Molecular Marker-Assisted Breeding

Jonathan H. Crouch

Head, Applied Genomics Laboratory
International Crops Research Institute for the Semi-Arid Tropics
Patancheru 502 324, Andhra Pradesh, India. J.H.Crouch@CGIAR.ORG
http://grep.icrisat.cgiar.org/homepage/genomics/output.asp

Introduction

The potential value of genetic markers, linkage maps and indirect selection in plant breeding has been known for over 80 years. However, it was not until the development of DNA marker technology in the 1980s, that a large enough number of environmentally insensitive genetic markers could be generated to adequately tag a range of important agronomic traits. Since this time DNA marker technology has promised to dramatically enhance the efficiency of plant breeding as molecular biology has already revolutionized research in the life sciences. Yet it is only now as we enter the new millenium that advances in automated technology present the convenience, speed and level of throughput that can finally offer relevance to modern plant breeding programs.

The theoretical basis for molecular marker-assisted breeding is well established but still rapidly evolving with a wide array of published examples covering most crops of major economic importance. Meanwhile, dramatic advances are being made in applied genomics, which will undoubtedly fuel the development of knowledge-led breeding schemes. However, beyond these scientific developments there is a particular paucity of studies addressing the practical and economic benefits of molecular breeding.

DNA marker-assisted approaches must undergo a more wide-ranging scrutiny beyond the purely scientific issues, which will define the pace and extent of their transition from the research laboratory to the plant breeding arena. Here practical concerns and cost-benefit analysis become all important. This paper attempts to review some of the more important issues related to the multifaceted evaluation of DNA marker-assisted approaches that leaders of breeding programs must address before committing to such new endeavours.

The use of DNA markers for indirect selection offers greatest gains for quantitative traits with low heritability as these are the most difficult characters to work with in the field through phenotypic selection. However, this type of trait is also amongst the most difficult to develop effective marker assisted selection systems. This is largely due to the effects of genotype-by-environment (GxE) interaction and epistasis. Precise phenotypic evaluation in several locations and seasons is an important means to measure these effects and estimate the relative

contribution and stability of component quantitative trait loci (QTL). However, there is an increasing support for the idea that the use of much larger populations is an even more important factor. The dissection of quantitative traits using DNA markers has forced an increasing dependence on ever more complex biometric tools to facilitate interpretation and manipulation of the underlying genetic factors. Here the techniques which have allowed traditional plant breeders to deal with complex phenotypes are increasingly important in the new field of molecular breeding.

Current status of molecular breeding research and development

Methodologies

The basis of all genetic maps (and the markers on which they are based) relies on the theory that Mendelian genetic factors which exist close together on the same chromosome have a high probability of being co-transmitted from the parent to progeny. This concept was developed using flower characters in pea and further defined using eye and wing characters in *Drosophila*. These characters were convenient for early studies but are relatively infrequent markers that interact with the environment. A second generation of markers evolved through the use of isozymes, which are rarely associated with undesirable phenotypic effects and are more abundant than morphological markers. However, of the 3000 or so plant enzymes known, less than 60 have been assayed for isozyme polymorphism and only ten to twenty isozyme loci are commonly found to be polymorphic in most breeding populations.

Molecular markers have no phenotypic effect, are not affected by the presence or absence of other loci and a relatively large number can be detected in most breeding populations (Arus and Moreno-Gonzalez 1993). The concept of genetic maps based on molecular genetic markers was developed with reference to the human genome. Based on this development, molecular approaches have rapidly proven to be powerful tools for indirect selection and for studying the organization and behaviour of plant genomes (Dear 1997).

The primary resource of plant breeding programs is the genetic variability available within germplasm closely related to the crop of interest. However, the success of crop improvement programs is highly reliant on the power and efficiency with which this genetic variability can be manipulated. DNA marker technologies offer plant breeders the potential of making genetic progress more precisely and more rapidly than through phenotypic selection. Genetic markers also offer the possibility of addressing previously unattainable goals. This is now equally true for both temperate and tropical crops. In particular, progress in model systems offers the possibility of supporting both substantial and rapid developments in tropical crop improvement, which would not be conceivable through traditional methods.

There is now a wide array of DNA marker assays, each having a different set of advantages in any particular application (reviewed in Table 1, for further information see http://www.nal.usda.gov/pgdic/tutorial/lesson4.htm). Restriction fragment length polymorphism (RFLP), although used extensively in early mapping studies, particularly of cereal crops, are not amenable to applications in marker-assisted selection. RFLP marker analysis is too labour intensive and has too low a throughput potential for routine screening of

breeding populations. However, with the development of the polymerase chain reaction (PCR) there is now a range of assays which have great potential for molecular breeding.

Table 1. Major classes of genetic markers

Morphological traits: such as seed or flower colour are seriously limited in number while dominance, late expression, deleterious effects, pleiotropy and epistasis frequently reduce the usefulness of such markers.

Proteins: analysis of isozymes has had limited success due to the low number of available markers. However, new techniques which simultaneously assay more than 50 seed storage proteins and structural proteins etc. provide a very cost effective means of screening variation in expressed traits which may be particularly powerful for distinctness, uniformity and stability (DUS) testing.

Restriction fragment length polymorphism (RFLP): requires hybridization of probe DNA with plant DNA and although provides high quality data has a severely limited throughput potential.

Random amplified polymorphic DNA (RAPD): was the first of a new generation of markers based on the polymerase chain reaction (PCR). This technique uses arbitrary primers for initiating amplification of random pieces of plant DNA. This technique requires no knowledge of the genome to be screened but suffers inconsistencies between populations and laboratories.

Simple sequence repeat length polymorphism (SSRLP): also known as microsatellite, variable number of tandem repeats (VNTR) or sequence tagged microsatellite site (STMS) markers. This technique provides high quality, highly consistent results and remains the assay of choice for marker-assisted selection. However, these markers are expensive to develop as they require extensive sequence data from the species of interest.

Amplified fragment length polymorphism (AFLP): in this approach the sample DNA is enzymatically cut up into small fragments (as with RFLP analysis) but only a fraction of fragments are studied following selective PCR amplification. Although this assay provides a great quantity of marker information, it is not particularly well suited to high throughput marker assisted selection. However, techniques for converting AFLP marker bands to simple PCR tests have been reported.

Expressed sequence tag (EST): the development of EST markers is dependent on extensive sequence data of regions of the genome which are expressed. However, once developed they provide high quality, highly consistent results and because they are limited to expressed regions of the genome, markers themselves are directly associated with functional genes. EST markers are likely to be less polymorphic than SSR markers.

Single nucleotide polymorphism (SNP): the vast majority of differences between individuals are point mutations due to single nucleotide polymorphisms. As such, there are a vast number of potential SNP markers in all species. Considerable amounts of sequence data are required from parental genotypes to develop SNP markers, however, their great advantage lies in the potential to screen them using methods which do not involve electrophoresis, such as microarrays.

Random amplified polymorphic DNA (RAPD) was the first PCR-based assay to receive wide spread attention in plant fingerprinting and mapping studies. However, problems of reproducibility within and between populations and laboratories have largely cast this assay to

the history books. Microsatellite or simple sequence repeat (SSR) markers based on variable number of tandem repeats (VNTR), are much more reliable PCR-based markers as they are based on stringent amplification of known DNA sequences. Microsatellite markers are highly polymorphic co-dominant assays and remain the assay of choice for marker-assisted selection systems. Unfortunately, microsatellite markers are expensive and time consuming to develop and consequently as yet have only realized their full impact in the advanced crop systems.

Amplified fragment length polymorphism (AFLP) offers stringent amplification of a very large number of alleles in a single assay. This is particularly useful for germplasm fingerprinting but not appropriate for routine marker-assisted selection. AFLP assays are generally considered to provide dominant marker information. However, through the use of high quality DNA and stringent reaction conditions, it is possible to achieve a high proportion of co-dominant data based on fragment density measurements (KeyGene, unpublished data).

PCR-based markers for known genomic sequences will dominate marker-assisted selection systems of most crops for the foreseeable future. This is due to the ease with which PCR-based marker analysis can be highly automated. On this basis it is possible to screen a limited number of genomic regions across vast populations. In contrast, amplified fragment length polymorphism (AFLP) and protein-based approaches will be more appropriate for certain other molecular breeding applications where population sizes are smaller but genome coverage is more important. While for introgression breeding a complementary approach may be most efficient. In this scenario, the trait of interest may first be selected with the aid of PCR-based markers and then the optimum genome composition selected using AFLP markers.

Although it is possible to convert RFLP markers into simple PCR assays, this very often requires screening for new restriction polymorphism within the PCR amplified fragment amplified by the PCR assay. It is also possible to convert AFLP marker bands into simple PCR assays. The development of sequence tagged site (STS) tests from AFLP markers has been reported in several systems. However, the procedure is particularly difficult in crops that have a high level of duplicated genomic regions. In these cases, more intensive strategies must be adopted. In general the conversion of most AFLP bands results in the development of a STS test providing only dominant marker information. To put this in perspective, KeyGene offer a service for the conversion of AFLP marker bands with an average 50% success rate of developing a co-dominant STS test in around six months at a cost of around $30,000.

Most recently, the abundance of sequence information in the model systems has led to the development of two new types of markers. Expressed sequence tag (EST) markers can be generated in vast numbers through random sequencing of cDNA libraries. In general EST markers are not likely to be as polymorphic as SSR markers. However, through the use of differential libraries it is possible to develop markers based on genomic sequences expressed only under certain environmental conditions of interest, such as drought or disease stress.

The vast majority of genetic differences between individuals are due to single nucleotide polymorphisms (SNP). Although an almost limitless number of SNP markers may be available, it is the detection of such polymorphisms which currently presents the greatest

technological challenge. In this respect, micro-array or DNA chip technology may offer the most cost effective solution. To date this technology has largely been utilized for studying expression patterns in functional genomics research. However, recent developments in the use of micro-arrays for SNP detection suggest that in future it may be possible to simultaneously screen for a vast number of genomic marker loci in a single assay (Pastinen et al. 2000). In addition, micro-array analysis requires only minute reaction volumes, thus there is also a massive reduction in unit costs of marker-assisted selection systems based on this technology. Combining these two factors clearly offers the potential for developing completely new paradigms in cost effective molecular breeding strategies.

Applications

There are many applications for the use of DNA markers in breeding programs (reviewed in Table 2 and Mohan et al. 1997) which may be arranged in four broad groups based on the justification for the intervention:

Enhanced knowledge of breeding material and systems
Rapid introgression or backcross breeding of simple characters
Early or easy indirect character selection
New goals not possible through traditional breeding

Table 2. Primary applications of DNA markers in plant breeding

Improved access and utilization of germplasm resources – DNA marker analysis for defining the genetic structure of plant populations, species, genera and families in order to optimize the acquisition, management and utilization of germplasm collections.

Genetic analysis of breeding populations - For many crops, particularly tropical vegetatively propagated crops, the current genetic and cytogenetic knowledge restricts crop improvement efforts. Molecular markers are contributing to a substantial resurgence of progress in these areas.

Parental selection and predicting progeny performance - based on genetic diversity estimated by DNA marker analysis.

Marker-assisted selection – indirect selection of traits which are difficult to score (technically or due to environmental-specific expression), expressed late in the growth season and/or traits which are a primary selection criterion but occur infrequently in breeding populations. The benefits of this approach are compounded when multiple traits can be simultaneously selected.

Marker-accelerated backcross breeding – when introgressing traits from exotic germplasm, DNA markers can be used for indirect selection of that trait plus simultaneous selection of offspring with the least amount of other genomic material from the exotic parent.

Pyramiding genes from diverse sources – it may not be possible to identify different sources of resistance to the same disease through field evaluation. However, it is useful to combine different sources of resistance in the same variety in order to reduce the chance the pathogen will evolve mechanisms to breakdown this resistance. Similarly, many genes may contribute to important

agronomic characters such as yield but it may not be easy to identify the presence of individual genes through field evaluation.

Fingerprinting for impact assessment and protection of plant breeders' rights – by identifying unique DNA marker fingerprints, elite lines can be identified in farmers' fields and in new varieties.

Comparative mapping – Recent studies on cereal crops have shown a high level of similarity of certain genes and the position of those genes in the genome across members of this diverse group. These developments will allow the considerable progress in model systems to be increasingly utilized in related and unrelated species and genera.

Gene isolation, function and manipulation – Based on dense DNA marker maps, scientists can move onto the isolation and characterization of single genes and whole genomic regions. From this point, rapid progress can be made in determining gene function or transferring important genes across species barriers.

Fingerprinting pests and pathogens – DNA marker analysis for phytosanitation screening or monitoring changes in pest populations in order to predict the breakdown of current sources of resistance to viruses, bacteria, fungi, nematodes, arthropods and insects etc.

Evaluating the cost-benefit ratio of DNA-marker assisted breeding

Despite a vast array of publications reporting the identification of DNA markers for specific agronomic traits, there is a particular paucity of reports evaluating the application of such markers in real breeding programs. However, a few reports from the model cereal crop systems (maize and rice) plus simulation studies begin to provide some insight for practitioners.

It is notable that currently successes in cost effective DNA marker-assisted approaches are based on the enhancement of backcross methods. In the past, backcross breeding strategies have not been popular with many breeders due to the lengthy process of recovering the recurrent parent phenotype. DNA markers offer approaches to improve breeding lines for individual characters which may be both quicker and cheaper than traditional methods. The value of time is well appreciated in commercial breeding companies, where the shortening of crop breeding cycles can be worth tens of millions of dollars within individual country markets (Pandey and Rajatasereekul 1999).

The duration to product release will also become important in public breeding programs as activities become increasingly driven by national government and development investor priorities through short-term projects. Moreover, with increasing emphasis upon intellectual property rights, it is becoming increasingly difficult to use proprietary germplasm protected through essential derivative statements. Marker-assisted backcross breeding can help to identify breeding products beyond the bounds of essential derivative clauses.

Cost-benefit analysis will need to be carried out on a case by case basis, as is typically carried out by breeders when considering any component change in their program. However, absolute cost savings are not the only criteria in defining plant breeding strategies, increased timeliness within the whole breeding program may also be highly valuable. For example, interventions

which spread the work load beyond the intensive period from harvest to planting offer considerable benefits in terms of managing the overall efficiency of the breeding program which can not be easily quantified. Conversely, DNA marker techniques may allow breeders to address new goals not previously possible through traditional approaches. For example, the pyramiding of disease resistance genes with indistinguishable phenotypes may facilitate the development of durable sources of resistance, thus extending the life span of varieties.

As DNA marker-assisted approaches become more widely integrated into plant breeding programs their impact will compound and their unit costs will decrease. Similarly, as functional genomics provides ever more detailed understanding of the systems that breeders are manipulating, DNA markers will provide the tools to develop a new knowledge-led paradigm in plant breeding. With increasing demand for these technologies, competition in the market place will force a reduction in equipment and reagent costs which will in turn fuel increasing adoption of these approaches. Already, the large multinational plant breeding companies have justified large capital commitments to these technologies, based on the potential for more rapid product development which they feel will return their investment several times over.

In the following section case studies are used to demonstrate the direct financial benefit of single trait marker-assisted selection interventions. While in the subsequent section experimental projects are used to show the future potential of molecular breeding.

Marker-accelerated backcross breeding of quality protein trait in maize

Maize varieties tend to be low in two essential amino acids, lysine and tryptophan. Over the past 30 years, scientists at CIMMYT have developed high yielding maize lines with enhanced levels of these amino acids together with dramatically higher total protein levels. This advance has recently been awarded the World Food Prize for 2000. The demand for transferring the quality protein trait into new elite locally adapted lines of maize is likely to be very high.

At the core of this development is a recessive mutation known to result in a significant rise in the levels of lysine and tryptophan but this also has several deleterious pleiotropic effects, which can be corrected through the use of modifier genes. This trait is a particularly suitable target for marker-assisted breeding as it is only expressed late in the growth season and the phenotype can often only be assessed through laboratory procedures that cannot identify heterozygotes.

Through the application of co-dominant DNA marker-assisted selection using microsatellite markers, progeny with the desired homozygous genotype can be selected during the vegetative phase and thus only a small proportion need be grown to flowering, crossed and resown in the following season. Cost-benefit analysis of this system indicates that for even moderate population sizes, DNA marker-assisted selection is less expensive than phenotyping. This approach, therefore, offers considerable additional savings in terms of labour for crossing, seed production and preparation, and field costs in the subsequent season. The costs of phenotypic and marker-assisted selection options are reviewed in Table 3.

Table 3. Relative cost of different approaches to marker-assisted selection (MAS) and phenotyping for breeding quality protein trait in maize (derived from Dreher et al. 2000).

	Cost per sample
Wet chemistry phenotyping	$3.64
ELISA phenotypying	$2.62
Single trait MAS	$2.13
Genome background MAS	$68.73

In this study the cost of genome background MAS appears relatively high. However, this approach enables the simultaneous selection of individuals that have the maximum amount of recurrent parent genome together with loci for the quality protein trait. In this way marker-assisted selection can reduce by half the number of backcross cycles required (Ribaut and Hoisington 1998). The resultant large savings in field costs and implications of faster product development can warrant this expenditure. However, there are also several ways of reducing the cost of marker-accelerated backcross programs. Automation of multiplexed microsatellite marker assays would improve the cost-benefit ratio. More dramatic reductions in cost could be achieved by using an alternative assay such as AFLP or protein markers in a complimentary two stage approach. First microsatellite marker analysis would be used to select individuals with the appropriate alleles for the trait of interest. Second, this subset alone would be screened with AFLP, for example, to select those individuals with the most appropriate genomic composition.

In this study, the unit cost of biochemical phenotyping alone was higher than a single DNA marker assay. Thus, the cost-benefit analysis of this approach clearly favours marker-assisted selection. However, costing all aspects of traditional and new approaches provides a number of important insights. The capital set-up costs for molecular breeding are often considered prohibitively high. Yet the actual costs of precision drills, tractors, tillage equipment and combine harvesters is also extremely high. On this basis, investment in molecular breeding may be less than that required to expand field operations or replace farm machinery to maintain the existing level of field operations. Similarly, labour demands are high in traditional breeding programs while reagent costs are high in molecular breeding programs. Yet automated multiplexed genotyping in marker-assisted approaches can dramatically reduce unit costs whereas it is difficult to automate many of the labour intensive aspects of traditional approaches.

This study was carried out in Mexico where labour costs are relatively low while laboratory reagent costs are relatively high as compared to those experienced by temperate breeding programs. In this respect, the study is highly relevant for tropical breeding programs. In contrast, this study was carried out in an international research institute and thus can not reflect the range of capital depreciation procedures and land cost scenarios experienced by commercial plant breeding companies. For this reason it is important to realize that any case study can only provide an impression of cost-benefit ratios. The final determination for any institution will depend upon the balance of variable costs which will vary considerably based

on size of breeding program, level of commitment to DNA markers plus labour and reagent pricing structures in the host country.

Other confounding factors in this type of analysis arise from breeding plots being used for multiple purposes which complicates efforts to estimate costs of individual trait evaluations. Similarly, farm equipment may be used across many trials and even shared between breeding programs. Conversely, unit costs of molecular marker analysis will be severely biased by capital depreciation costs at lower throughput levels while field trial equipment is generally already in place and will only incur capital costs under replacement or expansion scenarios.

In the real world breeding schemes are designed as a compromise between costs, speed and precision of all components. In this study, a number of alternative scenarios were evaluated, ranging from inexpensive rapid systems which may not always work (1), through time consuming expensive but reliable traditional strategies (2), to expensive, rapid and reliable marker-assisted approaches (3 and 4). Some aspects of the cost-benefit analysis for these different breeding scenarios are reviewed in Table 4.

Table 4. Relative cost of different scenarios for breeding quality protein trait in maize, including high throughput (HTP) marker-assisted selection (derived from Dreher et al. 2000).

	Conventional Breeding		***Molecular Breeding***	
	1	*2*	*3*	*4*
Number of seasons	7	15	7	6
Reliability	Low	High	High	High
Cost (current)	$975	$4,367	$2,761	$5,084
Labour costs	59.7%	62.2%	39.4%	28.6%
Reagents	19.7%	13.5%	34.4%	51.7%
Cost (HTP)	$975	$4,367	$970	$1,241

Conventional approach (1) uses selection based on assessment of ear characters known to be associated with the quality protein trait (and biochemical analysis of selected individuals). Success of this approach is dependent upon the skill of those scoring the ears and whether or not the morphological marker is expressed in that cross or environment. Thus conventional approach (2) represents the most appropriate point of reference being based on biochemical analysis of the quality protein trait of all individuals at each generation. Molecular approach (3) represents simple marker-assisted selection for the target allele while approach (4) offers the greatest potential for progress by simultaneously selecting for the target allele and the background genotype.

With increasing emphasis on marker techniques, the proportion of costs allocated to labour decreases while the proportion for reagents increases. However, the actual costs of marker analysis can be dramatically reduced through the implementation of automated high throughput technologies and complementary approaches combining different types of assays.

Marker-assisted selection of salinity tolerance in rice
Phenotypic evaluation of salinity tolerance in rice is time consuming and expensive. First plants must be evaluated in controlled environment chambers using culture solution systems. Results are then confirmed under glasshouse pot experiments using saline irrigation. Finally, the selected advanced generation lines are then evaluated in replicated field trials costing around $30 per genotype. Most significantly, the success of field evaluation is highly unpredictable, particularly in terms of the level of salinity stress applied.

PCR-based markers have been identified for genes underlying salinity tolerance in rice. Using these markers for indirect selection of this trait costs less than 10% of phenotypic evaluation and allows a magnitude more plants to be screened in a season. In the IRRI breeding program, the use of this approach is reducing the breeding cycle by many years. Costs for traditional and marker-assisted selection are compared in Table 5.

Table 5. Molecular breeding of salinity tolerance in rice (IRRI unpublished data).

	Traditional Selection	*Molecular Breeding*
Screening cost		
Per plant	$30	$2
Per 1000 plants	$30,000	$2,000
Capacity	100/season	200/week
Time required	5 years	2 months

In this analysis, no account is taken of capital costs. However, even taken as a single component intervention, the capital cost of establishing a basic molecular marker screening facility could be recovered during the screening of less than 3,000 plants. More important perhaps is the increased opportunity to make intensive progress in this trait using DNA markers as compared to the very limited numbers that can be dealt with through traditional approaches.

Marker-assisted selection of late season traits in tomato and pepper
AFLP marker bands have been converted to simple PCR tests for the development of more than 50 marker-assisted selection systems in vegetable and field crops (KeyGene, unpublished data). Commercial plant breeding companies have contracted these molecular breeding projects for one of three main reasons:

Cost	Characters with expensive field or glasshouse evaluations
Precision	Characters with lower phenotypic accuracy
Linkage drag	To efficiently remove donor genome from around introgressed gene

Simple PCR tests have been developed for traits expressed late in the growth season of tomato and pepper. Once developed, the unit cost of marker-assisted selection was around 10% of the cost of phenotypic evaluation. In addition, indirect selection could be carried out one to two months prior to phenotypic evaluation. Clearly the cost of developing such a marker system is

high. However, in-company cost-benefit analysis showed that the initial investment was completely recovered during the first year of implementation. This scenario is representative of many biochemical trait assessments or pest and disease evaluations which may cost $10 to $30 per individual.

Table 6. Molecular breeding of tomato (KeyGene, unpublished data).

	Traditional Selection	*Molecular Breeding*
Identifying marker	-	$30,000
Developing assay	-	$30,000
Screening cost		
Per plant	$20	$2
Per 3000 plants	$60,000	$6,000
Total cost	$60,000	$66,000

As this study relates to contracted molecular marker analyses, there is no need to consider capital costs etc. and provides an opportunity to estimate costs of developing marker assays as this component has not been assessed in either of the cost-benefit studies carried out at CIMMYT and IRRI.

When using AFLP for marker-assisted selection, there are two components in the development of an appropriate assay. In the first instance, marker-trait associations must be identified. Assuming the need to screen around ten AFLP assays across a mapping population of 200 individuals, this might cost in the region of $30,000. Based on a somewhat optimistic scenario that this would provide sufficient candidate bands for successful conversion to a simple PCR assay, this might then cost a further $30,000. Despite this high developmental cost, in-house cost-benefit analysis indicated that this financial investment was fully recovered during the first year of implementation.

Using protein markers to improve the efficiency of seed production systems

In many countries, plant breeders' rights and essential derivation clauses of patents require specific levels of distinctness from new varieties entered for registration. Traditionally this has been based on morphological, physiological and biochemical characteristics. On this basis it is necessary to devote considerable resources during seed production to ensure new materials do not fail variety registration tests for distinctness, uniformity and stability (DUS). Such testing procedures add greatly to the cost of variety development and varietal registration trials.

DNA marker analyses may be warranted in order to precisely fingerprint new lines for plant variety protection. However, protein marker analysis is a particularly cost effective alternative for routine DUS testing as there is a need to study populations in order to estimate distinctness, uniformity or stability in seed production systems.

A proprietary electrophoresis technique has been developed that can simultaneously assay polymorphisms at more than 60 loci (Proteios, unpublished data). The use of protein markers has the theoretical disadvantage that expression may be affected by the environment and only expressed loci are considered. However, the simultaneous detection of such a large number of loci by a single low cost assay offers considerable practical advantages when attempting to gain a general and rapid reflection of genome composition.

Uniformity testing to confirm homogeneity of inbred variety stocks used for seed production, can be carried out by Proteios based on 96 individuals per line, for as little $250. This is considerably cheaper than characterizing lines based on phenotypic traits. Moreover, this analysis can be carried out on seed lots prior to sowing enabling selection and field establishment of just those lines or seed lots showing acceptable uniformity. Pollination control and seed production systems are extremely expensive. Thus marker-assisted selection methods which facilitate the establishment of just a proportion of lines offer considerable cost savings. Furthermore, the cost of carrying out a similar diversity analysis using DNA markers would be five to ten times higher.

This type of uniformity test is also critically important for assessing purity of inbred parents and of hybrid seed production. A similar approach can also be followed by companies producing tissue culture propagules for clonal production of horticultural or plantation crops. The high cost of such planting material leads to a high expectancy of quality amongst growers. Thus, it is important for micropropagation companies to develop screens to minimize the sale of mutants and contaminants.

The potential of molecular breeding

Pyramiding disease resistance genes in rice and millet

New plant types of rice are being developed which may have a yield potential 20% higher than current varieties. However, these lines are generally susceptible to bacterial blight. A number of bacterial blight resistance genes have been mapped using RFLP markers (Huang et al. 1997). Simple PCR tests were developed from these markers based on the DNA sequence of the respective RFLP probes. These tests were then used for indirect selection of resistance genes during phenotypic selection of the new plant type in backcross generations. On this basis an array of near-isogenic lines (NIL) were developed in just three generations.

Markers for bacterial blight resistance genes are also being used by several national programs within the Asian Rice Biotechnology Network (ARBN). Marker-accelerated backcross breeding has been used in China to transfer bacterial blight resistance into a popular restorer line and hybrid products are already being tested. Meanwhile, in India commercial cultivars and a line containing three bacterial blight resistance loci have been used in a similar marker-accelerated backcross program. Products from this initiative are already being evaluated in multilocational on-farm trials. The potential return of deploying bacterial blight resistance varieties is estimated at $5 to $6 million in each of the countries involved in the ARBN.

A similar approach has been used for the backcross transfer of QTL for downy mildew resistance in pearl millet (Witcombe and Hash 2000). Here a limited number of RFLP probes have been used directly for marker-assisted selection to improve disease resistance in both parent lines of a popular hybrid variety. Despite the labour intensive nature of this approach and the resultant limitation in population size in a given generation, good progress has been made and field evaluation of the finished projects is underway just four years after initiation of the project. Clearly, where breeding goals can not be achieved through traditional approaches, there is considerable scope for the use of molecular markers at almost any stage of development. Here the limitation is not the facilitating technology but the imagination and motivation of the facilitating scientists. Although this clearly demonstrates that marker-assisted selection can work with any type of marker, RFLP markers do not offer a viable scenario for plant breeding companies wishing to entry this field as the cost of establishing facilities for radioactive labeling work would be prohibitively high.

Molecular breeding of drought tolerance in maize

Drought is an important constraint to production in many tropical regions. However, drought tolerance is a highly complex character and it's precise evaluation is confounded by unpredictable environmental conditions and the time consuming and expensive nature of assessing components trait. In addition, the efficiency of selection is lower under drought conditions than well-watered conditions, due to a decrease in the heritability of grain yield.

Drought tolerance in maize has been dissected into several component traits and respective QTL mapped (Ribaut et al. 1997). Using combinations of different QTL, the CIMMYT scientists showed that just as breeders use a selection index combining different traits, that similarly the most effective molecular breeding strategy must combine selection of QTL for a range of key traits.

Diversity analysis and marker-assisted germplasm enhancement

Breeders have traditionally been reticent about the use of wild germplasm in their breeding programs due to complex, long-term and unpredictable outcomes, particularly in crops where quality traits are important market criteria. DNA marker assisted approaches will provide breeders with the tools to effectively unleash the vast resources held in germplasm collections.

DNA marker-based diversity analysis will enable gene banks to define core collections, which will provide a user friendly entry point for breeders to access large and varied germplasm collections. This analysis will also greatly aid selection of genotypes for broadening the genetic base of breeding populations and for the development of heterotic populations for breeding F_1 hybrid varieties.

It is clear that different assays screen genomes in differential ways, thus diversity estimates based on single assays may be significantly different. On this basis, accurate diversity analysis becomes an intensive multi-component endeavour, which may be best carried out by international institutions in their role as custodians of international germplasm collections.

New paradigms in plant breeding using DNA markers

SINGLE LARGE-SCALE MARKER-ASSISTED SELECTION IN MAIZE

In order to fully capitalize on the power of DNA marker selection to the greatest effect, it will often be necessary to redesign breeding schemes. For example, a new approach for optimum combination of traditional and molecular breeding approaches has recently been proposed in maize breeding (Ribaut and Betran 1999). This suggests that a single large-scale marker-assisted selection could be carried out just once at the beginning of the breeding cycle. On this basis, progeny would be selected with a favourable fixed genetic background at specific loci throughout the genome while maintaining the maximum possible level of allelic variation across the remainder of the genome. As the aim is to carry out marker-assisted selection only once, large populations must be used to counter the large selection pressure thereby applied. Traditional breeding approaches may then be used in subsequent generations to select for adaptation to the local environment.

Simulation studies of a similar approach, although based on some over-simplified scenarios, provides some interesting insight into optimum number of locations, replications and population size in molecular breeding programs (Moreau et al. 2000). Most significantly, this simulation analysis suggests that even when GxE interactions are considered, it is optimum to perform only one replication per trial. Conversely, large population sizes are required for efficient marker-assisted selection. This suggests that it is more efficient to increase the sample size than to increase the accuracy of phenotypic evaluation. Similarly, the number of trial locations required is also low but for practical reasons of trial failure and environmental dependent character expression, it is not appropriate to follow this advice. A further extension of this is that QTL-marker validations are more valuable when carried out on a separate population than through extending the original population or evaluation system (Melchinger et al. 1998). In these simulation studies, marker-assisted approaches remained efficient for QTL with even very low heritabilities (0.15).

ADVANCED BACKCROSS BREEDING IN TOMATO

Tomato has represented the model system for molecular marker research and breeding since the advent of RFLP technology. Most recently, molecular breeding research in this crop has been used for the development of a new paradigm in development and application of marker-assisted selection systems. This approach involves the simultaneous discovery and transfer of important QTL from unadapted germplasm into elite breeding lines (Tanksley and Nelson 1996).

The advanced backcross approach has been successfully used to identify markers for QTL contributing to fruit size, shape, colour and firmness together with soluble solids and total yield. On this basis, QTL marker associations were identified in one backcross generation and immediately applied in the subsequent backcross generation some six months later (Tanksley et al. 1996). This is a critically important development as all the case studies detailed above only consider the cost-benefit of applying marker-assisted selection. However, the cost of

developing marker assays and the cost of identifying markers flanking loci contributing to the trait of interest can be very high.

Tomato researchers are also leading the way in terms of defining synteny relationships between broad leaf crops and the model plant species *Arabidopsis* (Ku et al. 2000). By defining these relationships it will be possible to utilize progress in the model system to quickly identify markers for genes of parallel function in the crop species of interest.

Conclusions from case studies

The molecular breeding projects presented above are from model systems where the application of molecular markers has been shown to offer real advantages over conventional approaches. The cost-benefit equation is particularly impressive for traits where traditional phenotyping is expensive and only possible late in the growth season. However, achieving new goals such as pyramiding disease resistance genes may offer even greater financial rewards through extending the life span of new varieties. Marker-accelerated backcross programs may also provide similar financial rewards through more rapid product deevelopment. Nevertheless, it is in the area of breeding complex traits, particularly those that are highly sensitive to environmental conditions, that molecular breeding promises greatest gains.

Clearly, cost-benefit analysis will need to be carried out on a case by case basis, as is already typically carried out by breeders when considering any component change in the program. However, the cost-benefit ratio for marker-assisted breeding increases rapidly as several distinct traits are simultaneously screened using multiplexed assays. More difficult to estimate is the value of reduced generations to varietal release and increased timeliness in the whole breeding program that marker-assisted approaches offer.

As DNA marker-assisted approaches become more widely integrated into plant breeding programs their impact will compound and their unit cost will decrease. Similarly, as functional genomics provides ever more detailed understanding of the systems that breeders are manipulating, DNA markers will provide the tools to develop a new knowledge-led paradigm in plant breeding. With increasing demand for these technologies, competition in the suppliers market place will force a reduction in equipment and reagent costs which will in turn fuel increasing adoption of these approaches. Already, the large multinational companies have justified the large capital expenditure for highly automated high throughput molecular marker screening systems, which they feel will ultimately return their investment several times over. However, traditional approaches will remain cost effective for many traits and replicated multilocational evaluation will always be a necessary precursor to varietal release.

Establishing a molecular breeding program

There are many levels at which companies can enter molecular breeding and many marker-assisted selection strategies which can have cost effective implications for small to medium sized companies. Except for large multinational seed companies, the adoption of new technologies such as DNA markers must progress slowly at first whilst important cost-benefit

decision are made and traditional breeding programs orientated towards the most effective use of this new source of information. For most small to medium-sized companies (SMC), contracting out DNA marker screening is a cost effective option at least during the initial phase of adopting these new techniques. This minimizes the burden of human and capital investment and offers a realistic insight into unit costs. For the purpose of this paper it is assumed that SMCs have an average annual turnover in the range of $1M to $50M and may thus have profits available for reinvestment in the range of $0.1M to $1M.

CONTRACTING MOLECULAR MARKER SERVICES

The cost of contracting small molecular breeding projects gives some indication of the operational funds required for this type of work. The following two examples present current costs for AFLP-based contract analysis at KeyGene.

Genetic distance analysis

10 assays across 90 genotypes	$13,000
Total time	3 months

Marker-accelerated backcross breeding

BC_1: 8 assays across 90 genotypes	$11,000
BC_2: 4 assays across 90 genotypes	$6,000
Total cost (small genome crops*)	$17,000
Total time	3 months
Total cost (large genome crops*)	$23,000
Total time	5 months

*crops with haploid chromosome number up to 12 (small genome) more than 12 (large genome)

Genetic distance analysis can assist in the selection of new germplasm to enter into the breeding program and for parental selection in F_1 hybrid variety breeding programs. Through the use of AFLP-based marker-accelerated backcross breeding, finished products can be generated in two to three generations (depending on crop genome size) as compared to five to six generations based on traditional approaches.

ESTABLISHING IN-HOUSE CAPACITY

With increasing emphasis upon DNA marker methods there will be tendency to look towards the development of in-house capacity as a means of reducing costs and increasing flexibility. Moreover, the importance of developing in-house expertise should not be underestimated in terms of the resultant synergies within the entire breeding system. In particularly, the maximum efficiency of DNA marker approaches is only realized upon the development of new breeding schemes based on a complement of traditional and novel techniques. It is difficult to make rapid progress in this critical developmental step without having some in-house expertise. Nevertheless, the optimum model for most small to medium-sized companies

will consist of a compliment of in-house projects, contracted services, cooperative sponsored projects and collaborations with public sector research labs.

For most SMCs, particular where basic technical support staff costs are low, the most appropriate entry to molecular breeding is through the establishment of a modest low technology manual system for basic PCR-based marker screening. The cost of equipping a basic PCR-marker screening lab. (excluding lab. structure) is in the range of $50,000 to $75,000. Although with some imagination considerable savings can be made at all levels (Harris 1998).

The capital cost of automated technologies can only be justified where staff costs are high or throughput level is high. In the latter case, automated technologies offer the potential of reducing reaction volumes and multiplexing assays thereby dramatically reducing unit costs. However, it is unlikely that capital investments in automated technologies can be justified for throughput scenarios below half a million samples per year. This critical threshold is clearly much higher than the determining level for developing in-house capacity for biochemical analyses for example. This is as much a function of the competitive rates offered by contract service companies as it is a function of the high capital investment required for automated technologies.

The future of molecular breeding is likely to rely on sequence-based markers. The large-scale development of this type of marker is highly demanding on human and capital resources. Laboratory-based methods for microsatellite marker development may cost up to $1,000 per marker. There is an alternative in crops where considerable amounts of sequence data have been placed in the public domain. In these cases in-house or contracted bioinformatics groups can screen the databases for suitable sequences for microsatellite marker development. In the more popular crops a large number of microsatellite markers are already in the public domain or available through commercial vendors. However, in lesser studied crops the development of such markers, in the short-term at least, is likely to remain beyond the realm of most SMCs. Marker development in public programs and through contracted projects funded by consortiums of seed companies is likely to remain the model for some time to come.

SCALING-UP TO HIGH-THROUGHPUT THROUGH AUTOMATION

It is now feasible to consider partial or complete automation of all steps of the DNA marker screening process. Currently, the automation of PCR product separation and data capture remains the most expensive intervention. Thus, there remains considerable interest in the development of plus-minus tests, which would eliminate the need for this step and thereby offer considerable reductions of unit costs for individual DNA marker interventions. However, in the longer term breeding programs are likely to want to develop DNA-assisted approaches for multiple traits. In this situation, electrophoresis of multiplexed assays will also offer dramatic reductions in unit costs. Highly automated high throughput marker screening facilities are likely to cost in excess of $500,000.

DNA Extraction
All post-DNA extraction steps are readily automated through multiples of 96-sample systems. Robotic systems are available for 96-well DNA extraction, which work well on bacterial and animal cells. However, DNA extraction from plant cells, particularly certain tropical crop species, presents additional problems. In particular, cell disruption to release DNA may need to be particularly violent in the case of plant preparations. With the development of relatively inexpensive 96-well grinding systems ($10,000), reasonably high throughput can be obtained.

PCR set-up
The development of robotic systems for liquid handling offers several advantages beyond rapid and 24-hour operation. In particular, robotics offer continuous consistency and the ability to reduce reaction volumes and thereby unit costs. Fully integrated systems capable of DNA quantification and equilibration followed by pipetting of PCR components cost around $100,000.

Plus-minus tests
These approaches rely on the development of PCR assays which only produce one amplification product from one genotype and no amplification product from the alternative genotype at that locus. PCR products can then be processed manually or assayed fluorometrically for rapid identification of desired genotypes. Complete automation of this approach can be achieved through the use of reporter systems coupled with fluorescence reading thermocyclers.

PCR product separation and data capture
Although automated gel-based fragment analysis/sequencers have been available for many years, this technology was neither truly high throughput nor largely hands-off. The development of automated multiple capillary electrophoresis systems has opened the door to 24-hour automation with minimal manual support. The most advanced of these systems can handle over 1000 samples per day generating 5000-20,000 data points depending on assay type and application. Costs range from $100,000 to $300,000 depending on throughput potential.

Data manipulation and presentation
The automated systems described above offer dramatic increases in the potential for data generation. This in turn shifts the rate-limiting step to the area of data manipulation, presentation and interpretation. Here computational methods will provide critical links in maintaining a flow of decision making based on the data generated.

Outlook for the future

New paradigms for the molecular breeding of complex characters
Many traits of agronomic and economic importance exhibit continuous variation due to an underlying array of polygenes termed quantitative trait loci (QTL). The rate-limiting factor for developing effective marker-assisted selections systems for these characters remains the confounding effects of genotype-by-environment interaction and epistasis. This has resulted in a rapidly evolving array of computational methods aimed at addressing this issue. To date

there is a wide range of publications in this area using simulation studies but very few based on empirical data. This is partly due to the very large population sizes that are required for effectively mapping complex traits. The development and widespread adoption of automated high throughput instrumentation will undoubtedly result in rapid advances in empirical developments for QTL mapping.

Marker-accelerated backcross approaches form the basis of the leading successes in molecular breeding of simple traits. However, this breeding scheme is not appropriate for improvement of complex characters. Beyond the theoretical problems involved in QTL mapping discussed above, new paradigms in molecular breeding will be required to effectively utilize QTL markers in crop breeding.

Rapid developments are taking place in the model crop systems fueled by automated technologies, advanced biometrics and functional genomics. In these systems, a knowledge-led breeding paradigm is evolving that will facilitate rapid progress in allied yet less studied crop systems. Developments in molecular breeding of maize over the past decade provide an indication of the future for all crops. At the beginning of the 1990's it was noted that there had been a six-fold increase in maize yields during the preceding 60 years, yet little was known about the actual genetic basis of these improvements. Now not only have many of the underlying characters been dissected and mapped but scientists are already reporting the knowledge-led manipulation of these components through molecular breeding.

Modern plant breeding is evolving into a highly complex multifaceted high technology business. In this new era, successful plant breeding programs will be characterized by dynamic, holistic approaches led by functional multidisciplinary teams. A high level of synergy between team members will be a vital element in product-led innovation and problem solving for commercial success. In particular the role of computational methods (including biometrics and bioinformatics) will be the leading force behind realizing the full potential of DNA marker-assisted approaches, as indeed it is for the entire genomics revolution.

Acknowledgements

The author gratefully acknowledges the assistance of David Hoisington, Jean-Marcel Ribaut and Michael Morris (CIMMYT), Rodomiro Ortiz, Tom Hash and Hutokshi Crouch (ICRISAT), Gurdev Khush and Glenn Gregorio (IRRI), Hei Leung and Casiana Vera Cruz (IRRI-Asian Rice Biotechnology Network), Johan Peleman and Jeroen Rouppe van der Voort (KeyGene), Jan Tamboer (Proteios), and, Yogesh Prasad (Labindia/Applied Biosystems) for their contributions and discussions during the development of this paper.

References

Arus, P., Moreno-Gonzalez J. 1993. Marker-assisted selection. In: Hatward M.D., Bosemark N.O. and Romagosa I. (eds.), Plant breeding, Principles and prospects. Chapman Hall, London. Pp. 314-331.

Dear, P.H. (ed.) 1997. Genome mapping, a practical approach. Oxford University Press.

Dreher, K., Morris M., Khairallah M., Ribaut J.-M., Pandey S., Srinivasan G. 2000. Is marker-assisted selection cost-effective compared to conventional plant breeding methods? The case of quality protein maize. In:

Proceedings Fourth Annual Conference of the International Consortium on Agricultural Biotechnology Research (ICABR), 'The economics of agricultural biotechnology' Ravello, Italy, 24-28 August 2000.

Harris, E. 1998. A low-cost approach to PCR. Oxford University Press.

Huang, N., Angeles E.R., Domingo J., Magpantay G., Singh S., Zhang G., Kumaravadivel N., Bennett J., Khush G.S. 1997. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95:313-320.

Ku, H.M., Vision T., Liu J., Tanksley S.D. 2000. Comparing sequenced segments of the tomato and Arabidopsis genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc. Natl. Acad. Sci. USA 97:9121-9126.

Melchinger, A.E., Utz H.F., Schon C.C. 1998. Mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383-403.

Mohan, M., Nair S., Bhagwat A., Krishna T.G., Yano M., Bhatia C.R., Sasaki T. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3:87-103.

Moreau, L., Lemarie S., Charcosset A., Gallais A. 2000. Economic efficiency on one cycle of marker-assisted selection. Crop Science 40:329-337.

Pandey, S., Rajatasereekul S. 1999. Economics of plant breeding: the value of shorter breeding cycles for rice in Northeast Thailand. Field Crops Res. 64:187-197.

Pastinen, T., Raitio M., Lindroos K., Tainola P., Peltonen L., Syvanen A.-C. 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Research 10:1031-1042.

Ribaut, J.-M., Jiang C., Gonzalez-de-Leon D., Edmeades G.O., Hoisington D.A. 1997. Identification of quantitative trait loci under drought conditions in tropical maize. 2. Yield components and marker-assisted selection strategies. Theor. Appl. Genet. 94:887-896.

Ribaut, J.-M., Hoisington D. 1998. Marker-assisted selection: new tools and strategies. Trends Plant Science 3:236-239.

Ribaut, J.-M., Betran J. 1999. Single large-scale marker-assisted selection (SLS-MAS). Molecular Breeding 5:531-541

Tanksley, S.D., 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. Theor. Appl. Genet. 92:191-203.

Tanksley, S.D., Grandillo S., Fulton T.M., Zamir D., Eshed Y., Petiard V., Lopez J., Beck-Bunn T. 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor. Appl. Genet. 92:213-224.

Witcombe, J.R., Hash C.T. 2000. Resistance gene deployment strategies in cereal hybrids using maker-assisted selection: gene pyramiding, three-way hybrids and synthetic parent populations. Euphytica 112:175-186.