CELLULAR & MOLECULAR BIOLOGY LETTERS
Volume 7, (2002) pp 695 – 702
http://www.cmbl.org.pl

Received 3 March 2002
Accepted 8 June 2002

THE IMPACT OF MOLECULAR MARKERS ON THE WHEAT BREEDING PARADIGM

ROBERT KOEBNER[1]* and RICHARD SUMMERS[2]
[1]John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK, [2]Monsanto UK Ltd., The Maris Centre, 45 Hauxton Road, Trumpington, Cambridge CB2 2LQ, UK

Abstract: We briefly review the limited application of marker assisted selection in past wheat breeding programmes, and contrast the current situation, where increasingly it has become feasible to tag almost any gene with a microsatellite assay. Although this capability is having an impact on the conduct of large breeding programmes, a much more profound change in breeding strategy will become possible when SNP technology has matured sufficiently so that the throughput of molecular marker-based genotyping will be able to keep pace with the numbers of plants that breeders can handle in the field. We discuss the considerations that will need to be addressed in the generation of a new breeding paradigm to take advantage of the genomics revolution.

Key Words: Marker-Assisted Selection (MAS), Molecular Markers, Wheat Breeding

INTRODUCTION

One of the most critically anticipated and most often cited benefits of genetic markers for plant breeding has been their use to facilitate marker-assisted selection (MAS), i.e. their exploitation as indirect selection tools in crop breeding programmes. The rationale is clear – to allow the breeder to achieve early generation selection for a trait (or combination of traits) either for which single plant selection is ineffective, a situation which applies because

* Corresponding author
Abbreviations: AFLP - amplified fragment length polymorphism; HMW-GS – high molecular weight glutenin subunits; MAS – marker aided selection; QTL -quantitative trait locus; RAPD – random amplified polymorphic DNA; RFLP – restriction fragment length polymorphism; SNP – single nucleotide polymorphism; STS – sequence tagged site; STMS – sequence tagged microsatellite site.

environmental variation has a significant influence over trait expression, and/or because the trait concerned is under complex genetic control; or when field trials are difficult or expensive, as is the case for a number of diseases. Enabling favourable allele frequency to be increased early in the breeding process would deliver substantial efficiency gains - instead of having to carry forward a small number of large populations over several generations, the MAS breeder selects among a larger number of smaller sized populations, each of which has been pre-screened to remove altogether, or at least to reduce the frequency of unfavourable alleles at the maximum number of agronomically important loci, which may be either single genes or QTL (quantitative trait loci). In the past, the application of MAS has been technically limited by a lack of suitable markers. However, the rapid - and continuing - technological development of DNA-based diagnostic assays has reached the stage where, in principle, this limitation no longer exists. The issue has evolved into one where cost and practicality have to be set against the competitive advantage that the expected increase in breeding efficiency should deliver. In this paper, we explore the past, the present and the future of MAS breeding in winter wheat, *Triticum aestivum*, particularly in the context of the genomics revolution that is pervading much of modern biology.

MAS IN WHEAT – THE PAST

The history of MAS in wheat is one of the successful implementation of a very small number of non-DNA based assays. The primary example remains the widespread use of a series of correlations established between breadmaking quality and the presence of particular high-molecular-weight glutenin subunits (HMW-GS) among the seed storage protein complex. The genes encoding HMW-GS occur as tandem pairs at a homoeologous series of loci. One (or both) of these two genes may be silent in any given variety, and thus, in principle, the number of expressed HMW-GS in a homozygous individual varies between zero and six, although it most commonly lies between three and five. Since the two genes at each locus are so tightly linked with one another, each locus behaves genetically as a single gene. A significant level of well-catalogued allelic variation is present at each of these three loci. Comparisons of the quality characteristics of a considerable number of varieties and segregants with their HMW-GS profile have allowed a ‘quality score’ to be associated with each allele [1], and these scores have been retained, with only slight modifications, for 20 years. The power of the assay is that it is performable on single seeds, whereas assessing dough quality directly requires quantities of seed that are not available until the later generations of a pedigree selection programme. Thus, although the test is only partially predictive – allelic variation in the HMW-GS is responsible for only a fraction of the overall variation in dough quality found in segregating populations (eg 20%, as reported in [2]) – the simplicity of the assay and its adaptability to early generation selection means that its use has been responsible for significant genetic advance with respect to wheat dough quality. Substantially as a result, the UK import requirement for quality wheat (generally

from Canada) has fallen dramatically in the years since the introduction of HMW-GS MAS.

The second, and only other widely used example of a protein-based marker used in wheat MAS, is the endopeptidase (EP) allele *Ep-D1b* which predicts the presence of *Pch1*, a gene conferring resistance to the eyespot disease (causative agent *Pseudocercosporella herpotrichoides*). The resistance originates from the wild relative *Aegilops ventricosa*, and its transfer into wheat is in the form of a large segment of the relevant *Ae. ventricosa* chromosome. As a result of the inhibition of recombination between this segment and its normal wheat equivalent, *Pch1* is closely linked to a number of markers which, in wheat, are separated by significant map distances; in particular, the isozyme locus *Ep-D1* displays almost total linkage with *Pch1* [3]. The EP assay requires isoelectric focussing, a less robust and technically more demanding procedure than the SDS-PAGE used for HMW-GS testing, but it continues to be used in a number of breeding programmes. Although it can be applied to single seeds, in this format it is usually a destructive assay, as the level of enzyme activity in the endosperm is much lower than in the embryo/scutellum. Seed assays are therefore usually applied as a bulked progeny test, delaying selection by one generation. EP deployment has been less widespread than HMW-GS because the target gene is only useful in production situations where eyespot disease presents a significant threat to yield.

Current usage of these two assays in the Monsanto winter wheat programmes is quite extensive, running to (respectively) 50,000 and 10,000 individuals screened per year. The cost of each assay lies in the range US$0.15 – 0.30, depending on the assumptions made in generating this estimate.

CURRENT MARKER TECHNOLOGY RELEVANT TO MAS

Markers which rely on a direct DNA assay, rather than for example on a simply inherited phenotypic trait or a protein-based assay of a gene product, have been a central theme of wheat genetics research since the development of RFLP technology in the mid 1980's. Despite the considerable efforts to generate RFLP-based genetic maps of wheat, it has been recognised that RFLPs are unsuitable for large-scale MAS because of a shortage of relevant polymorphism in adapted germplasm pools, but critically also because of the high cost implications of their application to large numbers of individual plants. PCR has provided a crucial technical breakthrough, in that it relieves the requirement for the isolation of large amounts of purified DNA. However, most of the numerous marker types based on PCR are still impractical as MAS tools, either because they are too complex to allow automation (eg AFLP), because they are prone to non-reproducibility (eg RAPD), or because the level of polymorphism that they uncover is inadequate (eg STS). However microsatellites (or sequence tagged microsatellite sites, STMS), which assay variation in and around short repetitive sequences, have emerged as a feasible MAS tool, and considerable resources are being devoted, in both the public and the private sector, towards expanding the

genome coverage of these markers. Recently a substantial number of STMS loci have been merged into the wheat RFLP-based genetic map [4]. Although the STMS assay still requires gel electrophoresis for separation, its advantageous features are that (1) an expanding number of loci have been marked, helped by the discovery process becoming more streamlined, and especially by the fall in the cost of DNA sequencing; and (2) there is potential for multiplexing assays, either by combining assays which deliver distinct amplicon size and/or by using different fluorescent dyes for each assay so that the output signals from each do not interfere with one another. Nonetheless, there remains the problem of a shortage of marker polymorphism between the parents of best x best adapted crosses, which are typically used by breeders to maintain adapted gene complexes.

MAS IN WHEAT – THE PRESENT

An increasing number of agronomically significant genes have been tagged with linked STMS assays. Most of these are resistances to diseases, since single gene control of this class of character is widespread. A major MAS focus in the Monsanto winter wheat programmes is on the tracking of resistance to Fusarium head blight (causative agent *Fusarium graminearum* and/or *culmorum*), which is a difficult disease to handle by conventional pathology testing. The source of this resistance is the Chinese wheat variety 'Sumai 3', in which a significant proportion of its resistance has been attributed to three QTL, mapping to different chromosomes, with each of the relevant genomic sites tagged with microsatellite loci [5]. Other current MAS targets in these programmes are *Lr37* (a single gene encoding resistance to leaf rust, (causative agent *Puccinia recondita*), the wheat/rye translocation 1BL.1RS, and resistance to BYDV (barley yellow dwarf virus) introgressed from the related grass species *Thinopyrum intermedium* [6]. Future targets will include a range of disease resistance and quality traits. The cost of these assays is two to three times that of the protein-based assays, but anticipated efficiency gains flowing from automation of the whole MAS process – template acquisition, genotyping and data capture and analysis – are expected to reduce this cost differential.

FUTURE MARKER TECHNOLOGY RELEVANT TO MAS

The primary candidate for the next generation of marker is the single nucleotide polymorphism (SNP). At its simplest, a SNP consists of a single base difference within a given segment of DNA between two individuals. In any but the most closely related varieties, the potential number of such markers is enormous; while human STMS loci have been estimated to occur on average every few tens of kbp, SNP frequency averages one every 100-300bp. The overwhelming attractions of SNPs are two-fold: (1) they offer the potential for a high density of markers. The relevance of this to MAS is that it should be possible to find an informative marker in the right region in any segregating situation, even if the probability of finding polymorphism at any one SNP locus is low; and (2) the

SNP assay is not gel based, an important consideration, since the processing of samples by electrophoresis has become the most rate-limiting step in marker analysis, and is the step least easy to automate. A number of alternative SNP assays have been described in recent years [7], but as yet no industry standard has emerged.

All the major SNP detection techniques rely on an initial PCR amplification of the target DNA segment. The SNP can be targeted internally within the amplicon, or alternatively at, or immediately downstream, of the 3' end of one of the amplification primers. In the former case, the presence of a SNP can be detected by testing whether the SNP alters a restriction enzyme recognition sequence, in which case post-PCR digestion with the relevant enzyme will reveal whether the variant allele is present. Alternatively, non-gel based assays can assembled by the inclusion, along with the standard PCR primers, of a fluorescently-labelled oligonucleotide probe whose sequence incorporates that surrounding the SNP site. In this case, one primer is designed to recognise the variant sequence, while a second one recognises the standard type. If these two primers are labelled with different fluorochromes, then the amplicons derived from the alternative templates can be distinguished by the colour of their fluorescence. SNP discovery is a costly process, and is in its infancy in wheat; but the potential is enormous, considering (1) the essentially unlimited number that can be uncovered (as of February 2002, the public human SNP database contained >4 million entries, [8]), and (2) the non-requirement for electrophoresis: unlike the STMS, for which allelic difference is essentially a quantitative character (variation in amplicon size), allelic variation at a SNP locus is qualitative – the allele is defined by the identity of a particular base in the sequence. As a consequence, the number of possible alleles for a true SNP is limited to four (in practice less than this, as transitions C→G, A→T are much more frequent than transversions A→G etc.), which limits potential polymorphism at each SNP locus (in contrast to the situation for STMS loci); however the orders of magnitude greater frequency of SNP occurrence more than compensate for this.

HOW DOES/CAN/WILL THE LIFTING OF THE TECHNICAL LIMITATION ON POTENTIAL MARKER NUMBER INFLUENCE THE CONVENTIONAL BREEDING PARADIGM?

The genomics revolution is only now beginning, and it is timely therefore to consider how it can and should impact upon the crop plant breeding paradigm. Anticipating a lifting of the technical limitations to molecular marker based genotyping, the question of how to bolt on a small element of MAS to a conventional breeding approach changes into one which asks in what way the field-based activities that characterise the present paradigm will (or should) be adjusted to take advantage of MAS.

A key consideration in this debate is to ask how many *significant* factors (genes) are likely to be segregating in a population bred from a typical, rather narrowly

based cross between two elite genotypes. Some of these may be known genes, such as *Rht* (the determinant of semidwarf habit), or genes which are critical in the determination of flowering time, like the photoperiod sensitivity genes *Ppd* and the vernalisation genes *Vrn*; others are not known, but are selected by their effect on phenotype (yield, ideotype etc.). A minimal estimate of the number of such factors can be obtained from a consideration of the proportion of individuals from a typical breeder's F_2 population that reaches variety trials. Many such crosses produce no transgressive selections and are discarded; while from successful crosses, the small number of late generation selections trace back to just one or two F_2 individuals. Using 2,000 individuals as the size of a typical F_2 population, and assuming for simplicity that selection is based on homozygosity at the transgressive loci, a single effective factor should generate 2000/16 (=120) selections; two factors, 30 selections; three factors, eight selections; four factors, two selections; and five factors one selection. For a situation where a larger number of critical genes differed in allelic state between the parents, as would be the case in a wide cross, the size of F_2 needed (and the number of marker assays required) to select for fixation rises rapidly: for ten unlinked genes, the frequency of a homozygous individual is (0.25^{10}), or one in approximately one million. Naturally, a prudent breeder would seek to make more than a single F_2 selection, since later generations of selection will be applied to exploit variation at unmarked genes; this serves only to magnify yet further the MAS effort.

The above calculation is naturally a simplification. In reality it is unlikely, and probably unnecessary (although clearly desirable) for the breeder to fix allelic constitution at the earliest possible generation. The frequency of the heterozygote is much more favourable – in contrast to the example above calculated for 10 genes, the expectation of frequency of the multiple heterozygote is 0.75^{10} (rather than 0.25^{10}), or one in just seventeen. A rational strategy for MAS would therefore be to fix only a small number of loci at each generation, thereby keeping population size and MAS assay number per year to a manageable level.

This analysis begs a major question: how many breeders' traits are (even partially) determined by allelic variation at a small number of loci? Naturally, the more complex the genetic determination of a trait, the harder it will be to define the contribution of any particular allele at one of the contributing loci; this requires some sort of an *a priori* QTL analysis, which no breeder is likely to undertake. Furthermore, multiple gene control of a single trait makes it all the harder to realistically apply MAS, because the greater the number of loci that need to be fixed, the greater the number of MAS assays that need to be performed. Clearly also, the number of loci concerned will depend heavily on the genetic distance between the parents of a cross, since this will determine the proportion of the critical loci at which there is no allelic variation to be exploited.

Since the practicalities of early generation MAS are fast becoming a resource question (rather than a feasibility one), the value of the end product becomes highly relevant. An illustrative contrast can be made between a commodity crop like wheat, and a high value crop like tomato (*Lycopersicum esculentum*). What becomes critical therefore is the balance between added value and additional cost. In wheat, added value is much more likely to come from quality (rather than production quantity) traits, and a current illustration of this lies in the use of PCR-based markers to fix null alleles at the *Wx* genes. Seed of triple null *Wx* varieties produce a starch deficient in amylose [9], which is particularly suited to a range of end uses, including food stabilizers, thickeners and emulsifiers, as well as non-food uses in the gum and paper industries, and thus represents an escape from downward pressure on price that is imposed by the commodity market.

The success of MAS to improve the efficiency of conventional breeding systems is thus largely dependent on the predictable development of automatable assays, and an ever wider number of assayable targets. What, however, are the foreseeable consequences of heavy investment in MAS on introgression of new germplasm in a breeding programme? Such introgression is necessary to maintain genetic diversity and to avoid the plateauing of performance that would be inevitable where no introgression occurs. If MAS breeding is truly more efficient than conventional breeding, then the need for introgression will become more pressing more rapidly than it is at present. Marker assays of course need validation before they can be applied, so the inclusion of new germplasm will require prior effort to first define which loci bring in useful variation, and subsequently to tag them. Such a validation exercise will represent a substantial change in paradigm, and could act as a dangerous disincentive to introgression. However, the demonstration that even the most exotic materials can carry cryptic favourable QTL [10] should encourage more wide-crossing than is presently undertaken.

MAS IN WHEAT – THE FUTURE

MAS is a capital-intensive endeavour. Looking to the future, the expectation is of a continued centralisation of breeding into large units, which alone will be able to sustain the necessary levels of investment. Added to this economic pressure will be the real competitive pressure on the remaining industrial players to adopt MAS for fear of being ‘left behind’. One can anticipate the rapid development of DNA assays, particularly SNPs, along with an acceleration in the design and deployment of automation platforms, both aimed at the assays themselves and at the pre-assay (particularly DNA acquisition) and post-assay (data point collection) stages. Overall these will serve to reduce the unit cost per assay, and so allow an increase in the number of assays possible.

MAS will be increasingly applied (1) for traits which are difficult to manage via phenotype, because of low penetrance and/or complex inheritance; (2) for the maintenance of recessive alleles in backcrossing pedigrees; (3) for the

pyramiding of disease resistance genes; and (4) for making choices of parent in crossing programmes, in order to assure minimal levels of duplication of alleles across sets of genes targeted for selection, and fixation in those for which no variation is preferred.

REFERENCES

1. Payne, P.I., Holt, L.M., Thompson, R.D., Bartels, D., Harberd, N.P., Harris, P.A. and Law, C.N. The high molecular weight subunits of glutenin: classical genetics, molecular genetics and the relationship of breadmaking quality. **Proc. 6th Int. Wheat Genetics Symp**., Kyoto 1983, 827-834.
2. Kolster, P., Vaneeuwijk, F.A. and Vangelder, W.M.J. Additive and epistatic effects of allelic variation at the high-molecular-weight glutenin subunit loci in determining the bread-making quality of breeding lines of wheat. **Euphytica** 55 (1991) 277-285.
3. Summers, R., Koebner, R., Hollins, T., Förster, J. and Macartney, D. The use of an isozyme marker in breeding wheat (*Triticum aestivum*) resistant to the eyespot pathogen (*Pseudocercosporella herpotrichoides*) **Proc. 7th Int. Wheat Genetics Symp**., Cambridge 1988, 1195-1197.
4. Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P. and Ganal, M.W. A microsatellite map of wheat. **Genetics** 149 (1998) 2007-2023.
5. Anderson, J.A., Stack, R.W., Liu, S. , Waldron, B.L., Fjeld, A.D., Coyne, C., Moreno-Sevilla, B., Mitchell Fetch, J., Song, Q.J., Cregan, P.B. and Frohberg, R.C. DNA markers for Fusarium head blight resistance QTLs in two wheat populations. **Theor. Appl. Genet.** 102 (2001) 1164-1168.
6. Stoutjesdijk, P., Kammholz, S., Kleven, S., Matsay, S., Banks, P.M. and Larkin, P.J. PCR based molecular marker for the *Bdv2 Thinopyrum intermedium* source of barley yellow dwarf virus resistance in wheat. **Aust. J. Agric. Res.** 52 (2001) 1383-1388.
7. Koebner, R.M.D., Powell, W. and Donini, P. The contribution of current and forthcoming DNA molecular marker technologies to wheat and barley genetics and breeding. **Plant Breed. Rev.** 21 (2001) 181-220.
8. http://www.ncbi.nlm.nih.gov/SNP/get_html.cgi?whichHtml=overview
9. McLauchlan, A., Ogbonnaya, F.C., Hollingsworth, B., Carter, M., Gale, K.R., Henry, R.J., Holton, T.A., Morell, M.K., Rampling, L.R., Sharp, P.J., Shariflou, M.R., Jones, M.G.K. and Appels, R. Development of robust PCR-based DNA markers for each homoeoallele of granule-bound starch synthase and their application in wheat breeding programs. **Aust. J. Agric. Res.** 52 (2001) 1409-1416.
10. Tanksley, S.D. and McCouch, S.R. Seed banks and molecular maps: Unlocking genetic potential from the wild. **Science** 277 (1997) 1063-1066.