Molecular Plant Breeding: Principle Methods and Applications
Eds : R.K. Singh, Rajesh Singh, Guoyou Ye, A. Selvi and G.P. Rao
Studium Press LLC, Texas, USA, 2008, pp. 1-34

CHAPTER

1

Advances in Molecular Breeding: An Introduction

GYAN P. MISHRA[1], R.K. SINGH[2], PRADEEP K. NAIK[3], RAGHWENDRA SINGH[1] and SHASHI BALA SINGH[1]*

ABSTRACT

Many interesting developments in the field of molecular breeding are currently ongoing which are accompanied with new and exciting challenges, demanding involvement of expertise from other disciplines, including statistics, bioinformatics and economics. The main focus of research conducted these days is using molecular technologies to more efficiently develop improved breeds of commercial varieties. These markers have been used to construct linkage maps, gene tagging, gene search etc. Research is being done on candidate gene markers for different traits. Markers are being used in a number of breeding applications and for addressing issues of gene resource management. Considerable research is being carried out in collaboration with industry partners on characterizing a number of quantitative trait loci (QTL), determining their numbers and locations on the genomes and the size of their effect. Current research is concentrating on validating QTL and testing their usefulness for the marker-aided selection of plants within a breeding program. An expanding research area is determining genes controlling different traits using genomics approach. Genes are now, being studied for purposes of developing early flowering and sterility in many plants. Gene transformation systems are being developed in many plants. Rice, Maize, *Arabidopsis* etc. are being developed as a model species for molecular breeding.

Key Words: Molecular Markers, Gene tagging, QTLs, Marker aided selection

[1]*Field Research Laboratory, DRDO, C/o 56 APO, Leh, India*
[2]*Indian Institute of Sugarcane Research, Lucknow, UP, India*
[3]*JUIT, Waknaghat, Solan, HP, India*
**Corresponding author e-mail:gyan.gene@gmail.com*

INTRODUCTION

Plant breeding has been in rapid transition as more molecular tools are applied to commonly accepted field techniques. Predictably, the scientific development of this field is headed toward sequence-based knowledge which should improve both reliability and adoption in agriculture. This chapter describes the application of molecular technologies to plant breeding with more emphasis on marker-assisted breeding. One weakness in the application of molecular breeding is the general lack of better phenotypic measurement tools for components of yield in agronomic crops. This weakness occurs, in large part, because detailed, precise phenotypic measurements are not available in agronomic crop species for many traits (e.g., drought tolerance, standability), and it represents a severe limitation to the application of molecular breeding at this time. With the development of molecular markers, genetic maps have been constructed in plant species which allow for the localization of major loci and QTLs controlling agronomical trait variation. Molecular markers have been widely used for the introgression of major loci, but marker-assisted selection for quantitative trait breeding is used less.

Molecular markers systems: It reveals variation in genomic DNA sequence. The first generation of molecular marker, RFLP, was based on DNA-DNA hybridization and was slow and expensive. The invention of the polymerase chain reaction (PCR) to amplify short segments of DNA gave rise to a second generation of faster and less expensive PCR-based markers (Table 1). Among the techniques that are particularly promising are Amplified Fragment Length Polymorphism (AFLPs), Random Amplified Polymorphic DNA (RAPDs), Microsatellites, Sequence Characterized Amplified Regions (SCARs) or Sequence Tagged Sites (STS) etc.

Table 1. Type of different markers and their behavior.

Marker	PCR Based	Polymorphism	Dominance
RFLP	No	Low-medium	Co-dominant
RAPD	Yes	Medium-high	Dominance
SSR	Yes	High	Co-dominant
AFLP	Yes	High	Dominance / Co-dominance
SNP	Yes	Extremely high	Co-dominance / Dominance

Molecular markers are being used extensively to investigate the genetic basis of agronomic traits and to facilitate the transfer and accumulation of desirable traits between breeding lines. A number of techniques have been particularly useful for genetic analysis. These markers often used in conjunction with bulked segregant analysis and detailed genetic maps,

provide a very efficient method of characterizing and locating natural and induced mutated alleles at genes controlling interesting agricultural traits. Markers have also been used to identify the genes underlying quantitative variation for height, maturity, disease resistance and yield in virtually all major crops. In particular, the PCR-based techniques have been useful in the assessment of biodiversity, the study of plant and pathogen populations and their interactions; and identification of plant varieties and cultivars. Amplified DNA techniques have produced sequence-tagged sites that serve as landmarks for genetic and physical mapping. It is envisioned that emerging oligonucleotide-based technologies derived from the use of hybridization arrays, the so-called DNA chips and oligonucleotide arrays, will become important in future genomic studies. However, many of these are still under development, are proprietary, or require the use of expensive equipment, and are therefore neither suitable nor cost-effective for adequate transfer to developing countries. However, techniques are continuously changing and evolving, so technology transfer needs to keep pace with current developments in genomics.

In the last decade, the use of DNA markers for the study of and improvement of traits has become routine, and has revolutionized the crop breeding strategies. Increasingly, techniques are being developed to more precisely, quickly and cheaply improve the genetic structure of different crops. These techniques have changed the standard equipment of many labs, and most plant breeders are expected to be trained in DNA data generation and interpretation. As we enter the post-genomics era, the need for genetic markers does not diminish, even in the species with fully sequenced genomes.

Overview of Molecular Technologies

Molecular markers should not be considered as normal genes, as these do not have any biological effect, and instead are landmarks in the genome. These rely on DNA assay, in contrast to morphological markers, based on visible traits, and biochemical markers based on proteins produced by genes. Molecular markers are generally employed to detect and understand the naturally occurring polymorphism in DNA, which forms the basis for designing the strategies to exploit it for applied genetic purposes. Due to the rapid developments in the field of molecular genetics, a variety of different techniques have emerged to analyze genetic variation during the last few decades (Whitkus *et al.* 1994; Karp *et al.* 1996, 1997a,b; Parker *et al.* 1998; Schlötterer 2004). These genetic markers may differ with respect to important features, such as genomic abundance, level of polymorphism detected, locus specificity, reproducibility, technical requirements and financial investment. The main marker technologies that have been widely applied during the last decades are summarized in Table 2, and briefly

Table 2. Main characteristics of major types of molecular markers.

Characteristic	RFLP	SSR	AFLP	RAPD	SNP
Type of visualization	Single locus	Single locus	Multi-loci	Multi-loci	Single locus
Allelism	Co-dominant	Co-dominant	Dominant	Dominant	Co-dominant
Type of polymorphism	Sequence	No. of repeats	Sequence	Sequence	Sequence
Level of polymorphism	Good	Excellent	Excellent	Good	Excellent
Polymorphism at the locus	2 to 5 alleles	Multiple alleles	Presence/ absence	Presence/ absence	2 alleles
Quantity of DNA needed	Large	Small	Small	Small	Small
Quality of DNA needed	Good	No restrictions	Good	Good	Good
Reproducibility	Good	Good	Good	Low	Good
Time	Long	Fast, once markers are developed	Fast	Fast	Fast, once markers are developed
Cost	Expensive	Average	Average	Average	Expensive[a]
Technical difficulty	High	Low	Medium	Medium	High[a]

outlined below, together with their strengths and weaknesses. The information provided by the markers for the breeder will vary depending on the type of marker system used. Each one has its advantages and disadvantages and in near future other system will also likely to be developed. Various types of molecular markers are utilized to evaluate DNA polymorphism. Table 3 gives the details of four major classes of DNA polymorphism. No marker is superior to all others for a wide range of applications. The most appropriate genetic marker will depend on the specific application, the presumed level of polymorphism, the presence of sufficient technical facilities and know-how, time constraints and financial limitations. The markers can be used for both qualitative and quantitative traits for the improvement of trait of interest (Fig. 1).

General Applications of Different Marker Systems: Molecular markers have been looked upon as tools for a large number of applications ranging from localization of a gene to improvement of plant varieties by marker-assisted selection. These have also become extremely popular markers for phylogenetic analysis adding new dimensions to the evolutionary theories. Genome analysis based on molecular markers has generated a vast amount of information and a number of databases are being generated to preserve and popularize it. In general it is used for different and diverse purposes

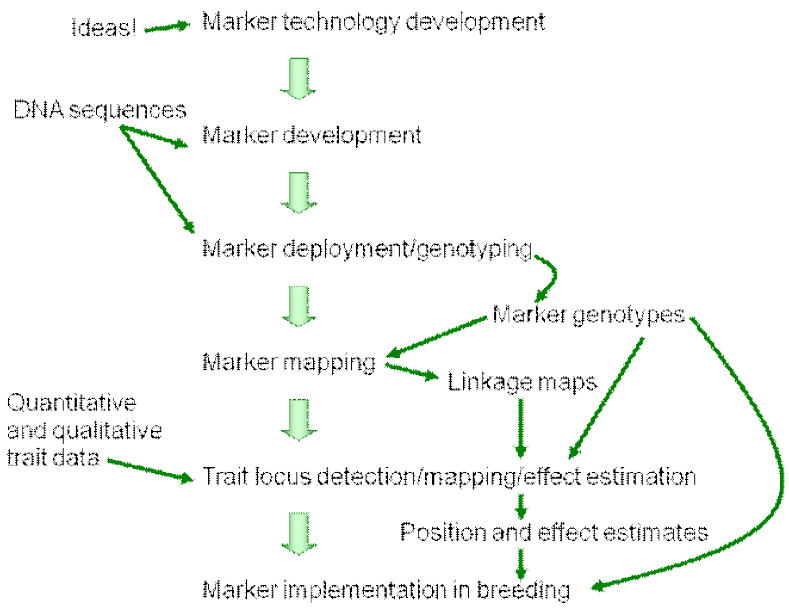

Fig 1. Schematic presentation of markers in plant breeding

Table 3. Four classes of DNA polymorphisms

Class	**Size of locus**	**Number of Alleles**	**Number of loci In population**	**Rate of Muta-tions**	**Use**	**Method of detection**
SNP	Single base pair	2	100 million	10^{-9}	Linkage mapping	PCR followed by ASO hybridization of primer extention
Microsatellite	30-300 bp	2-10	200,000	10^{-3}	Linkage mapping	PCR and gel electrophoresis
Multilocus minisatellite	1-20 kb	2-10	30,000	10^{-3}	DNA finger-printing	Southern blot and hybridization
Small changes in DNA content (Deletions and Duplications	1-100 bp	2	N/A	$<10^{-9}$	Linkage mapping	PCR and gel electrophoresis

like formation of comparative maps, framework maps, genetic maps, breeding, varietal/line identification (multiplexing of probes necessary), marker-assisted selection, F_1 identification, diversity studies, novel allele detections, gene tagging, bulk segregant analysis, map-based gene cloning, seed testing, fingerprinting, framework/region specific mapping, novel allele detections, high-resolution mapping, very fast mapping, region-specific marker saturation, alien gene introduction etc.

These techniques help in direct selection of many desired characters simultaneously using F_2 and back-cross populations, near isogenic lines, doubled haploids and recombinant inbred lines. Molecular markers have already played a major role in the genetic characterization and improvement of many crop species. They have also contributed to and greatly expanded our abilities to assess biodiversity, reconstruct accurate phylogenetic relationships, and understand the structure, evolution and interaction of plant populations. Molecular markers are required in a broad spectrum of gene screening approaches, ranging from gene-mapping within traditional 'forward'-genetics approaches through trait identification studies to genotyping and haplotyping studies. The availability of genetic markers is fundamental within plant biology and plant breeding. There are a wide range of uses and applications for molecular markers, but most are associated with the map-based cloning of individual genes, the characterization of quantitative multi-gene traits etc. With the subsequent association of genes and their related markers they additionally become valuable within the context of both genotyping and haplotyping.

MARKER ASSISTED SELECTION: Marker assisted selection is perhaps the first direct application of molecular markers for the improvement of any crop. High density linkage map and markers tightly linked to the gene/loci affecting the trait of agronomic importance are the two major pre-requisites for this to happen. Molecular marker assisted selection, often simply referred to as marker assisted selection (MAS) involves selection of plant carrying genomic regions that are involved in the expression of traits of interest through molecular markers.

Advantages: DNA markers are not affected by the environmental factors, heritability of traits, number of genes involved and gene interaction. Hence these are more reliable. The markers can be scored at seedling stage. This is especially advantageous when selecting for traits which are expressed only at later stage of development. By selecting at seedling stage considerable time labour and space can be saved.

Applications: Linkage is sought between DNA marker and agronomical important traits such as resistance to pathogens, insects, nematodes, tolerance to abiotic stress and quality parameters. Instead of selecting for a trait, the breeder can select for a marker that can be detected very easily

and early in selection scheme. The use of DNA markers for the indirect selection offer the greatest benefits for quantitative traits with low heritability as these are the most difficult characters to asses in the field experiments. Obviously the development of marker assisted assay for such traits is extra difficult and most costly due to the expensive phenotypic assays. With the development and availability of an array of molecular markers and dense molecular maps in crop plants, MAS has become possible for traits governed by major genes as well as quantitative trait loci (QTLs). Thus marker assisted selection has become a promising and potent approach for integrating biotechnology with conventional and traditional breeding.

MOLECULAR MARKERS FOR DIVERSITY STUDIES

Molecular markers can be used to:

- Estimate genetic diversity between varieties or between populations.
- Compare populations with reference points (e.g. groups of lines developed by classical breeding programmes, sets of local commercial varieties, or reference core collections).

This information can be used in programmes for population improvement through recurrent selection to maximize diversity within populations by crossing lines that are genetically the most distant; to manage pedigrees; or adjust the long-term strategies of improvement programmes.

Number of Markers and Its Density for Diversity Study: No definitive rule exists for choosing the number of markers to use in a diversity study, but for rice, between 12 and 24 are normally used. This is a safe range according to Barry (2000), who studied the diversity of rice varieties in Guinea and showed that similar structural patterns were obtained from as many as 12 to as few as six microsatellite markers. One marker per chromosome or chromosome arm is a reasonable marker density for ensuring the best possible coverage of the genome. Such density will also prevent bias due to the presence of genes involved in gametic or zygotic selection, or controlling traits under selection in a specific chromosome. Among the possible markers for each arm, those that reveal the largest number of alleles in the studied background should be selected as priority. Those with a very high mutation rate should be avoided as they may blur results.

POSSIBLE USE IN POPULATION IMPROVEMENT PROGRAMMES

Choosing parents and developing the base population: The choice of parents to constitute a gene pool depends on the objectives and strategy of each breeding programme. Both the number and nature of the parents are important. A large number of parents is not, in itself, an indicator of a broad

genetic base because some may share part or most of their genomes. The genetic relatedness between parents is, as a result, an important, although complex, aspect to take into account. So far, genetic relatedness has been evaluated on the basis of line pedigree (Graterol, 2000). Rice is strongly structured, and varieties belonging to the same varietal group are genetically closer than those varieties belonging to different groups (Glaszmann, 1987). Equal contribution from both parents is seldom true in a context of selection where plant breeders favour an ideotype.

Molecular markers are more informative. By determining the alleles present at different markers for the set of varieties to compare, we can establish a matrix of similarities that offers a truer picture of the relationships between varieties. The classic indices of similarity for comparisons between varieties are those of Dice and Jaccard for co-dominant markers, and Sokal and Michener for dominant markers (Perrier *et al.*, 2003). After transformation of the similarity matrix to obtain distances, we can produce a geometric representation of the distance between varieties through various clustering techniques such as the unweighted pair-group method, using the arithmetic average (UPGMA), or the neighbour-joining tree method (NJ tree). Results are usually represented as dendrograms. Lorenzen *et al.* (1995) provide recent examples of evaluating relatedness based on molecular markers between genotypes used in classic breeding programmes for soybean; Soleimani *et al.* (2002) for durum wheat; and Lima *et al.* (2002) for sugar cane.

Optimizing the Number of Recombination Cycles: An important question that arises when constituting a gene pool or, later in the recurrent selection process, when crossing selected progenies, is how many intercrossing cycles are needed to better recombine the parents' alleles. The objective in this case is panmixia. On a theoretical basis, Hanson (1959) concluded that three crossing cycles were needed. Marín-Garavito (1994), working on a recurrent selection population, did not find differences among the phenotypes of the first and third cycles of recombination. Such a question can also be answered by using co-dominant molecular markers. Badan *et al.* (1998) and Geraldi and Souza (2000), using means and variances of several rice populations for several traits, demonstrated that a population of 200 plants was adequate.

For each marker, the allelic and genotypic frequencies in the populations and Nei's diversity index (Nei, 1972) can be computed. These data can be used to analyse deviations from panmixia, using the Wright fixation index. To further check the efficiency of recombination in breaking unfavourable linkages, one strategy would be to select additional markers close to the markers under study. The segregation of this specific group of markers can then be compared in successive recombination cycles. The evolution of linkage disequilibrium between markers of increasing physical distances

can be evaluated over several cycles. The recombinations, if they occur normally, should progressively break linkages, and linkage disequilibrium should decrease from cycle to cycle.

***Effect of Population Improvement on Genetic Diversity*:** Several projects of recurrent selection, based on evaluations of $S_{0:2}$ families, are being carried out to improve various traits of agronomic importance. Examples include resistance to blast (Guimarães, 2000), tolerance of acid soils (Ospina *et al.,* 2000) and yield. Progress made for the trait or traits under selection is usually evaluated after a few cycles of recurrent selection. This is done by comparing phenotypes of the initial population with those of the progeny from the last cycle. Deliberate selection reduces genetic diversity to a level that depends on the initial variability and intensity of selection. Such reduction of diversity may not be harmful if the alleles eliminated are those that are unfavorable.

Molecular markers can be used to evaluate the degree of reduction of total diversity by comparing the diversity between successive recurrent cycles of the population. The markers can be chosen according to the allelic diversity in the subspecies used to construct the population (*indica* for tropical countries and *japonica* for temperate ones).

Locating the Male-Sterility Gene and Evaluating Linkage Drag: The male-sterility gene from 'IR36' has been used in most populations to facilitate crossing. Molecular markers can be used to locate the gene in the rice genome. Identifying a codominant marker closely linked to the male-sterility gene would allow distinguishing heterozygotes from homozygotes for the male-sterility gene among male-fertile plants extracted from populations. Such selection would accelerate line fixation during line extraction. The easiest method for finding such markers is through bulked segregant analysis (Michelmore *et al.,* 1991).

In this method, two separate DNA bulks are constituted: one from the pooled DNA of 10 male-sterile plants and the other from 10 male-fertile plants. The two pools are tested for allelic differences with a set of markers. Any method that enables rapid testing of a large number of markers is advisable.

Marker-Assisted Selection for Recurrent Selection Populations: One major use of molecular markers is to locate genes that control agronomic traits of interest on dense genetic maps and identify those markers closely linked with these genes. To facilitate mapping, specific bi-allelic populations, resulting from crosses between pure lines that correspond to the simpler genetic situation, are developed. De Koeyer *et al.* (2001) used RFLP markers to identify QTLs linked with yield, plant height, and flowering in a recurrent selection population of oat. The locations of the QTLs were consistent with

those detected in a classic population of recombinant inbred lines.

Once markers tightly linked with traits of interest are identified, then the interest is to manipulate them. Marker-assisted selection is based on linkage disequilibrium, which is artificially increased through hybridization. Intercrossing induces recombinations and reduces linkage disequilibrium in a recurrent selection population. Thus, the positions of the QTLs need to be regularly re-evaluated, not for every cycle but at least every 2 years, as suggested by Hospital *et al.* (1997).

The problem does not occur if the polymorphic marker is directly within the gene or genes of interest. But this implies extensive work to identify such genes through positional cloning or the candidate gene approach.

GENEBANK MANAGEMENT: The realization that the world was rapidly losing much of its agro-biodiversity led to a global effort to collect and conserve germplasm. An increasing awareness of the narrow genetic base of crops in advanced agriculture and potential susceptibility to crop failures (National Research Council 1972) further stimulated the efforts to collect, and a system of national and international Genebanks. A major question facing genebank managers concerns which material needs to be included in a collection to conserve a representative sample of the total genetic diversity range of a crop. Schoen and Brown (1993) document how molecular marker techniques provide the optimal strategy to sample materials from populations of wild relatives to maximize the number of alleles. Hamilton (1994) argues that simple marker diversity is an insufficient measure of maintenance of diversity, and suggests that more detailed studies of genetic correlations of quantitative genetic variation and gene and environment interaction are needed.

Relationship of Molecular Variation To Geography: The literature sometimes shows association of DNA-based relationships to geography and sometimes fails to show such associations. Whitkus *et al.* (1998) examined these alternative hypotheses using RAPD markers from a wide variety of geographically diverse wild and cultivated populations. Their results supported a single origin of modern cultivars in South America. However, they also discovered relationships of distinct populations in ancient Mayan groves in southern Mexico to wild Mesoamerican populations. Other studies showing a relationship of molecular variation to geography or ecology are Espejo-Ibañez *et al.* (1994), Garvin and Weeden (1994), Yang *et al.* (1996), Lanner *et al.* (1997), Nevo *et al.* (1997, 1998) Fahima *et al.* (1998), and Zerega (2004). However, there is often not a congruence of genetic distance and geographic distance as shown in Maass and Ocampo (1995), Lanner *et al.* (1996), Ursla *et al.* (1997), Varghese *et al.* (1997), Comes and Abbott (1998), Freville *et al.* (2001) and del Rio and

Bamberg (2002).

Marker Data for Germplasm Acquisition: Marker studies addressing the distribution of genetic variation within and between populations have been used to guide the acquisition of new material for germplasm conservation. Steiner *et al.* (1998) used RAPDs to measure genetic diversity of germplasm collections of *Trifolium incarnatum* L. In combination with data on pedigree, they documented low diversity in the existing cultivars and identified populations that need additional collecting. Other studies addressing the application of marker data for germplasm acquisition include Brown and Munday (1982), Murphy and Philips (1993), Lamboy *et al.* (1994), Tsegaye *et al.* (1996), Nebauer *et al.* (1999), Maquet *et al.* (1996, 1997) and Zoro Bi *et al.* (1998). In summary, molecular data are useful to guide genebank curators in decisions concerning acquisition in cases where molecular variation is associated with geographic variation.

CHARACTERIZATION OF GERMPLASM

Systematics: Taxonomic questions are addressed by almost every molecular marker class from microsatellites useful at the species level to DNA sequences useful at the generic and family levels. For example, Roa *et al.* (1997) used phenetic analyses of AFLPs to investigate the origin of cassava (*Manihot esculenta* Crantz) relative to four other wild species and two non-cultivated subspecies of *M. esculenta*. These subspecies were thought to be progenitors of the cultivars or escapes from cultivation. Molecular markers provide tools to test the taxonomic validity of species and can provide species-specific diagnostic markers. For example, Martin *et al.* (1997) used RAPDs to re-identify a collection of oat accessions. Lee *et al.* (1996) used RAPDs to discriminate *Brassica* L. varieties. Species-specific markers also have been found with minisatellites (Baurens *et al.* 1997, banana), and Zhou *et al.* (1997) used minisatellite sequences to discover genome-specific fragments in oat, but not cultivar-specific fragments. McGregor *et al.* (2002) used AFLPs to characterize 314 accessions of the wild potato subgroup *Solanum* L. series *Acaulia* Juz.

Fingerprinting studies: "Fingerprinting" is an attempt to discover cultivar-specific molecular markers that aid their identification. Certain cultivars are economically very important and form the backbone of regionally important industries. Potato is a clonal crop with many morphologically similar cultivars. Schneider and Douches (1997) were able to distinguish 24 of 40 potato cultivars with 7 SSR primer pairs. When the SSR data were combined with the tuber morphology, only five pairs of cultivars could not be distinguished. These many studies clearly document the use of SSRs for fingerprinting genebank accessions.

Putative natural hybrids: Hybridization is thought to be a major evolutionary force at both the diploid and polyploid levels (Rieseberg 1995). Rieseberg and Ellstrand (1993) showed that hybrids are no more likely to display intermediate character states than parental ones, and can additionally express an array of transgressive or novel traits. The utility of molecular markers to investigate hybrids decreases with time of divergence, as species-specific markers from both parents can be disrupted through recombination, and both can mutate to new markers. Molecular markers present powerful new tools to reinvestigate hypotheses of hybridization. One method is to search for markers in the hybrid that are found in both putative parents.

CROP BREEDING

Parental Contributions of Artificial Hybrids: DNA markers are very useful for confirming hybridity of artificial sexual hybrids or somatic fusion hybrids. Molecular markers are especially useful when hybridity is questioned by morphological reasons or for early screening of large putative hybrid populations. Thieme *et al.* (1997) used isozyme and RAPD data to screen hundreds of first generation somatic fusion hybrids and later sexual backcross generations between cultivated and wild species of potato, to distinguish hybrid from non-hybrid progeny. Lee *et al.* (1998) used RAPDs to disprove hybridity in putative *Saccharum* L. and *Erianthus* Michx. hybrids. Yamada *et al.* (1997) used cpDNA and mtDNA to investigate symmetric somatic fusion hybrids of cultivated potato and one of its wild relatives. The results suggested that chloroplast genomes of the respective parents segregated randomly while the mitochondrial genomes favoured the cultivated species parent.

Molecular Diversity and Heterosis: Heterosis, or hybrid vigour, originally referred to the selective superiority of heterozygotes regarding continuously variable characters of size, yield and vigour. Information on the genetic diversity in a crop species is important for selection of parental strains and in the prediction of hybrid performance especially in crops in which hybrids are commercially important. The various steps involved in hybrid breeding programs such as making several crosses and screening the combinations for superior performance and heterosis are very costly, laborious, and time consuming. Hence if heterosis can be predicted before making the crosses, then the number of crosses to be performed and the progeny to be screened can be considerably reduced. Smith *et al.* (1990) showed a strong correlation of genetic distance and heterosis in maize. Melchinger *et al.* (1994) reported that genetic distances of inbred lines of maize, as assayed via RFLP data, were not correlated with heterosis for yield. Paz and Veilleux (1997) reported that in crosses between a diploid cultivated potato species with diploid clones of other complex interspecific hybrids of potato, the greatest total tuber yield was associated with diverse

parents as measured by RAPDs. Manjarrez-Sandoval *et al.* (1997) compared RFLP similarity estimates of the parents, to yield in soyabean, and also found a positive correlation. The genetic distance estimated through molecular markers distributed throughout the genome could provide a mean to predicting hybrid performance prior to making and evaluating a actual cross. Various investigators are trying to correlate genetic diversity, as quantified by DNA markers, to predict hybrid performance, in various hybrid breeding programs because the level of genetic diversity between the parents has been proposed as a possible predictor of heterosis.

GENE SEARCH BASED ON MOLECULAR MARKERS

Complex traits such as yield are controlled by polygenes that are prone to influence the environment. Since each gene makes only a small contribution to the total variation of the trait, identifying individual polygenes has proved difficult in the past. Consequently, breeding for polygenic trait has been difficult. The past two decades have witnessed explosive developments in DNA based molecular markers technology. These markers are DNA sequences of specific size and/ or specific polynucleotides, and correspond to genomic sequences. They have fixed locations on the genome and follow patterns of Mendelian inheritance. Thus these markers can be regarded as specific phenotypes. Unlike the traditional traits, these markers are not influenced by the environment and hence can be easily identified at all stages. The major advantage of these markers is that they are not limited in number and can be used to mark every locus along the entire genome.

Tagging Traits With Molecular Markers: By studying the co-segregation of a trait and a molecular marker(s), one can identify the molecular marker(s) linked with the trait of interest. Thus the gene(s) controlling a trait can be tagged with specific molecular marker(s). Molecular tagging has come in handy to identify novel genes as well as to pyramid desirable genes. For example breeding for disease resistance, identification and tracking of different resistance genes for a given disease is difficult because presence of one resistance gene masks the effect of others. In earlier years special cytogenic stocks (e.g. monosomic lines in wheat) were used to identify and pyramid multiple disease resistance genes. However, such stocks are difficult to build and maintain, and were available only in few species. Using molecular markers different resistance genes can be tagged and followed during breeding. The recent plant breeding literature is replete with the reports of molecular markers linked to various traits such as disease resistance, male fertility restoration (of CMS lines), pest tolerance, quality traits etc. Thus molecular markers are aiding new discoveries and use of new genes. Another major advantage is that desirable genes can be assembled without evaluating for the trait. This is particularly advantageous where

phenotyping of the trait (e.g. disease or pest resistance, drought, salinity, yield) is difficult (Chopra, 2006).

Gene Search Based On Molecular Linkage Map: Detailed linkage maps have been constructed using molecular markers in different crops in recent years. Comparative analysis of the maps based on common markers has revealed highly conserved nature of marker positions and order among different species. This is termed synteny. For example, all members of the *Graminae* family share significant synteny, which is depicted in what is now termed Crop Circles. This synteny has emerged as an important tool for gene discovery. Finished rice genome sequence now serves as reference for discovery of corresponding new genes in other related species. For example Chen *et al.* (2003) found four QTLs for rice blast, which were also found in barley. Two of the barley genes had conserved isolate specificity while the other two had partial specificity. Similarly QTL for shattering resistance and plant height were found conserved among grasses. Synteny is not only seen at the macro level but also at the micro level i.e. in short stretches of a few hundred kb. Such microsynteny is particularly exploited for cloning genes through chromosome walking. For example verbalization genes *VRN1* and 2 of wheat were cloned using this strategy. The map based cloning is highly relevant in crops with large sized genomes such as barley, wheat and corn. For example stem rust resistance gene of barley *Rpg* 1 was cloned based on microsynteny of the region with rice although rice dose not have this gene (Kilian *et al.* 1997). The *Lr*21 gene of bread wheat was cloned by shuttle mapping between diploid and hexaploid wheat (Huang *et al.*, 2003). Recently, Gualtieri *et al.* (2006) reported microsynteny between rice chromosome 11 and apomictic species region of *Cenchrus ciliaris* and *Pennisetum squalatum*. Similarly, Calderini *et al.* (2006) found microsynteny between *Paspalum simplex* apomixis locus and rice chromosome 12. These findingings will speed up identification of gene(s) controlling apomixis. However, microsynteny is not strictly conserved and may show variation apparently at random. Therefore, a case-by-case examination is essential.

Gene Search Based on Sequence Information: A very large number of genes involved in plant growth, morphogenesis and development have been cloned. Similarly, genes mediating various metabolic pathways have been characterized using different approaches and in different organisms. The database of sequences of genes, cDNA etc. is growing rapidly and is available in public domain. Information about these sequences has become an important tool for gene discovery.

Gene orthologs, analogs and allele mining: Analysis of a number of genes sequences coding for proteins has revealed high degree of nucleotide/ amino acid homology among genes of related functions. In particular, conserved sequence domains are frequent among different classes of genes.

This is expected in the light of evolution of different life forms from common ancestors. Hence, using sequence information from one species, orthologous genes from related species have been cloned through PCR and other techniques.

Plant disease resistance genes fall under 5 categories, namely the NBS-LRR, protein kinase, eLRR-TM, LRR and toxin reductase. The largest group of plant defense genes belong to NBS-LRR class. These proteins have nucleotide binding site (NBS) and C-terminal leucin rich repeats (LRR). Using sequence information of these conserved domains, similar genes (analogs) can be easily isolated by PCR cloning and tested for their novelty. Such genes are termed as resistant gene analogs (RGAs). A number of such RGAs have been cloned in different crop species.

Similarly, using sequence information of specific genes, various allelic forms can be amplified and characterized. Since most of the inter-varietal differences originate due to allelic differences, allele mining has become a rapid method of search for new genes.

Component Search: So far we have considered gene search with reference to traits they confer on the individuals that carry them. However using knowledge of metabolic pathways, r-DNA and plant transformation technologies, and engineering novel traits has become feasible by putting together components from different, and even unrelated, sources. Therefore modern gene search is not only about full length genes (including regulatory elements) but also about components. The significance of this approach is best illustrated in Golden Rice story.

Rice seeds do not accumulate Vitamin A (or more precisely its precursor, beta carotene) even though a part of the synthetic pathway of Vit. A is present in rice in order to enrich rice grains with Vit. A, additional genes necessary for beta carotene gene were sourced from an unrelated plant (*psy* gene encoding for phytoene synthase from daffodil) and a bacterium (*Crt1* gene encoding carotene desaturase from *Erwinia uredovora*). Also a gene lycopene beta cyclase from *Narcissus pseudonarcissus* was introduced to convert lycopene to beta carotene. To ensure that beta carotene accumulation in the endosperm, these genes were linked to seed specific glutelin promoter. When this gene cassette was introduced into rice, the resulting seeds accumulated beta carotene in the seeds. Thus was engineered a new trait in rice (Ye *et al.*, 2000). Recently, further improvements have been made to elevate the expression levels of beta carotene from 1.6 µg/g to 30 µg/g of seed (Al-Babili and Beyer). This later development was made possible by replacing *psy* gene of daffodil with its counterpart from maize. This again illustrates the point made earlier of knowledge based gene search in modern plant breeding. Similarly, using bacterial RNAse gene bar, driven by an anther (tapetum) specific promoter, male sterility was engineered in brassica

for hybrid seed production (Mariani *et al.* 1990). There are several similar cases where metabolic engineering has provided novel compounds of pharmaceutical and industrial value.

This component search is highly knowledge intensive and requires convergence of interdisciplinary expertise and utilizes a range of modern tools and techniques related to gene expression. It is a telling example not only of the unification of entire biological system but also of how knowledge and interdisciplinary science can mould into societal gods of high value.

High Throughput Gene Discovery Base on Gene Expression Analysis: In recent years, complete genome sequences have become available for several organisms. Using bioinformatic tools and other data, probable coding regions have been specified. For example, in *Arabidopsis* a total of 25,498 genes have been predicted. Similarly, rice genome is estimated to code for a total of 37,544 genes. These coding region sequences can be arranged individually on a slide in arrays. Such microarray slides or gene chips are now commercially available. By hybridizing cDNA derived from contrasting treatments one can detect specific changes in gene expression. For example, in a study with rice, Rabbani *et al.* (2003) examined changes in gene expression in rice exposed to four different abiotic stresses, namely, drought, salinity, cold and abscisic acid treatments. In total 73 genes were found to be differentially expressed in response to these stresses. Of these, 58 cases were novel, not reported previously in rice. 15 genes were common to all the four stresses. Further, 51 of these genes were also known to be induced by similar stresses in Arabidopsis. This study demonstrates the power of new technology for gene identification. Also, the common gene expression responses of Arabidopsis and rice to abiotic stresses underscores the exportability of information gained from model species to the other species (Chopra, 2006).

Marker-Assisted Backcross and QTL Mapping

Marker-Assisted Backcross (MABC): The efficiency and complexity of MAS depend on the genetic nature of the trait (monogenic or polygenic). For monogenic traits, marker-assisted backcross (MABC) is the most straightforward strategy, whereas for polygenic traits various strategies are available. The principle of MABC for a single gene is quite simple. First, molecular markers tightly linked to the target gene must be identified, allowing assessment of the presence of the introgressed gene ("foreground selection"). Other markers are also used in order to accelerate the return to the recipient parent genotype at other loci ("background selection"). Background selection is based not only on markers located on the chromosomes carrying the gene to introgress (carrier chromosome), but also on non- carrier chromosomes. Markers devoted to background selection on

a carrier chromosome allow the identification of individuals for which recombination events took place on one or both sides of the gene, in order to reduce the length of the donor type segment of genome dragged along with the gene (Young and Tanksley, 1989). Background selection on non-carrier chromosomes was investigated by Hospital *et al.* (1992). Visscher *et al.* (1996) investigated both foreground and background selection. In three generations of MABC, isogenicity is higher than that obtained by classical methods. By comparison, traditional approaches would require approximately two more generations to obtain such an isogenicity (Hospital *et al.*, 1992).

QTL Mapping: Molecular markers can be used to locate important genes or QTLs in the genome. Once tagged by tightly linked markers, valuable alleles of the genes can be manipulated indirectly by selecting specific alleles at the flanking markers. Marker-assisted selection can help improve the efficiency of selection. Markers with known positions spaced at regular intervals throughout the genome are needed. For their technical simplicity, microsatellite markers are (again) currently the ones most chosen. AFLPs should be selected for quick mapping or saturating a map where microsatellites or RFLPs have left gaps. Once markers close to the gene or QTL of interest are identified, PCR-based markers are needed to manipulate them through marker-assisted selection, in which case, microsatellite markers are irreplaceable.

Theoretical Investigations on MAS for QTLs: Traits showing quantitative variation are usually controlled by a number of genes (quantitative trait loci, QTL), each with variable effect. Due to the genetic complexity of such traits, several QTLs with small effects must be simultaneously manipulated. As for monogenic traits, MABC is the most effective strategy when a small number of QTLs (less than five), coming all from the same parent, are transferred. In practice, the position of the QTL is not precisely estimated and the true position of the QTL is unknown, but is supposed to be within a confidence interval. From this confidence interval length, Hospital and Charcosset (1997) deduced the number of markers and their position relative to the estimated position of the QTL, in order to insure an optimal control of the QTL. However, as the minimum number of individuals that should be genotyped at each generation depends on (i) the confidence interval length, (ii) the number of markers and (iii) the number of QTLs, it seems illusive to transfer more than four or five QTLs with this simultaneous design unless a very large population can be considered, or the precision of the QTL location is very high.

Advanced Backcross QTL Analysis: This analysis is another strategy tailored for the discovery and transfer of valuable QTL alleles from unadapted donor lines into established elite inbred lines (**Tanksley** and

Nelson, 1996). The QTL analysis is delayed until an advanced generation (BC3 or BC4), and during the development of this generation, negative selection is pursued to reduce the frequency of deleterious donor alleles. The advantages of using BC3 /BC4 populations include reducing linkage drag and epistatic effects and decreasing the amount of time later needed to develop QTL-NILs (**Fulton** *et al.*, 1997).

Pyramidal Design: When the number of QTLs to introgress becomes important, **Hospital** and Charcosset (1997) proposed to use a **pyramidal design**. QTLs are first monitored one by one by MABC, to benefit from a higher background selection intensity, and then the selected individuals are crossed, to accumulate favorable alleles at the QTLs in the same genotype. According to predictions, the pyramidal design should provide approximately the same efficiency as the simultaneous design with almost one third of the individuals.

Index Method: When favorable alleles come from different sources, van **Berloo** and Stam (1998) proposed an **index method** to select among recombinant inbred lines for crossing, to obtain a single genotype containing as many favorable quantitative trait alleles as possible. The index is constructed for each pair of lines based on the genotype of the markers flanking the putative QTLs. Plants showing the optimal index are crossed together. The use of genetic markers to improve populations was proposed using a statistical approach based on an index combining phenotypic and marker information (Lande and Thompson, 1990).

MAP-BASED CLONING: A traditional strategy for cloning a gene requires prior knowledge about the biochemistry of a gene or its products including the gene's mRNA, a heterologous cloned copy of the gene from another species or differential nucleic acids preparations. Map-based cloning or positional cloning also known as reverse genetics is a strategy which does not require prior knowledge about the gene or its products; instead, this cloning strategy requires knowledge about the chromosomal location of the gene. The first step of map-based position cloning is to identify molecular marker that lies close to gene of interest. The identification of tightly linked markers and their relative order, in the vicinity of target gene, is feasible due to the availability of high density molecular maps and several regional targeting procedures. The next step is to screen a large insert genomic library (BAC or YAC) to isolate clones that hybridize with tightly linked marker probes. The markers most tightly linked to the gene are used to screen a large fragment library (YAC or BAC). The positive clones are supposed to contain the target gene. The step that follows termed as chromosome walking. This procedure involves creating new markers (usually sequences at the end of the clone) and screening segregating population with these new markers. Often this population is large (1000-

3000 individuals).The process is repeated with markers designed from different clones in order to achieve co-segregation of markers with the gene of interest. Co-segregation means that whenever one allele of the gene is expressed, the markers associated with that allele are also present and there is no recombination between gene and the markers. As target gene flanked on a single clone between two markers, the DNA fragments between the flanking markers are cloned. Bioinformatics analysis of the clone may reveal relevant candidates. To decide between the candidates, the most direct method is transformation, whenever possible. The steps for map based cloning can be summarized as follows:

- Identify a marker tightly linked to your gene in a "large" mapping population.
- Find a YAC or BAC clone to which the marker probe hybridizes.
- Create new markers from the large-insert clone and determine if they co-segregate with target gene.
- If necessary, re-screen the large-insert genomic library for other clones and search for co-segregating markers.
- Identify a candidate gene from large-insert clone whose markers co-segregate with the gene.
- Perform genetic complementation (transformation) to rescue the wild-type phenotype.
- Sequence the gene and determine if the function is known.

DEVELOPMENTS IN FUNCTIONAL GENOMICS

Allele mining: Identification and access to allelic variation that affects the plant phenotype is of the utmost importance for the utilization of genetic resources. Allele mining is a research field aimed at identifying allelic variation of relevant traits within genetic resources collections. For identified genes of known function and basic DNA sequence, genetic resources collections may be screened for allelic variation by e.g. the "tiling strategy" using DNA chip technology (e.g. Lemieux *et al.* 1998). In that approach the basic DNA sequence of a gene is spotted on a chip in the form of large series of sequence overlapping probes consisting of 15–20 bases. Each base position in a fluorescently labeled sample is then interrogated for the presence of point mutations by monitoring hybridization signals with the spotted probes. Because the sequence of samples is determined in comparison with the primary composition of a gene, this method is also known as "re-sequencing". With this method new point mutations, in relatively large DNA fragments, can be detected. As an example, the tiling strategy has been used by the International Rice Research Institute (IRRI)

to identify favorable alleles related to tolerance to biotic and abiotic stress factors in rice.

Association genetics: In allele mining studies as described in the previous section, allelic variation is analyzed for identified genes whose function and basic DNA sequence are known and whose map position in the genome will generally have been determined. On the contrary, in association genetic studies no prior information about the genes of interest is available, but associations between genetic markers and the considered traits are simply derived from observational research. Association genetics focuses on the identification of correlations between phenotypic traits and genetic markers with the aim to identify and locate the underlying genes in the genome (association mapping). Association genetics originated in human genetic studies focusing on the application of significant associations between marker data and diseases for mapping and diagnostic purposes (e.g. Peroutka 1997). Recently, association genetics has gained increasing interest from plant geneticists (e.g. Buckler and Thornsberry 2002; Rafalski 2002). The rationale behind association genetics is that, in general, alleles at different loci are expected to be randomly associated into genotypes, or in other words, to occur in linkage equilibrium. The more adjacent two loci are, the lower the probability of chromosomal recombination occurring between the loci during meiosis. However, given sufficient numbers of regeneration cycles, once established linkage disequilibria will eventually be disrupted by recombination even in the case of tightly linked loci. Linkage disequilibria are therefore expected to be readily lost from populations, the rate of decay being determined by the recombination fraction and the number of generations elapsed. However, selection tends to accumulate favourably interacting alleles and hence opposes the decay of linkage disequilibria.

Comparative genomics: Comparative genomics focuses on the integration of genome information derived from different species with the aim to obtain more insight in the genetic organization of traits through the identification of conserved mechanisms (e.g. Laurie and Devos 2002). This research field has emerged from previous findings of comparative mapping, by which conservation of tracts of collinear markers have been investigated not only among members of the same botanical family, but also among different species, and often led to better understanding of genome evolution. Comparative mapping has shown that a large proportion of the markers and genes are indeed located at comparable positions in the genome (synteny). Together with the availability of an increasing number of genome sequences, including those of known genes, the conservation of gene sequences and their functions among species have also been investigated and are used to further develop the knowledge obtained from previous genetic linkage maps for different species.

Conserved Orthologous Set (COS) markers are conserved markers that may serve as anchors for map development (Fulton *et al.* 2002).

Degenerate Oligonucleotide Primed-PCR (DOP-PCR) uses partially degenerated primers for polymorphism detection (Telenius *et al.* 1992). Because no prior sequence information is required, DOPPCR is considered useful when crops are involved for which no, or limited sequence information is available. Comparison of the genetic maps of different species will reveal information on chromosome evolution and the common genetic control of traits in different organisms. Comparative maps are being developed for various important crops and are expected to facilitate the understanding of the genetic organization of traits in less studied organisms. This may also reveal novel alleles for relevant genes that subsequently may be exploited in crop improvement.

Functional Genomics: The Force Behind the Future of Plant Breeding

Simply defined, functional genomics is a scientific approach that seeks to identify and define the function of genes, and uncover when and how genes work together to produce traits. Current structural genomic approaches (i.e., mapping) generally focus on traits controlled by one or only a few genes, and often they only provide information regarding the location of a gene or genes. Although obtaining location information is a critical first step, functional genomics goes further to examine the interrelationships and interactions between thousands of genes to determine when and why certain traits are expressed, which sets of genes are specifically responsible for that expression, and under what conditions. This information equips scientists to create varieties with exact combinations of traits. Whether the ultimate goal is improved drought tolerance, enhanced nutritional composition, or higher yield potential, functional genomics will play an increasingly importantly role in helping scientists achieve their aims.

FOUNDATIONS FOR MOLECULAR BREEDING

Integration of Genomics and Genetics: It is expected that the foundation for studying natural genetic variation at the DNA sequence level for traits in elite breeding germplasm pools and the design of successful molecular breeding strategies will involve integration of structural and functional genomics technologies within the framework of classical trait mapping methods. The availability of physical maps for a number of the important crop plants and the complete genome sequence for the model plant *Arabidopsis thaliana* (L.) Heynh. (The Arabidopsis Genome Initiative, 2000) and the crop plant rice (*Oryza sativa*; Goff *et al.*, 2002; Yu *et al.*, 2002) has enabled alignment of genetic maps with physical sequence and thus opportunity to target

sequence data from genomic regions for gene discovery and gene function analysis. Further advances in technologies for the study of gene expression and protein interactions have opened up glimpses of some of the gene networks that are involved in the relationships between inter individual DNA sequence variation and trait phenotypic variation in elite breeding populations.

Some Current Considerations: Much of the outcomes from plant genomics efforts to date have involved large-scale data generation of a narrow sample of genotypic variation and its systematic organization, combined with descriptive efforts to annotate the features observed in the organized data. Comparative approaches have been used to initiate candidate gene searches from model to target species. Moving from this broad comparative genomics view of gene discovery to a breeding strategy view that is focused on understanding the detailed organization of extant allelic variation for multiple traits within a selected elite germplasm pool, presents significant challenges.

Power of Selection: An important consideration in the design of any selection process, be it molecular or phenotypic, is that *you get what you select for*. Direct selection on the phenotype of the end-product traits of commercial significance is a relatively robust, albeit in some cases slow, approach when genetic variation exists for the target traits. Replacing or augmenting this system with a knowledge-based approach that targets selection at the level of DNA sequence variation will also rapidly bring about genetic changes. The rigor of the associations we develop between population level sequence variation and phenotypic variation will determine the robustness of this molecular breeding approach. We will be continually forced to refine our knowledge of trait genetics and gene-to-phenotype associations.

OUTCOME OF GENOMICS RESEARCH

Without appropriate investment into crop genomics research, we will always lack detailed knowledge in two areas critical for successful breeding: (i) the structural organization and functional properties of genetic variation for traits and (ii) the influences that plant breeding strategies have on the genetic variation that is widely used in agriculture. Without detailed knowledge in both of these domains, it is difficult to answer many of the important questions asked of breeding programs: (i) how sustainable is a breeding effort in the long term, (ii) how robust are the products of a given breeding program, and (iii) how important is it to and what are the appropriate procedures to maintain and utilize genetic resources?

Genomics gives us some new perspectives on genetic variation. With access to data and information at the sequence level, our views of what

contributes to the natural genetic variation that resides within the germplasm pools developed by breeding programs are changing. The traditional view of genetic variation as a function of loci with fixed effects acting in a predominantly additive manner is challenged by many of the properties of genes that are observed using genomics technologies. An important component of experimental evidence indicates that gene regulation is an important source of genetic variation. What appears to be linear when examined at the phenotypic level is not necessarily linear at the level of the gene network (Peccoud *et al.*, 2004). This creates a complex situation where many of the effects of genes can be highly context dependent. Therefore, the genetic background and environments within which the genes are studied will influence the estimates of the effects of the genes. Plant breeders have always been exposed to this phenomenon but have never had the tools to investigate its genetic basis. However, strategies based on manipulation of the genotype at the molecular level will only be able to utilize the currently available experimental information from statistically determined genotype-phenotype associations. Reminiscent of much of the debate around the effects of epistasis on response to selection, this forces the plant breeding community to ask specific questions of the implications of molecular breeding strategies for both short-term and long-term genetic improvement of complex traits.

Designing Molecular Breeding Strategies: Today we can consider coordinated development of three themes and associated research paths arising from research over the last 10 to 15 yr.

1. There is a growing knowledge base of the genetic architecture for some traits and how genetic variation is organized within unimproved and elite germplasm pools. Further work in this area will require the integration of genomics technologies with the study of genetic variation to conduct focused gene-to-phenotype studies. This will require fundamental questioning and in some cases refinement of our models of genetic variation. Development of our bioinformatics and computational modeling tools will be necessary.
2. High throughput genetic profiling of individuals for key regions of the genome is now feasible for elite and unimproved germplasm pools. High performance management, manipulation, analysis and interpretation of molecular and phenotypic data will continue to be areas of research priority.
3. Determining the power of molecular and conventional breeding strategies to achieve directed phenotypic changes for simple to complex traits.

High performance computing and simulation is being used to

complement theoretical and experimental investigations. There is a clear need for further research into the appropriate statistical and biological modeling procedures for determining and testing gene-to-phenotype associations for complex traits.

OTHER ASPECTS OF MOLECULAR BREDING STRATEGY

Gene Machines: More direct clues about the function of individual genes are derived from the large-scale insertional mutagenesis of genomes using well-defined transposons, the 'gene machines' (also known as reverse genetics). The 'machine' consists of an active and well characterized transposon (known DNA sequence), a large population of plants (approximately 40,000 individuals) segregating for the transposon, DNA samples and progeny of each individual in the population and a partial DNA sequence of the genes of interest (the targets of the new insertions by the transposon). The population is created in such a way that the transposon inserts into many new sites in the genome. Some insertions into or near genes will modify an aspect of gene expression, possibly affecting gene function and resulting in an observable change in phenotype (a mutant). Since the transposon's DNA sequence is known, the new site of insertion is effectively identified or 'tagged'. Using the DNA sequences of the transposon and the partial sequences of the gene(s) of interest, the entire population is surveyed by PCR and/or hybridization for new insertions into DNA sequences related to the gene(s) of interest and for evidence of cosegregation of the new insertions and a mutant phenotype. DNA fragments that cosegregate with the new mutants may be readily cloned, sequenced, analyzed and identified regarding their information content and relation to the change in phenotype. Many of the DNA fragments will be pieces of genes and will provide the first clues as to the role of that gene in the context of that organism. The details of various strategies for transposon tagging have been summarized (Sundaresan 1996; Maes *et al.* 1999) and the plans and background of a maize-specific transposon-tagging gene machine have been made available (http://www.zmdb.iastate.edu/zmdb/nsf_grant_online.html).

Amplification of Ancient DNA: The DNA persisting in specimens of paleontological, archaeological or forensic interest is inevitably damaged. D'Abbadie *et al.* (2007) described a strategy for the recovery of genetic information from damaged DNA. By molecular breeding of polymerase genes from the genus *Thermus* [*Taq* (*Thermus aquaticus*), *Tth* (*Thermus thermophilus*) and *Tfl* (*Thermus flavus*)] and compartmentalized self-replication selection, they evolved polymerases that can extend single, double and even quadruple mismatches, process non-canonical primer-template duplexes and bypass lesions found in ancient DNA, such as hydantoins and abasic sites. Applied to the PCR amplification of 47,000–60,000-year-old cave bear

DNA, these outperformed *Taq* DNA polymerase by up to 150% and yielded amplification products at sample dilutions at which *Taq* did not. These results demonstrate that engineered polymerases can expand the recovery of genetic information from Pleistocene specimens and may benefit genetic analysis in paleontology, archeology and forensic medicine.

***Cotton Breeding for Drought and Salinity*:** Plant stress caused by drought and salinity are among the major constraints on crop production and food security worldwide. Breeding programs to improve crop yield in dry and saline environments have progressed slowly due to our limited understanding of the underlying physiological, biochemical, developmental, and genetic mechanisms that determine plant responses to these forms of stress, as well as to technical difficulties in combining favorable alleles to create the improved high yielding genotypes needed for these environments. Cotton is an especially appropriate system for research into the molecular basis of plant response to water deficit and salinity, as it originates from wild perennial plants adapted to semi-arid, sub-tropical environments which experienced periodic drought and temperature extremes that are associated with soils with high salt content. The current primary molecular breeding approaches include transgenic modification and quantitative trait mapping with marker-assisted selection. The preliminary work in QTL mapping for drought response and the relationships of the QTLs with the drought-associated measurements is developing a foundation for understanding and using the molecular basis of drought tolerance. QTL mapping for salt tolerance is not moving apace. Using and/or regulating transgene effects on the plant responses to drought and salinity has shown success and will continue to increase our understanding of the complexity of plant's physiological pathways. Improvements in all areas of molecular breeding are almost certain, but the most effective improvements will come from exploiting our improved understanding of the genetic architecture (Jenks *et al*., 2007).

Tomato Breeding for Flavor Traits: Both theoretical and applied aspects of marker-assisted back-crossing of quantitative traits are studied by Lecomte *et al.* (2004) with special emphasis on a program of tomato breeding for flavor traits. The detection of QTLs controlling the variation of tomato quality traits was performed in the progeny of a cross between a cherry tomato chosen for its good flavor and a line with bigger but less tasty fruits. A number of QTLs were detected for the flavor traits, some with major effects. Co-localizations of QTLs controlling several traits were found. Most of the favorable alleles came from the cherry tomato parent for chemical and sensory traits, showing the potential usefulness of this line for tomato organoleptic quality improvement. A marker-assisted selection scheme was thus initiated in order to transfer into elite lines the five regions carrying

the most important QTLs involved in fruit quality. The backcross scheme was first optimized taking into account both theoretical and practical aspects. Three recurrent lines were chosen in order to study the effect of genetic background on QTL expression.

Molecular Breeding for Rainfed Lowland Rice: The conventional breeding program also takes at least 15 years for releasing new cultivars. New breeding strategy can be established to shorten period for cultivar improvement by using marker-assisted selection (MAS), rapid generations advance (RGA), early generation testing in multi-locations for grain yield and qualities. Four generation of MAS backcross breeding were conducted to transfer gene and QTL for bacterial blight resistance (BLB), submergence tolerance (SUB), brown plant-hopper resistance (BPH) and blast resistance (BL) into Kao Dawk Mali 105 (KDML105) a line from Thailand. Selected backcross lines, introgressed with target gene/QTL, were tolerant to SUB and resistant to BLB, BPH and BL. The agronomic performance and grain quality of these lines were as good as or better than KDML105. Approximately 15 populations based on crosses between donor lines selected for drought tolerance and recipient lines selected for yield and quality traits of acceptance to farmer/ markets are being developed in Thailand, Laos and Cambodia (Toojinda *et al.*, 2004).

PLANTMARKERS—A DATABASE OF PREDICTED MOLECULAR MARKERS

PlantMarkers is a genetic marker database that contains a comprehensive pool of predicted molecular markers. We have adopted contemporary techniques to identify putative single nucleotide polymorphism (SNP), simple sequence repeat (SSR) and conserved orthologue set markers. A systematic approach to identify as broad a range of putative markers has been undertaken by screening the available open Sputnik unigene consensus sequences from over 50 plant species. A web presence at http://markers.btk.fi/ provides functionality so that a user may search for species-specific markers on the basis of many specific criteria not limited to non-synonymous SNPs segregating between different varieties or measured polymorphic SSRs. Feedback forms are provided with all sequence entries to enable inclusion of, for example, map location for markers validated by the research community (Rudd *et al.*, 2005)

CONCLUSIONS

Marker based selection could be more efficient than purely phenotypic selection in quite large populations and for traits showing relatively low heritability. Regardless of the ultimate reason for the application of molecular markers for molecular breeding, there is a general need for both high-

density and uniform maps that represent whole genomes. While techniques for the classification and typing of alleles have become amenable to high-throughput screening, e.g. microarray-based genotyping, the more important marker discovery steps require significant investments in both time and money. Single nucleotide polymorphism (SNP) markers, for example, have enjoyed massive popularity through their high density within the genome and their ease of characterization. The identification of these markers, however, requires access to reliable DNA sequence from the complete range of plants strains/varieties or ecotypes that will subsequently be used. With the availability of PCR-based markers, which are simple to use, the cost of different marker techniques is progressively decreasing and may become affordable in the framework of breeding programmes. Molecular breeding applications must not only respond to the challenge of improving food security and fostering socio-economic development, but in doing so, promote the conservation, diversification and sustainable use of plant genetic resources for food and agriculture. Nowadays the biotechnology toolbox available to plant breeders offers several new possibilities for increasing productivity, crop diversification and production, while developing a more sustainable agriculture. Their use in applied breeding programmes can range from facilitating the appropriate choice of parents for crosses, to mapping/tagging of gene blocks associated with economically important traits (often termed "quantitative trait loci" (QTLs)). Through comparative genomics, molecular markers can be used in ways that allow us to more effectively discover and efficiently exploit biodiversity and the evolutionary relationships between organisms. Substantial progress has been made in recent years in mapping, tagging and isolating many agriculturally important genes using molecular markers due in large part to improvements in the techniques that have been developed to help find markers of interest.

New developments in genomic research have given access to an enormous amount of sequence information as well as new insights on the function and interaction of genes and the evolution of functional domains, chromosomes and genomes. In this context, functional and comparative genomics can help in comparative genetic mapping and linkage analysis of useful agricultural traits. Future DNA marker techniques, such as the use of oligonucleotide arrays, are likely to be sequence based. Comparative analyses of sequence information in the growing databases now publicly available on the World Wide Web will be an invaluable resource for formulation of different strategies for molecular breeding. This will have an increasingly important role in the evolution of newer ways for molecular breeding.

REFERENCES

Al-Babili, S. and Beyer, P. 2005. Golden Rice — Five years on the road — five years to go? Trends in Plant Sci., 10:565-573.

Badan, A.C. de Carvalho, Geraldi, I.O., Guimarães, E. P. & Ospina, Y. 1998. Estimativa do tamanho efetivo de amostra ideal para caracterizar uma população de arroz para as características peso de panículas e peso de 100 grãos. Genet. Mol. Biol., 21: 246.

Barry, M. B. 2000. Diversité génétique de variétés locales de riz en Guinée. Analyse de la variabilité intra-variétale au moyen de marqueurs microsatellites. Rennes, France, Université de Rennes. 30 pp. (DEA thesis)

Baurens, F., Noyer, J.L., Lanaud, C. and Lagoda, P. J. L. 1997. Assessment of a species specific element (Brep 1) in banana. Theoretical and Applied Genetics, 95:922–931.

Brown, A. H. D. and Munday, J. 1982. Population-genetic structure and optimal sampling of land races of barley from Iran. Genetica, 58:85–96.

Buckler, E.S. and Thornsberry, J. M. 2002. Plant molecular diversity and applications to genomics. Current Opinions in Plant Biolology, 5:107–111.

Calderini, O., Chang, S. B., de Jong, H., Busti, A., Paolocci, F., Arcioni, S., de Vries, S. C., Abma-Henkens, M. H., Klein Lankhorst, R. M., Donnison, I. S., Pupilli, F. 2006. Molecular cytogenetics and DNA sequence analysis of an apomixis-linked BAC in *Paspalum simplex* reveal a non pericentromere location and partial microcolinearity with rice. Theor. Appl. Genet. 112:1179-1191.

Chen, H. L., Wang, S. P., Xing, Y. Z., Xu, C. G., Patrick, M. H., Zhang, Q. F. 2003. Comparative analyses of genomic locations and race specificities of loci for quantitative resistance to *Pyricularia grisea* in rice and barley. Proc. Nat. Acad. Sci. USA, 100:2544-2549.

Chopra, V. L. 2006. The search of new genes then and now. 13th B. P. Pal Memorial Lecture, May 26, 2006. IARI. 6-11.

Comes, H.P. and Abbott, R. J. 1998. The relative importance of historical events and gene flow on the population structure of a Mediterranean ragwort, *Senecio gallicus* (Asteraceae). Evolution, 52:355–367.

D'Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., Woodgate, R., Pääbo, S. and Holliger, P. 2007. Molecular breeding of polymerases for amplification of ancient DNA. Nature Biotechnology, 25:939 – 943.

De Koeyer, D. L., Philips, R. L. & Stuthman, D. D. 2001. Allelic shifts and quantitative trait loci in a recurrent selection population of oat. Crop Sci., 41: 1228-1234.

Del Rio, A. H. and Bamberg, J. B. 2002. Lack of association between genetic and geographical origin characteristics for the wild potato *Solanum sucrense*. American Journal of Potato Research, 79:335–338.

Espejo-Ibañez, M. C., Sanchez, M. P., Sanchez, M. D. and Yelamo, M. D. 1994. Isoenzymatic variability in seeds of some Spanish common beans (*Phaseolus vulgaris* L. Leguminosae): relation to their domestication centers. Biochemical

Systematics and Ecology, 22:827–833.

Fahima, T., Roeder, M.S., Grama, A. and Nevo, E. 1998. Microsatellite DNA polymorphism divergence in *Triticum dicoccoides* accessions highly resistant to yellow rust. Theoretical and Applied Genetics, 96:187–195.

Freville, H., Justy, F. and Olivieri, I. 2001. Comparative allozyme and microsatellite population structure in a narrow endemic plant species, *Centaurea corymbosa* Pourret (Asteraceae). Molecular Ecology, 10:879–889.

Fulton, T. M., van der Hoeven, R., Eannetta, N. T. and Tanksley, S. D. 2002. Identification, analysis and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell, 14:1457–1467.

Fulton, T.M., Beck-Bunn, T., Emmatty, D., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D. and Tanksley. S. D. 1997. QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. Theor Appl Genet., 95:881-894.

Garvin, D. F. and Weeden, N. F. 1994. Isozyme evidence supporting a single geographic origin for domesticated tepary bean. Crop Science, 34:1390–1395.

Geraldi, I. O. & Souza, Jr, C. L. de. 2000. Muestreo genético para programas de mejoramiento poblacional. *In* E.P. Guimarães, ed. Avances en el mejoramiento poblacional en arroz, pp. 9-20. Santo Antônio de Goiás, Brazil, Embrapa Arroz e Feijão.

Glaszmann, J.C. 1987. Isozymes and classification of Asian rice varieties. Theor. Appl. Genet., 74: 21-30.

Goff, S.A. *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. spp. *japonica*). Science, 296:92–100.

Graterol, M. E. J. 2000. Caracterización de poblaciones e introducción de variabilidad genética para iniciar un programa de mejoramiento poblacional del arroz en Venezuela. *In* E.P. Guimarães, ed. *Avances en el mejoramiento poblacional en arroz*, pp. 87-103. Santo Antônio de Goiás, Brazil, Embrapa Arroz e Feijão.

Gualtieri, G., Conner, J.A., Morishige, D.T., Moore, L.D., Mullet, J.E., Ozias-Akins, P. O. 2006. A segment of the Apospory Specific Genomic Region (ASGR) is highly microsyntenic not only between the apomicts *Pennisetum squamulatum* and *Cenchrus ciliaris*, but also with a rice chromosome 11 centromeric-proximal genomic region. Plant Physiology, 140:936-971.

Guimarães, E.P. 2000. Mejoramiento poblacional del arroz en America Latina: Dónde estamos y para dónde vamos. *In* E.P. Guimarães, ed. *Avances en el mejoramiento poblacional en arroz*, pp. 299-311. Santo Antônio de Goiás, Brazil, Embrapa Arroz e Feijão.

Hamilton MB. 1994. *Ex situ* conservation of wild plant species: time to assess the genetic assumptions and implications of seed banks. Conservation Biology, 8:39–49.

Hanson, W.D. 1959. The breakup of initial linkage blocks under selection mating systems. Genetics, 44: 857-868.

Hospital, F. and Charcosset, A. 1997. Marker-assisted introgression of Quantitative Trait Loci. Genetics, 147:1469-1485.

Hospital, F., Chevalet, C. and Mulsant, P. 1992. Using markers in gene introgression breeding programs. Genetics, 132:1199-1210.

Hospital, F., Moreau, L., Lacoudre, F., Charcosset, A. and Gallais, A. 1997. More on the efficiency of marker-assisted selection. Theor. Appl. Genet., 95:1181-1189.

Huang, L., Brooks, S.A., Li, W., Fellers, J.P., Trick, H.N. and Gill, B.S. 2003. Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics, 164:655-664.

Jenks, M.A., Hasegawa, P.M. and Jain, S. M. 2007. Advances in Molecular Breeding Toward Drought and Salt Tolerant Crops. Dordrecht (Netherlands), Springer , Accession No**:** 415202, ISBN 978-1-4020-5577-5 **,** Call No**:** 631.5 J42 (RR) 775-796.

Karp, A., Edwards, K.J., Bruford, M., Funk, S., Vosman, B., Morgante, M., Seberg, O., Kremer, A., Boursot, P., Arctander, P., Tautz, D. and Hewitt, G.M. 1997a. Molecular technologies for biodiversity evaluation: opportunities and challenges. Nature Biotechology, 15:625–628.

Karp, A., Kresovich, S., Bhat, K.V., Ayad, W.G. and Hodgkin, T. 1997b. Molecular tools in plant genetic resources conservation: a guide to the technologies IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute, Rome, Italy.

Karp, A., Seberg, O. and Buiatti, M. 1996. Molecular techniques in the assessment of botanical diversity. Annals of Botany, 78:143–149.

Kilian, A., Chen, J., Han, F., Steffenson, B. & Kleinhofs, A. 1997. Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Molecular Biology, 35:187-195.

Lamboy, W.F., McFerson, J.R., Westman, A.L. and Kresovich, S. 1994. Application of isozyme data to the management of the United States national *Brassica oleracea* L. genetic resources collection. Genetic Resources and Crop Evolution, 41:99–108.

Lande, R. and Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 124:743-756.

Lanner, H.C., Bryngelsson, T. and Gustafsson, M. 1997. Relationships of wild *Brassica* species with chromosome number 2n = 18, based on RFLP studies. Genome, 40:302–308.

Lanner, H.C., Gustafsson, M., Falt, A.S. and Bryngelsson, T. 1996. Diversity in natural populations of wild *Brassica oleracea* as estimated by isozyme and RAPD analysis. Genetic Resources and Crop Evolution, 43:13–23.

Laurie, D.A. and Devos, K.M. 2002. Trends in comparative genetics and their potential impacts on wheat and barley research. Plant Molecular Biology, 48:729–740.

Lecomte, L., Hospital, F., Buret, M. and Causse, M. 2004. Recent Advances in Molecular Breeding: The Example of Tomato Breeding for Flavor Traits. Proc. XXVI IHC – Advances in Vegetable Breeding, Eds. J.D. McCreight and E.J. Ryder, Acta Hort. 637.

Lee, D.J., Berding, N., Jackes, B.R. and Bielig, L.M. 1998. Isozyme markers in *Saccharum*

spp. hybrids and *Erianthus arundinaceus* (Retz.) Jeswiet. Australian Journal of Agricultural Research. 49:915–921.

Lee, D.J., Reeves, J.C. and Cooke, R.J. 1996. DNA profiling and plant variety registration: 1. The use of random amplified DNA polymorphisms to discriminate between varieties of oilseed rape. Electrophoresis, 17:261–265.

Lemieux, B., Aharoni, A. and Schena, M. 1998. Overview of DNA chip technology. Molecular Breeding, 4:277–289.

Lima, M.L.A., García, A.A.F., Oliveira, K.M., Matsuoka, S., Arizono, H., Souza, Jr, C.L. de & Souza, A.P. de. 2002. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane. Theor. Appl. Genet., 104: 30-38.

Lorenzen, L.L., Boutin, S., Young, N., Specht, J.E. & Shoemaker, R.C. 1995. Soybean pedigree analysis using map-based molecular markers, I: Tracking RFLP markers in cultivars. Crop Sci., 35: 1326-1336.

Maass, B.L. and Ocampo, C.H. 1995. Isozyme polymorphism provides fingerprints for germplasm of *Arachis glabrata* Bentham. Genetic Resources and Crop Evolution, 42:77–82.

Maes, T., P. de Keukeleire and T. Gerats. 1999. Plant tagnology. Trends Plant Sci., 4:90-96.

Manjarrez-Sandoval, P., Carter, Jr. T.E., Webb, D.M. and Burton, J.W. 1997. Heterosis in soybean and its prediction by genetic similarity measures. Crop Science, 23:1443–1452.

Maquet, A., Zoro Bi, I.Z., Delvaux, M., Wathelet, B. and Baudoin, J.P. 1997. Genetic structure of a Lima bean base collection using allozyme markers. Theoretical and Applied Genetics, 95:980–991.

Maquet, A., Zoro Bi, I.Z., Rocha, O.J. and Baudoin, J.P. 1996. Case studies on breeding systems and its consequences for germplasm conservation. Genetic Resources and Crop Evolution, 43:309–318.

Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J., Goldberg, R.B. 1990. Induction of male sterility in plants by a chimaeric ribonuclease gene. Nature, 347: 737-741.

Marín-Garavito, J.M. 1994. Efecto de número de ciclos de recombinación en la variabilidad de poblaciones de arroz *(*Oryza sativa *L.)*. Palmira, Colombia, Facultad de Ciencias Agropecuarias, Universidad Nacional de Colombia. 50 pp. (BSc thesis)

Martin, C., Juliano, A., Newbury, H.J., Lu, B.R., Jackson, M.T. and Ford Lloyd, B.V. 1997. The use of RAPD markers to facilitate the identification of *Oryza* species within a germplasm collection. Genetic Resources and Crop Evolution, 44:175–183.

McGregor, C.E., van Treuren, R., Hoekstra, R. and van Hintum, T.J.L. 2002. Analysis of the wild potato germplasm of the series *Acaulia* with AFLPs: implications for *ex situ* conservation. Theoretical and Applied Genetics, 104:146–156.

Melchinger, A.E., Graner, A., Singh, M. and Messmer, M.M. 1994. Relationships among European barley germplasm. I. Genetic diversity among winter and

spring cultivars revealed by RFLPs. Crop Science, 34:1191–1199.

Michelmore, R.W., Paran, I. & Kesseli, R.V. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc. Natl. Acad. Sci. USA, 88: 9282-9832.

Murphy, J.P. and Philips, T.D. 1993. Isozyme variation in cultivated oat and its progenitor species, *Avena sterilis* L. Crop Science, 33:1366–1372.

National Research Council. 1972. Genetic vulnerability of major crops. National Academy of Sciences, Washington, D.C., USA.

Nebauer, S.G., del Castillo-Agudo, L. and Segura, J. 1999. RAPD variation within and among natural populations of outcrossing willow-leaved foxglove (*Digitalis obscura* L.). Theoretical and Applied Genetics, 98:985–994.

Nei, M. 1972. Genetic distance between populations. Am. Nat., 106: 283-292.

Nevo, E., Apelbaum Elkaher, I., Garty, J. and Beiles, A. 1997. Natural selection causes microscale allozyme diversity in wild barley and a lichen at 'Evolution Canyon', Mt. Carmel, Israel. Heredity, 78:373–382.

Nevo, E., Baum, B., Beiles, A. and Johnson, D.A. 1998. Ecological correlates of RAPD DNA diversity of wild barley, *Hordeum spontaneum*, in the Fertile Crescent. Genetic Resources and Crop Evolution, 45:151–159.

Ospina, Y., Châtel, M. & Guimarães, E.P. 2000. Mejoramiento poblacional del arroz de sabanas. *In* E.P. Guimarães, ed. *Avances en el mejoramiento poblacional en arroz*, pp. 241-254. Santo Antônio de Goiás, Brazil, Embrapa Arroz e Feijão.

Parker, P.G., Snow, A.A., Schug, M.D., Booton, G.C. and Fuerst, P.A. 1998. What molecules can tell us about populations: choosing and using a molecular marker. Ecology, 79:361–382.

Paz, M.M. and Veilleux, R.E. 1997. Genetic diversity based on randomly amplified polymorphic DNA (RAPD) and its relationship with the performance of diploid potato hybrids. Journal of the American Society of Horticultural Science, 122:740–747.

Peccoud, J., K., Velden, V., Podlich, D.W., Winkler, C.R., Arthur, W.L. and Cooper, M. 2004. The selective values of alleles in a molecular network model are context dependent. Genetics, 166:1715–1725.

Peroutka, S.J. 1997. The medical utility of genomics data in neuropsychiatry: mutational genetics versus association genetics. Current Opinions in Biotechnology, 8:688–691.

Perrier, X., Flori, A. & Bonnot, F. 2003. Les méthodes d'analyse des données. In P. Hamon, M. Seguin, X. Perrier, J.C. Glaszmann, eds. *Genetic diversity of cultivated tropical plants*, pp. 31-64. Enfield, NH, USA, Science Publishers and CIRAD.

Rabbani, M.A., Maruyama, K., Abe, H., Khan, M.A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M. 2003. Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses. Plant Physiolo., 133:1755-1767.

Rafalski, J.A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science, 162:329–333.

Rieseberg, L.H. and Ellstrand, N.C. 1993. What can molecular and morphological markers tell us about plant hybridization? Critical Reviews in Plant Science, 12:213–241.

Rieseberg, L.H. 1995. The role of hybridization in evolution: old wine in new skins. American Journal of Botany, 82:944–953.

Roa, A.C., Maya, M.M., Duque, M.C., Tohme, J., Allem, A.C. and Bonierbale, M.W. 1997. AFLP analysis of relationships among cassava and other *Manihot* species. Theoretical and Applied Genetics, 95:741–750.

Rudd, S., Schoof, H. and Mayer, K. 2005. PlantMarkers—a database of predicted molecular markers from plants. Nucleic Acids Research,33:D628-D632.

Schlötterer, C. 2004. The evolution of molecular markers—just a matter of fashion? Nature Reviews Genetics, 5:63–69.

Schneider, K. and Douches, D.S. 1997. Assessment of PCR-based simple sequence repeats to fingerprint North American potato cultivars. American Potato Journal, 74:149–160.

Schoen, D.J. and Brown, A.H.D. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proceedings of the National Academy of Sciences USA, 90:10623–10627.

Smith, O.S., Smith, J.S.C., Bowen, S.L., Tenborg, R.A. and Wall, S.J. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield, heterosis and RFLPs. Theoretical and Applied Genetics, 80:833–840.

Soleimani, V.D., Baum, B.R. & Johnson, D.A. 2002. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat. Theor. Appl. Genet., 104: 350-357.

Steiner, J.J., Piccioni, E., Falcinelli, M. and Liston, A. 1998. Germplasm diversity among cultivars and the NPGS crimson clover collection. Crop Science, 38:263–271.

Sundaresan, V. 1996. Horizontal spread of transposon mutagenesis: new uses for old elements. Trends Plant Sci., 6:184-190.

Tanksley, S.D. and Nelson, J.C. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theoretical and Applied Genetics, 92:191–203.

Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A.J. and Tunnacliffe, A. 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics, 13:718–725.

Thieme, R., Darsow, U., Gavrilenko, T., Dorokhov, D. and Tiemann, H. 1997. Production of somatic hybrids between *S. tuberosum* L. and late blight resistant Mexican wild potato species. Euphytica, 97:189–200.

Toojinda, T., Tragoonrung, S., Vanavichit, A., Siangliw, J. L., Pa-In, N., Jantaboon, J., Siangliw, M. and Fukai, S. 2004. Molecular breeding for rainfed lowland rice in the Mekong Region. New directions for a diverse planet: Proceedings of the 4th International Crop Science Congress, Brisbane, Australia, 26 Sep – 1

Oct 2004.

Tsegaye, S., Tesemma, T. and Belay, G. 1996. Relationships among tetraploid wheat (*Triticum turgidum* L) landrace populations revealed by isozyme markers and agronomic traits. Theoretical and Applied Genetics, 93:600–605.

Ursla, F.W.M., Hayward, M.D. and Kearsey, M.J. 1997. Isozyme and quantitative traits polymorphisms in European provenances of perennial ryegrass (*Lolium perenne* L.). Euphytica, 93:263–269.

Van Berloo, R. and Stam, P. 1998. Marker-assisted selection in autogamous RIL populations: a simulation study. Theor. Appl. Genet., 96:147-154.

Varghese, Y.A., Knaak, C., Sethuraj, M.R. and Ecke, W. 1997. Evaluation of random amplified polymorphic DNA (RAPD) markers in *Hevea brasiliensis*. Plant Breeding, 116:47–52.

Visscher, P.M., Haley, C.S. and Thompson, R. 1996. Marker-assisted introgression in backcross breeding programs. Genetics, 144:1923-1932.

Whitkus, R., de la Cruz, M., Mota Bravo, L., Gomez Pompa, A. and De la Cruz, M. 1998. Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. Theoretical and Applied Genetics, 96:621–627.

Whitkus, R., Doebley, J. and Wendel, J.F. 1994. Nuclear DNA markers in systematics and evolution. In DNA-based markers in plants (Advances in cellular and molecular biology of plants, vol. 1.) (RL Phillips and IK Vasil, eds.). Kluwer Academic Publishers, Dordrecht, The Netherlands,116–141.

Yamada, T., Misoo, S., Ishii, T., Ito, Y., Takaoka, K. and Kamijima, O. 1997. Characterization of somatic hybrids between tetraploid *Solanum tuberosum* L. and dihaploid *S. acaule*. Breeding and Science, 47:229–236.

Yang, W.P., De Oliveira, A.C., Godwin, I., Schertz, K. and Bennetzen, J.L. 1996. Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. Crop Science, 36:1669–1676.

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P. & Potrykus, I. 2000. Engineering provitamin A (b-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science, 287:303-305.

Young, N.D. and Tanksley, S.D. 1989. RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor.Appl. Genet., 77:353-359.

Yu, J. *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. spp. *indica*). Science 296:79–92.

Zerega, N.J.C., Ragone, D. and Motley, T.J. 2004. Complex origins of breadfruit (*Artocarpus altilis*, Moraceae): implications for human migrations in Oceania. American Journal of Botany, 91:760–766.

Zhou, Z., Bebeli, P.J., Somers, D.J. and Gustafson, J.P. 1997. Direct amplification of minisatellite-region DNA with VNTR core sequences in the genus *Oryza*. Theoretical and Applied Genetics, 95:942–949.

Zoro Bi, I., Maquet, A., Degreef, J., Wathelet, B. and Baudoin, J.P. 1998. Sample size for collecting seeds in germplasm conservation: the case of the Lima bean (*Phaseolus lunatus* L.). Theoretical and Applied Genetics, 97:187–194.