I.1 The Principle: Identification and Application of Molecular Markers

P. Langridge and K. Chalmers[1]

1 Introduction

Plant breeding is based around the identification and utilisation of genetic variation. The breeder makes decisions at several key points in the process. First in deciding on the most appropriate parents to use for the initial cross or crosses and then in the selection strategy used in identifying the most desirable individuals amongst the progeny of the cross. The efficiency of the breeding and selection process can be assessed in many different ways including the ultimate success of the varieties released and the frequency with which new varieties are produced. A major cost and logistical issue in plant breeding are the actual number of lines that need to be carried through the evaluation and selection phases of a program. Large breeding programs for annual crops may carry hundreds of thousands of lines to produce a new variety only once every few years. Field trials can be expensive and evaluation of some traits, such as quality and yield stability can be expensive to assess. Molecular markers have proved to be a powerful tool in replacing bioassays and there are now many examples available to show the efficacy of such markers.

The use of molecular markers to track loci and genome regions in crop plants is now routinely applied in many breeding programs. The location of major loci is now known for many disease resistance genes, tolerances to abiotic stresses and quality traits. Improvements in marker screening techniques have also been important in facilitating the tracking of genes. For markers to be effective, they must be closely linked to the target locus and be able to detect polymorphisms in material likely to be used in a breeding program. The prime applications of markers in most breeding programs have been in backcross breeding where loci are tracked to eliminate specific genetic defects in elite germplasm, for the introgression of recessive traits and in the selection of lines with a genome make-up close to the recurrent parent. In progeny breeding, markers have proved valuable in building crucial parents and in enriching F_1s from complex crosses. Markers have also improved the strategies for gene deployment and enhanced the understanding of the genetic control of complex traits such as components of quality and broad adaptation.

[1] Australian Centre for Plant Functional Genomics, University of Adelaide. Waite Campus, Glen Osmond, SA 5064, Australia

Biotechnology in Agriculture and Forestry, Vol. 55
Molecular Marker Systems (ed. by H. Lörz and G. Wenzel)

2 Status

Recent developments that have occurred in molecular markers for many crop species have major implications for the future of the technology. There are three key components that are particularly significant. First, for many species, we now have markers closely linked to many traits of importance in the breeding programs. Indeed, for major crop species, we have markers for more loci than can be screened in a conventional breeding program. Second, we have tools that allow marker scanning of the whole genome. Of particular importance has been the development of microsatellite or SSR markers that now form the basis for analysis and allow highly multiplexed SSR screens. This trend will continue as newer, cheaper marker screening based on SNPs become available. The technological advances have improved our capacity for whole genome screens. Third, through association mapping projects we have, or are in the process of developing, whole genome fingerprints for many key lines and varieties of importance in breeding programs. These studies are developing large databases of historic germplasm that should, over the next few years, start to reveal the ways in which breeding programs have selected for and against specific regions of the genome. We can see these developments, particularly in crops such as maize and barley, where markers for most of the major disease resistance clusters, for key components of feed or processing quality and for many loci conditioning tolerance to abiotic stresses are available.

The new marker systems have several important implications for the future of marker-assisted selection (MAS) and breeding strategies in general. Existing strategies for MAS were initiated with a view of markers as providing a rapid and cheap alternative to bioassays and they have largely been used in this role. While highly successful, this strategy does not fully exploit the technology. The key limitation to an expansion of the scale and complexity of marker use is the size of the populations that would be required if one were to try and select for alleles at a large number of loci simultaneously. A further important feature of recent advances has been related to how we best take advantage of the genome information that has been generated for major crop species. We know, for example that chromosome 2H in barley and group 7 chromosomes of wheat, carry clusters of genes, often in repulsion that we would like to break up. Again conventional use of markers has not been very effective in utilising such genome regions. Conversely, we know that there are some chromosomes where there is little allelic variation between lines and it is a waste of effort to try and break these up in a breeding program.

The key challenge of new work is to investigate strategies for whole genome breeding: how we can use genome-wide information in the form of graphical genotypes and known locations of key loci and marker tags for both desirable and undesirable alleles, to design optimal breeding strategies that integrate as much of the available information as possible.

3 Molecular Markers

Molecular markers have been taken, in recent years, to refer to assays that allow the detection of specific sequence differences between two or more individuals. However, it should be recognized that isoenzyme and other protein-based marker systems also represent molecular markers and were in wide use long before DNA markers became popular. One of the earliest type of DNA-based molecular markers, restriction fragment length polymorphisms (RFLPs), were based around the detection of variation in restriction fragment length detected by Southern hybridisation. The types of sequence variation detected by this procedure could be caused by single base changes that led to the creation or removal of a restriction endonuclease recognition site or through insertions or deletions of sufficient size to lead to a detectable shift in fragment size. This technique has been largely superceded by microsatellite or simple-sequence repeat (SSR) markers and is now rarely used in screening material for breeding programs, but it remains an important research tool. SSR markers detect variation in the number of short repeat sequences, usually two or three base repeats. The number of such repeat units has been found to change at a high frequency and allows the detection of multiple alleles. The large expansion of DNA, particularly EST, sequence databases has now opened the opportunity for the identification of single nucleotide polymorphisms, SNPs. These occur at varying frequencies depending on the species and genome region being considered. In Arabidopsis SNP frequencies of 0.007–0.0104 have been measured (Kawabe et al. 1997; Purugganan and Suddith 1998) while in maize a range of 0.00047–0.0037 has been measured (Hilton and Gaut 1998; Wang et al. 1999). SNPs are widely seen as providing the key advantage of multiple detection systems many of which, such as mass spectroscopy, offer high throughput at low detection cost. Importantly, new array based screening methods, such as DArT (Jaccoud et al. 2001) appear to offer still cheaper assays due to their very high multiplexing capability. Interestingly, molecular markers may be coming full circle with protein markers again being proposed as viable genetic markers for MAS. Mass spectrometric methods for mass fingerprinting of proteins and for the analysis of low molecular weight proteins, again opens the option for high throughput protein screening. In these cases, single amino acid changes in protein sequence can often be detected and this provides a means for revealing variation in the corresponding DNA coding sequence.

In each method, DNA sequence variation is being detected. However, each method analyses different aspects of DNA sequence variation and different regions of the genome. For example, RFLPs were detected using cDNA clones, namely coding sequence, but frequently detected variation that lay in regions flanking the genes. SSR markers have generally been from non-coding regions although the recent move to three-base repeats and the use of ESTs as the source of SSR markers is changing this. Other markers such as RAPD and

AFLP markers appear to frequently target repetitive regions of the genome. The stability of the sequence difference may also be an issue in some cases. SSRs are seen as being too unstable for some applications since the mutation rate may in some cases be high.

The decision about the most appropriate marker system to use will vary greatly depending on the species, the objective of the marker work and resources available.

4 Identifying Marker/Trait Associations

The most widely used methods for identifying marker/trait associations are based around the construction, phenotyping and genotyping, with molecular markers, of special populations. The steps in identifying marker/trait associations and developing the markers through to application are summarised in Table 1. The populations are generally constructed from two varieties that show a major difference in the traits targeted for mapping. The genetic structure of the segregating populations can be immortalised by producing double haploids or recombinant inbred lines. The populations produced then become a major resource for a wide range of studies. Many such populations have become international resources used by researchers around the world. The ITMI population used for wheat research is an example of this. The population made from a cross between the wheat variety Opata 85 × W7984 a synthetic wheat, has become the international reference for wheat genetic research (Langridge et al. 2001). New markers, such as SSR and SNP, are being placed on the population continually and the population has been screened for a wide range of disease, abiotic stress tolerance, physiological and quality traits. The beauty of these populations is that they continue to grow in value as they are more and more widely used. Such reference populations are now available for several crop and model plant species.

However, there are also problems with the use of such structured populations. Many of the reference populations were constructed to facilitate marker screening and were based on highly diverse parents, this was the case for the ITMI reference population

There are three important issues that will frequently impact on the most appropriate procedure to be used in finding marker/trait associations:

- There is a major cost in phenotyping. This clearly varies depending on the trait being analysed, but usually the more complex and expensive the phenotypic screening is, the more valuable will be markers for the trait. Costs of phenotyping can be particularly important for traits that require extensive field trials, such as yield or tolerance to some stresses, or require large amounts of material for analysis, such as malting quality in barley or animal feeding trials. Due to costs, the number of replicates and sites is often limited, reducing the sensitivity of some of the analyses.

Table 1. Steps in identifying marker/trait association

1. Defining the target
Decision about marker development
- Is the trait of importance to breeding program or to biological research?
- Is a molecular marker needed?
  - What is the cost of the bioassay relative to marker assay?
  - Is the trait dominant versus recessive? – recessive traits may be hard to identify in a bioassay and will be a prime target for marker development
  - Perhaps there is no alternative to marker use:
    - Quarantine trait – e.g., resistance to a disease not present in the country
    - Pyramiding resistances – accumulating multiple genes for resistance to protect against resistance breakdown
    - Map-based cloning of genes – high resolution map is needed to minimize region that needs to assessed
    - Gene deployment – where desirable alleles are available to several loci, but only one is really needed. How does one decide on the best one to use?

2. Identify germplasm for marker development
Available germplasm, with and without the trait

3. Population structures
Deciding on the best material to use for identifying the marker trait associations
- Knowledge of genetics
  - Is the trait simply inherited or multigenic?
  - What is the heritability?
  - If this information is not available, a trial experiment may be needed

  A simple cross can be constructed to measure segregation ratio and heritability
  Complex traits and traits of low heritability are often prime targets for marker development as they are hard to assay otherwise
- Decide on best population structure
  The structure of the population will be related to the trait and purpose
  Populations structure will differ between:
  - in-bred versus out-breeding species
  - long generation versus short generation plants
  - perennial versus annual plants
  - Doubled haploids – one meiotic event per line
  - F_2s – two-meiotic events per plant
  - Recombinant inbreds or single seed descent
  - Complex crosses between highly heterozygous parents

  Population size
  - For single gene 50 F_2s may be adequate
  - Map-based cloning over 1000 required

  Is an existing population already available?
  - Screen parents of existing crosses and mapped populations

4. Phenotypic evaluation
- Is phenotypic evaluation possible for single plants?
- For some traits a large number of seeds or plants may be required or field trials at multiple sites, e.g., quality and yield and traits of low heritability
- For association mapping phenotypic information can be collected from existing programs or lines pooled that have a common phenotypes, e.g., lines adapted to a common environment or of common quality ranking

Table 1. (Continue)

5. Genotyping
- Identify markers that detect polymorphisms between parents
- Screen population

Marker density will depend on objective
- Full maps
  Screen with sufficient markers to give good genome coverage, usually around 1 marker every 10 cM
- Bulked segregant analysis
  Bulk two extremes of phenotype and screen with markers
- Map-based cloning
  - high resolution in small region
  - Screen population with markers flanking target region
  - Identify recombination events between the flanking markers
  - Use the recombinant lines for the high-density marker screening
    Note: only the recombinant lines need to be phenotyped
- Association mapping
  Genotype multiple lines, for example in pedigree

6. Identifying marker trait associations
- Full maps
  Construct linkage map based around molecular markers
  Locate trait loci by regression analysis, interval mapping or related technique
- Bulked segregant analysis
  Screen pools of lines with and without trait with molecular markers
- Association mapping
  Measure rate of linkage disequilibrium between traits and markers
  Deviations from the expected frequency of alleles

7. Developing markers for application – Marker validation
- Test marker trait association in alternative populations and estimate reliability of marker in predicting phenotype
- Identify polymorphisms between lines used in breeding program
- Develop a palette of suitable markers with associated polymorphism data
  Usually around 10 markers within 10 cM of trait
  Provide protocols and polymorphism data to breeding programs

- The lines (varieties) used to construct the populations are often out-of-date by the time the marker/trait information is available. For example, many mapping programs are using populations constructed over a decade ago. This reduces the value of the information gathered and slows its implementation into active breeding programs.
- The structure of the populations limits the types of traits that can be mapped and many of the subtleties of adaptation can only be analysed with special populations.
- Generally, mapping is restricted to known traits for which a well-defined bioassay is available.

4.1 Full Linkage Maps

Complete linkage maps generated from screening the progeny of a cross have provided the basis for most early marker development work. However, this is difficult and labour-intensive, particularly in species with large numbers of linkage groups such wheat, where linkage maps must be constructed for 21 chromosomes or where chromosome numbers are large or variable, such as kiwi fruit (*Actinidia* sp.) or sugarcane. Usually, 10 to 20 markers are required for each chromosome to give reasonable genome coverage. For many species where the germplasm base is small, the level of polymorphism may be low so a large number of markers are required to detect sufficient polymorphisms for mapping. The key feature of such maps is that they provide considerable information on the genome structure of an organism and provide major resource for researchers even though their application to practical plant breeding is now becoming increasingly limited. As mentioned above, once established, a well-mapped population can be used for a wide range of genetic studies provided the individual lines can be maintained.

4.2 Bulked Segregant Analysis

Bulked segregant analysis (BSA) was described by Michelmore et al. (1991). It involves pooling individuals from the two phenotypic extremes of a segregating F_2, doubled haploid or similar population. DNA isolated from the two pools is then screened with DNA markers, usually SSR or AFLP, and polymorphic bands identified. Clear polymorphisms seen between the two pools will be derived from regions of the genome that are common between the individuals that made up the pools. The remainder of the genome will be randomly contributed by the parents and should show no polymorphisms between the pools.

This technique offers the important advantage of identifying markers associated with the trait without the need for full map construction. However, it requires markers that can be easily screened in mixed DNA preparation. This essentially means PCR-based marker assays. AFLP markers have proved particularly suitable for these assays. The method also allows the use of a wide variety of population structures and can often be applied to material produced within a breeding program.

The key disadvantage is that one is not provided with a genetic distance between the marker and the trait. Indeed, it is always necessary to confirm the marker/trait association by screening individuals from the population to confirm that the marker is a reliable predictor of the trait. This can usually be done by taking the individual plants used to construct the bulks and determining just how many of the lines actually show the expected marker pattern.

BSA is now widely used and in many marker development programs has become the major method for marker identification.

4.3 Association Mapping

Association mapping is becoming an increasingly important tool for marker analysis and application. Particular emphasis will be placed on this technique here since we believe it will replace alternative marker development procedures, will greatly facilitate marker delivery to breeding programs and will provide valuable insights into the genetics and evolution of crop plants.

Molecular markers offer an easily quantifiable measure of genetic variation within crop species. However, many species, such as wheat, display a low level of polymorphism (Chao et al. 1989; Lui et al. 1990) hampering the identification of markers linked to agronomically important traits and complicating the differentiation of varieties and the analysis of genetic variability.

While the low level of polymorphism may be problematic, the level of genome conservation observed between varieties offers an opportunity to identify markers associated with traits of interest (Paull et al. 1998). The rationale of the approach is based on 'linkage drag' (Hanson 1959; Stam and Zeven 1981), a feature of chromosome behaviour whereby flanking DNA surrounding the target gene diminishes at a much slower rate than unlinked regions. Since varieties of a particular crop species are often closely related, differences between related accessions may reflect differences in important agronomic traits. A comparison of the molecular marker profiles of accessions demonstrating a particular trait with those lacking the trait, facilitates the identification of linked markers. Loci linked to the trait of interest will show the same marker phenotype within each group while unlinked loci will show a random distribution of marker alleles. This approach is analogous to bulk segregant analysis (Michelmore et al. 1991), but is dependent on traits remaining as part of a larger linkage block during crossing and selection.

A further problem with the identification of marker/trait associations using defined populations such as F_2 populations, doubled haploids derived from F_1s or recombinant inbreds, is that they are usually built from only two parents and often do not reflect germplasm in an active breeding program. By the time the populations are ready for mapping they often involve germplasm that is out of date and no longer optimal for a pragmatic breeding program.

These limitations in existing mapping strategies can be addressed through association or linkage disequilibrium (LD) mapping. LD mapping is based on seeking associations between phenotype and allele frequencies. It is the basis for gene mapping in species where large mapping populations cannot be readily produced such as mapping in farm animals and humans. There are three advantages of this approach in mapping in crop species. Firstly, it provides a new perspective for trait mapping. This is because it uses different population structures (based largely around pedigree) and it uses a different set of phenotypic data. Consequently, we can expect to see new marker/trait associations, but more importantly, this technique will help identify targets for more detailed analysis. For example, we can expect to find genetic associ-

ations with lines or varieties that have performed particularly well at certain sites. In many cases, we will not be able to associate this with a specific aspect of the site environment, but we will have an indication of where to look for an environmental factor.

Secondly, LD mapping also provides detailed fingerprinting information on a large number of lines and varieties and this information will be valuable in several of the breeding strategies outlined below.

Thirdly, the LD method uses real breeding populations, the material is diverse and relevant and the most important genes (for example, for adaptation) should be segregating in such populations. The breeder is also integrally involved in the process and this may lead to improved rate and efficiency of validation and adoption. Many breeding programs are reluctant to grow and assess a huge number of lines with little or no potential for a direct commercial outcome. The advantage of LD mapping to the breeder is that mapping and commercial variety development can be conducted simultaneously.

When a novel mutation occurs at a locus determining the expression of a QTL, all other alleles of that locus are considered to be in complete linkage disequilibrium with the new mutant. However, as time goes by, the level of observed linkage disequilibrium will deteriorate as recombination between the mutant allele and other loci occurs. The level of LD observed will be in direct relationship to the distance from the mutant allele and also a function of the number of generations that have passed since the original mutation event. Several factors will also mediate this effect. In the case of outcrossing species there will be a relatively rapid breakdown of LD and in outcrossing species it is expected that LD will only be detected over a relatively short distance that may be measured in the region of a few tens of kb, although the actual extent of LD in maize still appears controversial (Remington et al. 2001; Ching et al. 2002). In these species it may be necessary to use direct DNA sequencing to identify and track linkage disequilibrium. However, for in-breeding species, LD breaks down relatively more slowly and is clearly detectable at the centiMorgan level. This offers the opportunity to use more conventional marker-based assay detection systems such as SSRs as an appropriate detection system. The existence of detectable levels of LD in in-breeding populations then offers the possibility of carrying out association mapping. This may be done by systematically screening molecular marker loci at defined intervals across the genome. In human and animal systems, this has readily been achieved using the abundant SNP resources available. However, in most crop species, such a resource is unlikely to become available in the near future. The question then is will the available marker systems, in particular SSRs provide sufficient genome coverage? This question is currently being addressed for several species and the results should start appearing in publications over the next few years.

5 Application of Molecular Markers

The following section aims to provide an overview of the current and predicted applications of molecular markers within plant breeding programs. The application of molecular markers to pedigree/progeny and backcross breeding is summarised in Table 2 for each stage in the breeding process. Although the focus here is on application of markers to practical breeding programs, it is important to remember that molecular markers have also become a critical tool for a wide range of genetic studies and are widely used in many aspects of genetic research from map-based cloning of genes through to the study of genome structure, organisation and behaviour.

5.1 Trait-Based Selection

One of the earliest demonstrations of the power of molecular markers was provided by Beckman and Soller (1986) for the indirect selection of genes in a breeding program. Molecular markers offer several key advantages over many bioassay systems:

- DNA can be extracted from tissue sampled from growing plants at very early stages of development. This allows sufficient time to use linked markers to identify heterozygous, backcross or topcross F_1 individuals prior to anthesis and further crossing. In contrast, optimum expression of many important phenotypes in a bio-assay, such as disease resistances, often occurs at development stages close to or after anthesis when crossing should take place. Plants in the bio-assay system are also frequently grown under suboptimum conditions for crossing and seed set.
- Markers can be used accurately on a single plant. For most bio-assays, many individual plants must usually be screened although this will vary with the bioassay. It is rare that a single plant assay will give a reliable phenotypic assessment. In contrast, the proportion of single plants incorrectly scored from a molecular marker assay is related to the closeness of linkage between the marker and the trait. For markers within 10 cM, the error is therefore less than 10%. The confidence level of single plant selection would increase even further with the use of flanking markers (Beckmann and Soler 1986).

An example of this application can be seen in the South Australian Barley Improvement Program. In a breeding strategy commencing in the spring of 1994, a single dominant resistance gene to cereal cyst nematode (*Ha2*) was transferred from a resistant to susceptible variety through three cycles of marker-assisted backcrossing using a single molecular marker. At no stage were more than four BC_xF_1 plants backcrossed. One hundred and twenty doubled haploid lines were produced from marker selected BC_3F_1 and 66% of the

Table 2. Role of marker in pedigree/progeny and backcross breeding for inbred annual crops

Phase	Pedigree/progeny		Backcrossing		Marker roles
Parental choice	A×B				Understand the genetic relationships between current and future members of the germplasm pool. In hybrids, estimate likely heterosis through diversity analysis
			A×B		Choose parents with small genetic distance to reduce number of backcrosses. Develop strategy for introgressing and selection of "quarantine traits" Transfer transgenes into elite lines
Crossing			BC1	F_1×A	Identify progeny
			BC2	F_1×A	• Carrier of trait
			BCn	F_1×A	• Low percentage donor genome • Small introgression segment
					Save one generation per backcross for recessive traits
	A×B				Enrichment of F_1s – markers used to characterise germplasm
	(A×B)×C				Enrichment of F_1s – markers used to characterise germplasm
	(A×B)×(C×D)				Enables more complex crossing strategies
Segregating generations	F_2 to F_n				High throughput markers required to select desired alleles Markers used to identify desired progeny in parent building schemes
			Fixation		Use markers to choose lines close to recurrent parent to "fast-track" line to release by reducing evaluation requirements as before
Evaluation of fixed lines	Year 1	Limited sites, limited replications			Limited role until QTL for adaptation/quality validated
	Year 2	More sites, more replications			
	Year 3	Regional trials			
	Year 4	National list year 1			
	Year 5	National list year 2			
			Year 1 limited sites and seasons Year 2 regional trials Year 3 national list		Whole genome marker analysis can identify individuals close to recurrent parent, thereby saving expensive yield and quality testing
Pure seed					Random genome survey plus key economic traits Markers used to compare reselections for bulking
Commercialisation					Use markers to provide evidence in "essentially derived" discussions Markers used to identify or compare new varieties against other varieties of "common knowledge"

regenerants were classed as resistant by phenotypic assay. A line selected from among these was commercially released 7 years from the time of the first cross. This compares to 14 years for a conventional breeding strategy for malting barley.

5.1.1 Enrichment of Complex F_1s

Plant breeders often use three- and four-way crosses since they allow an increase in the range of traits that can be simultaneously incorporated into elite progeny. However, the frequency of elite progeny from this type of cross is usually very low and this has reduced the application of this strategy. Breeders have tended to take the longer route of making simple crosses, fixing desirable alleles and then intercrossing selected, fixed lines. MAS offers a powerful alternative to increase the desirable allele frequency for each locus contributed from a quarter parent from 25% of progeny to 50% by screening the top cross F_1 or four-way cross F_1. This application has become the most common application of molecular markers in both wheat and barley breeding in southern Australia and its application continues to grow. Importantly, it is expected to increase in power through the development of new whole genome screening and selection strategies.

5.1.2 Early Generation Selection

The selection theory required to implement MAS in early generations is similar to other forms of selection although MAS is closer to 'simultaneous' rather than 'tandem' (or stepwise) selection, which is often a feature of early generation, phenotypic selection. In early generations breeders usually visually select traits of high heritability since complex traits such as yield cannot be effectively selected in rows or small plots. However, MAS is more effective than phenotypic selection when population sizes are large and heritability is low (Lande and Thompson 1990; Whittaker et al. 1997). Therefore, breeders and geneticists need to design and implement MAS strategies that allow selection of complex traits in early generations. The combination of MAS with techniques such as single seed descent and doubled haploid generation offers the option to address some of these difficulties. The testing of these strategies is now well advanced in many breeding programs, but assessment of the full impact of this strategy will, as with gene enrichment, only be possible in later generations. However, the efficient handling and management of single plants, a pre-requisite for single seed descent (and doubled haploid production), lend them to MAS.

5.2 Whole Genome Selection

5.2.1 Choice of Donor Parent in Backcrossing

Marker-based genetic diversity studies have generated large data sets that can be used to select from a number of possible donor parents for a desired trait. The objective is to identify the donor parent that is the minimum genetic distance from the recurrent parent. This should reduce the number of backcrosses required to recover the recurrent parent phenotype. Genetic distance estimates may also assist in assessing the suitability of prospective donor parents of unknown or diverse pedigree or where limited phenotypic information is available. Information gained from the routine DNA fingerprinting of potential parents could contribute to the broader information base used for selection of suitable donor parents including the selection of more diverse parents for speculative crosses.

5.2.2 Recovery of Recurrent Parent Genotype in Backcrossing

The key objective of a backcrossing strategy is to reduce the proportion of the donor parent genome by about 50 % at each generation of backcrossing. Until recently, most backcrossing has focused on this principle and ignored the variation in the proportion of the donor parent genome that exists around the expected mean. Molecular markers allow selection for the desired donor allele, and also for recombinant individuals that have a genome composition closer to the recurrent parent than would be predicted from theoretical expectations (Tanksley and Rick 1980). MAS against donor parent or for recurrent parent genome, provides a means of reducing both the time and number of generations required to adequately recover the recurrent parent genotype.

This strategy has been applied in the South Australian wheat and barley breeding programs. In general, BC_1 derived lines have been identified that carried a proportion of donor parent genome not significantly different from the mean of the BC_3 generation. Selecting these individuals saves two cycles of backcrossing. Similar results were found in simulation studies conducted by Hospital et al. (1992), Visscher et al. (1996), and Frisch et al. (1999) where they also found that at least two generations of backcrossing could be saved.

A major constraint to the adoption of recurrent parent background selection in a practical backcrossing program is the large number of polymorphic marker alleles required to cover the entire genome. New marker developments, such as DArT and SNP-based markers will help address this limitation. In their simulation study, Frisch et al. (1999) showed that the use of marker loci of known location was more efficient than random marker alleles. This means that markers that have not been localised to the genome, such as AFLP, are of less value in this strategy than SSR markers or DNA micro-array technologies.

Stam and Zevens (1981) estimated that the typical segment length of donor parent DNA retained after three backcrosses was surprisingly high at 51 cM in a 100-cM chromosome. There are now many examples of deleterious linkage drag in plant breeding to support this theoretically derived conclusion particularly for alien segment (Paull et al. 1994). Therefore, a more focused selection strategy may be more useful. The methods would be as follows:

- use flanking markers at 10–20 cM around the estimated position of the gene, to maintain the donor allele frequency during later generations of backcrossing
- use more distant (30–40 cM) flanking markers to select for small donor segment around the desired gene
- use carefully selected markers spaced 30–40 cM over the remainder of the genome to select against donor parent alleles.

5.2.3 Linkage Block Analysis and Selection

It has been clear to most breeders that certain chromosomal regions carry key clusters of genes which have been highly conserved through selection. The high degree of success of conservative breeding strategies around the world in the past and the apparent poor combining ability found between many germplasm pools suggests that major linkage blocks may also be important to breeding programs for most major crop species. For example, a study in barley indicated that conserved regions of the genome derived from a North African landrace introduction in the early 1970s appears to have been crucial for the improved adaptation of successful South Australian varieties released from that time. Regions on chromosome 2H appeared to be particularly strongly associated with improved adaptation. This region also showed significant associations with grain yield, grain yield stability, grain size and flowering time (all across a number of environments; Atmodjo et al., pers. comm.). The syntenous region in wheat also appears to carry a block of genes associated with adaptation including the photoperiod sensitivity genes, *Ppd1, Ppd2*, and *Ppd3* on chromosomes 2D, 2B and 2A respectively (Borner et al. 1998) and QTL associated with flag leaf area, flag leaf length, leaf width, plant height at stem elongation and anthesis, number of heads per square meter, and grain weight (Coleman et al. 2001). This type of information could lead to the development of specific targeted breeding strategies, where either allelic variation in these regions is actively sought or linkage blocks from superior adapted genotypes are actively conserved through marker assisted selection.

5.2.4 Key Recombination Events

The presence of major linkage blocks containing groups of genes that have been important for adaptation has several implications for breeding programs. For most species recombination is not evenly spread across the genome nor are target traits evenly distributed. This logically leads to the suggestion that specific recombinational events may be important in making major advances in breeding. This is particularly important for species where wild relatives are used as sources for new alleles, such as disease resistances. For example, the wild relatives of wheat have proved valuable sources of useful traits yet these alien segments have shown very low recombination with wheat even when derived from a close relative. Molecular markers are valuable in identifying the rare lines where recombination has occurred and in characterising the recombination events (Langridge et al. 2001). A major benefit arising from the various mapping initiatives has been the increased knowledge of the structure and behaviour of the genomes of crop species and the physical and genetic control of important traits. In very well studied species, such as barley, a comprehensive understanding of the genome has facilitated the design and recent adoption of complex crossing and key recombination event selection strategies using markers. As our understanding of the genetic control of important traits in crop species improves, so will the potential for us to apply specific targeted marker-assisted breeding strategies to compliment the broader, traditional approaches.

6 Directions

6.1 Quantitative Trait Loci

The agronomic performance of crop varieties is mainly influenced by complex quantitative traits, for example, components of yield and quality. Since the development of molecular markers, it has become feasible to identify and genetically localise the contributing genetic factors as quantitative trait loci (QTLs) and to utilise these QTLs for crop improvement. This has led to an increasing number of QTL studies, involving most agronomically important crop species.

Despite successes in mapping QTLs, the relevance of this information for breeding new varieties is limited. In most cases, the QTL analysis has been carried out in crosses utilising parents drawn from elite germplasm sources. Hence, in most cases the studies have been able to identify a limited number of alleles that are already present in the mainstream breeding programs and offer little opportunity for variety improvement.

6.2 Diversity

The risk most frequently raised for MAS is the temptation to use only parents for which either markers and/or polymorphic markers exist, thus further narrowing genetic diversity within breeding programs. In particular, this may concentrate the use of a few, well-characterised disease resistance genes to the exclusion of less well documented sources. This risk can be minimised by breeder discretion allocating a proportion of the program to 'new' or 'uncharacterised' sources as has already happened in some breeding programs. It might be more useful to think of this problem as a challenge of how marker technology be used to expand the useful gene pool.

The primary gene pools of many crop plants are so depleted in genetic variability that breeders have relied upon wild relatives for sources of disease resistance and other traits. Although crop germplasm collections contain many thousands of potentially useful wild accessions, their utilisation is sometimes hindered by hybridisation barriers preventing interspecific crosses and/or by undesirable characteristics inherent in exotic germplasm. Breeders have used exotic germplasm almost exclusively as a source of major genes for disease and insect resistances, and have mostly relied on repeated intercrossing of adapted elite genotypes for the improvement of quantitative traits, like yield and quality.

Tanksley and Nelson (1996) presented a novel alternative to the limitations inherent in conventional approaches of utilising exotic germplasm. By combining the introgression of novel QTL alleles from exotic sources of germplasm with QTL analysis and discovery, they have been able to demonstrate significant variation in the expression of a number of agronomically important traits. This procedure has been termed advanced-backcross-QTL analysis (AB-QTL). This utilises exotic germplasm as the genetic donor for the improvement of quantitative agronomic traits and combines marker and phenotype analysis in advanced backcross generations such as BC_2, BC_3 or more recently, crosses such as BC_2F_2.

To date, several reports on the application of the AB-QTL strategy are available for tomato, rice, wheat and barley. In each case, favourable exotic QTL alleles for important agronomic traits have been identified. It is proposed that the introgression of new exotic QTL alleles, will contribute to an increased level of genetic diversity in a range of cultivated species. An example of this has been demonstrated in fruit yield in cultivated tomato. Through the introgression of wild-species alleles from *Lycopersicon pimpinellifolium* and *L. peruvianum*, fruit yield was increased by up to 17 and 34%, respectively (Tanksley and Nelson 1996; Fulton et al. 1997). Similar results were seen from AB-QTL studies in rice. In this case, two wild-species QTL alleles have been associated with an increase of yield by 17 and 18% on rice chromosomes 1 and 11, respectively (Xiao et al. 1996, 1998). More recently, *Hordeum spontaneum* has been used as a source of novel alleles in barley and in wheat, synthetic hexaploids created from *Triticum dicoccides* and *Aegilops taushii*. In

both examples, alleles from the exotic germplasm were associated with a positive effect on agronomic traits.

An alternative approach to that of AB-QTL for utilising genetic variation within exotic germplasm is to attempt to identify alternative alleles by undertaking germplasm screens. The use of wild barley (*Hordeum vulgare* ssp. – *spontaneum*) as a source of novel malting quality alleles has been reported by Eglinton et al. (1998). One hundred and fifty four accessions of *H. spontaneum* were screened for β-amylase polymorphism and three novel alleles (*Bmy*1-Sd3, -Sd4, and -Sd5) were identified in addition to those detected in cultivated barley. The corresponding Sd4 and Sd5 enzymes exhibit intermediate levels of thermostability, similar to the Sd1 β-amylase. The Sd3 β-amylase from wild barley exhibits thermostability significantly greater than the other five allelic forms of β-amylase and as such provides for improved fermentability during processing.

6.3 Whole Genome Breeding

As a result of work carried out on many crop species over the past decade, markers are becoming available for a large number of important traits. We have good strategies in place to use these markers to accelerate a number of breeding techniques, in particular backcrossing. However, we do not have strategies to manage the introgression of more than about five loci simultaneously, whether by direct crossing or by merging crossing streams. A major problem is that our existing breeding and marker implementation strategies are based around the selection and monitoring of individual loci. However, we now have high-throughput marker screening techniques that allow us to monitor the entire genome. Indeed, since the major cost of marker screening is the DNA isolation, the more marker information that is gathered for each line, the greater the cost/benefit ratio. Can we move from a trait-based selection system to a recombination-based strategy where we manage the entire genome and select individuals with a particular genome configuration (based on recombination events)? This approach would offer major gains in the efficiency of breeding programs because it would allow a dramatic reduction in the sizes of populations needed to achieve a specific outcome. However, to do this effectively we need to understand and analyse the behaviour of the entire genome in a breeding program and the major genomic events that have led to adaptation to our environment.

7 Conclusions

Molecular markers are now well established as tools in plant breeding and genetics. They have also provided a major new impetus to plant breeding programs offering considerable improvements in the efficiency and sophistication of breeding. Their use as research tools is also well developed and they have played a key role in improving our understanding of genome organisation, structure and behaviour for many of our major crops. However, the application of molecular markers in practical plant breeding has been patchy. Marker resources and capabilities for marker implementation are largely unavailable for minor crops. Even for some of the major crops, such as wheat and rice, markers are not widely used in public breeding programs with a few notable exceptions. Given their huge potential, this slow acceptance and implementation is disappointing and is probably related to a lack of flexibility by many public breeding programs to make the structural and strategic changes needed for effective marker implementation. It may also be partly due to the lack of active participation of breeders in the marker development programs in some countries. As the results of marker application become more apparent and move from theory to released varieties, this attitude may change.

The key developmental challenges for molecular markers now lies in developing new breeding strategies where the objectives will be increasing the germplasm base and increasing the number of traits that can be effectively selected simultaneously. The new marker technologies that offer greatly reduced costs in marker screening and high multiplexing capabilities will be central to these developments. Essentially we will move to whole genome-based selection strategies where specific recombinational events are sought and changes will be assessed on a genome-wide scale. In this way we can look to better manage chromosome regions that may come from wild relatives or land races, track several traits at once and keep the population sizes as small as possible.

References

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms in plant genetic improvement of agricultural species. Euphytica 35:111–124

Borner A, Korzun V, Worland AJ (1998) Comparative genetic mapping of loci affecting plant height and development in cereals. Euphytica 100:245–248

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic linkage maps of wheat homologous group 7 chromosomes. Theor Appl Genet 78:495–504

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey SV, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 7:3–19

Coleman RK, Gill GS, Rebetzke GJ (2001) Identification of quantitative trait loci for traits conferring weed competitiveness in wheat (*Triticum aestivum*). Aust J Agric Res 52:1235–1246
Eglinton JK, Langridge P, Evans DE (1998) Thermostability variation in alleles of barley *beta*-amylase. J Cereal Sci 28:301–309
Frisch M, Bohm M, Melchinger AE (1999) Comparison of selection strategies for marker assisted backcrossing of a gene. Crop Sci 39:1295–1301
Fulton TM, Nelson JC, Tanksley SD (1997) Introgression and DNA marker analysis of *Lycopersicum peruvianum*, a wild relative of the cultivated tomato into *Lycopersicum esculentum*, followed through three successive backcross generations. Theor Appl Genet 95:895–902
Hanson WD (1959) Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing and selfing. Genetics 44:833–837
Hilton H, Gaut BS (1998) Speciation and domestication in maize and its wild relatives: evidence from the globulin-1 gene. Genetics 150:863–872
Hospital F, Chevalet C, Mulsant P (1992) Using markers in gene introgression breeding programs. Genetics 132:1199–1210
Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29(4):E25
Kawabe A, Innan H, Terauchi R, Miyashita NT (1997) Nucleotide polymorphism in the acid chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Mol Biol Evol 14:1303–1315
Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756
Langridge P, Lagudah E, Holton T, Appels R, Sharp P, Chalmers K. (2001) Trends in genetic and genome analyses in wheat: a review. Aust J Agric Res 52:1043–1077
Lui YG, Mori N, Tsunewaki K (1990) Restriction fragment length polymorphism (RFLP) analysis in wheat. I. Genomic DNA library construction and RFLP analysis in common wheat. Jpn J Genet 65:367–380
Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832
Paull JG, Pallotta MA, Langridge P, The TT (1994) RFLP markers associated with *Sr22* and recombination between chromosome 7A of bread wheat and the diploid species *Triticum boeoticum*. Theor Appl Genet 89:1039–1045
Paull JG, Chalmers KJ, Karakousis A, Kretschmer JM, Manning S, Langridge P (1998) Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor Appl Genet 96:435–446
Purugganan MD, Suddith JI (1998) Molecular population genetics of the *Arabidopsis* CAULIFLOWER regulatory gene: nonneutral evolution and naturally occurring variation in floral function. Proc Natl Acad Sci USA 95:8130–8134
Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresvick S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484
Stam P, Zeven AC (1981). The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227–238
Tanksley SD, Nelson JC (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 SD, Rick CM (1980) Isozymic gene linkage map of tomato: applications in genetics and breeding. Theor Appl Genet 57:161–170
Visscher PM, Haley CS, Thompson R (1996) Marker-assisted introgression in backcross breeding programs. Genetics 144:1923–1932
Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. Nature 398:236–239

Whittaker JC, Haley CS, Thompson R (1997) Optimal weighting of information in marker assisted selection. Gen Res Cam 69:137–144

Xiao J, Li J, Grandillo S, Ahn S, McCouch SR, Tanksley SD, Yuan L (1996) Genes from wild rice improve yield. Nature 384:223–224

Xiao J, Li J, Grandillo S, Ahn S, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait-improving QTL alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899–909

Young DA (1999) A cautiously optimistic vision for marker assisted breeding. Mol Breeding 5:505–510