIATIONS

AFLP  Amplified fragment length polymorphism
ASI   Anthesis-silking interval
BLUP  Best linear unbiased prediction
BSSS  Iowa (B) Stiff Stalk Synthetic
CHA   Chemical hybridization agent
CMS   Cytoplasmic male sterility
DH    Doubled haploid
DTF   Days-to-flowering
DTH   Days-to-heading
EGMS  Environment-induced genic male sterility
FNP   Functional nucleotide polymorphism
GCA   General combining ability
GD    Genetic distance
IBPGR International Board for Plant Germplasm Resources
IHO   Illinois high oil
3. MARKER-ASSISTED SELECTION STRATEGIES FOR BREEDING HYBRID RICE

I. INTRODUCTION

Exploitation of heterosis or hybrid vigor to increase crop yields started early in the twentieth century with maize. From inbreeding a number of crop plants including maize, George H. Shull developed a perspective on heterosis that he outlined in a 1908 publication entitled “Composition of a Field of Maize.” He argued persuasively that by inbreeding isolated homozygous lines, these lines could be crossed to capitalize on heterosis. East (1908) discussed the danger of inbreeding. Commercially feasible F₁ maize hybrids were developed following Jones’s 1918
proposition of the doubled cross. Average yields increased twofold after double crosses replaced the open-pollinate cultivars (OPCs) during 1935–1965 and increased another twofold after single crosses replaced the double crosses during 1966–1996 (Hallauer 1999). The T cytoplasmic male sterility (CMS) had been used to produce hybrid seeds during 1950–1970. Wide use of this single cytoplasm resource resulted in the epidemic of Southern leaf blight in 1970 caused by a new virulent Race T of Helminthosporium maydis, and an estimated 15% of the maize crop in the United States was destroyed. Hybrid maize is now based on hand detasseling, which avoids the use of CMS female parents. The success of hybrid maize is due to the ability of major seed companies such as Pioneer Hi-Bred International, Inc., to select for yield stability based on extensive testing and the efficiency of seed production that insures profitability. Now over 90% of maize in the United States is hybrid (Duvick 1999).

Hybrid rice was first commercialized in the early 1970s in China. Unlike maize, rice is self-pollinated and hermaphroditic. Hybrid rice breeding has been based on using CMS or environment-induced genic male sterility (EGMS). A breeding system using three-lines (CMS line, CMS maintainer, and CMS restorer) was established by using a male sterile plant discovered from a wild rice species (Oryza rufipogon Griff. or O. f. spontanea) in 1970 by Prof. Longping Yuan and his assistant (Li and Yuan 2000). A two-line hybrid rice system using EGMS was established by using a photoperiod-sensitive genic male sterility (PGMS) mutant discovered in 1973 by Mingsong Shi from a japonica cultivar ‘Nongken 58’ (Shi 1981). Now over 50% of rice in China is hybrid (Li and Yuan 2000). Hybrid rice is also commercialized in other Asian countries and hybrid breeding programs have been established in several South American and North American countries.

Hybrids provide many advantages in a crop production system. The principal benefit is increased yield. In open-pollinating species, one of the most often overlooked benefits is uniformity, an element which has allowed for the rapid expansion of production in many crop plants such as the vegetables. Additional benefits may include stress tolerance and pest resistance and other performance characteristics. Breeders of hybrid crops can react faster and with more options to meet changing markets, customer needs, and production demands. Other advantages of hybrids include the ability to combine useful dominant genes available in different inbred lines, to optimize the expression of genes in the heterozygous state, and to produce unique traits.

Yield increases accounted for about 92% of increased cereal production in the developing world between 1961 and 1990. The yield advantage of hybrids ranged from 15% (maize) to 50% (sunflower), compared
to non-hybrids (Duvick 1999). However, yield growth rates are stagnating in some areas and, in a few cases, falling. A slowdown in the rate of yield increase of major cereals raises concern because increased yields are expected to be the source of increased food production in the future (Reeves et al. 1999). Utilization of heterosis and the improvement of the efficiency of hybrid breeding is one of the ways by which we can attempt to lift the yield ceiling.

Producing F1 hybrids depends on the development of an efficient system for hybridization. There are several approaches to production of hybrids: CMS, EGMS including PGMS and thermo-sensitive genic male sterility (TGMS), protogyny, self incompatibility, chemical hybridization agent (CHA), and hand-emascula- tion systems. No matter which system is used, the central task is to develop and maintain two parents that have desirable traits. One serves as a pollen receptor and the other as a pollen donor for hybrid seed production. All desirable traits for hybrid seed production and heterosis itself, which are not required in pure line breeding, are difficult to measure, compared to other agronomic traits. Genomics in cereals, especially in rice, has created a substantial information base for their improvement. Marker-assisted selection (MAS), the first benefit that breeders can obtain from genomics, has been receiving a great deal of attention and will play an important role in hybrid breeding. The objective of this paper is to discuss the development of molecular marker strategies in breeding hybrid rice by drawing upon information obtained from maize and other crops. Information on the types of marker systems available and their relative merits is included, as is the use of markers in (1) evaluation and characterization of germplasm resources, (2) selection for different types of traits, (3) gene introgression, (4) prediction of hybrid performance, and (5) monitoring of seed quality in the seed production process.

II. FEATURES OF HYBRID BREEDING

A. Hybrid Prediction

The process of plant breeding has developed through several key phases, including unconscious selection in Neolithic times, empirical art during the development and expansion of agriculture, and a predictive science-based approach practiced today. In general, prediction in plant breeding started from the evaluation of progeny performance. Pedigree methods and use of statistical tools in assessing progeny performance brought about a way for plant breeders to quantitatively separate the heritable portion of variation from the non-heritable and thus make parental choices based on heritability (Goldman 2000). As a result, exploitation
of additive genetic variance during the inbreeding process, and dominance variance during the testcrossing phase when assessing performance across multiple testers, has become a standard for cross-pollinated crops. There are several breeding practices that also make plant breeding more predictive. The breeding procedure known as backcrossing makes the progeny highly predictive. Identification of general and specific combining ability in maize helps identify superior inbred lines that can be used more efficiently in breeding new inbreds and hybrids. Wide area testing of hybrids makes it possible to develop widely adapted hybrid maize and improves the efficiency of maize breeding operations.

Hybrid rice breeding has a very different story. Breeders of self-pollinating crops have been highly successful in breeding inbred cultivars. Although inbred breeding procedures were established and used for many years, hybrid rice breeding attempted a completely different approach. However, hybrid breeding procedures established for open-pollinated crops were unsuitable for rice. Testcrossing and wide area testing, which is not a problem for hybrid maize breeding, has been a bottleneck for hybrid rice. Rice breeders have been searching for more reliable prediction methods in their breeding programs.

High performance of hybrid plants results from the complementary action of both parents. Thus, parents with excellent performance per se may not produce desirable hybrids; superior hybrids may come from low-yielding parental lines. The evidence from maize indicates that inbred grain yield is not highly correlated with hybrid grain yield (Hallauer and Miranda 1988). Correlations between midparent values and hybrid means for the 91 crosses studied by Dudley et al. (1992) ranged from 0.46 for ear height to 0.71 for plant height. For grain yield, the correlation was 0.56. These correlations are too small to be of practical use in a breeding program. Despite their low values, the inbred-hybrid yield correlations were positive. They indicated a tendency for high-yielding inbreds to produce high-yielding hybrids. For a specific breeding program, however, the tendency does not make any kind of prediction meaningful.

The second reason for the unpredictability is the lack of a full understanding of the genetic basis of heterosis, which affects all aspects of hybrid performance. As noted in the detailed discussion in Section IX, hybrid performance is positively related to the genetic distance between the parental lines when comparison is based on crosses derived from the intra-heterotic group. This “prediction” does not work, however, when comparison is based on the crosses derived from parental lines belonging to different heterotic groups, ecotypes, or subspecies. Although many hypotheses, such as physiological complementary, overdominance, dominance, and partial dominance, have been proposed as the
genetic explanation of heterosis, “isolated” research conclusions cannot be extended to other populations from the same varietal group, let alone the hybrids derived from very different crosses or species. As a result, limited and less-reliable information often provides contradictory results. As will be discussed later, prediction of hybrid performance depends on a thorough understanding of heterosis from different aspects, including genetics, physiology, and developmental biology.

B. Selection for Hybrid Performance

Hybrid performance depends on genes, and their interactions and combinations. Selection for hybrid performance in breeding programs is based on testcrossing and progeny testing. That is, we breed hybrids through selection of parental lines with desirable agronomic traits. To associate the parental phenotype with hybrid performance, breeders have to cross their candidate breeding lines with several testers, and from the hybrid progeny, determine if the candidates contain the genes required for hybrids and whether the parental combinations produce useful hybrids. This indirect selection, based on testcrossing and progeny testing, is time-consuming and very expensive. Furthermore, the association between the parental line and hybrid from one cross cannot be used to make a prediction about other associations.

A cross of two extremely low-yielding inbreds can give a hybrid with good mid-parent or high-parent heterosis but poor performance, whereas a cross of two high-yielding inbreds might exhibit less mid-parent or high-parent heterosis but nevertheless produce a hybrid with good performance. High-yielding hybrids owe their yield not only to heterosis but also to other heritable factors that are not necessarily influenced by heterosis. For effective selection, one needs to know the relative importance of each genetic contribution—of heterosis and non-heterosis—in individual hybrids (Duvick 1999). Furthermore, when examining yield trends in a time series of successively released hybrids, breeders need to know what portion of the genetic yield gains (if gains are made) is due to increase in heterosis, and what portion to increases in non-heterosis (Duvick 1999).

C. Seed Production and Commercialization

In cultivar breeding, once an outstanding line is identified to be superior to the commercial control, it can be registered as a new commercial cultivar for production. In hybrid breeding, however, a superior hybrid requires not only a high yield potential and desirable agronomic traits such as disease resistance and good quality, but also an economically
viable seed production and maternal production system for each specific hybrid. Simply to exhibit good hybrid performance is insufficient. To be commercially viable, potential lines must produce excellent hybrids and have many traits desirable for seed production, such as high outcrossing rate and good flowering habits. This allows the cost of seed production to be low enough so that the seed producer benefits from the production and sale of hybrid seed.

Requirements for activities such as roguing, detasseling/emasculating, pollination assistance, and other special management systems for seed production and parent multiplication can affect the cost and quality of the final product. This is very important to crops such as rice and wheat where reducing seed cost becomes critical for successful commercialization. Thus, a good seed production system will determine the cost of seed production and therefore the final commercialization of the hybrid.

D. Grain Production

In hybrids, commercial grain is produced from hybrid plants, but hybrid seed is produced from inbred parents. The agronomic practice used for the two production systems could be very different. For example, hybrid rice may have a very different nitrogen response than the parental lines and thus different kinds of nitrogen management are required, depending on the environments in which the hybrid is growing. Low potassium can be a problem for hybrid rice in many areas, although this element may be sufficient enough for inbred lines (Yuan and Chen 1988).

Modern maize hybrids differ from open-pollinated and earlier hybrids primarily in their response to stress. New hybrids have improved water stress performance, are much less prone to silk delay, have significantly lower respiration rates during silking, have longer periods of grain fill, and are higher-yielding under both low- and high-input environments (Tomes 1998). Severe droughts in the American Cornbelt during 1934 and 1936 resulted in poor maize crops; however, hybrids often outperformed their open-pollinated counterparts under these conditions (Goldman 1999). This is due to the strong root system and lodging resistance of hybrids. These characteristics helped hybrid maize quickly replace OPCs. Contrary to maize, the first rice hybrids developed in China and the United States were tall and very prone to lodging if the same nitrogen rate as for inbred cultivars was used. As a result, a very different kind of nitrogen management is required for hybrids to maintain a good plant stand.

In conclusion, hybrid breeding consists of all the components of inbred line breeding plus the additional complexity of recombining
Plant breeders have continually looked to genetics for assistance (Duvick 1996), including quantitative genetics, cytogenetics, tissue culture, and mutation breeding. While utility has come from all new innovations, practical plant breeding based on hybridization and selection continues with little change in its basic structure. Now, a powerful new plant breeding technique, MAS, has appeared that promises to vastly increase selection efficiency. This review will explore the development and use of this MAS strategy in improving hybrid cereals, with an emphasis on rice.

III. COMPONENTS OF MARKER-ASSISTED SELECTION

Key components that are required for an efficient MAS system include (1) suitable genetic markers and their characterization, (2) high-density molecular maps, (3) established marker-trait associations for traits of interest, (4) high-throughput genotyping systems, and (5) functional data analysis and delivery systems.

A. Genetic Markers and Maps

Genetic polymorphism at morphological, cytological, and molecular levels can be used as markers to tag traits, chromosomes, or DNA fragments. Genetic markers have been used for several decades in linkage analysis, gene mapping, gene transfer, and as aids to selection. Many different types of markers have been developed, including morphological variants, protein polymorphisms such as variation at isozyme loci, and DNA polymorphisms. DNA-based markers received the most attention since the first genetic map was established using restriction fragment length polymorphism (RFLP) (Botstein et al. 1980). While DNA markers have many advantages over other types of markers, morphological and protein polymorphisms are still very useful, especially when marker-trait associations have been well established or when desirable agronomic traits are used as markers. There are examples using morphological markers for heterosis-related applications. In this review, the term “markers” includes all types.

DNA-based markers reflect genetic polymorphism at the DNA level, which result from any possible differences existing in nucleotides. Compared with other types of genetic markers, DNA markers have almost no practical limitation in numbers, often have no direct phenotypic effect, and are unaffected by environment. DNA markers can be classified into four different types based on the method used for polymorphism detection: (1) DNA-DNA hybridization such as RFLPs; (2) PCR-based markers
such as random amplified polymorphic DNA (RAPD) using arbitrary primers, or simple sequence repeat (SSR) and sequence tagged site (STS), both of which use specific primers; (3) combining use of PCR and restriction enzyme technique such as amplified fragment length polymorphism (AFLP) or transposon display (TD); and (4) single nucleotide differences that are referred to as single nucleotide polymorphisms (SNPs).

Desirable DNA markers for MAS should meet the following requirements: detection of high frequency of polymorphism, codominance, abundance, whole genome coverage, high duplicability, suitability for high-throughput analysis and multiplexing, technical simplicity, cost effectiveness, requirement of small amount of DNA, and user-friendly (such as suitability for different genotyping systems and facilities). Among all these requirements, codominance is the most important for characterization of F1 hybrids. The two parental inbreds and hybrid combinations can be distinguished unambiguously. For all types of DNA markers mentioned above, SSR satisfies all the requirements. As estimated from a draft rice sequence, the density of SSRs in the genome is approximately one SSR per gene. These markers can be shared internationally through Internet-distributed primer sequences. SSR markers can be genotyped manually using agarose or polyarylamide gels and ethidium bromide or silver staining, or in highly automated facilities using ABI3700 Sequencers. SSR markers can be multiplexed doing PCR and for multiple-sample loading on gels using fluorescent labeling (Coburn et al. 2002). DNA extracted from a small piece of leaf or from single dry seeds will be enough to run several hundreds of markers (Xu et al. 2002). SSR is a more mature technique than SNP, although the latter has great potential for super high-throughput analysis using chip technology.

As more and more genes are cloned, it will be possible to develop molecular markers based on sequences with the gene of interest. Intragenic markers provide several advantages over gene-linked markers. First, there is no recombination between the marker and the gene, or intergenic recombination. Second, multiple alleles can be tagged and distinguished. An example in rice is the SSR marker, RM190, derived from a microsatellite sequence with a splice site in the waxy gene, which is responsible for amylose synthesis, an important grain quality trait for hybrid rice.

Selection of target chromosomal regions based on associated markers (foreground selection) and selection of genetic background for one of the parental genomes (background selection) may need different markers. Markers for foreground selection must be genetically mapped and associated with agronomic traits. Genetic markers revealing multiple bands or representing multiple loci are usually difficult to trace back to the specific allele/locus known to be associated with the trait, particularly when
the population used for MAS is different from the mapping population. These types of markers include RAPDs and AFLPs, which are not well suited for foreground selection. To use marker-trait associations based on these markers, it is best to convert them into types of markers that are more locus-specific, such as STS or SSR markers. For background selection, any type of markers that detect a high rate of polymorphism is useful. Background selection does not require the use of mapped markers as long as they can reveal genome-wide polymorphism.

The efficiency of MAS largely depends on how well markers are linked to the target trait. Construction of a high-density genetic map using high-throughput molecular markers is the first step to a large-scale MAS program. A reference map is required for each crop based on a permanent segregating population that can be shared internationally such that it has the potential for continuous placement of additional markers by the entire community on the map later on. This map should be constructed using markers that are friendly to users. There are two reasons why we need a high-density molecular map. First, a minimum requirement for MAS based on marker-trait association includes a three-marker system: one marker cosegregating with the trait for foreground selection and the other two in each side of it for background selection. Since the target gene can be located in any region of the genome, a dense map is required for identifying this triplet at any position in the genome. Second, markers identified using mapping populations may not be polymorphic in breeding populations derived from other parental lines. To guarantee that the three-marker system will work for other breeding populations, many markers have to be identified around the target region.

To saturate the rice SSR map, an international effort has been initiated through the International Rice Microsatellite Initiative (IRMI) (McCouch et al. 2002), with a goal of characterizing and mapping 2000 microsatellite markers developed from the Monsanto rice sequence database. As a member of the IRMI group, RiceTec, Inc. has undertaken a genetic mapping effort to place IRMI markers onto the existing SSR map, which was constructed using a subset of doubled haploid lines derived from an indica × japonica cross, ‘IR64’/‘Azucena’, and having a framework map consisting of 432 RFLP and SSR markers. A reference genetic map consisting of a thousand SSR markers is now available to rice molecular breeding programs.

B. Marker Characterization

It is not enough to just have thousands of genetic markers in hand. To use molecular markers efficiently, they have to be characterized for many features, including number of alleles, polymorphic information content...
Characterization of molecular markers helps to identify markers close to the genes of importance to breeding programs and to evaluate germplasm and breeding materials. A core set of SSR markers evenly distributed on all rice chromosomes and suited for multiplexing has been established at RiceTec, Inc., and is currently used for evaluation of germplasm accessions, construction of heterotic pools, and MAS in hybrid rice breeding. In this section, marker characterization will be illustrated using an example taken from a rice research project at Cornell University (Xu et al. 1997a,b; Xu et al. 2004). A total of 236 cultivars were collected to represent world genetic diversity, which included two subsets, 125 U.S. rice cultivars and 111 collected from 22 other countries. All germplasm accessions were genotyped using 100 RFLP and 60 SSR markers. The results related to marker characterization are summarized below.

1. Allele Number. The number of alleles at a marker locus is related to the genetic diversity that can be revealed by that marker. The more alleles at a locus, the higher the degree of diversity that can be revealed and the more efficiently closely related lines can be distinguished. A total of 274 alleles were detected at the 100 RFLP loci, with an average of 2.7 alleles per locus. SSR markers detected a total of 714 alleles, with an average of 11.9 alleles per locus. The world collection embodied 99.3% of RFLP and 95.8% of SSR alleles, while the U.S. collection embodied 82% of RFLPs and only 56% of SSR alleles.

2. PIC Value. The relative informativeness of each marker can be evaluated based on its PIC value, which reflects the amount of polymorphism and is a function of the number of alleles and allele frequencies at any given locus. The average PIC value was almost twice as high for SSRs (0.66) as for RFLPs (0.36). Average PIC values for the 100 RFLP markers were 0.40 for the world collection, and 0.21 for the U.S. collection. Average PIC values for the 60 SSR markers were 0.74 for the world collection, and 0.50 for the U.S. cultivars.

3. Informative Markers. Based on PIC values and the number of alleles detected, a set of highly informative markers can be selected such that the same amount of genotyping information can be obtained by surveying fewer molecular markers. Selected markers should be evenly distributed throughout the genome. Based on the 236 × 160 rice dataset, a group of 24 RFLP/SSR markers were selected to represent all 12 chromosomes and
are recommended as a set of highly informative markers for preliminary fingerprinting of rice germplasm and breeding populations.

Molecular markers that are suited for MAS in hybrid breeding should have the same requirements as for other breeding projects. In addition to the requirements discussed in Section IIA for a marker system, there are some other requirements for a marker and a core set of markers. A useful marker should have many alleles per locus (>10 for SSRs), high PIC value (>0.8), suitable difference in allele sizes (4 to 10 bp between any two alleles for SSRs), strong signal for detection, less background or noise signal, and high replicability or reliability. A useful set of markers should provide whole genome coverage, even distribution on each chromosome, and high potentiality for multiplexing.

C. Marker-Trait Associations

Establishment of highly significant marker-trait associations is one of the prerequisites for MAS. Demonstrated linkages between target traits/genes and molecular markers are traditionally based on genetic mapping experiments, and it is important to confirm that these associations are consistent in mapping populations and breeding populations. For efficient MAS, marker(s) should co-segregate or be closely linked with the target trait, with a distance of 2 cM or less. Markers associated with major genes or quantitative trait loci (QTL) in one population may be used directly for MAS in other materials. For genes with relatively small effects, however, cross-population comparison of genes, alleles, and gene effects are required because of multilocus and multiallelic features that characterize most quantitative traits. To find tight marker-trait associations, a two-step process could be involved (especially for quantitative traits). The first step is based on a primary mapping population derived from very diverse parents, often with complicated genetic backgrounds. The second step is based on near isogenic lines that share a common genetic background and differ only at the target locus. There are many factors that are related to the detection of marker-trait associations and the efficiency of MAS. Marker-trait association or trait mapping has been discussed in detail elsewhere (Xu 1992; Tanksley 1993; Xu 1997, 2002; Liu 1998; Lynch and Walsh 1998; Paterson 1998; Flint and Mott 2001).

1. Genetic Backgrounds. Gene mapping requires the ability to extract a genetic signal from background “noise.” Sources of “noise” include variability in the external environments in which plant phenotypes are evaluated and variability due to differences in the internal genetic backgrounds of the individuals in a population. For accurate gene mapping,
the “noise” must be minimized or eliminated. “Controlled” environmental and/or genetic backgrounds are created to help filter the “noise.” Creation of homogeneous genetic backgrounds will help define marker-trait associations. Xu (2002) described five approaches for creation of homogeneous or isogenic backgrounds: backcross-derived near-isogenic lines (NILs), selfing-derived NILs, whole genome selection of permanent populations, mutation, and chromosome substitution. Genetic materials, such as NILs with homogeneous backgrounds, have been used in many different investigations. If NILs are used, interaction between the target gene and other major genes/QTL can be eliminated and only epistasis between multiple target genetic loci needs to be considered. With removal of noise from heterogeneous backgrounds, the proportion of variance explained by the target loci will increase and minor genes (genes with smaller effects) can be identified. By minimizing the disturbance from the genetic background, multiple loci in a single chromosomal region can be separated and their effect on the phenotype can be partitioned. When all genotypic variation comes from the target loci, environmental effects can be estimated.

Heterogeneous genetic backgrounds can also come from populations with different structures, such as F2, doubled haploids (DH), and recombinant inbred lines (RILs), but derived from the same cross, or come from various crosses derived from different cultivars, subspecies, species, and families. Genetic materials with heterogeneous genetic backgrounds can be used to estimate epistasis, detect non-allelic genes, discover multiple alleles, and identify paralogous and orthologous genes. As a contribution to complicated genetic backgrounds, many quantitative traits per se are a complex consisting of several components or subtraits. For example, polygenic sterility in rice can be partitioned into several components, including male and female sterility, or ovary and pollen abortion, so that polygenes can be divided into several components with different functions and, thus, can be handled more easily (Shen and Xu 1992). Genetic backgrounds in a population can also be complicated by the contribution of other related traits.

2. Alleles at Multiple Loci. When multiple loci control a trait, their alleles of positive or negative effect (increasing or decreasing trait value) tend to be dispersed between parents, each with positive alleles at one or some loci but negative alleles at others (Xu and Shen 1992). These dispersed alleles can be cryptic transgressive, which has been found in the parents with similar phenotypes (Xu et al. 1998). In genetic mapping, phenotypic difference between parents is unnecessary for detection of QTL. In many QTL studies, mapping populations were developed without consideration of phenotypic differences between the two parents and
QTL mapped without a statistical test of the parental difference. In most cases where no parental difference is found, QTL are still detected, which could be due to the complementary distributions of positive and negative allelic effects in the parents.

As observed in QTL mapping, on average, about four QTL are identified for each trait in rice (Xu 2002), the same as the average obtained for 176 trial-trait combinations as reviewed by Kearsey and Farquhar (1998). When QTL identified for the same trait are summarized over different projects/populations, this number becomes much larger. For example, plant height has been mapped using 13 populations with 63 QTL reported. Some of the QTL are allelic to each other, that is, they were mapped to the same chromosomal region or intervals of less than 15 cM. After elimination of possible allelic QTL, the total number of QTL for plant height is reduced to 29, with up to five QTL existing on one chromosome (Xu 2002). In contrast, over 50 independent single gene mutations have been identified so far for plant height in rice, as summarized by Kinoshita (1998). Some plant height QTL were co-located with major plant height loci, suggesting that the gene controlling quantitative variation may be the same as those associated with macromutations. This has been demonstrated based on high-resolution mapping and cloning of a QTL for plant height in rice (E. M. Septiningsih and S. R. McCouch at Cornell, pers. commun.). To date, QTL alleles that have been cloned in rice all correspond to previously identified single gene mutants. For example, a photoperiod sensitivity QTL, \( H_d1 \), is allelic to the major gene, \( Se_1 \) (Yano et al. 2000). QTL allelism tests and determination of major-gene and QTL correspondences are facilitated by the availability of high-density molecular maps with a common set of markers shared among researchers.

3. Multiple Alleles at a Locus. Two-parent derived populations in diploid crops have only two alleles segregating at each locus. Identification of multiple alleles requires comparison of populations derived from different crosses. To distinguish alleles identified in one cross from those in another, all alleles must be accurately sized and documented. Rice amylose content, mainly controlled by the \( wx \) gene, is a good example for multiple alleles at a locus. A polymorphic microsatellite was identified in the \( wx \) gene (Bligh et al. 1995) located 55 bp upstream of the putative 5'-leader intron splice site. A total of 16 \( wx \) microsatellite alleles were identified in worldwide rice germplasm (Ayres et al. 1997; Zeng et al. 2000). Now the question is whether the multiple alleles identified at the waxy locus can be associated with specific quantitative effects, including developing an understanding of how each allele interacts with other genes/alleles in the genetic background.
and in response to environment. Using the best-characterized examples, such as wx, the challenge will be to extend this kind of analysis to other traits or genetic loci.

Using gene-based molecular markers that have multiple alleles in gene mapping could help identify multiple alleles at a locus. Genetic mapping studies using different populations have identified some common major genes and QTL. It is necessary, however, to further clarify whether common or different functional alleles were identified at those loci. Reporting the sizes of associated (closely linked) alleles and using allele-rich markers in marker-trait association studies will provide a baseline of information required for this clarification, with the assumption that each marker has a corresponding allele at the trait locus.

There are many reasons why close marker-trait associations are required: (1) chromosomal location associated with the trait must be reduced to a manageable piece of DNA if cloning of specific genes is necessary; (2) to identify all the related genes for a specific trait, a high-density genetic map is required because the fewer markers are used, the smaller proportion of genetic factors contributing to that trait will be sampled; (3) large genetic distances between markers and target traits will contribute to the rapid decrease of MAS efficiency after several successive cycles of selection; and (4) to minimize linkage drag involved in gene introgression, closely linked markers around the target region are needed.

QTL mapping presumes accurate phenotypic scoring methods, something that can be difficult to optimize and even more difficult to keep consistent for months or years. Just a few misscored individuals can totally confound QTL discovery and placement (Young 1999). This is also true for fine mapping of major genes for map-based cloning, where misscoring of several plants in a population with thousands of individuals will result in a large error (up to one cM) in estimating genetic distances. High levels of accuracy are required to dissect a chromosomal region associated with a given trait and narrow down the candidate region to a single contig, that is, a set of clones that can be assembled into a linear order.

D. Genotyping and High-throughput Genotyping Systems

To make marker-based technology practical for breeding applications, an automated genotyping system is required. Such an automated system using SSRs to genotype rice germplasm and breeding populations was developed, through improved DNA extraction and loading with multiple-PCR products. This system brings the cost per data point down to as low as $0.30 with daily data output up to 4608 data points per ABI
3100 Sequencer (Xu et al. 2002). This level of efficiency makes it possible to genotype thousands of individual plants with a panel of eight SSR markers in a week using two sequencers.

SNPs have gained wide acceptance as genetic markers for use in linkage and association studies, especially for human genetics. High-throughput SNP genotyping has great potential for many applications, including MAS on the basis of whole genome approaches. This has led to a requirement for high-throughput SNP genotyping platforms. Development of such a platform depends on coupling reliable chemical assays with an appropriate detection system to maximize efficiency with respect to accuracy, speed, and cost. Current technology platforms are able to deliver throughputs in excess of 100,000 genotypes per day, with an accuracy of >99%, at a cost of 20–30 cents per genotype (Jenkins and Gibson 2002). In order to meet the demands of the coming years, however, genotyping platforms need to deliver throughputs in the order of one million genotypes per day at a cost of only a few cents per genotype. In addition, DNA template requirements must be minimized such that hundreds of thousands of SNPs can be interrogated using a relatively small amount of genomic DNA. Jenkins and Gibson (2002) predicted that the next generation of high-throughput genotyping platforms would exploit large-scale multiplex reactions and solid phase assay detection systems. Released genomic sequences of rice and Arabidopsis can be used to develop gene-based SNPs for other related species.

E. Data Management and Delivery

To handle the daily data flow from the lab to the breeder and integrate information from molecular markers, genetic mapping, and phenotyping, many informatics tools are needed. For efficient data management and delivery, it is important for all researchers to follow general rules through all these procedures. A standard reporting system is also critical for comparative genomics, QTL allelism tests, data sharing and mining, and the association between major genes and QTL. As discussed by Xu (2002), a standard system for marker-trait association should include associated alleles and allele characterization such as allele sizes, gene effects, variation explained by each gene or all genes in the model, gene interaction if more than one gene is identified, and genotype × environment interaction if more than one environment is involved. Genetic information should be shared and combined with data generated in plant breeding, for example, germplasm diversity, mapping populations, pedigrees, graphical genotypes, mutants, and other genetic stocks.

With several thousand data points flowing out of the laboratory every day, timely scoring and delivery of the results to breeders are basic
requirements for a high-efficiency breeding system. Well-trained assistants for genotyping and scoring, coupled with research scientists who can analyze data in meaningful ways, are the key components for a data management and delivery system. A laboratory with well-equipped facilities has to be also well equipped with qualified personnel and software required for data integration, manipulation, analysis, and mining. Timely delivery of data to the breeder is also equally important, because in many cases the time window the breeder can use for selection is very limited. With the high-throughput genotyping and data management systems currently available, it takes about a week to generate and analyze data for a breeding-related population consisting of several hundred individuals. This includes activities ranging from leaf tissue harvesting to DNA extraction, genotyping, data scoring, analyzing, summarizing, and reporting.

IV. GERMPLASM EVALUATION

Germplasm resources represent the genetic variability required for continuous improvement of crop plants. The old paradigm for evaluation and utilization of germplasm involves looking for a clearly defined character by screening entries from a genebank. This approach works well when the trait of interest is controlled by one or few genes. For traits such as yield, genetically controlled by many genes, it is impossible to identify all these genes phenotypically because each gene has a relatively small but similar effect. As a result, exotic germplasm, which is perceived to be a poor bet for the improvement of most traits based on phenotypic examination, may contain some favorable genes (alleles) that lie buried amidst the thousands of accessions maintained in genebanks (de Vicente and Tanksley 1993; Tanksley and McCouch 1997; Xiao et al. 1998). The new paradigm involves looking for genes using molecular markers and/or the integrative power of QTL analysis, which can be used to extract superior genes (alleles) from the inferior germplasm accessions. Molecular marker-assisted germplasm evaluation aims to complement phenotypic evaluation by helping define the genetic architecture of germplasm resources and by identifying alleles that are associated with key phenotypic traits. Molecular markers may allow for characterization based on gene, genotype, and genome, which provide more accurate and detailed information than classical phenotypic or passport data. Many features revealed by molecular markers, such as unique alleles, allele frequency, and heterozygosity at marker loci, mirror the genetic loci for heterosis and the traits of agronomic importance. On a more fundamental level, molecular marker information may lead
to the identification of useful genes contained in collections and transfer of those genes into well-adapted cultivars. Bretting and Widrlechner (1995) comprehensively reviewed genetic markers and their application in plant genetic resource management, including procedures related to acquisition/distribution, maintenance, and utilization. In this section, discussion will be focused on the aspects more related to breeding applications.

A. Assessing Collection Redundancies and Gaps

With a large number of germplasm accessions available for each cultivated plant, it is likely that many represent duplicate or nearly identical samples of the same cultivar, while others embody rare alleles or highly unusual allele combinations, with many genes or alleles still missing in current collections. According to the International Board for Plant Genetic Resources (IBPGR), over 3.6 million germplasm accessions for different crop species are conserved at international and national genebanks, which include 90,000 for rice, 120,000 for wheat, and 25,000 for maize (Iwanaga 1993). Evaluation of genetic diversity will help in the understanding of genetic structure of existing collections and design acquisition strategies. In particular, calculation of genetic distance (GD) can be used to identify particularly divergent subpopulations that might harbor valuable genetic variation that is underrepresented in current holdings.

Redundancy germplasm accessions exist in many germplasm collections because of different names for the same cultivars or duplicate samplings of the same accessions. Pedigree-related cultivars, siblines, and NILs may represent another type of redundancy because they are genotypically duplicated at most of the genetic loci. For example, U.S. rice cultivars ‘M5’, ‘M301’, ‘M103’, ‘S201’, ‘Calrose’, ‘Calrose 76’, ‘CS-M3’, and ‘Calmochi-202’ shared the same panel of alleles at all 100 RFLP loci tested by Xu et al. (2004). All these cultivars can be traced back to a common ancestor, ‘Caloro’. No genetic polymorphism could be detected either at any of the 60 SSR loci between ‘Calrose’ and ‘Calrose 76’ (Xu et al. 2004), because they are isolines, with ‘Calrose 76’ representing a variant derived from ‘Calrose’ via chemical mutagenesis (Rutger et al. 1977). Using 15 SSR markers, Dean et al. (1999) assayed 19 sorghum [Sorghum bicolor (L.) Moench] accessions identified as “Orange” presently maintained by the U.S. National Plant Germplasm System (NPGS). They found most accessions are genetically distinct, but two redundant groups were found. The variance analysis also indicated that it should be possible to reduce the number of Orange accessions held by NPGS by almost half without seriously jeopardizing the overall amount
of genetic variation contained in these holdings. Chavarriaga-Aquirre et al. (1999) evaluated genetic diversity and redundancy in a cassava core collection. The core collection (630 accessions) was selected from the base collection (over 5500 accessions) on the basis of diversity of origin (country and geographic), morphology, isozyme patterns, and specific agronomic criteria. A small number (1.34%) of potential duplicates were identified from the core collection through isozyme and AFLP profiles.

Different germplasm collections can be compared for the frequencies of alleles at all marker loci so that distinctive alleles, allele combinations, and allele frequency patterns can be identified. Chromosomal regions containing loci that show the greatest changes in allele frequency between the collections can be located. The rationale for this analysis is to define the genomic regions where selection under the environment had given rise to allele combinations or allele frequency patterns that distinguished a group of accessions with less diversity from more diverse accession groups. In the rice example (Xu et al. 2004), alleles at two RFLP loci (a 6.5 kb allele at CDO686/HindIII and a 6.0 kb allele at BCD808/XbaI) and alleles at six SSR loci were represented at frequencies of 17.1% to 33.6% in the world collection, but had been completely lost in the U.S. cultivars. When low-frequency or underrepresented alleles are defined as those that occur in four or fewer U.S. cultivars but are very frequent (i.e., >17% for SSR and >30% for RFLP) in the world collection, alleles underrepresented in the U.S. collection were found at 19 RFLP and 18 SSR loci. Three U.S. cultivars, ‘Della’, ‘Rexmont’, and ‘Caloro’, retained 34 of the 37 low-frequency alleles. ‘Della’ alone retained 24 (64.9%) of them, which could be traced back to two of its ancestral cultivars, ‘Rexoro’ and ‘Delius’. The U.S. rice cultivars that were developed from a small set of germplasm introductions help explain why these cultivars retained an unusually large number of alleles that show a decline in frequency among later developed U.S. cultivars.

B. Monitoring Genetic Shifts

Maintaining genetic diversity and preventing genetic drifts is one of the most important objectives for germplasm conservation. In open-pollinated species, deviations from random mating, primarily in the form of assortative or consanguineous matings, needs to be monitored during germplasm regeneration. In maize, deviations from random mating were widely studied with emphasis on detailed multilocus isozyme analyses of one or two synthetic or open-pollinated maize cultivars (Kahler et al. 1984; Pollak et al. 1984; Bijlsma et al. 1986). In general, levels of selfing did not exceed those expected under random-mating mod-
els, but significant deviations were caused by temporal variation in the pollen pool or by gametophytic selection.

The genetic profiles of germplasm accessions can change during the course of medium- or long-term storage. Storage effects fall into three broad categories: (1) the occurrence of mutations, (2) the occurrence of chromosomal aberrations, and (3) shifts in gene frequencies resulting from differential genotypic viability in heterogeneous populations (Roos 1988). After a comprehensive review of storage effects on seeds, Roos (1988) found little evidence for heritable changes in germplasm attributable to storage-induced chromosomal aberrations, and noted “little need for concern about mutation as a significant factor in altering the composition of germplasm collection.” However, differential seed longevity can markedly reduce genetic variability over time (Bretting and Widrlechner 1995). This is well documented by experiments involving mixtures of eight bean lines (Roos 1984) and four seed storage protein genotypes within a cultivar of wheat (Stoyanova 1991).

Genetic shifts can be caused by in vitro culture. The genetic stability of germplasm maintained in tissue culture (in vitro) has generally been monitored with karyotypic markers such as chromosome number and morphology (D’Amato 1975), because cytological variability has been considered a primary cause of somaclonal variation. Lassner and Orton (1983) reported that isozymatically identical in vitro cultures of celery were markedly variable cytologically. This finding should reinforce the concept that the genetic stability of in vitro cultures should be monitored with a battery of different genetic markers, particularly those DNA markers that collectively span the whole genome (Bretting and Widrlechner 1995).

A certain level of heterogeneity could exist in germplasm accessions that are mainly self-pollinated, which provides a buffer for maintaining genetic diversity and preventing genetic drifts. Monitoring heterogeneous accessions will help develop strategies for regeneration of germplasm samples without loss of the allelic diversity provided by heterogeneity. In general, traditional cultivars had a higher level of heterozygosity, as reported in rice by Olufowote et al. (1997). Genetic diversity resulting from the heterozygosity was also found within inbred lines from different sources in rice (Olufowote et al. 1997) and maize (Gethi et al. 2002). As reported by Xu et al. (2004), a total of 120 (50.6%) of the 236 rice accessions was found heterozygous at one or more RFLP or SSR loci, and the number of heterozygous loci detected in a single rice accession ranged from 0 to 39 (25.3% of the 160 loci). These heterozygous allele patterns could indicate either seed mixtures or true heterozygosity remaining in these cultivars despite the fact that all accessions had been purified.
C. Identifying Unique Germplasm

Progress in hybrid breeding is demonstrated by the development of new parental lines and hybrids that are superior for one or more characteristics once a commercialized hybrid system has already been established. Progress depends on (1) discovery and generation of genetic variation for heterosis and agronomic traits and (2) accurate selection of rare genotypes that possess new or improved attributes due to superior combinations of alleles at multiple loci. Over the past century, the development and successful application of modern breeding methodologies has produced the high-yielding cultivars and hybrids on which modern farming is based. As the demand for uniform performance and grain quality has increased, new cultivars and hybrids are increasingly derived from adapted, genetically related, and elite modern cultivars/hybrids, while genetically more variable, but less productive, primitive ancestors are excluded from most breeding programs (Tanksley and McCouch 1997). In a study of pedigree relationships among 140 U.S. rice accessions, Dilday (1990) concluded that all parental germplasm in public cultivars used in the southern United States today could be traced back to 22 plant introductions in the early 1900s, and those used in California could be traced to 23 introductions. The same situation is true for soybean and wheat. Virtually all modern U.S. soybean cultivars can be traced back to a dozen strains from a small area in northeastern China, and the majority of hard red winter wheat cultivars in the United States originated from just two lines imported from Poland and Russia (Tanksley and McCouch 1997). To broaden the genetic base of specific cultivated species, the genetic diversity within collections must be assessed in the context of the total available genetic diversity for each species. With the use of DNA profiles, the genetic uniqueness of each accession in a germplasm collection or in a population can be determined, and the identity and frequency of individual alleles can be clearly described and characterized (Brown and Kresovich 1996; Smith and Helentjaris 1996).

The sampling of exotic germplasm should emphasize the genetic composition rather than the appearance of exotic accessions. Accessions with DNA profiles most distinct from that of modern germplasm are likely to contain the greatest number of novel alleles. Assuming that most marker alleles having the same molecular weight are likely to be common by descent in a specific varietal group, we are able to trace alleles that are frequent in one specific collection but existed in low frequency or not at all in the other. Examination of the chromosomal distribution of the loci harboring underrepresented alleles indicated that underrepresented alleles in the U.S. rice cultivars as discussed in previous sections tended to cluster on 11 chromosomal fragments (Xu
et al. 2004). This raises the question about what genes are located in these regions and whether U.S. breeders have consciously or unconsciously narrowed the range of genetic variation in these regions. It also suggests that molecular marker analysis could be used to identify parents harboring rare or novel alleles in these regions so that the functional significance of the resident genes could be determined using both traditional crossing and sequence-based genomics approaches. Considering the allele frequency profiles across all cultivars will give us some idea of which cultivars may retain or contain the rare genes/alleles and whether these alleles may be important to our future breeding programs.

The germplasm that holds unique alleles may contain unique genetic variation required for trait improvement. In the rice example (Xu et al. 2004), 15 (6.4%) of the rice accessions in the 236 × 160 data set had unique alleles for at least one RFLP locus, with a total of 21 unique RFLP alleles were found. Eighty-one (34.3%) rice accessions had unique SSR alleles with a total of 153 unique SSR alleles identified. The germplasm accessions identified with unique alleles have unique geographical origins with high genetic diversity and could have potential use in the exploitation of heterosis.

Genetic similarity between any two cultivars can be calculated as the proportion of shared alleles. Theoretically, shared allele frequency (SAF) is positively related to the number of cultivars in the analysis and negatively related to the informativeness of the markers. The most similar accessions should share alleles at almost all marker loci, while the least similar accessions should have none of the alleles in common. When evaluating genetic similarity, SAFs are averaged over all possible pairs of cultivars. A smaller average similarity indicates a greater genetic difference with respect to the rest of the cultivars in the collection. Based on the averaged SAF, the most diverse accessions can be selected to represent cultivars that host the least-frequent alleles and are genotypically most different from other accessions. From 236 rice cultivars, Xu et al. (2004) selected the 16 most diverse accessions (with SAF < 50%) based on RFLP markers and 49 accessions based on SSR markers. Most of these selections, such as ‘Caloro’, ‘Cina’, ‘Badkalamkati’, ‘DGWG’, and ‘TN1’, were ancestral cultivars that had been used as parents in breeding programs over 40 years ago, and none of the selections is from the U.S. collection that has a much narrower genetic basis.

D. Construction of Core Collection

According to its original definition (Frankel 1984), a core subset of a germplasm collection contains, with minimal redundancy, most of the entire collection’s genetic diversity. Several different methods have
been used to construct core collections (Crossa et al. 1995; Hamon et al. 1995; Schoen and Brown 1995; van Hintum 1995). Construction of a core collection has a target of selecting approximately 10% of the germplasm accessions to represent at least 70% of the genetic variability (e.g., Brown 1989a,b). In addition to phenotypic evaluation, molecular marker technology provides us with a new tool to construct a core collection that can represent most of the genetic diversity at molecular level. Combining the use of different types of markers that reveal different levels of genetic diversity will help select a core collection to better represent genome-wide diversity. Shared allele frequency and frequency for hosting rare alleles are two important criteria that can be used to construct the core collection.

In the rice example given in Section IIIB, accessions were selected based on the frequency of unique RFLP and SSR alleles and shared allele frequency. Subsets of various sizes were selected (representing 5% to 50% of each of the U.S. and world collections), using random selection as a control. For each sample size, 200 replications were analyzed through a resampling technique and the number of alleles in each subgroup was compared to the total number of alleles identified in the larger collection from which the subsets were sampled. The following conclusions were obtained (Xu et al. 2004): (1) more samples are needed to represent a diverse germplasm collection (the world collection) than a germplasm collection containing more pedigree-related cultivars (the U.S. collection), (2) combining use of shared allele frequency and unique alleles improves representativeness of a core collection, (3) core collections selected on the basis of shared allele frequency require much fewer samples than random selection for the same level of representativeness, and (4) more samples are needed to represent genetic diversity at marker loci that reveal a higher level of polymorphism. A 31-cultivar subset (13% of the entire collection), selected on the basis of both shared allele frequencies and number of unique alleles detected, represented 94.9% of RFLP alleles and 74.4% of SSR alleles. It can be expected that selection criteria based on additional sources of information will further improve the value and representativeness of core collections.

When molecular markers are developed from DNA sequences with unknown function, identical marker alleles in two collections may not mean that the two collections share identical genes that are linked to the marker locus. There is a potential risk that genetic variation for important phenotypic traits could be lost if core collections are based only on those markers. As the genome sequence is deciphered and the function of many genes is determined, intragenic markers will become available for many genes. Core collections of germplasm constructed using intra-
genic markers and functional alleles will represent a “core collection” of genes. As gene structure-function relationships are clarified with greater precision, it will be possible to focus attention on genetic diversity within the active sites of a structural gene or within key promoter regions. This will make it productive to screen large germplasm collections for functional nucleotide polymorphism (FNP), targeting the search for alleles that are likely to be phenotypically relevant at specific loci. From a primary collection, a user who had identified an accession or accessions of interest would move to the next level of information, where clusters of germplasm known to represent a broader spectrum of diversity within a specific gene pool or a specific trait could be defined. The second level of investigation could be conducted using carefully designed sets of molecular markers known to target specific traits or regions of the genome. The construction of core collections may help establish heterotic groups and choose parents for establishing base populations.

In the age of genomics, the context of germplasm resources should be expanded to include whole plants, seeds, plant parts, tissue, and clones, from distinct species and synthetic germplasm and all types of mutants. An extreme example is DNA or protein sequence. The ultimate goal of germplasm conservation is to maintain the genes and gene combinations. With a suitable vehicle, germplasm could be maintained in the form of a gene sequence. Genetic manipulation allows gene flow across the reproductive barriers existing among distant species. So germplasm evaluation is not necessarily defined for each species or crop (Xu and Luo 2002). The issues discussed in this section, currently suitable for each crop, can be extended to all vehicles of genetic information such as tissues, clones, genes, and sequences.

E. Germplasm Genotyping Database

Plant geneticists and breeders may use the data from a germplasm evaluation project as a guide in choosing the most efficient crosses for genetic studies and breeding. For instance, it provides preliminary polymorphism data for many pairwise combinations of parents. Theoretically, the dataset with 236 rice cultivars provides polymorphism surveys for $236 \times 235/2 = 27,730$ possible cross combinations, including thousands of indica/indica, indica/japonica, and japonica/japonica crosses. With an increasing number of markers surveyed on a variety of germplasm accessions and as more data flows into the database from multiple sources, it is increasingly possible to determine the genetic constitution and genetic relationships among a wide range of parental lines, cultivars,
and wild relatives. This also provides the foundation for developing hypotheses based on association genetics to relate agronomically important phenotypes to the presence or absence of specific molecular marker alleles.

A comprehensive DNA fingerprinting of crop gene pools, including as many cultivars, hybrid parents, and progeny as possible, is the first step for using MAS in hybrid breeding. These data can be integrated with both phenotypic information and pedigree information. A database of DNA marker alleles for the elite gene pool of a crop provides information on specific DNA polymorphisms that is needed to design, execute, and analyze genetic mapping experiments, targeted at specific traits or specific crosses. The same database serves as a classification tool, describing the overall levels and patterns of variation within the crop gene pool and illustrating subdivisions within a gene pool such as heterotic groups. Such information is useful in making predictions about the performance of new cultivars and hybrids, or selecting parents for crosses that are likely to yield new gene combinations, or afford an optimal degree of heterosis.

An efficient approach to the screening of germplasm involves the ability to rapidly create a nested series of core collections, based on information about geographical, phenotypic, and genotypic diversity stored in a database. The construction of such a system would require a large-scale effort to provide genotypic information using a standard set of markers that could serve as a reference point. As new markers and marker systems were developed, they could be overlaid onto the essential framework of diversity established previously. An increasingly powerful information system could be developed if data models were made explicit and the data structures were modular so that new types of genetic information could be readily incorporated as they became available. As we have seen from the rice example (Xu et al. 2004), RFLP and SSR markers provide complementary information about overall genetic diversity, but one marker system may have specific advantages that recommend it for a particular type of study. By accumulating historical information in a systematic way, germplasm collections would rapidly gain in value because they could be screened computationally for essential molecular and phenotypic characteristics of interest.

Databases for whole genomic sequences for several important species, both dicots and monocots, are available, with more being added, allowing directed discovery of genes in higher plants and classification of alleles present in a wide range of breeding germplasm. As indicated by Sorrells and Wilson (1997), identification of the genes controlling a trait and knowledge of their DNA sequence would facilitate classification of variation in a germplasm pool based on gene fingerprinting or charac-
terization of variation in key DNA sequences. Classification of sequence variants within genes such as SNPs at a large number of targeted loci would substantially reduce the amount of work required to determine their relative breeding value and lead to the identification of superior alleles.

V. TRAITS REQUIRING TESTCROSSING OR PROGENY TESTING

Improvements in inbreds per se will play an increasingly important role in breeding hybrid crops. Bringing dominant genes for non-heterotic traits such as disease resistance into inbreds will enhance the overall performance of hybrids. MAS for non-heterotic traits should be performed the same as those in breeding cultivars. Although there is every reason to believe that plant breeding in the 21st century will still depend, to a great extent, on conventional methods for phenotypic selection, molecular biology could help identify and manipulate favorable alleles and select the traits that are not measurable under normal environments with conventional methods. Using molecular markers in plant breeding has been discussed elsewhere (Beckmann and Soller 1986; Paterson et al. 1991; Dudley 1993; Stuber 1994; Xu 1994; Xu and Zhu 1994; Lee 1995; Paterson 1996; Hospital and Charcosset 1997; Mackill and Ni 2001; Xu 2002). Xu (2002) described six situations that are suitable for MAS with the current knowledge available. These include selection without testcrossing or a progeny test; selection independent of environments; selection without laborious fieldwork or intensive laboratory work; selection at an earlier breeding stage; selection for multiple genes and/or multiple traits; and whole genome selection. Selection for traits requiring testcrossing and/or a progeny test will be discussed in this section.

In breeding hybrids, many traits need testcrossing and progeny testing for unambiguous identification. Typical examples for all hybrid crops include testcrossing for screening of heterosis and combining ability. For hybrid crops based on a cytoplasmic male sterility system, male-sterility restoration is at the top of the testcross list. For hybrid crops that use distant crosses with hybridization barriers, for example, partial sterility in \textit{indica} \times \textit{japonica} crosses of rice, a testcross is required to find genes related to the hybridization barrier.

In testcrossing, each candidate plant or family will be crossed to testers and then its genotype will be inferred from a progeny test in the next season. Each candidate must be harvested and maintained separately and only the plants/families with the target trait will advance to
the next procedure. Testcrossing may continue for several generations until the selected plants reach a certain level of homozygosity. A successful breeding program would take more than 50% of the breeders’ effort for testcrossing and progeny testing. Using MAS, one might reduce or eliminate testcrossing and/or progeny testing for traits controlled by major genes.

### A. Fertility Restoration

Many important crop species, including rice, sorghum, and sunflower, depend on CMS and its fertility restoration for hybrid seed production. As indicated above, a large amount of testcrossing and progeny testing is involved in breeding CMS lines and their restorers. Testcrossing can start as early as with the F$_2$ generation. F$_2$ plants will be selected first for other agronomic traits, and selected plants are testcrossed to maintainer lines for CMS maintaining ability or to restorer lines for restorability. The testcross progeny will be planted the following season for fertility observation. Only the plants with complete sterile testcross progeny (for CMS) or completely fertile testcross progeny (for restorability) will be moved to the next breeding procedure. MAS can be used to replace testcross and progeny testing if markers closely linked to fertility restoration are identified. Table 3.1 lists crops and their restorability genes that have been associated with molecular markers. A total of 12 crop species were reported with molecular markers identified for fertility restoration, including maize, rice, sorghum, wheat, barley, rye, sunflower, oilseed rape, sugar beet, and onion. Genes for fertility restoration in rice have been found each on chromosomes 1, 2, 3, 4, and 5, and two on chromosome 10. Many markers are closely linked with fertility restoration and can be directly used for MAS. In rice, an RFLP marker RG140 with PvuII digestion linked with the $Rf_3$ on chromosome 1 was useful for increasing screening efficiency for restorers (Virmani 2002). Although many genes for fertility restoration have been reported as QTL without genetic distances available, associated markers still provide useful information for MAS. In several cases, RAPD markers have been converted to STS or sequence characterized amplified region (SCAR) markers that are more suitable for MAS.

### B. Outcrossing

Evolutionary change in plant mating systems from outcrossing (cross-pollination) to inbreeding (self-pollination) has occurred frequently throughout the history of flowering plants and has been described as the
### Table 3.1. Molecular markers associated with fertility restoration in crops.

<table>
<thead>
<tr>
<th>Species</th>
<th>CMS</th>
<th>R gene</th>
<th>Marker type</th>
<th>Linked markers</th>
<th>Chromosome or Distance (linkage group)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Allium cepa</em></td>
<td>S</td>
<td></td>
<td>RFLP</td>
<td>AOB210, API65</td>
<td>(B)</td>
<td>King et al. 1998</td>
</tr>
<tr>
<td><em>Beta vulgaris</em></td>
<td>Owen</td>
<td>X</td>
<td>RFLP</td>
<td>pKP1238</td>
<td>(3)</td>
<td>Pillen et al. 1993</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td></td>
<td>RFLP</td>
<td>K11-1000, pKP753</td>
<td>4, 5.2, 1.7</td>
<td>Laporte et al. 1998</td>
</tr>
<tr>
<td><em>Brassica napus</em></td>
<td>Owen</td>
<td>2QTL</td>
<td>RFLP</td>
<td>OPC02-1150, RFLP</td>
<td>3, 4</td>
<td>Hjerding-Panagopoulos et al. 2002</td>
</tr>
<tr>
<td><em>Helianthus annuus</em></td>
<td>PET1</td>
<td>Rf1</td>
<td>RAPD</td>
<td>OPC07-900, OPD02-1000</td>
<td>(6)</td>
<td>Ji et al. 1996</td>
</tr>
<tr>
<td></td>
<td>PEF1</td>
<td></td>
<td>RFLP</td>
<td>OPC02-1150, OPC03-150</td>
<td>4, 10.8, 5.4</td>
<td>Quillet et al. 1995</td>
</tr>
<tr>
<td><em>Raphanus sativus</em></td>
<td>Rapha</td>
<td>RFLP</td>
<td>RAPD</td>
<td>OPC02-1150, OPC03-150</td>
<td>1.2</td>
<td>Murayana et al. 1999</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>BT</td>
<td>Rf1</td>
<td>RFLP</td>
<td>C2155, C1361</td>
<td>10</td>
<td>Kurata et al. 1994</td>
</tr>
<tr>
<td></td>
<td>BT</td>
<td>Rf1</td>
<td>SSR</td>
<td>OSRRf</td>
<td>3.5, 3.9</td>
<td>Akagi et al. 1996</td>
</tr>
<tr>
<td></td>
<td>HL</td>
<td></td>
<td>SSR</td>
<td>RM258</td>
<td>7.8</td>
<td>Huang et al. 1999</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td></td>
<td>RFLP</td>
<td>RG69a, RG413</td>
<td>3, 10.4, 5.4</td>
<td>Li et al. 1996</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td></td>
<td>RFLP</td>
<td>C22, RG4449d</td>
<td>4, 10.4, 5.4</td>
<td>Li et al. 1996</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td></td>
<td>RFLP</td>
<td>RG69A-RG413</td>
<td>3, 10.4, 5.4</td>
<td>Li et al. 1996</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td></td>
<td>RFLP</td>
<td>C22, RG449D</td>
<td>4, 10.4, 5.4</td>
<td>Li et al. 1996</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td></td>
<td>RFLP</td>
<td>RG435-RG172A</td>
<td>5, 10.4, 5.4</td>
<td>Li et al. 1996</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>Rf3</td>
<td>RAPD, RFLP</td>
<td>RG532, RG140/RG458</td>
<td>1, 1.4, 1.9</td>
<td>Zhang et al. 1997</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Species</th>
<th>CMS</th>
<th>Marker type</th>
<th>Linked markers</th>
<th>Chromosome or Distance (linkage group)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Oryza sativa</em></td>
<td>WA</td>
<td>RFLP</td>
<td>RG532, R173</td>
<td>1</td>
<td>6.0, 18.4</td>
</tr>
<tr>
<td>(cont.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yao et al. 1997</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>RFLP</td>
<td>G403, C234</td>
<td>10</td>
<td>3.3, 19.1</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>RFLP</td>
<td>C1361, S11019</td>
<td>10L</td>
<td>Tan et al. 1998</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>RFLP</td>
<td>R2309, RG257</td>
<td>10S</td>
<td>Tan et al. 1998</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>SSR</td>
<td>RM171, RM228</td>
<td>10L</td>
<td>3.7, 3.4</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>SSR</td>
<td>RM244</td>
<td>10S</td>
<td>Jing et al. 2001</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>RFLP</td>
<td>RG374, RG394</td>
<td>1</td>
<td>Shen et al. 1998</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>SSR</td>
<td>RM244</td>
<td>1</td>
<td>Jing et al. 2001</td>
</tr>
<tr>
<td></td>
<td>WA</td>
<td>SSR</td>
<td>RM244</td>
<td>1</td>
<td>Jing et al. 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFLP</td>
<td>R335F/UBC487</td>
<td>(K)</td>
<td>He et al. 1995</td>
</tr>
<tr>
<td><em>Phaseolus vulgaris</em></td>
<td>Ci</td>
<td>RAPD,</td>
<td>Bng228,</td>
<td>4.5, 0</td>
<td>Jia et al. 1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFLP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Secale cereale</em></td>
<td>G</td>
<td>RAPD,</td>
<td></td>
<td>4RL</td>
<td>Börner et al. 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFLP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>RFLP</td>
<td>PSR596-PSR634</td>
<td>1RS</td>
<td>Miedaner et al. 2000</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>RFLP</td>
<td>PSR899-MWG573</td>
<td>4RL</td>
<td>Miedaner et al. 2000</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em></td>
<td>A1</td>
<td>AFLP/</td>
<td>Xtxa2582, Xtxp18, Xtxp250</td>
<td>(H)</td>
<td>Klein et al. 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A3</td>
<td>Rf4 ST/CAPS</td>
<td>LW7, LW8</td>
<td>(E)</td>
<td>Wen et al. 2002</td>
</tr>
<tr>
<td><em>Triticum aestivum</em></td>
<td>T</td>
<td>RFLP</td>
<td>Xabc249, Xcdo442</td>
<td>1BS</td>
<td>Ma and Sorrells 1995</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>RFLP</td>
<td>Xbcd156, Xcdo388</td>
<td>1BS</td>
<td>Kojima et al. 1997</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>RFLP</td>
<td>Xksug48</td>
<td>6BS</td>
<td>Ma et al. 1995</td>
</tr>
<tr>
<td><em>Zea mays</em></td>
<td>T</td>
<td>RFLP</td>
<td>umc97, umc92</td>
<td>3S</td>
<td>Wise and Schnable 1994</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>RFLP</td>
<td>umc153, sus1</td>
<td>9L</td>
<td>Wise and Schnable 1994</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>AFLP</td>
<td>Arf-8</td>
<td></td>
<td>Wise et al. 1999</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>RFLP</td>
<td>whp1, bnl17.14</td>
<td>2L</td>
<td>Kamps and Chase 1997</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>RFLP</td>
<td>NP1114A</td>
<td>8S</td>
<td>Sisco 1991</td>
</tr>
<tr>
<td><em>Hordeum vulgare</em></td>
<td>msm1</td>
<td>RAPD/STS</td>
<td>OPI-18/900, MWG2218</td>
<td>6H</td>
<td>Matsui et al. 2001</td>
</tr>
</tbody>
</table>
most common evolutionary trend in angiosperm reproduction (Stebbins 1957, 1970). For example, wild rice is frequently cross-pollinated, while cultivated rice is self-pollinated. Many characters involved in mating system evolution, such as sizes of floral organs or amount of pollen produced, are quantitative in nature. Hybrid seed production depends on the improvement of outcrossing-related traits and for self-pollinated crops, it might involve a reconstruction (or recovery) of the outcrossing mating system.

Various techniques to produce hybrids have been developed depending on the crop, including hand emasculation, roguing of staminate plants in dioecious lines, use of gynoecious or highly female lines, CMS and genetic male sterility, protogyny, or self-incompatibility (Janick 1998). The rate of outcrossing is often the limiting factor determining whether a hybrid has potential for commercialization: Seed cost and price are both largely dependent on how easy it is to produce high-quality hybrid seed that both seed providers and farmers accept. Maize was particularly suitable for hybrid breeding because of monococism and the simple emasculation techniques practiced in breeding that allowed for easy inbreeding and outbreeding (Simmonds 1979). The necessity of high seeding rates in highly self-pollinated crops such as rice and wheat introduces an economic problem: Seed production costs must be low enough and yield of hybrids in the farmers’ fields must be high enough that farmers can profit from purchase and use of hybrid seed and companies can profit from their production and sale (Goldman 1999). Demands for low-cost seed dictates that seed yields be increased. As with sorghum and maize, some of the best new parents in sunflower are clearly more vigorous and high yielding than their predecessors. Sunflower is unique among the hybrid crops in that females are single headed but males have multiple heads, a recessive trait. Presence of multiple heads in the male ensures a long period of pollen availability, and better seed yield on the female, but it also hinders visual estimates of yield of the line per se (Duvick 1999).

Yield of hybrid seed is determined by many variables, both genetic and environmental. In productive, favorable environments, seed yield from seed set through cross-pollination can approach those of conventional self-pollinated cultivars in wheat (Lucken 1986) or might be up to 80% of inbred lines in rice (Yuan and Chen 1988; Lu et al. 2001). The breeder’s approach to high, stable seed production is, first, to identify those plant and flower features that affect cross-pollination; second, to find variation for these traits; and third, to incorporate genes for favorable expression of traits into parental lines (Lucken 1986). Considering all hybrid cereal crops with the CMS system, measurements for
increased outcrossing rate will include choice of favorite climate conditions for seed production; ensuring flowering synchronization of the two parents; providing a suitable pollen source; developing male sterile lines with desirable outcrossing traits; supplementary pollination; and adjustment of flowering habit and stigma characteristics using growth regulators such as gibberellic acid.

Many cereals are naturally self-pollinated. Their floral structure is adapted for inbreeding. Breeding parental lines may need to completely convert the floral structure and make them suitable for outcrossing. Outcrossing in rice depends on the capacity of stigmas to receive alien pollen and the capacity of anthers to emit much pollen to pollinate other plants in the proximity (Oka 1988). Genetic variation of floral traits in wild rice provides opportunities to modify floral structure for hybrid rice parents through the breeding process (Xu and Shen 1987, 1988a,c; Xu et al. 1988). For example, a wild rice, *Oryza longistaminata*, has big and exerted stigmas with a long life for outcrossing, which can be used for improving the female parent. Although genetic difference of stigma exertion between cultivated rice and a wild species is controlled by a major gene (Xu and Shen 1987), traditional breeding using *Oryza longistaminata* as an exerted stigma source has not been successful. Linkage between long exerted stigma and undesirable agronomic traits in wild rice species is quite strong and needs to be broken to incorporate these traits into selected genotypes. Using the gene *eui* (elongated upmost internode) to correct the panicle enclosure associated with CMS has been used in China for high-yielding seed production with the minimized gibberellic acid application, but has not been terribly successful. In addition, many cultivation practices have been used to facilitate outcrossing in rice by clipping flagleaf, applying gibberellic acid, supplementary pollination, and adjustment of flowering dates (Yuan and Chen 1988).

The situation described for rice is also applicable to wheat and other self-pollinated species, especially those for which hand-emasculatation is impractical. Although Wilson (1968) considered that the floral structure of wheat is oriented toward cross-pollination, a close examination of its floral traits clearly indicated that wheat is less suited, in its present form, to cross-pollination than crops such as maize, sorghum, and rye (Wilson and Driscoll 1983). A spike of cereal rye has ten times as many pollen grains as wheat, its flowers open considerably wider, and it has the ability to cross-pollinate under adverse weather conditions (De Vries 1971). While a great deal of variability for floral characteristics exists in wheat, the possibility of introducing additional characters favoring cross-pollination from wheat’s ancestral relatives should be explored.
Miller and Lucken (1976) reported that environmental variation resulted in a sixfold difference in grain yield on a male-sterile wheat line grown at five locations in North Dakota over three years. After review of the status of hybrid wheat, Lucken and Johnson (1988) indicated the need for acquiring more knowledge about genetic variation of floral biology, including (1) spike and flower morphology; (2) pollen dispersal, buoyancy, durability, and vigor; (3) stigma accessibility, receptivity, and durability; and (4) development of selection screens for these traits. Genetic mapping of genes for restoration, pollen shedding, anther extrusion, seed set, female receptivity, combining ability, yield, quality, and disease resistance could revolutionize hybrid wheat breeding (Jordaan et al. 1999). Openness of the wheat flower and longevity of the stamen have been found to be under genetic control and can be improved through selection. Selection for better pollen quality and greater quantities thereof produced by the male will help improve the outcrossing. Technology is also available to transfer these characteristics from rye or even triticale (*Triticosecale*) to wheat.

Many factors affecting outcrossing provide opportunities for MAS. However, there are very few investigations on genetic mapping of traits related to outcrossing. Grandillo and Tanksley (1996) examined anther length in a backcross between *Lycopersicon esculentum* and *L. pimpinellifolium*. They found two QTL affecting this trait, on chromosomes 2 and 7, which accounted for only 24% of the phenotypic variation. Georgiadis et al. (2002) investigated traits that distinguish outcrossing and self-pollinating forms of currant tomato, *L. pimpinellifolium*. Five QTL total were found involving four traits: total anther length, anther sterile length, style length, and flowers per inflorescence. Each of these four traits had a QTL of major (>25%) effect on phenotypic variance. In rice, some genetic mapping projects have been undertaken that target outcrossing. It is anticipated that MAS will provide a powerful tool to help fix the outcrossing-related issues in crops that are naturally self-pollinated but have great potential in hybrid breeding. Linkage drag associated with the introgression of outcrossing-related genes from wild species or distant cultivars may be overcome with marker-assisted background selection. Testcrossing-required traits, such as stigma longevity and receptivity, and labor-intensive traits, such as pollen load, can be selected much more easily through linked markers.

C. Wide Compatibility

Hybridization barriers exist in distant crosses of many crop species to some extent. Because the parents are not genetically compatible, hybrids
derived from intersubspecific crosses such as *indica × japonica* in rice are partially or completely sterile with seed set less than 30%, in addition to other unfavorable agronomic traits such as tall plant height and long days-to-heading (DTH) (Wang et al. 1991). Some intermediate cultivars have little or no hybrid barrier with either *indica* or *japonica*. The “intermediate type” was studied as early as in the 1930s, but Ikehashi (1982) first proposed the “wide compatibility” trait, which can be defined as the ability to make intersubspecific hybrids fertile. Discovery of wide compatibility in rice offers an opportunity for overcoming the reproductive barrier exhibited in hybrids between *indica* and *japonica*, and thereby for using the strong heterosis derived from intersubspecific crosses.

To identify wide compatibility and transfer the related genes to other genetic backgrounds, testcrossing and progeny testing are required, as for fertility restoration. Several sets of testers were carefully selected for this purpose. IRRI evaluated wide compatibility using ‘Akihikari’, ‘Toyonishki’, and ‘Taichung 65’ as *japonica* testers, and ‘IR36’, ‘IR50’, and ‘IR64’ as *indica* testers. The China National Two-line Hybrid Rice Research Cooperative Group selected ‘Youmangzaoshajing’, ‘Banilla’, and ‘Akihikari’ as *japonica* testers and ‘Nantehao’, ‘Nanjing 11’, and ‘IR36’ as *indica* testers. Cultivars or individuals in breeding populations are considered wide compatible if the pollen fertility and seed set of the hybrids between them and all six testers are over 70%. A nation-wide effort in China for screening wide compatibility cultivars identified 51 cultivars, 49 of which had an average F₁ fertility of over 80% when crossed with the six testers (Gu and Tang 2001). A lot of work is involved in testcrossing and progeny test to find out the cultivars or plants with wide compatibility. Molecular marker-assisted identification of wide compatibility genes will accelerate and facilitate the breeding process by eliminating or minimizing testcrossing and progeny testing.

A genetic model was proposed by Ikehashi and Araki (1986) to account for wide compatibility. According to this model, there are three alleles at the *S5* locus: a neutral allele, *S5n*, an *indica* allele, *S5i*, and a *japonica* allele, *S5j*. A zygote from *S5n* allele with either of the other two alleles, *S5nS5i* and *S5nS5j*, would be fully fertile, while a zygote genotypically *S5iS5j* would be partially sterile. Using morphological markers, *S5* was found to be closely linked with genes *C* (chromogen for apiculus pigmentation) and *wx* (waxy endosperm) on chromosome 6 (Ikehashi and Araki 1987). This chromosomal location was confirmed in several studies using isozymes (Li et al. 1991) and RFLP markers (Liu et al. 1992; Zheng et al. 1992; Yanagihara et al. 1995). A more precise
location of the S5 locus was provided by K. Liu et al. (1997), which was about 1 cM from an RFLP marker R2349.

Many other wide compatibility cultivars have been identified (Xu et al. 1989; Gu and Tang 2001) and five new loci, $S_7$, $S_8$, $S_9$, $S_{15}$, and $S_{16}$, were located on chromosomes 4, 6, 7, 12, and 1, respectively, by Yanagihara et al. (1992) and Wan et al. (1993). Allelic interaction at these loci can cause hybrid sterility, independently of each other, and neutral alleles to overcome this problem have been identified in different rice cultivars (Ikehashi and Wan 1998). These neutral alleles at different loci plus $S_5^w$ are extremely important for enhancing the level of heterosis in rice. Except for these major genes, wide compatibility can be attributed to multiple loci with small effects (Shen and Xu 1992). QTL analysis of a wide compatibility cultivar ‘Dular’ identified five loci, located on chromosomes 1, 3, 5, 6, and 8, which jointly explained 55% of the variation for fertility (Wang et al. 1998). Wide compatibility has been selected using associated SSR markers in our breeding program.

VI. ENVIRONMENT-DEPENDENT TRAITS

Plant populations used for gene analysis can be evaluated in either natural or controlled environments or both. Controlled environments can be compared with each other or with natural environments. If two environments mainly differ in one macro-environmental factor, they are considered contrasting or near iso-environments (NIEs), and the standard plot-to-plot variation and other residual micro-environmental effects can be neglected (Xu 2002). If the two environments are from experiments in different years or locations, it is assumed that location and year effects do not confound the effect of the macro-environmental factor.

Some traits need to be measured under NIEs, where plants respond differently. In such cases, one environment imposes much less stress on plants than the other, for example, two environments with normal and high temperatures, respectively. The effect of the stress environment can be measured by comparing it to a much-less-stress or non-stress environment. A relative trait value is then derived from two direct trait values measured in each environment to ascertain the sensitivity of plants to the stress (Ni et al. 1998). If different plants have an identical phenotype under the much-less-stress environment (this is not true for a segregating population in most cases), the direct trait value in the stress environment can be used to measure sensitivity. When both environments impose little stress on plants and the plants respond differently, however, one should use relative trait values.
A. Photoperiod/Temperature Sensitivity

A typical example for environment-dependent traits is photoperiod sensitivity that can only be measured in NIEs, one with short daylength and the other with long daylength. Plants start to flower when specific photoperiod and/or temperature conditions are met. Because of the complementary action of genes from two parents, hybrids could have very different photoperiod/temperature responses so that it is difficult to predict the flowering date of hybrids from those of their parents without a complete understanding of all related genes in the parents. Flowering synchronization of two parents is one of the factors influencing hybrid seed production and thus the economic advantage over the inbred lines/cultivars. To understand photoperiod and temperature responses, hybrids and their parents must be planted in a variety of environments or NIEs. Genetic study of these responses will finally characterize the parental photoperiod-thermo response pattern and its effect on their hybrids and thus make hybrid photoperiod-thermo response predictive. Once molecular markers have been associated with photoperiod- and thermo-sensitivity, MAS will help minimize the requirements for multi-environment/location tests, which will reduce the breeding cost and shorten the breeding cycle.

Using a rice DH population between ‘Zhaiyeqing 8’ and ‘Jingxi 17’, DTH and photo-thermo sensitivity (PTS) were investigated in two environments (Beijing and Hangzhou, China) that differ mainly in daylength and temperature (Xu et al. 1997c, Xu 2002). Four chromosomal regions were significantly associated with DTH in either or both locations, whereas a different locus on chromosome 7 (G397A-RM248) was significantly associated with PTS, indicating that the PTS QTL was independent of the QTL for DTH. By evaluating days-to-flowering (DTF) of individual ‘CO39’/ ‘Moroberekan’ RILs under 10 h and 14 h daylengths and greenhouse conditions, Maheswaran et al. (2000) identified 15 QTL for DTF. Only four of them were also identified as influencing response to photoperiod.

Different QTL have been identified using direct and relative trait values, and in rice, DTH and photoperiod are often controlled by different QTL. On the other hand, direct and relative traits could share some QTL. That means DTH and photoperiod sensitivity are genetically related to some extent because both traits are related to the basic vegetative growth that rice plants must achieve to flower. There are QTL mapping studies undertaken in NIEs, but QTL were mapped using trait values scored in each environment rather than using relative measures. The traits themselves were mapped rather than the relative response
measured under the NIEs. In rice, numerous QTL for days to heading or flowering have been mapped using molecular markers but very few of them have been tested under both long- and short-day conditions. Using an F2 between *japonica* ‘Nipponbare’ and *indica* ‘Kasalath’, Yano et al. (1997) identified two major and three minor QTL for heading date. Three of them (*Hd1*, *Hd2*, and *Hd3*) were identified later as photoperiod-sensitivity genes by test of the QTL-NILs under different daylengths (Lin et al. 2000), and one of them (*Hd1*) was cloned (Yano et al. 2000). In hybrid breeding, it is very important to understand what genes are involved in each parent for flowering and how they are working together in their hybrids. To synchronize flowering time of both parents for seed production, both parents should have the genes (of the same type) related to heading date and photoperiod and temperature sensitivity. Otherwise, two parents have to be planted separately to make them flower about the same time. Using MAS, days to flowering for the hybrid could be predicted from the parental genotypes and specifically designed from different combinations of parental genes.

B. Environment-induced Genic Male Sterility

Male sterility can be induced by specific environmental factors. An EGMS was first discovered in rice by Shi (1981) from ‘Nongken 58’, a *japonica* cultivar. The mutant ‘Nongken 58S’ is sterile when the days are long (>13.5 h) but becomes fertile when days are short (<13.5 h). Thus, fertility conversion is triggered by the length of photoperiod. This EGMS is called PGMS. Investigations on some derivatives from ‘Nongken 58S’ and other EGMS lines indicated that male sterility was affected by both photoperiod and temperature, and some EGMS lines were influenced more by temperature (and thus can be called TGMS), which reverted to partial or full fertility under certain temperature regimes (Zhou et al. 1988; Yang et al. 1989; Wu et al. 1991; Virmani and Voc 1991; Lu et al. 1994). Based on the EGMS system, Yuan (1987) proposed a strategy of hybrid rice breeding that did not involve a maintainer line and hence was called the two-line method. So far, more than ten EGMS lines have been used in China to breed two-line hybrid rice (Chen 2001; Lu et al. 2001; Lu et al. 2002), and over 20 two-line hybrids have been released to farmers; the area of two-line hybrid rice increased to 2.67 million ha in 2001 (Ma and Yuan 2002). EGMS has also been reported in pepper, tomato, wheat, barley, sesame, pea, rape, and soybean (Li and Yuan 2000).

Compared with the CMS-based three-line method, the two-line method has several advantages. First, the EGMS parent can be used to produce hybrids with any rice cultivar, which removes the restriction of restorer
genes. Second, it is easier to introduce wide compatibility gene(s) into EGMS lines than into CMS lines, which makes intersubspecific hybrid breeding more feasible. Third, negative effects inherited from the male sterile cytoplasm can be overcome. However, the dependency of male sterility on temperature or photoperiod-temperature interaction requires two different environments in the breeding and selection process. Breeding populations have to be planted in one environment where the plants will be sterile to make sure of the presence of sterility genes and in another where the plants will be fertile to confirm the fertility conversion and produce seeds. Using associated molecular markers, confirmation of fertility conversion involving two environments can be avoided. So far three genes, pms1, pms2, and pms3, conferring PGMS in ‘Nongken 58S’ and its derivative ‘32001S’, were associated with RFLP markers on chromosomes 3, 7, and 12, respectively (Table 3.2). Although most associated markers are RFLPs, it is easy to convert them into PCR-based markers. Molecular markers on five chromosomes (2, 6, 7, 8, and 9) were associated with TGMS, most of which were microsatellite markers. A gene for reverse TGMS, which is fertile under high temperature but sterile under low temperature (contrary to the original TGMS), was associated with molecular markers on chromosome 10 (Table 3.2). These studies have laid a foundation for MAS in breeding EGMS lines. To facilitate incorporation of the tms2 gene, a SSR marker, RM11, located on chromosome 7, was identified and found to be useful in identifying heterozygous fertile plants in F2 populations and F3–F4 progenies for selection of progenies in advance (Lu et al. 2002). Lang et al. (1999) reported that PCR-based markers were 85% accurate in identifying tms3 in the juvenile stage.

C. Biotic and Abiotic Stresses

Breeding of insect and disease resistance and tolerance to abiotic stresses has become a worldwide issue. To identify insect/disease resistance, plants must be inoculated artificially or naturally, or in specific environments where the stress exists. Artificial inoculation is impractical when the insects/diseases are under quarantine control. On the other hand, evaluation of plant response to different insects/diseases or different biotypes/strains/races of the same stress agents is very difficult if not impossible using traditional screening methods. Using molecular markers associated with the stress response will help select for resistance without inoculation or creating or finding specific environments. Plant response to multiple biotic stresses can be screened simultaneously using molecular markers associated with these stresses. In traditional breeding programs, selection for tolerance to abiotic stress such as salinity, drought
Table 3.2. Marker-trait associations identified in rice for photoperiod-sensitive genic male sterility (PGMS), thermo-sensitive genic male sterility (TGMS), and reverse TGMS (RTGMS).

<table>
<thead>
<tr>
<th>Trait</th>
<th>Gene</th>
<th>Source</th>
<th>Chromosome</th>
<th>Linked markers</th>
<th>Distance (cM)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGMS</td>
<td>pms1</td>
<td>32001S</td>
<td>7</td>
<td>RG477, RG511/RZ272</td>
<td>3.5, 15.0</td>
<td>Zhang et al. 1994b</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RG477/R277, R1807</td>
<td>0.1, 6</td>
<td>Liu et al. 2001</td>
</tr>
<tr>
<td></td>
<td>pms2</td>
<td>32001S</td>
<td>3</td>
<td>RG191, RG348</td>
<td>7.0, 10.6</td>
<td>Zhang et al. 1994b</td>
</tr>
<tr>
<td></td>
<td>pms3</td>
<td>Nongken 58S</td>
<td>12</td>
<td>RZ261/C751, R2708</td>
<td>5.5, 9.0</td>
<td>Mei et al. 1999</td>
</tr>
<tr>
<td>TGMS</td>
<td>tms1</td>
<td>5460S</td>
<td>8</td>
<td>RZ562, RG978</td>
<td>1.7</td>
<td>Wang et al. 1995</td>
</tr>
<tr>
<td></td>
<td>tms2</td>
<td>Lorin PL12</td>
<td>7</td>
<td>(R643A, R1440)</td>
<td>1.7</td>
<td>Yamaguchi et al. 1997</td>
</tr>
<tr>
<td></td>
<td>tms3</td>
<td>IR32364S</td>
<td>6</td>
<td>OPAC3-640, OPAA7-550</td>
<td>7.7, 10.0</td>
<td>Subudhi et al. 1997</td>
</tr>
<tr>
<td></td>
<td>tms5</td>
<td>Annong S-1</td>
<td>2</td>
<td>RM174, R394</td>
<td>0, 2.5</td>
<td>Jia et al. 2000</td>
</tr>
<tr>
<td></td>
<td>tms-vnl</td>
<td>TGMS-VN1</td>
<td>2</td>
<td>E5/M12-600, E3M16-4003</td>
<td>3.3, 28.8</td>
<td>Dong et al. 2000</td>
</tr>
<tr>
<td>RTGMS</td>
<td>rtms1</td>
<td>J207S</td>
<td>10</td>
<td>RM222, RG257</td>
<td>11.8, 4.6</td>
<td>Jia et al. 2001</td>
</tr>
</tbody>
</table>
and submergence tolerance, and lodging resistance can only be done at specific environments that are either present at specific locations or created at well-controlled environments. Selection for these traits is considered most difficult in breeding programs. To measure responses to agrochemicals, such as herbicides and plant growth regulators, such chemicals must be applied to plants at the right stage under suitable environments. MAS has made it possible to perform indirect selection for all these traits using tightly linked or intragenic markers.

There are several successful examples using MAS for biotic stress in rice. Hittalmani et al. (2000) used MAS to combine three rice blast resistance genes (Pi1 on chromosome 11, Piz-5 on chromosome 5, and Pi-ta on chromosome 12) into a single genotype. For Piz-5, a single marker was used, whereas flanking markers were used for the other two. MAS was efficient in developing gene pyramids and the line containing all three resistance genes had a broader resistance spectrum than lines with only one of them. Huang et al. (1997) pyramided four bacterial blight resistance genes, Xa4, xa5, xa13, and Xa21, using PCR-based markers. Sanchez et al. (2000) transferred three bacterial blight resistance genes into susceptible rice lines possessing desirable agronomic characteristics. For efficient MAS, sequences from RFLP or genomic clones, linked to the resistance genes (xa5, xa13, and Xa21), were converted to STS markers. This work showed the effectiveness of using markers linked to recessive genes in a backcrossing program, particularly in the presence of a dominant resistant gene. In an F2 population, selection efficiency was as high as 95% for xa5 and 96% for xa13.

Breeding for drought tolerance is becoming a top priority in many countries, as less and less water is available for agriculture. Current knowledge on physiology suggests that drought tolerance in rice depends on one or more of the following components: (1) the ability of roots to exploit deep soil water to provide for evapotranspirational demand, (2) the capacity for osmotic adjustment that allows plants to retain turgor and protect meristems from extreme desiccation, and (3) control over nonstomatal water loss from leaves (Nguyen et al. 1997). These components are generally applicable to other cereal crops. There is a large body of literature published in China indicating that hybrid rice has a strong root system characterized by more and longer roots, with better distribution in the soil as compared to its parents and inbred lines (Yuan and Chen 1988). In 1999, Zhang et al. (1999) summarized the QTL identified for drought-tolerance components in rice. More than 100 QTL had been identified for osmotic adjustment, dehydration tolerance, abscisic acid accumulation, stomatal behavior, root penetration index, root thickness, total root number, root length, total dry root weight, deep root dry weight, and root pulling force.
In maize, grain yield under drought stress is negatively correlated with the anthesis-silking interval (ASI), the difference in days between pollen shedding and silk emergence. A short ASI means rapid silk extrusion because time to anthesis is little affected by drought. Five QTL that were stable over stressed environments were identified under several levels of drought. A backcross MAS scheme to improve the drought tolerance of an elite but drought-susceptible inbred line, CML 247, has been successfully completed, using PCR-based markers as a preselection tool (Ribaut et al. 1999). A second MAS experiment used molecular markers to select changes in the frequency of alleles at loci having a known association with drought tolerance as a result of recurrent selection in an open-pollinated population. Plant selection based on the presence/absence of those alleles whose frequency changed could give increased drought tolerance in less time than recurrent selection. Ribaut et al. (1999) concluded that new breeding schemes involving optimal combinations of MAS and conventional selection to improve drought tolerance in maize hold considerable promise for the future.

In rice, the International Rice Research Institute (IRRI) has several drought-tolerance breeding programs using identified QTL and MAS. QTL affecting root parameters were identified using a rice DH population derived from the cross ‘IR64’/‘Azucena’. A marker-assisted backcross program was started to transfer the alleles of ‘Azucena’ (upland rice) at four QTL for deeper roots on chromosomes 1, 2, 7, and 9 from selected DH lines into ‘IR64’ (elite rice cultivar) (Shen et al. 2001). The backcross progeny were selected strictly on the basis of their genotype at the marker loci in the target regions up to the BC3F2, from which BC3F3 near-isogenic lines (NILs) were developed and compared to ‘IR64’ for the target root traits. Of the three tested NILs carrying target 1 (QTL on chromosome 1), one had significantly improved root traits over ‘IR64’. Three of the seven NILs carrying target 7 (QTL on chromosome 7) alone, as well as three of the eight NILs carrying both targets 1 and 7, showed significantly improved root mass. Four of the six NILs carrying target 9 (QTL on chromosome 9) had significantly improved maximum root length. The results indicate that MAS can be used to select the traits requiring specific environments for phenotyping.

VII. QUALITY TRAITS

Many important traits in crops are phenotypically invisible or unscorable and must be measured in the laboratory using sophisticated equipment or facilities. Some analyses require a large amount of seeds so that they cannot be measured until late generations when a relatively
large amount of seed becomes available for each selection entry. Grain chemical properties such as starch and protein content, and physical properties such as milling quality, are important quality traits for most cereal crops that fall under this category.

A. Seed Traits

As a major storage organ of cereal seeds such as rice, endosperms provide humans with proteins, essential amino acids, and oils. An understanding of the inheritance of endosperm traits is critical for the improvement of seed quality. Genetic behavior in triploid endosperms is much different from that of the maternal plants that supply assimilates for grain growth and development. Thus, methods suited for genetic analysis of traits in maternal plants (diploids for most cereal crops) cannot directly be used for endosperm traits (Xu 1997). Any genetic analytical method for endosperm traits needs to combine a genetic method developed for diploid maternal plants with a triploid model proposed for conventional genetic analysis.

The genetic system controlling endosperm traits may be much more complicated than that controlling traits of the plant per se. Because the plant provides seeds with a portion of their genetic material and almost all the nutrients required for growth and development, seed traits are genetically affected by both the seed nuclear genes and maternal nuclear genes. In addition, cytoplasmic genes may also affect some seed traits through their indirect effects on the biosynthetic processes of chloroplasts and mitochondria. To understand endosperm traits with biological accuracy, one should take into consideration maternal genetic effects and cytoplasmic effects along with the direct genetic effects of seeds. As seeds initiate a new generation that differs from their maternal plants, some seed traits should be considered as one generation advanced over their maternal plants. Genetic analysis of endosperm traits should be based on the DNA extracted from both maternal plants and endosperm tissues in order to understand the relative contribution of the different genetic factors to the variation of endosperm traits (Xu 1997). Currently all endosperm traits have been treated the same as other traits of the plant, with few reports (Tan et al. 1999) that considered the generation advancement issue.

B. Hybrid Seed Traits

Although F₁ plants are uniform, seeds borne on them represent the F₂ seed generation and are expected to segregate for some grain characteristics. Major determinants of grain quality in cereal crops are milling;
grain size, shape and appearance; and cooking and eating characteristics. Some grain tissues are of maternal origin and some result from fertilization and union of genetically diverse gametes. For example, the lemma and palea of the rice hull are maternal tissues. Seed size and shape are determined by the shape and size of hulls and the latter is determined by the genotype of F₁ plants. As a result, all F₂ seeds borne on F₁ plants have nearly identical dimensions even though the parents could have very different seed sizes. Endosperm is triploid tissue resulting from the union of one male nucleus with two female nuclei. If the parents differ in endosperm traits, the F₂ grains on F₁ plants show clear-cut segregation (Kumar and Khush 1986; Tan et al. 1999). Single seed analysis of a hybrid rice, ‘Shanyou 63’, indicated that the amylose content for seeds on a F₁ plant could range from 8% to 32% when two parents had 15.8% and 27.2% amylose content, respectively. A similar situation was reported for barley. If the parents differ significantly in malting quality characters, the grain produced by barley hybrids will be heterogeneous and heterozygous for characters critical to the malting process (Ramage 1983).

The cooking and eating quality of milled rice is related to starch properties such as amylose content, gelatinization temperature, and gel consistency. These three traits are highly correlated and the latter two are controlled by a locus located in the same chromosomal region as amylose content (Tan et al. 1999). When both parents are non-glutinous but differ widely in amylose content, the raw F₂ grains are distinctly classifiable into different amylose categories. But, when cooked in bulk, the grains do not vary in cooking and eating characteristics (Khush et al. 1988). That might explain why the hybrid rice ‘Shangyou 63’, which contains 8% to 32% amylose contents among different grains, has been acceptable for many years in China. Although genetic heterozygosity of hybrids may not impair grain quality in terms of physical and chemical characteristics as long as one of the parents is not glutinous or poor in grain quality (Khush et al. 1988), amylose difference among grains may contribute to unstable milling quality and uneven endosperm traits such as chalk. For hybrid rice breeding programs in countries demanding high-quality standards, both parents should be equally good if hybrids are to meet the standard quality of inbred cultivars. To develop glutinous rice hybrids, both parents must have glutinous grains. To develop hybrid rice possessing premier grain quality like that of basmati, both parents must possess basmati quality (Virmani and Zaman 1998). The same situation may be true for other cereal hybrids. It is strongly suggested that to develop hybrid crops with grain quality as good as the best inbred lines/cultivars, both parents must have the best quality. Use of parental lines with equally good quality will help minimize this effect, but might
result in low heterosis because of less genetic diversity for grain traits between the parents.

Genetic contribution to quality comes from both parents, but one of them could be more important than another in some specific situations. Endosperm properties might be affected more by female parents due to the maternal effect, or more by male parents due to the xenia effect (Xu and Shen 1988b). The composition and development of the kernels can be changed by the nature of pollen. This was first shown by Kiesselbach (1926) as the change of a sweet corn endosperm into a starchy endosperm after pollination of a sweet corn female by a flint endosperm male. Large xenia effects were observed for sorghum malt quality in the F1 but this was entirely lost in the F2 generation (Wenzel and Pretorius 2000). Curtis et al. (1956) observed that the germ is markedly influenced in weight, oil and protein content by both the seed parent and the pollen parent of corn, with a pronounced maternal effect.

C. Selection of Quality Traits

Many quality traits are genetically controlled by multigenic loci, or by multiple alleles at a locus because of the triploidy of the endosperm. As a result, the same phenotypes may come from different genetic factors or different alleles from the same locus. Phenotypic selection for the same trait values may not result in the same alleles or genes fixed in parents. MAS will help distinguish different genetic loci that contribute to the same quality traits to avoid the problem that the two parents used for hybrid production have different alleles or alleles from different loci.

Almost all quality traits are only measurable at or after the reproductive stage and they are good candidates for MAS. MAS can be made at any stage and in any generation so that breeders do not need to maintain a large number of candidate plants generation after generation. Methods for non-destructive extraction of DNA from single dry seeds provide an opportunity for selection of seed traits so that selection can be processed before planting. Early-stage selection also provides more opportunities for selection of traits with relatively low heritability. MAS could be used for early-stage quality tests or DNA-based quality tests, whereas such tests would be delayed in a conventional breeding program because a relatively large amount of seeds is required.

Malt quality traits of barley, which are influenced by many loci, are suitable candidates for MAS. The genetic make-up of modern malting barley cultivars has been achieved by carefully balancing these traits through lengthy breeding processes. In a MAS project conducted among the progeny from the same cross for which the QTL had been mapped,
Igartua et al. (2000) confirmed the presence of QTL on chromosome 7 affecting all traits. MAS was effective in selecting phenotypically superior lines and the magnitudes of the combined effects for these regions were close to the estimates calculated in the mapping experiment. The use of marker information in the selection did not eliminate the need to gather reliable phenotypic data but it should permit breeders to allocate resources to the evaluation of progeny that are more likely to carry favorable alleles.

Rice amylose content, mainly controlled by the \textit{wx} gene, is a good example of MAS. Ayres et al. (1997) determined the relationship between polymorphism at that locus and variation in amylose content. Eight \textit{wx} microsatellite alleles were identified from 92 long-, medium-, and short-grain U.S. rice cultivars. When used as predictors of amylose content, these eight alleles explained an average of 85.9\% of the variation. The amplified products ranged from 103 bp to 127 bp in length and contained (CT)\textsubscript{n} repeats, where \textit{n} ranged from 8 to 20. Average amylose content in cultivars with different alleles varied from 14.9\% to 25.2\%.

Using more diverse rice germplasm accessions (\textit{n}=243), Zeng et al. (2000) identified 15 alleles at the \textit{wx} locus using microsatellite class and G-T polymorphism, resulting in a total of 16 alleles identified so far. Although the microsatellite marker was located in the intron of the waxy gene, a complete association between marker alleles and amylose contents still depends on fully understanding other genes involved in the starch synthesis.

VIII. GENE INTROGRESSION AND WHOLE GENOME SELECTION

In previous sections, several types of special traits that are important to hybrid breeding and most suitable for MAS were discussed. Except for quality traits discussed in Section VII, there are many other traits that need laborious fieldwork or intensive laboratory work. Xu (2002) fully described all the trait categories most suitable for MAS, and Dekkers and Hospital (2002) also listed the opportunities for the use of molecular data. In this section, some general considerations for all trait categories will be discussed.

A. Gene Introgression

Emerging MAS technology should provide the vehicles for using markers to expedite the acquisition of important genes from exotic populations or from wild species. Gene introgression involves the introduction
of a target gene into a productive, recipient line or cultivar. Gene introgression can be used in both backcrossing and intercrossing programs. By using DNA markers to identify recombinants, introgressed chromosome segments might be “trimmed” to minimal size, reducing the extent to which the recurrent genotype is disrupted by undesirable alleles closely linked to the target trait (Tanksley and Rick 1980).

It is critical in plant breeding that allelic substitution be precise so that only the target gene and the shortest possible segment of the linked chromosome are transferred from the donor parent to the recipient parent, the latter of which is usually a cultivar or inbred line with very good combining ability. To reduce false positives in MAS, markers must be tightly linked to the target trait, and flanking markers or multiple markers around the region could be used simultaneously. A three-marker system, with three markers located on a chromosome block of a few (≤5 cM), will be desirable in this case (Zhang and Huang 1998). The marker in the middle, preferably intragenic or cosegregating with the gene, will be used to indicate the presence of the target gene in the selection process. The marker on each side will be used to indicate the absence of the chromosome segment from the donor parent (negative selection), that is, selection for recombination between the target gene locus and the marker locus. As more and more genes have been cloned, the marker in the middle could be developed from the cloned gene or gene sequence, as discussed in Section IIIA. This system will be very useful when the target gene is only available in a wild species and linkage drag is proven to be associated with the chromosome segment to be transgressed.

In rice, a series of advanced backcrossing populations have been developed through collaborations between Cornell University and breeders around the world to identify and introgress trait-enhancing alleles from wild species into high-yielding elite cultivars. The first such study employed a cross between the wild rice relative *Oryza rufipogon* and the Chinese *indica* hybrid ‘V20’/‘Ce64’ (Xiao et al. 1998). Although the *O. rufipogon* accession was phenotypically inferior for all 12 traits studied, transgressive segregation was observed for all traits, and 51% of the QTL detected had beneficial alleles from *O. rufipogon*. By MAS and field selection, an excellent CMS restorer line (‘Q661’) carrying one of the QTL for yield components has been developed. Its hybrid, ‘J23A’/‘Q661’, outyielded the check hybrid by 35% in a replicated trial for the second rice crop in 2001 (Yuan 2002). A second QTL study used an advanced backcross population between the same *O. rufipogon* accession and the upland *japonica* rice cultivar ‘Caiapo’, and identified beneficial QTL alleles from *O. rufipogon* for 56% of the trait-enhancing QTL detected (Moncada et al. 2001). A third study employed the *O. rufipogon* in a cross with the long-grain ‘Jefferson’, a U.S. tropi-
cal *japonica* cultivar, and the *O. rufipogon* allele was favorable for 53% of the yield and yield component QTL (Thomson et al. 2003). There are several ongoing projects to introgress these favorable alleles from the *O. rufipogon* accession into cultivated rice.

Information from the mapping studies reported may not be used directly in designing MAS experiments in breeding programs because different genetic backgrounds and environments have been involved. Other than epistasis and G × E interaction, many important questions regarding genetic regions tagged by molecular markers remain unknown. For instance, the number of polymorphic genes and the number of functional alleles at each of these loci in the gene pool are generally unknown for most traits. Cross-population comparison of genetic mapping studies using biparental materials can only provide limited information. As a result, many questions remain to be answered before any MAS experiment for major gene/QTL transfer can be used: (1) information on genetic parameters of a locus such as location, effect, and linked markers; (2) number of loci and number of alleles at each locus and variation explained by each locus and by all loci; (3) background effect on genes and gene interaction; and (4) linkage drag and its genetic basis. A reasonable strategy for molecular breeding is to do simultaneous gene identification and introgression such as advanced backcross QTL analysis as suggested by Tanksley and Nelson (1996). A comprehensive backcrossing breeding program through international breeding activities has been initiated for integration of DNA markers with phenotypic selection (Li 2001), which focuses on improvement of inbred cultivars and the parental lines for hybrid rice.

**B. Whole Genome Selection**

MAS can also be practiced at the whole genome level. DNA marker-based whole genome selection or background selection can be used to accelerate recovery of a recurrent genotype in the backcrossing process for breeding parental lines. Compared to a backcross program that usually takes five to seven generations to recover most recurrent parental backgrounds, MAS may save two to four backcross generations in the transfer of a single target allele (Tanksley et al. 1989; Hospital et al. 1992; Fisch et al. 1999). Combined with selection for target traits, whole genome selection allows the breeder to simultaneously transfer multiple traits through backcrossing.

When single chromosomes are distinguishable, partial genome selection or whole chromosome selection are alternatives to whole genome selection so that the other chromosomes remain unchanged. MAS could be focused on a chromosomal region/arm if it is separable from the rest
of the genome. Genes controlling the same traits or trait category may cluster in some specific chromosomal regions, which are called gene blocks. Regional mapping strategies (Xu 1997; Monna et al. 2002), combined with a high-density genetic map, can help construct high-density regional maps that target gene blocks for separation of closely linked genes.

As genetic mapping information accumulates from different mapping populations, it will be possible to establish a complete profile for all the genes associated with a specific trait or trait category. Whole genome selection can be used to select the best trait/gene combinations based on selection for each of the target loci whose position in the genome is known. It is possible to select the best cassette for any traits and/or trait combinations.

To transfer the bacterial blight resistance gene \( \text{Xa21} \), 128 RFLP markers, evenly distributed on the 12 rice chromosomes, were used to recover the genetic background of ‘Minghui 63’, a widely used parent. Plants containing \( \text{Xa21} \) in the \( \text{BC}_{3}F_{1} \) generation were screened with markers covering the genome, and those homozygous for ‘Minghui 63’ alleles were saved. The improved version of ‘Minghui 63’ and its hybrid with ‘Zhenshan 97’ showed the same resistance spectrum as the resistant donor. Field examination of a number of agronomic traits showed that the improved ‘Minghui 63’ and its hybrid were identical to the originals except for their resistance to bacterial blight (Chen et al. 2000). MAS was also used by the same group to improve ‘6078’, an elite restorer line with high yield potential, by transferring \( \text{Xa21} \) from IRBB21 (Chen et al. 2001). Background selection in this study took place in the \( \text{BC}_{1}F_{1} \) and also in the \( \text{BC}_{2}F_{1} \) population using AFLP markers with unknown positions. It is not clear exactly how much of the genetic background of the recurrent parent was recovered in the individual obtained in the final selection (because of unknown chromosomal distribution of the 129 polymorphic bands), but the agronomic performance and combining ability of the selected line was very similar to the original.

C. Selection for Multiple Genes/Traits

MAS provides opportunities for simultaneous selection of multiple traits/genes. In some cases, multiple pathogen races or insect biotypes must be used to identify plants for multiple resistances, but in practice this may be difficult or impossible because different genes may produce similar phenotypes that cannot be distinguished from each other. Marker-trait association can be used to simultaneously select multiple resistances from different disease races and/or insect biotypes, and pyramid them into a single line through MAS as discussed in Section VIC.
To find a CMS restorer in rice through testcrossing and progeny test, a
candidate male plant has to be testcrossed with a CMS line to find out
if it has fertility restorability based on the fertility of testcross progeny.
However, sterility in testcross progeny could result from the absence of
either restorability genes or wide compatibility genes or both when an
intersubspecific cross is involved. MAS could be used to distinguish the
two different types of sterility.

Consider phenotypic selection for multiple traits in rice, such as
TGMS, amylose, and wide compatibility. Candidate plants must be
tested in two different environments where TGMS can be identified.
Each plant must be testcrossed with wide compatibility testers, follow-
ing up with a progeny test in the next season. At the same time, a rela-
tively large amount of seed must be harvested for amylose measurement.
While conventional selection methods require a delay until a large num-
ber of seeds are available and a reasonable level of homozygosity is
reached, in MAS only a leaf harvested at any growth stage in any segre-
gating population is required.

Hybrid rice provides an advantage over inbred cultivars because dom-
inant genes and/or QTL with additive effects from both parents can be
integrated into one hybrid. An integrated breeding program including
MAS was initiated in China to improve an elite hybrid rice, ‘Shanyou
63’, a cross between ‘Zhenshan 97’ and ‘Minghui 63’. Xa21, a wide-
spectrum bacterial blight resistance gene, was introduced into the
restorer ‘Minghui 63’ by MAS and a Bt gene that is toxic to stem borer
was introduced into ‘Minghui 63’ through transformation. An allele at
the Wx locus from ‘Minghui 63’ was transferred by MAS to ‘Zhenshan
97’ to improve cooking and eating quality of the hybrid, resulting in a
new version of ‘Zhenshan 97’ with medium amylose content, soft gel
consistency, and high gelatinization temperature. The pyramiding of Bt,
Xa21, and wx genes created an improved ‘Shanyou 63’ (He et al. 2002).

D. Integrated Genetic Mapping and MAS

MAS programs involve genetic mapping to identify genes of interest fol-
lowed by transfer to another genetic background. This procedure was
used to enhance ‘B73’ and ‘Mo17’ inbreds in maize (Stuber et al. 1999).
The procedure per se is inefficient because it requires identification of
the targeted segments (containing the putative genetic loci) prior to
transfer to the recipient line. In many cases, genetic mapping results
obtained from specific crosses cannot be used for MAS for the same traits
in different crosses. There are three reasons for this phenomenon. First,
quantitative traits are usually controlled by many genes. Genes are only
segregating at the loci where two parents are genetically different and
thus can be mapped using the population derived from these two parents. For a randomly selected mapping population, the parents will have a strong chance to share identical alleles at some of the genetic loci. There is a high probability that segregating genes in any breeding population could be different from the genes already mapped. Second, multiple alleles at a locus work in the same way to complicate MAS, because mapping parents could have alleles that are different from those of breeding populations. Interaction among these multiple alleles will modify marker-trait associations when different allele combinations are considered. Third, G × E interaction could make the establishment of marker-trait association depend on specific environments.

One of the best ways to avoid these limitations is to integrate genetic mapping with MAS, that is, marker-trait associations identified from a breeding population will be used for MAS of the same population. This is critical for quantitative traits that are genetically controlled by many genes and interact with environments. Advanced backcrossing QTL analysis, proposed by Tanksley and Nelson (1996) to accelerate the process of molecular breeding, is one of the approaches that can be used for this purpose. Stuber et al. (1999) discussed their effort to test a marker-based breeding scheme for systematically generating superior lines without any prior identification of genes in the donor sources. The identification of and mapping of genes in the donor is a bonus obtained when the derived NILs are evaluated. This method is somewhat similar to advanced backcross-QTL analysis. Other approaches include using associations identified in F2 populations to select the subsequent self-pollinated populations.

E. Response to Selection

MAS for traits controlled by major genes will receive a strong response. However, the response to selection for quantitative traits using associated molecular markers will depend on several factors: linkage between markers and genes, trait heritability, gene effects, gene interactions, population size, the number of plants selected, and the breeding scheme. In this section, response to MAS of QTL will be discussed.

An example in Lande and Thompson (1990) demonstrated that on a single trait the potential selection efficiency by using a combination of molecular and phenotypic information, compared to standard methods of phenotypic selection, depends on the heritability of the trait, the proportion of additive genetic variance associated with marker loci, and the selection scheme. The relative efficiency of MAS is greatest for traits with low heritability if a large fraction of the additive genetic variance is associated with marker loci. Limitations that may affect the potential
utility of MAS in applied breeding programs include (1) the level of linkage disequilibrium in the populations, which affects the number of marker loci needed; (2) sample size needed to detect trait loci with low heritability; and (3) sample errors in the estimation of relative weights in the selection indices.

In theory, MAS is proposed to be more efficient than phenotypic selection when the heritability of a trait is low, where there is tight linkage between QTL and DNA markers (Dudley 1993; Knapp 1998), with larger population sizes (Moreau et al. 1998), and in earlier generations of selection before recombinational erosion of marker-trait associations (Lee 1995). Edwards and Page (1994) proposed that the distance between markers and QTL was the factor that most limited gains from MAS. Yousef and Juvik (2001) reported an empirical experiment that provided equivocal results regarding the relative efficiency of MAS and phenotypic selection in enhancing economically important quantitative traits in sweet corn. MAS and phenotypic selection were applied to three F$_{2:3}$ base populations with either the sugary 1 (su1), sugary enhancer 1 (se1), or shrunken 2 (sh2) endosperm mutations. One cycle of selection was applied to both single and multiple traits such as seedling emergence. Selection efficiencies were evaluated on the basis of gains over one cycle. Among 52 paired comparisons between MAS and phenotypic selection composite populations, MAS resulted in significantly higher gain than phenotypic selection for 38% of the comparisons, while phenotypic selection was significantly greater in only 4% of the cases. The average MAS and phenotypic selection gain, calculated as percent increase or decrease from the randomly selected controls, was 10.9% and 6.1%, respectively.

Recognizing that small mapping populations are not adequate for QTL mapping is the first and most important realization needed in the research community (Young 1999). Scientists must understand that simply demonstrating that a complex trait can be dissected into QTL and mapped to approximate genomic regions using DNA markers is not enough. Projects need to utilize better scoring methods, larger population sizes, multiple replications and environments, appropriate quantitative genetic analysis, various genetic backgrounds, and, whenever possible, independent verification through advanced generations or parallel populations. Only then will sufficient experimental evidence be in place for a successful MAS program.

“What if we knew all the genes for a quantitative trait in hybrid crops?” This was asked by Bernardo (2001), who has been working on the prediction of hybrid performance through computer simulation. With maize as a model species, he found through trait and gene best linear unbiased prediction (TG-BLUP) that gene information is most useful in selection
when few loci (e.g., 10) control the trait. With many loci (50), the least square estimates of gene effects become imprecise. Gene information consequently improves selection efficiency among hybrids by only 10% or less, and actually becomes detrimental to selection, as more loci become known. Bernardo further indicated that increasing the population size and trait heritability to improve the estimates of gene effects also improves phenotypic selection, leaving little room for improvement of selection efficiency via gene information. He thought genomics is of limited value in selection for quantitative traits in hybrid crops. Epistatic interactions, which were assumed absent in his study, would make the estimation of gene effects even more difficult. It is unknown whether methods other than TG-BLUP or multiple regression would substantially enhance the usefulness of gene information in selection.

Response to phenotypic selection can be evaluated using molecular markers. The Illinois long-term selection experiment on maize oil and protein contents (Dudley and Lambert 1992; Dudley 2004) and marker-assisted evaluation (Goldman et al. 1993) is an example of marker-assisted evaluation of the response to selection. This selection experiment was initiated in 1896, and by 1989 had experienced 90 generations of selection, increasing oil content from 4.7% in the original population to 19.3% in the Illinois High Oil (IHO) strains. In contrast, 87 generations of selection for low oil concentration reduced oil content from 4.7% to <1% in the Illinois Low Oil (ILO) strains. Mean protein concentration for 76 generations were 25% for the Illinois High Protein (IHP) strains and 4% for the Illinois Low Protein (ILP) strains (Dudley and Lambert 1992). This long-term divergent selection response can be attributed to the accumulative action of alleles with similar effect that had been dispersed among the individuals of the original population (Xu 1997), although de novo mutations may be an alternative explanation for this divergence, as indicated by selection for bristle number in Drosophila (Mackay 1995). The selection strains offer a unique opportunity to investigate the genetic basis of kernel chemical traits, and have been used to produce maize populations to map the QTL responsible for the selection response (Goldman et al. 1993). By using 90 genomic and cDNA clones distributed throughout the maize genome to detect RFLPs between IHP and ILP strains, 22 loci distributed on 10 chromosome arms were significantly associated with protein concentration, and clusters of three or more significant loci were detected on chromosome arms 3L, 5C, and 7L, suggesting the presence of QTL with large effects at these locations. A multiple linear regression model consisting of six significant loci on different chromosomes explained over 64% of the total variation (Goldman et al. 1993). These significant QTL associations can be used to account for the long-term selection response and the protein
content difference between the IHP and ILP strains. It can be expected that the longer the selection proceeds, the bigger the difference of protein content will be in the resulting selection strains, and thus the potential to detect additional QTL, as long as the populations continue to respond to selection. This expectation can be tested by QTL mapping using the crosses from the IHP and ILP strains derived from different cycles of selection.

We would expect genetic fixation with long-term selection programs. However, selection experiments discussed above for maize for high and low protein or oil and in *Drosophila* for bristle numbers (Yoo 1980) show no indication of genetic fixation from long-term selection resulting in remarkable changes in phenotype. Frequent identification of large-effect QTL, as reviewed by Tanksley (1993), Kearsey and Farquhar (1998), and Xu (2002), makes steady and sustained selection response puzzling: Alleles of large effects should be fixed rapidly, after which no further response would be seen. Barton and Keightley (2002) named two factors that might explain this apparent paradox. First, QTL-mapping experiments underestimate the number of QTL and overestimate their effects. Second, mutation generates alleles of large effect, which can be picked up quickly enough by selection to sustain a continuing selection response. Several mechanisms have been described that can create de novo variation, including intragenic recombination, unequal crossing over among repeated elements, transposon activity, DNA methylation, and paramutation (R. L. Phillips, pers. commun.). Barton and Keightley (2002) listed several factors that make it difficult to estimate the true numbers and effects of loci influencing a quantitative trait. Hyne and Kearsey (1995) pointed out that in a typical experiment (heritability ~40%, ~300 F₂ individuals), no more than ~12 QTL are ever likely to be detected, which is supported by empirical data on the numbers of QTL detected in plants (Tanksley 1993; Kearsey and Farquhar 1998; Xu 2002). Beavis (1994) indicated that unless samples are large (>500, for example), the effects of statistically significant QTL are substantially overestimated.

**IX. PREDICTION OF HYBRID PERFORMANCE AND HETEROSIS**

**A. Combining Ability and Heterosis**

There are two types of combining ability, general and specific. General combining ability (GCA) is defined as an attribute of an inbred line and is measured as the average performance of all hybrids made with that inbred line as a parent. The higher the GCA of an inbred, the higher the average performance of its hybrids. Specific combining ability (SCA) is
defined for specific combinations of parents and is measured by the deviation of the hybrid performance from the expected performance as estimated from the GCA of the parents. As a result, hybrid performance is determined by its parents’ GCA and the cross’s SCA. The widely grown rice hybrid in China, ‘Shanyou 63’, is an example of this. The female parent (CMS line), ‘Zhenshan 97’, and the male parent, ‘Minghui 63’, have been used to produce many successful hybrids with various fertility restorers and CMS lines. Heterosis is an important cause of the increasingly high yields of rice, maize, grain sorghum, and oil sunflower, although it is not the only cause. Improvements in GCA, in additive genes as well as in dominant, overdominant, or epistatic gene combinations, have been crucial to the improvement of hybrids in all four crops. Heterosis will continue to be a very important cause of hybrid superiority in yield and yield stability. SCA—specific combinations of inbred lines with good GCA—will remain the essential requirement for production of superior new hybrids (Duvick 1999).

The use of marker-aided prediction of advanced generation combining ability on the basis of data from early generation testcrossing was evaluated by Johnson and Mumm (1996). Using a total marker score generated from regressing $F_3$ line testcross yield on the corresponding $F_3$ line RFLP genotypes, $F_5$ testcross yields were predicted with more accuracy than they would have been with $F_3$ testcross yields alone as predictors. It was concluded that marker-assisted prediction of advanced generation performance from early generation testcrosses was effective. These results were in agreement with those reported by Eathington et al. (1997).

B. Genetic Basis of Heterosis

Heterosis is a complex physiological phenomenon affected by many factors. Yield is the most important trait in crop-based heterosis analysis. In many investigations, genes for yield per se and genes for yield-related heterosis have been confounded with each other. Several different hypotheses have been proposed for the explanation of heterosis but none of them is widely accepted so far. Among these hypotheses, arguments have focused on the dominance hypothesis (Davenport 1908) and the overdominance hypothesis (East 1908; Shull 1908), both of which are based on describing the genetic effects of single loci. Recent studies, as will be discussed later, have indicated that epistasis plays an important role in genetic control of both quantitative traits and heterosis. The dominance hypothesis proposes that heterosis results from the cancellation of effects from deleterious recessive alleles, contributed by one parent, by dominant alleles contributed by the other parent in the heterozygous $F_1$. This hypothesis emphasizes the contribution of the
dominance to heterosis. The overdominance hypothesis assumes that a specific heterozygous combination of alleles at a single locus is superior to either of the homozygous combinations of the parental alleles at that locus. Two alleles with different effects are presumed to interact physiologically such that heterozygotes show superiority in viability and adaptability compared to either of the homozygotes, and the degree of superiority increases with the degree of allelic heterozygosity. Supporters of the overdominance hypothesis offer two main objections to the dominance hypothesis. First, it should be possible to accumulate, by selection, all the favorable dominant alleles into one homozygous strain and obtain inbreds that are as vigorous as hybrids. Second, F2 distributions should be skewed because of the $\frac{3}{4}$ dominants to $\frac{1}{4}$ recessive segregation (Lamkey and Edwards 1999). But in 1916, Jones noted that with linkage the two hypotheses became indistinguishable and Collins pointed out in 1921 that with a large population size the F2 distribution is essentially symmetrical. From the results of Gardner (1963) and Moll et al. (1963), the clear conclusion was that statistical overdominance in the early generations was a consequence of linkage disequilibrium, that is, favorable dominants linked to deleterious recessives. For self-pollinated species, the genetic basis for heterosis may be different because of purging of deleterious recessive alleles by long-time inbreeding (Crow 1999). As a complex trait, heterosis is likely to be conditioned by numerous genetic factors functioning and responding to a wide variety of interacting situations, and epistasis must also play an important role in the genetic control of heterosis. Crow (1999, 2000) provided a historical review on the dominance and overdominance hypotheses.

**1. Evidences from Genetic Mapping Using Molecular Markers.** Xiao et al. (1995) investigated the genetic basis of heterosis in rice by backcrossing an *indica/japonica* RIL population to its two parental cultivars. For 12 agronomic traits, a total of 27 QTL could be detected in either backcross population. At about 82% of these QTL, heterozygotes had higher phenotypic values than their corresponding homozygotes. Ten QTL influencing grain yield components detected in both backcross populations were completely or partially dominant. RILs having phenotypic values superior to the F1 hybrid between the parental lines were found for all traits evaluated. They concluded that dominance complementation is the major genetic basis of heterosis in rice. This conclusion was strengthened by the finding that there was no correlation between most traits and overall genome heterozygosity and that there were some RILs in the F8 population having phenotypic values superior to the F1 for all of the traits evaluated—a result not expected if overdominance were a major contributor to heterosis. Digenic interaction was not evident in this study.
Stuber et al. (1992) analyzed the genetic mechanism of heterosis in a maize single-cross hybrid, ‘Mo17’ × ‘B37’, using RFLP markers. They found that most yield-QTL heterozygotes had higher phenotypic values than either homozygote class and they concluded that overdominance was the major genetic basis of heterosis in maize. This overdominance hypothesis was also supported by the results from wheat (Sun et al. 1997), Arabidopsis (Mitchell-Olds 1995), and juvenile aspen (Li and Wu 1996). However, maize researchers were careful to point out that estimates of dominance variance exceeding that for straight dominance could be due to either overdominance or linkage disequilibrium of linkage loci with favorable alleles in repulsion phase (pseudo-overdominance). Subsequently, fine mapping experiments demonstrated that the two major QTL in maize that had been identified with overdominance effects by Stuber et al. (1992) acted in a dominant manner with alleles in repulsion phase linkage by fine mapping (Graham et al. 1997). Thus, the effects at these two QTL strongly support the dominance theory of heterosis in this chromosomal region of maize.

Epistasis represents non-allelic interaction, including additive-by-additive interaction, additive-by-dominance interaction, and dominance-by-dominance interaction. Because almost all heterotic traits are genetically controlled by the combined effect of many QTL, epistasis is expected to play an important role in the genetic control of these complex traits, as it does for many other quantitative traits. The first evidence for epistasis in rice is provided by Yu et al. (1997). They located QTL for yield and its components using F3 families derived from a rice hybrid, ‘Shanyou 63’. A total of 32 QTL were detected for four traits; 12 were observed in both years and the remaining 20 were detected in only one year. Overdominance was observed for most of the yield QTL and also for a few yield-component QTL. Correlations between marker heterozygosity and trait expression were low, indicating that the overall heterozygosity made little contribution to heterosis (yield). Digenic interactions, including additive-by-additive, additive-by-dominance, and dominance-by-dominance, were frequent and widespread in this population. The interactions involved large numbers of marker loci, most of which individually were not detectable on a single-locus basis; many interactions among loci were detected in both years. The authors concluded that epistasis plays a major role as the genetic basis of heterosis.

The three reports discussed above for dominance, overdominance, and epistasis hypotheses were all based on the use of yield and yield components per se to measure hybrid performance without use of parental lines as a control to derive values for the midparent or better-parent heterosis. The method of measurement will identify genes for yield and yield components rather than genes for heterosis. For open-
pollinated species like maize, which has severe inbreeding depression, it is very difficult (if not impossible) to do side-by-side comparisons of the F₁ hybrids with their parents. But, theoretically, this comparison is absolutely necessary if heterosis rather hybrid performance needs to be measured.

Several recent investigations have been reported for genetic analysis of heterosis per se in rice. Zhang et al. (2001) designed a mating scheme that generated a fixed or “immortalized” F₂ population, using a population of 240 RILs derived from the ‘Zhenshan 97’/’Minghui 63’ cross. In this design, crosses were made between the RILs chosen by random permutations of the 240 RILs. In each round of permutation, the 240 RILs were randomly divided into two groups and lines in the two groups were paired at random without replacement to provide parents for 120 crosses. The procedure can be repeated as many times as desired and each round of permutation will pair parents for 120 crosses. This population provides opportunities for genetic mapping of heterosis per se rather than analyses based on measurements of trait performance, as long as the hybrids and the parents for each cross are planted side by side in the field, allowing measurement of heterosis for each cross. Three rounds of such random permutations, including 360 crosses, resulted in two conclusions (Zhang and Li 2002). First, all kinds of genetic effects, including single-locus heterotic effects caused mostly by overdominance, and all three forms of digenic interactions (additive by additive, additive by dominance, and dominance by dominance) appeared to play a role in the genetic basis of heterosis in the “immortalized F₂” population. However, the QTL were not fine mapped, leaving open the possibility that, as in maize, the single-locus effects were due to pseudo-overdominance, rather than true overdominance. Second, single-locus heterotic effects and dominance-by-dominance interaction could, together, adequately account for the genetic basis of heterosis in the F₁ hybrid.

Z.-K. Li et al. (2001) investigated the genetic basis of heterosis in rice using 254 RILs derived from a cross between ‘Lemont’ (japonica) and ‘Teqing’ (indica) and two BC and two testcross populations derived from crosses between the RILs and their parents plus two testers (‘Zhong 413’ and ‘IR64’). For both grain yield per plant and biomass per plant, there were significant negative effects associated with homozygous loci detected in the RIL population and a high level of heterosis in each of the BC and testcross hybrid populations. Epistasis was found for a large number of QTL pairs and a few QTL with significant main-effects were identified. Together, these QTL were responsible for over 65% of the phenotypic variation of biomass and grain yield in each of the populations, with epistatic effects explaining a much greater portion of the
variation than the main effects. As a result, most QTL associated with decreased grain yield and biomass, or with heterosis in rice appeared to be involved in epistasis, and about 90% of the QTL contributing to heterosis appeared to be overdominant. These results and the results from the genetic analysis of yield components (Luo et al. 2001) indicate that epistasis and overdominance, rather than dominance, were the major genetic basis of heterosis in rice.

2. Conclusions. As a complex character involving yield and yield components, heterosis should be genetically controlled by many genes. Although genetic study of quantitative traits has identified a limited number of QTL, each explaining a relatively large proportion of genetic variation, much more QTL could be found when multiple populations are considered. For example, a total of 63 QTL for plant height have been identified from 13 rice mapping populations and these QTL distributed on 29 chromosomal regions (Xu 2002). For a specific hybrid, heterosis is more likely genetically controlled by a relatively small number of genes; for explanation of heterosis involved in all hybrids derived from a species, a large number of QTL will be needed.

Heterozygosity and its related gene interactions are the primary genetic basis for explanation of heterosis because the hybrid is heterozygous across all genetic loci that differ between the parents. Thus, the degree of heterosis depends on which loci are heterozygous and how within-locus alleles and inter-locus alleles interact with each other. Interaction of within-locus alleles results in dominance, partial dominance, or overdominance, with a theoretical range of dominance degree from zero (no dominance) to larger than 1 (overdominance). Interaction of inter-locus alleles results in epistasis. Genetic mapping results have indicated that most QTL involved in heterosis and other quantitative traits had a dominance effect. As statistical methods that can estimate epistasis more efficiently become available, epistasis has been found more frequently and proven to be a common phenomenon in the genetic control of quantitative traits including heterosis. With so many genetic loci involved, it is unlikely that there is no interaction at all between any pair of them.

When a trait is controlled by multiple QTL, their alleles of positive or negative effects (increasing or decreasing trait value) tend to be dispersed among cultivars that are used as parents for developing mapping populations (Xu and Shen 1992). Considering all genetic loci controlling a specific trait, a specific genotype (cultivar) usually has alleles of increasing effect at some loci but alleles of decreasing effect at the others, which is called allele dispersion or repulsion-phase linkage if the related loci link to each other. Identification of numerous QTL from
crosses derived from two parents with the same phenotype strongly support the hypothesis of allele dispersion. Accumulation of alleles with similar effects from allele-dispersed cultivars will result in transgressive segregation, which has been proven in two rice crosses where transgressive segregation for tiller angle has been fixed in the inbred lines selected for extreme phenotypes (large tiller angle). When similar alleles dispersed in four rice cultivars were accumulated, the fixed transgressive selections showed tiller angle three times larger than that of the original parents (Xu et al. 1998). The accumulation of dispersed alleles should add to the list of the factors that can be used to explain the response of long-term selection for oil and protein contents in maize (Dudley 2004; Walsh 2004). Allele dispersion or repulsive linkage can result in pseudo-overdominance when two such loci closely linked to each other and they are not separable by linked markers. As an example, the overdominance identified in the maize cross Mo17 × B37 by Stuber et al. (1992) has been proven to be pseudo-overdominant later (Graham et al. 1997). As reported by Z.-K. Li et al. (2001), most (~90%) QTL contributing to heterosis in rice appeared to be overdominant. It is very unlikely that each of these overdominant QTL is due to pseudo-overdominance from the repulsive linkage of two completely or partially dominant QTL, although the genetic map used for heterosis mapping was less saturated.

It can be concluded that two different types of allele interaction, both within-locus and inter-locus, each should play an important role in the genetic control of heterosis. Contribution of a specific locus to heterosis could be due to any single type of these interactions. When multiple loci are involved that were not taken into account in the early 1900s, various combinations of within-locus and inter-locus interactions (especially dominance-by-dominance interaction) could contribute to the genetic control of heterosis. For a specific cross and specific trait, heterosis might be explainable by any single type of these interactions. For different crosses, species, or traits, however, their heterosis has to be explained by the dominance of different degrees in combination with all possible inter-locus interactions, as indicated by Goldman (1999). Current available information on heterosis has been provided by individual studies using specific crosses, most of which were designed for hybrid performance rather than heterosis per se. As a result, none of the mechanisms discussed above contributing to heterosis can be completely excluded. A full understanding of heterosis will depend on cloning and functional analysis of all genes that are related to heterosis. This process would be very similar to that for understanding disease resistance genes that functionally appear much simpler than heterosis.
C. Construction of Heterotic Groups

1. Germplasm Classification. Understanding the genetic structure and diversity in the gene pools with which breeders work is critical for the choice of parents, construction of heterotic groups and, thus, for the development of hybrids. Germplasm accessions can be classified based on morphological traits, geographic distribution, evolutionary and breeding history, pedigree, and/or genotypic diversity at molecular marker loci.

Geographic distribution and breeding process are straightforward criteria for germplasm classification. Rice is an example. Cultivated rice, *O. sativa*, originated in South and Southeast Asia between 10,000 and 15,000 years ago (Chang 1984) and was dispersed throughout the world by human migration and international trade. It is believed that at least two independent domestication events gave rise to the two major subspecies of *O. sativa*, namely the long-grained, tropical “*indica*” rice, and the short-grained, temperate “*japonica*” rice (Oka 1988). In addition, a third group of cultivars is recognized as “*tropical japonica*” or “*javanica*” rice, characterized by bold grains, adaptation to dry, or upland, growing conditions in the tropics (Chang 1976). Rice culture was successfully introduced into the United States around 1685, when a tropical *japonica* cultivar, ‘Carolina Gold’, was carried to Charleston Harbor on a ship that came from Madagascar (Wilson 1979). Tropical *japonica* rice became the mainstay of rice production along the Atlantic Coast for 200 years and later spread to the tidal wetlands and prairie land along the Gulf Coast (Wilson 1979). The short-grained, temperate *japonica* types grown in California are derived largely from introductions from Japan, Korea, and China (Wilson 1979; Rutger and Bollich 1991).

Germplasm accessions can also be classified based on phenotypic and/or genotypic similarity. There are several genetic distance (GD) or similarity indices that have been used as criteria for classification. A similarity index can be constructed using various types of original data. Both categorical and quantitative data have been used for phenotype-based classification. However, genotype-based classification or clustering is usually based on binary (categorical) data by scoring the presence and absence of molecular marker alleles. This scoring method has two shortcomings. First, the score does not generally reflect differences in allele sizes. For example, a difference of 2 bp and a difference of 200 bp at a SSR locus are reflected simply as a polymorphism, losing all size-related information. Second, the binary scoring method expands the data size by as many times as the number of alleles. When a clustering program has a limitation on the size of the dataset, the data has to be cut to fit the program. A better scoring method retains information about allele sizes, and similarity is then calculated for each pair of samples based on allele
sizes per se or shared allele frequency. Allele sizes can be easily converted into allele presence-absence, whereas the latter is not convertible.

In rice, a genetic similarity index was computed based on SAFs (S) between each pair of rice accessions, and then 1-S was used as the GD to construct the dendrograms depicting genetic relationships among these rice accessions (Y. Xu and S. R. McCouch, Cornell University, unpubl. data). Different types of molecular markers are suited for different levels of classification. RFLP marker-based analysis of the whole collection of 236 rice cultivars identified two major groups corresponding to *indica* and *japonica*, respectively, whereas most of the U.S. cultivars belonged to a different group. SSR markers that can distinguish closely related accessions were then used to subgroup the U.S. cultivar collection. Two subgroups were identified, representing two different types of grain shapes, long grain and medium- and short-grain, with average grain lengths 9.7 mm and 8.4 mm, respectively. The major long-grain cultivars released in Texas tended to cluster in the long-grain subgroup while the medium- and short-grain cultivars released in California formed a cluster in the short/medium subgroup. Within each subgroup, some cultivars were closely clustered. They were either closer in pedigree or more similar in morphology. Many rice accessions released in Texas clustered on the top of the long-grain subgroup. There is very limited heterosis in the hybrids derived from cultivars within this subgroup.

Southern U.S. cultivars are mostly derived from tropical *japonica* germplasm that traces its ancestry to Indonesia. This ancestral source, coupled with selection during breeding, makes the U.S. cultivars distinct from other *indica* or *japonica* cultivars, though they share many genes from these groups. For example, these U.S. cultivars may have long grains as typical of *indica* rice, although many have intermediate or short grains. Also akin to the tropical *japonica* germplasm native to Indonesia, many of these cultivars were found compatible with both typical *indica* and typical *japonica* without much hybridization barrier (Xu et al. 1989; Gu and Tang 2001). The intermediate-type cultivars and the cultivars with wide-compatibility genes, as discussed in Section VC, have been frequently used to overcome the *indica-japonica* crossing barrier, resulting in the diffusion of genes and disappearance of reproductive isolation between the two subspecies. These wide-compatible cultivars belong to the third major group or subspecies of rice cultivars. Representation of this group will greatly affect germplasm classification and heterosis studies.

Using male sterility in hybrid seed production provides the opportunity to produce large amounts of seed. Simultaneously, it raises concern that cytoplasm uniformity of CMS lines could result in genetic vulnerability, as happened in maize in 1970 because of the extensive use of
T-CMS. In rice, sorghum, and rye, only one or few CMS sources for each crop have been widely used for hybrid breeding. Genetic characterization and classification of alternative CMS sources will help establish diverse CMS pools and avoid CMS uniformity. To facilitate this, marker profiles could be used to classify genotypes. Molecular marker alleles specific to cultivar groups could be identified and then used to classify other germplasm accessions such as those containing different sources of CMS. Marker information could be exploited better for germplasm classification if it were integrated with phenotypic information, including information about CMS.

2. Concept of Heterotic Groups. The concept of heterotic groups or heterotic pools was first developed in maize based on the observation that inbreds selected out of certain populations tended to produce better performing hybrids when crossed to inbreds from other groups (Hallauer et al. 1988). This recognition resulted from the systematic crossing of thousands of inbred lines from different source populations and evaluation of the hybrids (Havey 1998). In the review of capturing heterosis in forage crop cultivar development, Brummer (1999) indicated that the key to successful semihybrid production is to keep heterotic groups separate, only intercrossing them for testing and release. Merging of heterotic groups into a larger breeding population results in the loss of interacting linkats (alleles), limiting potential yield advance. Breeding highly heterotic hybrids largely depends on selection of desirable parents as a prerequisite for most hybrid breeding programs, and thus depends on genetic diversity in the germplasm resources available to plant breeders. For example, commercial maize hybrids are typically made between inbreds from opposite, complementary heterotic groups. Therefore, construction or development of heterotic groups has been one of the key components in hybrid breeding for many crops. Introgressing exotic germplasm is often suggested as an approach to increase genetic differences between opposing heterotic populations, thereby potentially increasing heterotic response. An understanding of heterotic relationship between populations is needed to exploit exotic germplasm intelligently. Melchner and Gumber (1998) reviewed the development of heterotic groups in five major crops with different pollination systems: allogamous—maize and rye, partially allogamous—faba bean and oilseed rape, and autogamous—rice. In this section, discussion will be devoted to construction of heterotic groups based on both conventional methods and molecular marker information.

A possible explanation for heterotic groups is that populations of divergent genetic backgrounds have unique allelic diversity that may have arisen from founder effects, genetic drift, or the accumulation of
unique allelic diversity by mutation or selection. Significantly greater heterosis could result from this genetic diversity by specific interallelic interactions (overdominance), repulsion-phase linkage among loci showing dominance (pseudo-overdominance) (Havey 1998), and/or inter-locus interaction (epistasis). Apparently, the most obvious potential heterotic groups are either geographically separated populations or separate subspecies and ecotypes. Melchinger and Gumber (1998) recommended the following criteria for the identification of heterotic groups and patterns in descending order of importance: (1) high mean performance and large genetic variance in the hybrid population to ascertain future selection response; (2) high per se performance and good adaptation of both or at least one of the parental heterotic groups; (3) low inbreeding suppression in the source materials for the development of inbreds; and (4) a stable CMS system without deleterious side effects, as well as effective restorers and maintainers, if hybrid breeding is based on CMS.

3. Construction of Heterotic Groups Based on Hybrid Performance. With large numbers of inbred or open-pollinated lines or populations available, it is unfeasible in most crops to make diallel crosses and produce sufficient F₁ seed for multi-environment field-testing. Therefore, Melchinger and Gumber (1998) suggested a multi-stage procedure to identify heterotic groups, which consists of the following steps: (1) grouping the germplasm based on genetic similarity; (2) selection of representative genotypes (e.g., two or four lines or one population) from each subgroup for producing diallel crosses; (3) evaluation of diallel crosses among the subgroups together with parents in replicated field trials; and (4) selection of the most promising cross combinations as potential heterotic patterns using the identification criteria. If established heterotic patterns are available, using selected elite genotypes from them as testers for the production and evaluation of the germplasm to be classified is recommended. Based on the testcross performance, populations or lines having similar combining ability and heterotic response could be merged to constitute a new independent heterotic group, if they behave differently from the existing heterotic groups; however, if their behavior is similar to an existing heterotic group, they could be merged with it to enlarge its genetic base. Heterotic patterns in many crop species have been established solely based on the large numbers of test-crosses and breeding experience, without the use of DNA-based markers.

Ron Parra and Hallauer (1997) reviewed heterotic patterns used in the major maize production regions of the world. Some patterns have had importance in specific production regions. Others have been exploited on several continents, for example, the heterotic patterns based on ‘Reid Yellow Dent’ (RYD) and ‘Lancaster Sure Crop’ (LSC) from the
temperate United States, and ‘Tuxpeño’ and Estación Tulio Ospina’ from tropical Mexico and South America. Two heterotic groups from which inbreds commonly are selected and used to produce superior maize hybrids are Iowa (B) Stiff Stalk Synthetic (BSSS) and derivatives of LSC (Darrah and Zuber 1986; Gerdes and Tracy 1993). Although both populations are primarily comprised of southern dent germplasm, LSC has more northern flint germplasm than BSSS (Smith 1986; Gerdes and Tracy 1993). With genetically balanced sets of crosses, inter-group hybrids out-yielded the respective intra-group hybrids by 21% in RYD × LSC crosses (Dudley et al. 1991) and by 16% in Flint × Dent crosses (Dhillon et al. 1993). In both studies, the percentage of increase in heterosis for yield of inter-group over intra-group crosses was about twice as large as for the hybrid yield itself. Intrapopulation maize hybrids have been developed and commercialized in tropical maize. Some of the earlier hybrids were intrapopulation hybrids. The first hybrid (‘Suwan 2301’), developed by Kasetsart University in Thailand, was an intrapopulation interline hybrid derived from ‘Suwan-1’. Data suggested that interpopulation interline hybrids are generally superior (Han et al. 1991). Even when the populations are not heterotic, the interpopulation interline hybrids give superior performance. However, populations such as ‘AED’ × ‘Tuxpeno’ (P44), ‘Tuxpen’ (P21), and ‘La Posta’ (P43) have produced outstanding intrapopulation interline hybrids. It is possible to produce good hybrids from the same population provided it has high per se performance and high general combining ability (Vasal et al. 1999).

The systematic search for suitable heterotic patterns by Hepting (1978) laid the foundation for hybrid breeding in rye. Based on the hybrid performance and heterotic derivation in a complete diallel among seven open-pollinated populations, he found that cross combinations involving populations from the two most widely used germplasm groups, Petkus and Carsten, were most promising for grain yield. In fact, all rye hybrids released in Germany since 1985 are based on the Pampa CMS system and are of the Petkus × Carsten type.

In oilseed rape, Grant and Beversdorf (1985) evaluated diallel crosses among six OPCs of European and Canadian spring rapeseed and suggested European spring × Canadian spring rapeseed as a promising heterotic pattern. Similarly, Lefrot-Buson et al. (1987) found crosses between winter rapeseed of European × Asian origin to be most productive. The results of diallel crosses among seven spring rapeseed cultivars of European, Canadian, and Asian origin confirmed that crosses between European × Asian and Canadian × Asian spring rapeseed exhibited higher heterosis than crosses between parents originating from the same region (Brandle and McVetty 1990).
Four distinct heterotic groups within sunflower are now being utilized by breeders throughout the world (Vear and Miller 1993). OPCs developed in Russia are used in deriving female maintainer inbred lines. The U.S. restorer group, derived by crossing wild annual species of sunflower with cultivated lines, is a distinct source of disease resistance and fertility restorer genes. Romanian female lines, along with their South African derivatives, are used throughout the industry. Also used are the Argentinean INTA OPCs for developing female lines (Miller 1999).

Rice might be the only crop where hybrids are widely grown but very few studies on heterotic groupings have been reported. Heterosis in rice has been utilized largely through CMS. Fortunately, rice breeders in China identified the restorers for CMS from geographically distant rice cultivars from southeastern Asia and used them in hybrid rice breeding. This resulted in high levels of heterosis among intra subspecies \((\textit{indica} \times \textit{indica})\) hybrids. A large-scale screening of diverse CMS maintainers and restorers provided some clue as to heterotic pattern. Three ecotypes from different subspecies, \textit{indica}, \textit{japonica}, and \textit{javanica}, have different morphological and physiological characteristics and ecogeographical distribution and, therefore, serve as a basis for defining distinct heterotic groups. As summarized by Yuan (1992a), heterosis for grain yield in crosses among the three rice ecotypes has the following trend: \(\textit{indica} \times \textit{japonica} > \textit{indica} \times \textit{javanica} > \textit{javanica} \times \textit{japonica} > \textit{indica} \times \textit{javanica} \times \textit{japonica}\). This mirrors the current situation of heterotic pools in rice. It is well known to hybrid rice breeders that a high level of heterosis results from crosses between CMS lines bred in China and restorer lines derived from southeast Asian \textit{indica} cultivars, which is the heterotic pattern for \(\textit{indica} \times \textit{indica}\) hybrids.

Wheat breeders lag much behind their colleagues in other crops in establishing heterotic pools (Jordaan et al. 1999). Heterotic groups have not been well described in vegetables either.

4. Construction of Heterotic Groups Using Molecular Marker Information. DNA-based markers may be used to classify parental lines into different heterotic groups, each with a high level of similarity in genetic backgrounds. This reveals genetic diversity at the whole genome level, and helps identify effects of selection, genetic drift, and mutation. Molecular markers have been playing an increasingly important role in the construction of heterotic groups since the 1990s. Most reports are focused on maize, wheat, barley, and canola. Because marker-based groupings reflect the genetic differences among parental lines, they can contribute to parental improvement and to effective selection for heterotic hybrids. In general, heterotic groups constructed on the basis of
marker information match up very well with pedigrees, but have the advantage that missing historical information, such as the incomplete pedigree information or ambiguous pedigree, will not affect the marker-based method.

Using RFLP markers, Meng et al. (1996) grouped 46 cultivars of *Brassica napus*, originating from China, into six groups, and a very significant difference was found between the cultivars from China and a group of six cultivars from *B. napus* L. var. *oleifera* subvar. *biennis* originating from Europe. They suggested that the latter could be used to broaden the genetic diversity of breeding populations of oilseed rape.

In maize, different types of molecular markers have been successfully used to differentiate heterotic groups with results that are consistent with pedigree-based grouping (Mumm and Dudley 1994; X. Liu et al. 1997; Peng et al. 1998; Wu et al. 2000). Based on heterosis and combining ability analyses using cultivars from different heterotic groups, Peng et al. (1998) proposed seven heterotic patterns for the utilization of maize heterosis. Divergence at molecular marker loci has been useful in assigning maize inbreds to known heterotic groups previously established in breeding programs and the molecular information agreed with pedigree information (Lee et al. 1989; Melchinger et al. 1991; Messmer et al. 1993).

In rice, Zhang et al. (1992) analyzed 12 *indica* and 14 *japonica* cultivars using RFLP markers. The average GD measured by RFLPs between *indica* lines was three to four times higher than that between *japonica*, which confirmed the results based on the morphological studies. Using 160 RFLP markers and 21 wide-compatibility cultivars and three *indica* and three *japonica* cultivars, Zheng et al. (1994) constructed a dendrogram tree and discussed the potential of wide compatibility in hybrid breeding using *indica/japonica* crosses. Based on diallel crosses among eight *indica* lines representing the parents of the best-performing commercial rice hybrids grown in China, Zhang et al. (1995) studied molecular divergence and hybrid performance. Their results suggest the existence of two heterotic groups within *indica*, one comprised of rice strains from southern China and the other comprised of strains from Southeast Asia. Using two types of molecular markers, RFLPs and AFLPs, Mackill et al. (1996) obtained similar grouping results. Using RAPD and SSR markers, Xiao et al. (1996a) separated the ten parental lines into two major groups that correspond to *indica* and *japonica* subspecies. These researches indicated that molecular markers are useful tools in detection of genetic diversity between parental cultivars.

The results from barley (Melchinger et al. 1994) and wheat (Sun et al. 1996; Ni et al. 1997) also supported the conclusion that DNA markers are very useful tools for construction of heterotic groups.
5. Future Directions. Heterotic groups are the backbone of successful hybrid breeding. Decisions regarding the definition and utilization of heterotic groups are of fundamental importance and must be made at the beginning of a crop improvement program. In most cases, breeding for heterosis without knowledge of heterotic patterns has proven to be a hit-or-miss approach (Jordaan et al. 1999).

It is evident from the review of various studies that adapted populations, isolated either by time and/or space, are the most suitable candidates for promising heterotic patterns. Genetic diversity can be related to geographic origin of parental lines. The geographical variation can be related to ecological and environmental variations that, in turn, dictate survival fitness, created by spontaneous and induced genetic variation in natural and directed-selection situations. Consequently, the parental lines derived from different geographic origins are considered to have more genetic diversity than those derived from the same geographic origin.

International breeding efforts through various collaborative breeding programs in rice, wheat, and maize have been very successful in breeding both hybrid and non-hybrid cultivars. However, internationalization of plant breeding efforts and massive exchange of unimproved and improved germplasm throughout the world have altered the genetic structure and adaptation of germplasm accessions with which breeders have been working. As a consequence, differences in geographic origin of the parental lines may not always reflect genetic diversity among them. On the other hand, extensive hybridization practiced in several international and national crop breeding programs has created new forms of genetic diversity, and one can expect to find substantial genetic diversity among parents from the same geographical origin (Virmani 1996). The negative effect of using distant crosses is the confusion of heterotic groups existing among cultivars of different geographic origins. For example, breeding wide-compatible inbred cultivars as a bridge for harnessing indica/japonica heterosis in rice has reduced heterosis compared to what would be expected from crosses between typical indica and japonica cultivars. Therefore, it is important to keep in mind that we are not disturbing the current heterotic groups, which have been established either naturally or creatively, when we use distant crosses in hybrid breeding programs.

Heterotic groups should not be considered as closed populations, but should be broadened continuously by introgressing unique germplasm to warrant medium- and long-term gains from selection. Heterotic groups consisting of poorly utilized and unadapted germplasm should be enhanced through joint public-private breeding ventures. Different phenotypes may or may not reflect divergent genetic backgrounds.
Phenotypically different populations may possess the same genetic background, and divergent phenotypes may be conditioned by allelic differences at relatively few loci (Havey 1998). MAS can be useful in creating, maintaining, and improving heterotic groups. As discussed above, marker-based grouping of germplasm and breeding populations will help establish heterotic groups that hold maximum genetic diversity between groups but minimum diversity within groups. Identification of marker alleles that are specific to each heterotic group will help keep them genetically separated. MAS can be used to improve the existing heterotic groups through introgressing target genes from one heterotic group or outsource germplasm to another with minimum linkage drag from the donor. As we discussed previously, MAS will help breeders realize their goals without linkage drag and unwanted genetic background.

D. Hybrid Prediction

1. Reasons for Hybrid Prediction. Hybrid breeding includes two major procedures: breeding parental lines and selection for the best hybrids from the cross combinations of those parental lines. These procedures involve a large amount of work for field evaluation, testcrossing, and progeny tests. Breeders continually have to decide which experimental single crosses to test, which advanced hybrids to recommend for further testing or commercialization, and which inbred parents to cross to form new base populations for inbred/population development (Bernardo 1999). Suppose a breeder has 100 inbreds from heterotic group 1 and 100 inbreds from heterotic group 2. There are 10,000 possible (group 1 × group 2) single crosses. For developing new hybrids, there are 495,000 possible (group 1 F2) × (group 2 tester) combinations, and 495,000 possible (group 1 tester) × (group 2 F2) combinations, if testcrossing starts from the F2. Due to limited resources, breeders are unable to test all combinations in all environments of interest but may test a limited set of single crosses and F2 × tester combinations. Typically, <1% of the maize single crosses tested by a breeder eventually become commercial hybrids (Hallauer 1990). Therefore, predicting hybrid performance has always been a primary objective in all hybrid-breeding programs. Methods for predicting the performance of single crosses would greatly enhance the efficiency of hybrid breeding programs. Development of a reliable method for predicting hybrid performance and/or heterosis without generating and testing hundreds or thousands of single cross combinations has been the goal of numerous studies using marker data and combinations of marker and phenotypic data, particularly in maize and rice.

It is reasonably believed that heterosis originates, in some way, from the genetic differences or heterozygosity between the parents. Theoret-
ically, hybrid performance is equal to the average parental performance plus heterosis. In the past several decades, hybrid prediction has been largely based on the evaluation of genetic diversity among parental lines. It has been expected that understanding the relationship between heterozygosity/parental difference and heterosis would help predict hybrids. The development of molecular marker techniques has provided new tools for hybrid prediction and DNA markers have been used extensively in investigating correlations between parental GD and hybrid performance.

2. Statistical Methods for Hybrid Prediction. The best linear unbiased prediction (BLUP) procedure has been used for decades for evaluating the genetic merit of animals, especially dairy cattle. Intrapopulation, additive genetic models have traditionally been used for BLUP in animal breeding (Henderson 1975). During the last several years, Bernardo has attempted to use BLUP in maize breeding with interpopulation genetic models that involve both GCA and SCA (Bernardo 1994, 1996). Results have indicated that BLUP is useful for routine prediction of single-cross performance. The predicted performance of single crosses may subsequently be used to predict the performance of \( F_1 \times \) tester combinations, three-way crosses, or double crosses. Along with the pedigree relationship, the BLUP method can use trait data, or both trait and marker data, for prediction.

3. Genome-wide Heterozygosity and Hybrid Prediction. In most cases, isozyme-based GD estimates are positively associated with hybrid performance for grain yield (for a review, see Stuber 1994); however, it remained unclear whether the low correlations between both measures observed in most cases were due to poor coverage of the genome or whether other causes were involved. The availability of a large number of DNA-based markers provided the opportunity for genome-wide surveys of heterozygosity and laid the foundation for a better understanding of the relationship between heterozygosity at marker loci and heterosis.

The relationship between parental genetic divergence and hybrid performance was first studied in maize. Variability for molecular markers generally agreed with pedigree information and assignment (based on hybrid performance) to known heterotic groups (Smith et al. 1990; Dudley et al. 1991; Melchinger et al. 1991); however, variability at molecular marker loci was ineffective in predicting specific hybrid performance from crosses among maize inbreds (Lee et al. 1989; Melchinger et al. 1992). Some reports indicated high correlation between hybrid performance/heterosis and parental GDs or the degree of heterozygosity (Smith et al. 1990; Lee...
et al. 1989; Stuber et al. 1992), while others revealed very weak correlations (Godshalk et al. 1990; Dudley et al. 1991). Considering all these and other studies in maize, the correlation depends largely on the plant materials and their origins and suggests that specific loci are involved in heterosis rather than heterozygosity per se. Correlations between single-cross performance and molecular marker diversity for unrelated parental inbreds have been too low to be of any predictive value (Godshalk et al. 1990; Melchinger et al. 1990; Dudley et al. 1991). Molecular-based GD estimates also failed to predict superior hybrid performance in oat (Moser and Lee 1994), soybean (Gizlice et al. 1993), and chickpea (Sant et al. 1999).

4. Hybrids Are More Predictable Within than Between Heterotic Groups. Let us first consider intra-heterotic crosses. Correlations between heterozygosity/GD and hybrid performance/heterosis varied for hybrids between lines that belong to the same heterotic group (within-group hybrids). In maize, correlations of GD with $F_1$ performance and heterosis were significant and positive for all traits of within-group hybrids, flint × flint crosses (Boppenmaier et al. 1993). This was supported by Benchimol et al. (2000) using 18 tropical maize inbred lines where correlations of parental GDs with single crosses and their heterosis for grain yield were higher for line crosses from the same heterotic groups. The same conclusion was reported by Zhao et al. (1999) using diallel crosses derived from 11 elite rice lines where the correlations of marker heterozygosity with hybrid performance and heterosis were high in crosses within subspecies. In other cases, however, weak or no correlation was found for within-group hybrids. Examples include weak or no significant associations of GD with $F_1$ performance and mid-parent heterosis in soybean (Cerna et al. 1997), wheat (Martin et al. 1995), and U.S. long-grain rice cultivars (Saghai Maroof et al. 1997). These results may be due to the low levels of heterosis in these varietal groups.

Weak or no correlation was found for hybrids between lines that belong to different heterotic groups (between-group hybrids). In maize, correlations of GD with $F_1$ performance and heterosis were not significant for the subsets of flint × dent and dent × dent crosses (Boppenmaier et al. 1993). Using 18 tropical maize inbred lines, Benchimol et al. (2000) found that correlations of parental GDs with single crosses and their heterosis for grain yield were low for line combinations from different heterotic groups. In rice, Xiao et al. (1996a) reported that yield potential and its heterosis showed significantly positive correlations with GD for $indica \times indica$ or $japonica \times japonica$ crosses, but the correlations were not significant for $indica \times japonica$ crosses. This was confirmed by Zhao et al. (1999) that very little correlation was detected in intersubspecific crosses using diallel crosses derived from 11 elite rice cultivars.
A theoretical study conducted by Charcosset and Essioux (1994) supported the reports discussed above. The correlation between heterozygosity at marker loci and heterosis was investigated for (1) hybrids between lines that belong to the same heterotic group (within-group hybrids); (2) hybrids between lines that belong to different groups (between-group hybrids); and (3) all hybrids, both within- and between-groups. Within a group, significant correlations may be detected because of linkage disequilibrium generated by drift. At the between-group level, no correlation is expected since linkage disequilibrium should differ randomly from one group to another, which is consistent with the results of recent experiments. When all hybrids are considered simultaneously, divergence of allelic frequencies among groups for the markers and the QTL produces a correlation between heterosis and heterozygosity at marker loci. This correlation increases with the number of markers that are considered.

Based on results from various studies in maize, Melchinger (1993) summarized the relationship between parental GD and mid-parent heterosis (MPH) in a schematic representation. For crosses among related lines, there exists a tight association between GD and MPH for yield characters because both measures are a linear function of coancestry, $f$, and, thus, decrease with increasing $f$. For intra-group crosses, the correlation $r$ (GD,MPH) is generally positive, too. This can be explained by hidden relatedness between some parents considered to be unrelated based on their pedigree, and the presence of the same linkage phase between QTL and marker loci in the maternal and paternal gametic arrays of intra-group hybrids, which results in a positive covariance between GD and MPH (Charcosset et al. 1991). In contrast, no significant association between both measures exists for inter-group hybrids. In this case, the maternal and paternal gametic arrays may differ in the linkage phase for many QTL-marker pairs; as a consequence, positive and negative terms cancel each other in their net contribution to covariance (GD,MPH), resulting in a low or zero correlation (Charcosset and Essioux 1994). High estimates of $r$ (GD,MPH) could be expected if correlations are calculated across different types of crosses due to group effects. This is because both GD and MPH are expected to increase from crosses among related lines to intra-group crosses and further to inter-group crosses (Melchinger 1999).

5. Heterosis-associated Markers and Hybrid Prediction. It has been common practice in most studies to determine GD or heterozygosity estimates from a set of DNA markers chosen for good coverage of the entire genome but not for linkage to genes influencing heterosis of the target trait. Theoretical investigations (Charcosset et al. 1991) and computer
modeling (Bernardo 1992) demonstrated that with intra- and inter-group crosses the correlation between GD and MPH is expected to decrease if genes influencing heterosis are not closely linked to markers used for calculation of genetic estimates and vice versa if markers employed for calculation of GDs are not linked to genes controlling the trait. Hence, increasing the marker density alone will not necessarily improve the ability to predict MPH by GD estimates; rather, markers must additionally be selected for tight linkage to genes affecting heterosis of the target trait in the germplasm under study. This is corroborated by comparison of results obtained with 209 AFLPs vs. 135 RFLPs (Ajmone Marsan et al. 1998) and a study by Dudley et al. (1991). Using these associative loci will help establish strong correlations between heterozygosity and heterosis. However, allelic differences at marker loci do not assure allelic differences at linked loci for heterosis. For a limited number of markers to be useful as predictors for hybrid performance, the effects of alleles at the loci linked to specific marker alleles must be ascertained (Stuber et al. 1999).

Zhang et al. (1994a) proposed two statistical parameters: general and specific heterozygosity, to measure genotypic heterozygosity. The former is the heterozygosity calculated from the GDs between the parents using all possible markers, and the latter is that from using marker loci that are significantly associated with the traits of interest revealed by single factorial analysis of variance. The results from rice indicated that there was a weak correlation between general heterozygosity and heterosis but a significant correlation between specific heterozygosity and heterosis for yield and biomass. As a result, heterosis could be predictive to some extent (Zhang et al. 1994a, 1995).

Joshi et al. (2001) studied the correlations between GD, hybrid performance, and heterosis in rice using the markers associated with grain yield and other traits reported by Xiao et al. (1996b) and McCouch et al. (1997) and 21 F1 hybrids derived from two CMS lines and 14 restorer lines. There was no significant correlation for grain yield. Two factors contributed to this result: loose marker-trait associations and different genetic systems for controlling heterosis and grain yield in the materials used by different researchers. As a result, markers identified for grain yield from one population cannot be used to predict heterosis for another. Successful heterosis prediction based on molecular markers will depend on how many genes have been identified from different genetic backgrounds and how these genes interact with each other to control heterosis. The complexity of these interactions explains why there are no successful examples of heterosis prediction.

Attempts to identify QTL-marker associations by use of multiple regression techniques, in which observed phenotypic values for a fixed
set of hybrids are regressed on their coded marker genotypes, must be regarded with great caution, because the number of variables from which the regressors are selected is usually as large or even greater than the number of phenotypic observations. Moreover, comparison of QTL mapping results from different populations in maize (Stuber 1995; Lübberstedt et al. 1998) and rice (Xu 2002) suggested that QTL regions affecting a given trait are not necessarily consistent across different germplasm accessions.

6. Favorable Allele Combination and Hybrid Prediction. Although the importance of genetic diversity to heterosis has been known for several decades, heterogenic gene combinations may not always lead to heterosis and heterosis is ultimately dependent upon the balance between favorable and unfavorable interactions of genes. It is reasonably inferred that heterosis could be caused by specific gene combinations derived from the two parents. Those genes may simultaneously produce different genetic effects in different genetic backgrounds. So, for parental improvement and hybrid prediction, investigating the specific gene combinations that contribute to heterosis should be more important than studying any single gene or QTL. Using 99 half-diallel rice hybrids derived from nine CMS lines and 11 restorer lines, Liu and Wu (1998) found that four favorable alleles and six favorable heterotic patterns on the parental lines significantly contributed to the heterosis of their hybrids for grain yield, whereas six unfavorable alleles and six unfavorable heterotic patterns significantly reduced heterosis. They suggested that optimal hybrids with superior grain yield could be developed by assembling those favorable alleles into, and removing the unfavorable alleles from, their parental lines.

7. Conclusions. There are several conclusions that can be drawn from the numerous investigations on the relationships between heterozygosity and GD with hybrid performance and heterosis. First, the higher the heterozygosity between the parents, the stronger the heterosis. Second, using more markers alone will not improve the prediction. Third, prediction is possible using markers known to be associated with hybrid performance or heterosis if the association is used to predict performance of a hybrid derived from the same heterotic pattern. Fourth, genetic variation (the presence of heterosis) is a prerequisite for prediction. Fifth, the relationship of heterozygosity with heterosis and with hybrid performance will be different if the two involve different genes. The last conclusion was supported by results of Zhu et al. (2001) that heterosis was highly significant but hybrid performance was not when 57 rice accessions from six ecotypes and their hybrids were genotyped
by 48 SSR and 50 RFLP markers. In conclusion, heterozygosity at general marker loci can be used to construct heterotic groups. It is anticipated that prediction could be possible if heterozygosity is derived from specific marker loci that are associated with heterosis and hybrid performance and all possible associated loci have been identified and their effects and interactions clearly defined.

Considering the fact that only heterotic crosses are of commercial importance and of interest to the breeder, the practical value of the genetic distance approach for prediction of heterosis and hybrid performance is limited (Vuylsteke et al. 2000). This is true for some crop species like maize. For rice, however, the reproductive barrier between the two subspecies, indica and japonica, has enforced a limitation on the utilization of indica/japonica heterosis, although the use of the wide-compatibility gene(s) has had a great impact on the limitation. Hybrid breeding for indica rice has been based on crosses within the indica group. The strong relationship between the heterozygosity at marker loci and heterosis within the indica group as reported before (Xiao et al. 1996a) indicates that GD estimates based on molecular markers could be very useful in assigning indica cultivars into different subgroups for hybrid indica rice development.

Screening for heterosis-related molecular markers as suggested by Melchinger et al. (1990), using specific heterozygosity proposed by Zhang et al. (1994a), and identifying favorable combinations of allele and heterotic patterns (Liu and Wu 1998) are among the approaches that could be exploited further to improve the prediction of hybrid performance/heterosis using molecular markers. Understanding genetic variation among cultivars to be tested and identifying markers associated with heterosis and heterosis-related traits are two important components in hybrid prediction. We should keep in mind that marker-heterosis associations identified in one cross may not be suitable for selection in others because heterosis could be controlled by many genes and each cross has different genes and gene combinations in action.

Despite their low values, the inbred-hybrid yield correlations were positive. They indicated a tendency for high-yielding inbreds to produce high-yielding hybrids. Hybrid breeding is always accompanied by the improvement of parental lines. Modern maize inbreds, grown at today’s high density, can yield nearly as much as hybrids of the 1930s (Duvick 1984; Meghi et al. 1984). Duvick (1999) has suggested that if as much effort had been put into improvement of OPCs as has been devoted to hybrid improvement over the years, the gap between the best hybrids and the best OPCs might be less than what it currently is. Some authors even argue that OPCs might be superior to hybrids (Lewontin and Berlan 1990), but their assumption is not backed up by data. No experiments
have been reported that test the theory that OPCs could be improved at the same rate as hybrids (or at a higher rate), if equal effort were expended on each kind of breeding (Duvick 1999). This assumption, if true, would greatly support the dominance theory of heterosis.

In conclusion, the potential application of DNA markers in hybrid breeding depends very much upon whether divergent heterotic groups have been established or not and upon crop species. If well-established heterotic groups are unavailable, marker-based GD estimates can be used to avoid producing and testing crosses between closely related lines. Furthermore, crosses with inferior MPH could be discarded prior to field-testing based on prediction. Another potential application exists. If new lines of unknown heterotic pattern or inbreds developed from crosses between parents from different heterotic groups (e.g., commercial hybrids) are to be evaluated for testcross performance, GD estimates could assist the breeder in the choice of appropriate testers for evaluating the combining ability of the lines. However, with regard to the typical situation of hybrid breeding, in which crosses are produced between lines from genetically divergent heterotic groups, GD estimates based on an unselected set of DNA markers alone are not promising for predicting hybrid performance or any of its components.

X. SEED QUALITY ASSURANCE

Production of high-quality hybrid seed is crucial to the success of a heterosis-breeding program. Two major factors influencing seed quality are purity and viability. Hybrid seed do not show large differences from non-hybrids in viability despite heterosis for some traits such as germinating ability or disadvantages such as short seed life for hybrid rice because of the partially enclosed hull. In this section, only seed purity will be discussed. Two major factors contribute to the purity of hybrid seed: off-types and false hybrids. Off-types includes non-hybrid seed resulting from non-parental plants in parental multiplication and hybrid seed production, and false hybrids includes non-hybrid seed resulting from the selfing of the female parent due to instability of male sterility if a male sterile line is involved in seed production.

A. Off-type

Off-type in hybrid seed and grain production is a potential quality problem. Traditionally, off-types are defined as individual plants that are phenotypically different from the plants developed by breeders. They may result from mechanical mixtures, outcrossing, mutation, or residual
genetic variation. Off-types may be found in a parental line or only visible in the hybrid. The presence of off-type plants will reduce the uniformity of the crop and thus reduce its productivity and quality. Phenotypic off-types can be easily rogued if there are not too many. Besides the off-types that are visible phenotypically, many off-types are genetically different from the typical plants and they are hard to distinguish visually. Genotypic off-types may impose more severe effects on hybrids, which could be one of the major reasons for reduced performance of hybrid cultivars or inbred lines. They also impose an effect on the parent, which could be exaggerated by multiplication of the parents themselves and their progeny. Molecular marker technology provides a powerful way for distinguishing both the phenotypic and genotypic off-types from hybrid plants and their parents. Marker-trait associations and high-resolution molecular markers such as SSRs could be used to distinguish two plants with very similar genetic backgrounds. With ten or more SSR markers, rice breeders can identify distinct off-types from their breeding populations and hybrid seed bulks and obtain detailed genotypic information such as where the off-type genotypes come from and what the proportion of the off-types is to the typical plants. A selection and purification decision can be made to refine their breeding materials and hybrids.

B. False Hybrids

False hybrids come from the selfing of the female parent because a male sterile line becomes fertile due to instability of male sterility. This has happened to two-line hybrid rice when the EGMS female parent becomes fertile during seed production. As discussed previously, a TGMS line can serve as a sterile line under one environmental condition and can propagate itself under another. The ability to maintain sterility makes TGMS lines practicable as a female to cross with any other commercial rice cultivars. However, most TGMS lines require a specific temperature to maintain their sterility. Abnormal weather can bring the temperature down below the critical temperature that is required for conversion of TGMS lines from sterility to fertility (or simply called fertility conversion), which makes TGMS lines fertile or partially fertile in locations where they are sterile in normal years. This results in self- or sib-seed contamination, a potential problem for seed production of two-line hybrid rice. The mixture of real hybrids with selfed seeds from the EGMS line cannot be used in rice production, resulting in a great loss to seed producers or rice growers. As a way of insuring the seed production, marking the seeds from EGMS lines using genetic markers can help identify and remove the false hybrids from the mixture.
1. Approaches to Eliminating False Hybrids. To eliminate pseudo hybrids from seed production, three strategies can be used: (1) breeding herbicide mutants and use of the herbicides to kill false hybrids, (2) use of morphological markers (such as purple leaf mutants) to distinguish the false hybrids and remove them by hand, and (3) use of mutants with faded green leaves to get rid of false hybrids.

Morphological and chemical markers have been investigated as a way to identify specific seeds/plants from a mixture. In rice, several morphological markers, such as purple leaves (Mou et al. 1995) and pale green leaves (Dong et al. 1995), have been used for marking EGMS lines. These markers can be used to distinguish real F₁ hybrids from selfed seeds (false hybrids) at the seedling stage. However, removing the purple false hybrid seedlings must be done by hand, which is labor-intensive and there is no assurance that all false hybrids will be completely eliminated. Faded green mutants with chlorophyll deficiency during the seedling stage cannot compete with normal plants and will die in the densely planted seedling nursery but will grow well in a less dense planting.

Selective responses of plants to herbicides can be used to identify a specific type of plant from a mixture and the response can be exploited as a “marker.” There are two different types of herbicides that can be used as selective chemicals to remove false hybrids in rice and other crops: herbicides that can selectively kill a crop and ones that are safe to the crop. Prerequisites for use of herbicide mutants in hybrid seed production include availability of herbicides (registered chemicals), low cost of herbicides, response to low concentration of herbicides, long sensitivity duration, minimal effect of environmental factors/weather, and easy identification of the herbicide response.

2. Herbicide Mutants in Rice. In rice, there are several herbicide mutants reported, including IMI mutants such as IMI-Cypress, resistant to imidazone; Bar (phosphinothrin acetyltransferase), resistant to Basta (Liberty); EPSP synthase, resistant to Roundup; and Norin 8 mutant, recessively sensitive to bentazon at the flowering stage.

Zhang et al. (1998) transferred the bar gene into a rice restorer line to make it resistant to the herbicide Basta, which can be used to selectively kill false hybrids. When a herbicide is safe to use with normal rice cultivars, recessive sensitivity to the herbicide can be used to selectively kill the seedlings from the selfed seeds of the female once it possesses the sensitivity. The herbicide-sensitive rice mutant obtained by Mori (1984) through radiation was lethal to bentazon, which was controlled by a recessive gene.
To prevent seed production from selfing contamination, Zhang et al. (2002) developed a system to secure seed purity using a herbicide sensitive TGMS mutant, M8077S, obtained by radiation. Genetic analysis using the F1, F2, F3 populations derived from this mutant and other normal cultivars revealed that bentazon lethality/sensitivity was controlled by a single recessive gene, bel. The mutant can be killed at the seedling stage by bentazon at 300 mg/L or higher, a dosage that is safe for its F1 hybrids and all other normal cultivars. This mutant is also sensitive to all the tested sulfonylurea herbicides. The response of segregating plants to these two types of herbicides indicated that the sulfonylurea sensitivity was also controlled by bel. By crossing this mutant with Pei-Ai 64S (a TGMS line), the bel locus was located on chromosome 3 with 7.1 cM from the closest microsatellite marker RM168. Phenotypic analysis indicated that the bel gene had no negative effect on agronomic traits in either homozygous or heterozygous status.

Because the M8077S mutant is lethal in the response of bentazon, homozygous plants at the mutated locus will be killed by the herbicide and selection in breeding programs must be made without application of the chemical. Associated microsatellite markers can be used to select these homozygous plants without spraying the herbicide. These markers have also been used to identify heterozygous plants during continuous backcrossing procedures for gene transfer from one genetic background to another. After two cycles of MAS, a new version of Pei-Ai 64S (a widely used EGMS cultivar) with bentazon sensitivity was obtained and is ready for seed production (Zhang et al. 2002).

The Bel gene can also be used to test for purity in seed production. Hybrid rice seeds must be tested for purity before release to rice producers. Traditionally seed samples were planted and evaluated for purity after flowering, when they could be distinguished from the false hybrid plants based on distinct agronomic traits. In order to obtain purity test results before the next planting season, seed samples are usually sent to a location such as Hainan, China, where rice can be planted in the winter. Nevertheless, this is labor-intensive and also very expensive. Using bentazon sensitivity, false hybrids can be detected by spraying at the 2- to 3-leaf stage. A seedling tray in a greenhouse or in a growth chamber will be enough for a purity test required for any sample of hybrid seeds, which functions the same way as molecular markers used to identify off-types and false hybrids.

In seed production, sterile plants are currently planted with pollinators in alternative rows. This system requires a great amount of manpower for transplanting and harvesting if planting is not mechanized. It is very difficult to use manpower to produce F1 seeds in countries where labor is limited and/or very expensive. Mixed planting of sterile plants
with pollinators could make mechanized seed production easier. This could also ensure a higher ratio of seed set on male sterile plants because the average distance between the pollinators and male sterile flowers becomes closer than under the alternative-row planting (Zhang et al. 2002). As discussed by Maruyama et al. (1991), incorporating a herbicide-sensitivity gene to a pollinator, that could be killed by spraying a specific herbicide right after pollination, could ensure that all the harvested seeds are F₁ hybrids only. Using PGMS and TGMS lines transformed with the herbicide-resistant gene Bar, the seeds of the herbicide resistant PGMS and TGMS lines were mixed with the pollen parent at the ratio of 4:1. The herbicide Basta was sprayed after natural pollination ended. The pollen parent died and the hybrid seeds could be harvested from the herbicide-resistant PGMS and TGMS lines. The natural outcrossing rate without artificial pollination ranged from 10.6% to 24.5% (Suh et al. 2002).

Combining the use of a herbicide-sensitive mutant with a herbicide-resistant mutant could be helpful not only for mixed plantings in hybrid seed production but also for removal of selfings. When a pollinator has a sensitivity gene to herbicide A and a resistance gene to herbicide B, it can be killed after pollination by spraying the herbicide A, while the plants from false hybrids can be killed by spraying the herbicide B. M8077S reported by Zhang et al. (2002) can be used as a gene donor for breeding herbicide sensitivity, while the bar gene as reported by Zhang et al. (1998) can be used for breeding a herbicide-resistant pollinator.

XI. GENERAL DISCUSSIONS

A. Economic Consideration

The value of MAS depends on several factors. Acceleration of breeding programs and shortening the breeding cycle would be the first advantage of MAS, because the benefits of releasing new hybrids more quickly can be substantial, particularly in competitive markets. The economic merit of MAS could include situations in which molecular costs are more than offset by savings in phenotypic evaluation. If molecular costs are in addition to, not in place of, phenotypic costs, the economic merit of MAS will become questionable and more difficult to evaluate. In other cases, the ability to select early offsets the extra costs that are associated with MAS.

The economic story from DNA sequencing may tell us what we can expect in terms of cost reduction in marker genotyping. Sequencing cost per finished base was $10 in 1990, but was reduced to $1 in 1996 and $0.1 in 2002. With further technology development, it is anticipated that the cost will be $0.01 in 2006, which is a thousand times cheaper
than in 1990. The cost of genotyping using molecular markers depends on marker type and its capacity in high-throughput analysis. For example, the lowest cost of SNP analysis is now about 20 to 30 cents per genotype, and a cost of only a few cents per genotype is expected in the coming years (Jenkins and Gibson 2002). With well-established marker systems and sequencing facilities, genotyping with SSR markers costs about 30 to 80 cents per data point, depending on marker multiplexing and the number of markers genotyped for each sample (Xu et al. 2002).

There are several ways to reduce the MAS cost. First, high-throughput analysis using automated genotyping and data scoring systems will help increase the daily data output. Second, using the same sample for selection of multiple traits will reduce the trait-based cost. Third, selection at an early stage of plant development and an early stage of the breeding process will minimize the number of plants that need to be retained so that the overall breeding cost will come down. Fourth, optimization of MAS systems, including facilities and personnel, will result in less cost per data point. However, there are very few experiments that have been done with the aim of cost reduction. A comparison between MAS and conventional greenhouse screening of common beans for resistance to common bacterial blight showed that the cost of MAS is about one-third less than that of the greenhouse test (Yu et al. 2000).

Today, a database literature search from CAB Abstracts for “marker-assisted selection” provides about a thousand hits, but, in most cases, MAS is mentioned only as a future perspective. Others have evaluated the potential of MAS using computer simulation. Overall, there are still few reports of successful MAS experiments. Apparently, almost all MAS projects in plant breeding that have been reported received special monies for demonstration of MAS applications rather than for pure breeding. Several private companies have been routinely using MAS in breeding programs, benefiting from their long-term basic research programs and the availability of all the components of MAS, as discussed in Section III. It is certainly a big investment for a breeding company/institution to start from scratch to run a MAS-based breeding program. Further work on the economic evaluation and optimization of strategies for the use of MAS in breeding programs is required. The economically optimal use of MAS will likely necessitate a complete re-designing or at least some modification of breeding schemes (Dekkers and Hospital 2002).

B. Bioinformatics and Breeding Database

Enabling infrastructure provides the raw materials and the data and the methods for collating and verifying information within the appropriate biological context. The infrastructure typically includes cDNA and
genomic sequence data, genetic maps of mutants, DNA markers and maps, candidate genes and quantitative trait loci, physical maps based on chromosome breakpoints, and libraries of large inserts of DNA such as bacterial artificial chromosomes and radiation hybrids. Information flow from molecular markers to genetic maps to sequences and to genes has been established. Apparently, however, there is a gap between the sequence-based information and breeding-related information such as germplasm, pedigree, and phenotype. We will depend on phenotyping as the basis for functional analysis of about 40% of genes even though a complete sequence is available. Integration of breeding-related information with a genomics database is required for genomics-based breeding programs.

MAS can be considered the first benefit that breeders can obtain now from genomics. The explosion of interest in marker-trait association studies has led to numerous reports in plants, each based on its own experimental population(s). Each experiment is limited in size and usually restricted to a single population or a cross planted in a specific environment. As suggested by Xu (2002), it is important for all researchers to follow general rules for reporting about genes and traits. One direction for the use of this database is to combine information from several studies, for example, by meta-analysis of results of QTL studies (Goffinet and Gerber 2000) or joint analysis of the raw data (Haley 1999). Extension of current databases to include raw data from gene mapping projects will stimulate this effort. On the other hand, many permanent populations have been shared internationally for genetic mapping. The raw data should be shared, too. A rice RFLP map constructed by using IR64/Azucena DHs has been saturated with about a thousand SSR markers (Chen et al. 1997; Temnykh et al. 2000; McCouch et al. 2002; Y. Xu, RiceTec, Inc., unpubl. data). Researchers involved in QTL mapping, however, have been using the first version of the molecular map consisting of only 175 RFLP markers. Sharing marker and phenotype information through a well-established database will make all sources of data more valuable.

C. Opportunities and Challenges

Genetic improvement through artificial selection has contributed greatly to advances in productivity that have been achieved over the past century in crops. Plant breeding has generally accounted for one-half of the increases in productivity of the major crops and the future will continue to depend on its advances. Most of the traits that are selected are complex quantitative traits. So far, most selection has been on the basis of observable phenotypic variation, which represents the collective effect
of all genes and the environment. Advances in genomics and MAS provide breeders both opportunities and challenges for the improvement of crops through manipulation of both qualitative and quantitative traits.

Well-established physical maps and the publicly available DNA sequence for a whole genome, as is currently available for rice, will help eliminate the many fine mapping steps that are required now to narrow down the candidate genome region to a kilobase resolution. Once identified, a target region could be associated with a contig that has been well located on the physical map and defined by molecular markers. On the other hand, high-throughput analysis combined with highly informative molecular markers enables us to manage populations with thousands of plants and thousands of markers in fine mapping. Recent advances in genomics have made it possible to map and determine the magnitude of the effect of individual loci controlling both qualitative and quantitative traits. Positional cloning has been successfully used to clone several QTL with relatively large effects (Frary et al. 2000; Fridman et al. 2000; Yano et al. 2000; Takahashi et al. 2001). Since heterosis and other quantitative traits are usually controlled by many genes, each with a relatively small effect, it will be a great challenge for molecular geneticists to verify the effects of minor genes (Xu 1997). If many genes are known, and favorable alleles are present in different lines or cultivars, MAS can be used to design new genotypes that combine favorable alleles at all loci.

A large number of molecular polymorphisms such as SSRs or SNPs and small or large insertions/deletions, discovered with genome sequencing, provide an opportunity for identifying the nucleotide change associated with quantitative trait variation. The nucleotide change that contributes to quantitative variation has been referred to as quantitative trait nucleotide (QTN) (Lyman et al. 1999) or FNP. Fine mapping combined with sequence analysis can quickly narrow the chromosomal region associated with quantitative variation (QTL) down to a specific nucleotide change.

Rice as a major food source for human beings has been moving ahead of other crop plants in terms of genome sequencing. Any success in this crop will benefit others in certain ways. Rice has become the model species for the cereals; chromosomes of other cereals such as maize, sorghum, sugarcane, millet, oats, wheat, and barley share much similarity to each other and to rice. The complete DNA sequence of rice will allow the tracking of genes/traits from rice to other grass species. It has become apparent that the differences between species of plants appear due to novel allelic specifications and interactions other than novel genes.

Can molecular genetics improve our understanding and manipulation of heterosis? Plant complexity and numerous interactions with the envi-
ronment will make it difficult or impossible for molecular genetics to identify universal aspects of heterosis suitable for alternation. Elements of a crop’s heterotic response are interwoven with many aspects of plant metabolism and development. We may not find a biologically based unifying mechanism or pathway of heterosis. Certainly, molecular genetics alone will not accomplish this, especially if there are not adequate human resources to analyze, explore, and integrate new sources of information while maintaining a highly effective infrastructure of plant breeding (Lee 1999). Without a deeper understanding of the physiological determinants of yield potential, molecular approaches that seek to empirically concentrate “yield genes” or heterosis genes are likely to fail. Instead, molecular geneticists must actively collaborate with crop physiologists, agronomists, and plant breeders so that genetic differences in yield potential and heterotic loci can be properly measured and identified. Additionally, measurement of yield-determining traits should be made on a time-series of hybrids grown at yield-potential levels, to determine if the changes in hybrid characteristics in breeding history have contributed to an increase in yield potential (Duvick and Cassman 1999). As our awareness and sophistication grows, with the new information, it should be possible to develop more rational and information-driven assessments and strategies.

Molecular-marker technology has revolutionized our understanding of quantitative traits in different backgrounds at different levels of genomics. Considering interactions of all types that could happen among numerous genes with minor effects, and interaction with external environments that could change from time to time and from location to location, we need to develop much more efficient systems for separating, pyramiding, and packaging all the genes related to hybrid performance and heterosis and make them functional at full scale in specific environments. These systems would be too complicated to be practical. Increasing efficiency should be a major objective for future generations of geneticists and plant breeders. Highly informative markers, isogenic mutation libraries, high-throughput technology, availability of full sequences from several model plants and genomics software are key tools for genetic manipulation of genes for quantitative traits. With all these developments, MAS could be an important component in plant breeding, especially for heterosis and hybrid performance.

The accepted dogma has been that recent advances in molecular genetics promises to revolutionize agricultural practices. However, as Lande and Thompson (1990) point out, there are several reasons why molecular genetics can never replace traditional methods of agricultural improvement, but instead should be integrated to obtain the maximum improvement in the economic value of domesticated populations. The
analytical results, as well as the more recent computer simulations and the limited empirical results (Stuber et al. 1999), are encouraging and support the use of genetic markers to achieve a substantial increase in the efficiency of hybrid breeding.

LITERATURE CITED


Fehr, W. R. (ed.). 1984. Genetic contribution to yield grains of five major crop plants, Special Publ. No. 7. CSSA, Madison, WI.


3. MARKER-ASSISTED SELECTION STRATEGIES FOR BREEDING HYBRID RICE


3. MARKER-ASSISTED SELECTION STRATEGIES FOR BREEDING HYBRID RICE


3. MARKER-ASSISTED SELECTION STRATEGIES FOR BREEDING HYBRID RICE


3. MARKER-ASSISTED SELECTION STRATEGIES FOR BREEDING HYBRID RICE


3. MARKER-ASSISTED SELECTION STRATEGIES FOR BREEDING HYBRID RICE


