21st century wheat breeding: plot selection or plate detection?

Robert M.D. Koebner[1] and Richard W. Summers[2]

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

The publicly reported limited application of marker-assisted selection (MAS) in wheat breeding programmes to date is reviewed and contrasted with the current situation, in which it has increasingly become technically feasible to tag almost any gene with a microsatellite assay. Although this capability is starting to have an impact on the conduct of large breeding programmes, a much more profound change in breeding strategy will become possible when single nucleotide polymorphism (SNP) technology has matured sufficiently so that the throughput of molecular marker-based genotyping is able to keep pace with the numbers of plants that breeders routinely handle in the field. We explore the extent to which the genomics revolution might generate a change in the conventional breeding paradigm, which has operated with such success for the best part of the 20th century, and identify the need for a substantial reduction in assay price before MAS is likely to make more than a marginal impact on present practice.

A much anticipated and frequently cited benefit of genetic markers for plant breeding relates to their suitability as indirect selection tools in crop breeding programmes. The three main advantages of marker-assisted selection (MAS) are clear. First, it becomes possible to select, on a single plant basis, for a trait (or combination of traits) in situations in which this is conventionally ineffective – either because environmental variation exerts a significant influence over trait expression or because phenotypic-based assessment is difficult or not cost-effective. Both these scenarios are commonplace in crop breeding programmes. Second, for traits that are under multigenic control, the individual genetic components – each of which individually might make only a relatively modest contribution to the overall determination of the phenotype – can be maintained and ultimately fixed in the homozygous state at the end of the breeding process. Third, both recessive genes, and those not readily amenable to phenotypic selection, can be maintained in segregating generations without the need for validation at each generation via a progeny test. This is of particular relevance in backcross programmes, in which an otherwise elite genotype (the recurrent parent) is corrected for a single fault (for example, susceptibility to a disease) by introducing the minimum amount of genetic material from the donor of the target trait. The process is achieved by repeatedly crossing the hybrid generations with the recurrent parent and selecting at each generation for segregants heterozygous for the target gene(s) – a strategy that is particularly cumbersome in cases in which the gene acts recessively.

Enabling favourable allele frequency to be increased early in a pedigree-based programme (see Box 1) 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-selected to remove altogether (or at least to reduce the frequency of) unfavourable allele(s) at as many agronomically important loci as possible – these could be either single genes or quantitative trait loci (QTL). However, until now, the widespread application of MAS has been technically limited by a lack of both suitable markers (see Box 2) and of high-throughput analytical platforms. However, the rapid, and continuing, development of DNA-based diagnostic assays is now reaching the stage at which, in principle, both these limitations will disappear (see Box 3). The evolving issue is becoming one in which cost and practicality need to be set against the competitive advantage that the expected increase in breeding efficiency should deliver. Three years on from a 'cautiously optimistic vision' for MAS [1], and one year on from a report of a modest scale of its application to wheat breeding in Australia [2], we explore its potential to alter the conventional paradigm of wheat breeding, particularly in the context of the accelerating genomics revolution that is pervading much of modern biology.

MAS in wheat – the past

The history of MAS in wheat, which goes back <20 years, is one in which a rather small number of non-DNA-based assays has been implemented to fix favourable variants at specific loci. The uptake of the method has been low despite the establishment of a growing number of gene tags for both specific genes (most commonly determining resistance to disease) and, more recently, for quantitative traits, which has been enabled by the development of molecular marker technology and the use of these markers for genetic mapping. The primary example remains the widespread use of a series of correlations established in the early 1980s between bread-making quality and the presence of particular variants among endosperm storage

Corresponding author: Robert M.D. Koebner (robert.koebner@bbsrc.ac.uk).

 0167-7799/03/$ - see front matter PII: S0167-7799(02)00036-7

Box 1. The conventional winter wheat breeding paradigm: pedigree selection

The pedigree method (so-called because the ancestry of all selections can be traced back, generation by generation, to a specific hybrid plant) developed piecemeal in Europe and Australia during the late 1800s [a] but was first formally described, at least in the English-speaking literature, in 1927 [b].

The main principle of the method is to apply individual plant selection in the early generations developed from a cross and line selection once a sufficient level of genetic homogeneity has been achieved within lines. A common model has been elaborated by Bingham and Lupton [c]. Since the early 1900s, it has been widely used for the breeding of self-pollinating crop species. The rationale is to generate a genetically diverse population from a varietal hybrid and then to rigorously select for superior genotypes over the course of subsequent generations (Fig. I). In its simplest form, the two parental genotypes are chosen to be complementary for valuable attributes (e.g. one parent might be high yielding but unsuitable for bread making whereas the other might yield below par but give flour of excellent baking quality – thus high yielding, good quality selections can be obtained). In the first segregating generation (F_2), each plant is genetically unique and segregation takes place at every gene for which the parents differed in allelic state. Selection is made on an individual plant basis and is thus restricted to those genes for which expression is materially unaffected by environmental factors and the effect of which is largely qualitative (some examples include sensitivity to certain diseases, time taken to reach flowering and dwarfness). To capture a good representation of the total variation possible in the F_2, several thousand progeny per parental combination are commonly grown as spaced plants. In subsequent generations, selection is applied for characters that are not assessable without some level of replication (this includes most of the production and end-user traits, in particular yield and final end-use quality), and tends, especially at the later generations, to be increasingly directed to lines rather than individual plants. Out of consideration for effective resource use, a heavy attrition needs to occur in the early generations, and it is in this phase – particularly at the F_2, in which the numbers are at their largest – that MAS has its greatest potential.

In a substantial breeding programme, in the order of 1000 crosses are processed in this way each year, generating an initial pool of some 2 000 000 genetically distinct individuals. From generation to generation the level of genetic fixation increases asymptotically, so that by F_6, when yield trialling is initiated (Fig. II), genotypes are on average 97% homozygous, whereas by the time the variety is released for cultivation, it is essentially true-breeding.

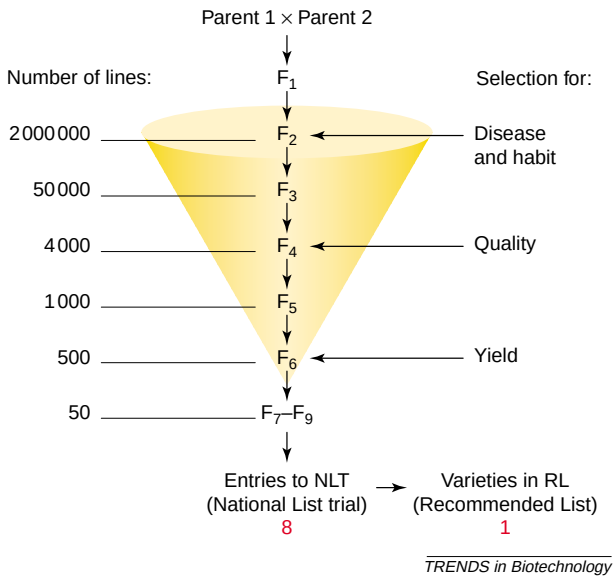

Fig. I. The selection funnel in a pedigree wheat-breeding programme. Representative numbers of selections taken forward in each generation are shown along the left edge of the triangle. Winter wheat generation time is typically a full calendar year.

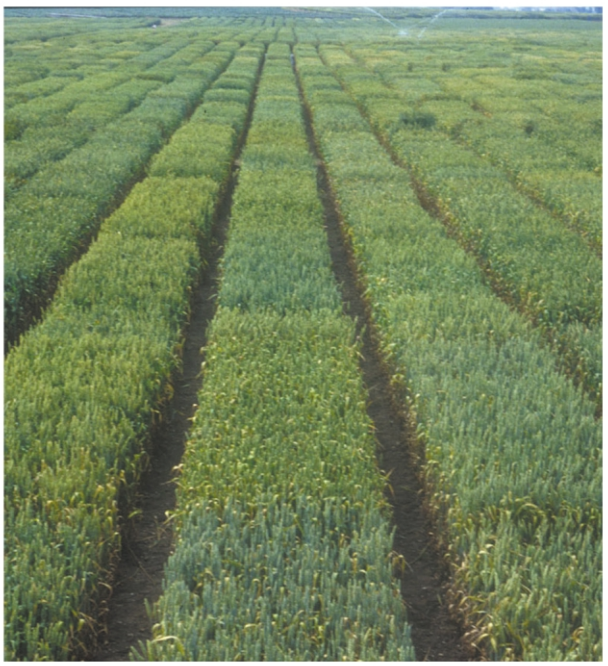

Fig. II. Late generation yield trial plots in a conventional winter wheat breeding programme.

References

a Lupton, F.G.H. (1987) History of wheat breeding. In *Wheat Breeding – Its Scientific Basis* (Lupton, F.G.H., ed.), pp. 51–70, Chapman & Hall, London & New York

b Love, H.H. (1927) A program for selecting and testing small grains in successive generations following hybridization. *J. Am. Soc. Agron.* 19, 705–712

c Bingham, J. and Lupton, F.G.H. (1987) Production of new varieties: an integrated research approach to plant breeding. In *Wheat Breeding – Its Scientific Basis* (Lupton, F.G.H., ed.), pp. 487–538, Chapman & Hall, London & New York

protein subunits, which are readily visualized in electrophoretic separations of crude protein extracts from single seeds. Comparisons of the quality characteristics of varieties and segregants with their protein profile allowed a 'quality score' to be associated with each allele [3] and these scores have been retained, with only slight modifications, for 20 years. The power of the assay lies in that it can be performed on individual seeds, whereas assessing bread-making quality directly requires quantities of seed that only become available in the later generations of a pedigree selection programme. Thus, although the test is only partially predictive – allelic variation at the relevant loci being responsible for only a fraction of the overall variation in dough quality found in segregating populations (e.g. 20%, see Ref. [4]) – the simplicity of the assay and its adaptability to early generation selection have made it an attractive and cost-effective procedure in breeding bread-making quality wheat varieties worldwide.

Box 2. Current marker-assisted selection (MAS) technology

Markers fall into three broad classes: those based on visually assessable traits (morphological), those based on gene product (biochemical); and those relying on a DNA assay (molecular). Morphological markers are simple but rarely useful because few phenotypes are determined by allelic variation at one locus (particularly in a polyploid such as wheat). By contrast, biochemical markers are single-locus-based but are not numerous because their number is limited by the availability of a specific assay. The first molecular markers (restriction fragment length polymorphisms, RFLPs) relied on the DNA:DNA hybridisation assay but, despite considerable efforts to generate RFLP-based genetic maps of essentially all the major crops, RFLPs have proven unsuitable for large-scale MAS mainly because of the high cost implications of analysing large numbers of individual plants. In wheat, RFLP has been unable to expose significant levels of polymorphism, particularly within adapted germplasm pools, although this has not been a problem for other crop species, notably maize.

PCR relieves the requirement for the isolation of large amounts of purified DNA but most of the numerous PCR-based marker types are still impractical as MAS tools, either because they are too complex to allow automation, because they are insufficiently robust, or because the level of polymorphism that they uncover remains inadequate. However, microsatellites (or sequence tagged microsatellite sites, STMS), which assay variation both within and in the flanking regions of 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 [a]. Although the STMS assay still requires gel (or capillary) electrophoresis for separation, its advantageous features are that (1) an expanding number of loci has been marked, helped by the discovery process becoming more streamlined, and especially by the fall in the cost of DNA sequencing; (2) there is potential for multiplexing assays, either by combining assays that 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; and (3) they can be processed on automated or semi-automated capillary electrophoresis machines, so can take advantage of the technical improvements driven by the push for greater throughput in DNA sequencing.

References

a Röder, M.S. *et al.* (1998) A microsatellite map of wheat. *Genetics* 149, 2007–2023

The second, and only other widely used example of a protein-based marker used in wheat MAS, is the product of the endopeptidase (EP) allele *Ep-D1b*, which predicts, as a result of tight genetic linkage, the presence of *Pch1*, a gene conferring resistance to the fungal pathogen that is the causative agent of the stem disease eyespot [5]. Direct selection for resistance to this disease is difficult because the pathogen grows slowly *in planta* and damage to the plant is not only at ground level but also, more importantly, is not restricted to the plant surface, making the extent of fungal penetration difficult to assess. In addition, single plant selection is seldom effective, owing to a significant degree of non-genetic interference. The EP assay requires isoelectric focusing, a less robust and technically more demanding procedure than the SDS–PAGE used for endosperm storage protein profiling but it continues to be used in several breeding programmes. When applied to single seeds, the EP test is usually a destructive assay because the level of enzyme activity in the endosperm is much lower than in the embryo or scutellum. Seed assays are therefore usually applied as a bulked progeny test, delaying selection by one generation. EP deployment has been less widespread than seed storage protein profiling because eyespot disease only represents a significant threat to yield in cool wet climates such as in NW Europe and NW USA.

Current usage of these two assays in the Monsanto winter wheat programme (which currently captures ~20% of the Northern European varietal market) runs to tens of thousands of individuals screened per year. The cost of each assay falls in the range US$0.15–0.30, depending on the financial assumptions made.

MAS in wheat – the present

Increasing numbers of agronomically significant genes have been tagged with linked microsatellite assays in recent years [6–8]. Most of these are resistances to diseases because single gene control of this class of character is widespread. Some disease resistances are

Box 3. The future of marker assisted selection (MAS) technology

The primary candidate for the next generation of marker is the single nucleotide polymorphism (SNP). At its simplest, an 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; whereas 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 output is of the binary type and this presents an easier target for automated data interpretation than the length-based outputs that are typical of STMS. In addition such data need not be generated by electrophoresis, giving the potential for simpler and cheaper analytical platforms. Several alternative SNP assays have been described in recent years [a], although no industry standard has as yet emerged. The large amount of sequence data emerging from extensive expressed sequence tag programmes is likely to speed SNP discovery in wheat, especially given the concerted international collaborative effort currently being mounted to achieve this (see http://wheat.pw.usda.gov/ITMI/2002/WheatSNP.html).

SNP discovery is a costly process and is in its infancy in wheat; but the potential is enormous, considering the essentially unlimited number that can be uncovered (as of June 2002, the public human SNP database has submissions for 4 275 093 SNPs; see http://www.ncbi.nlm.nih.gov/SNP/get_html.cgi?whichHtml = overview), and 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.

References

a Koebner, R.M.D. *et al.* (2001) *The contribution of current and forthcoming DNA molecular marker technologies to wheat and barley genetics and breeding Plant Breeding Reviews*, John Wiley & Sons Ltd, pp. 181-220

under oligogenic control, as is the resistance to Fusarium head blight (FHB), which is a MAS focus in the Monsanto winter wheat programme. For practical reasons, FHB is a difficult disease to handle by conventional pathology testing, and genetic analysis of the primary source of resistance has shown that a significant proportion of the effect can be attributed to three QTL, mapping to different chromosomes, with each of the relevant genomic sites tagged with a microsatellite locus [9]. Other current MAS targets in these programmes include single genes associated with durable resistance to rust diseases, a wheat/rye translocation chromosome, and resistance to a virus transmitted by aphids. Future targets are expected to include a range of disease resistance and quality traits. Current cost estimates of capillary sequencing platform-based genotyping put the price of one microsatellite data point at ~US$0.40, excluding the cost of both the PCR itself and the acquisition of DNA template. The cost of these assays is thus approximately three times that of the protein-based assays but the hope is that the anticipated efficiency gains flowing from automation of the whole MAS process (DNA template acquisition, genotyping, data capture and data analysis) will succeed in reducing this differential.

The effect on the conventional breeding paradigm of a non-limiting supply of markers

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

How many genes determine variation in major breeders' traits?

A key question to ask is how many significant factors (genes) are likely to be segregating in a standard breeder's population, which is typically derived from a cross between two elite, rather closely related genotypes. Some of these genes might be known, such as *Rht* (the determinant of semi-dwarf habit), or those with major effects on flowering time via either photoperiod sensitivity (*Ppd*) or vernalisation requirement (*Vrn*); others are not known but have been selected on the basis of a visible effect on phenotype (yield, ideotype, etc.). A minimal estimate of the number of such genes can be made 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; from successful crosses, the small number of late-generation selections generally trace back to just one or two F_2 individuals. Using 2000 individuals as the size of a typical breeder's F_2 population, and assuming initially (for simplicity) that selection is based on homozygosity at transgressive loci, a single effective gene would be expected to generate 2000/16 (= 120) F_2 selections; two genes, 30 selections; three genes, eight selections; four genes, two selections; and five genes one selection. This implies that the likely number of relevant differences between elite parents is of the order of five or more, and that unsuccessful crosses fail either because the number of such genes is higher, making the F_2 population size too small; or because the two parents were genetically too similar to one another. For a situation in which a larger number of critical genes differ in allelic state between the parents, as would be the case in a cross between diverse parents, the size of F_2 needed (and hence 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 from a single cross, because later generations of selection will be applied to exploit variation at unmarked genes; this serves only to magnify the MAS effort yet further. Of course, this calculation is 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 the frequency of the multiple heterozygote is 0.75^{10} (rather than 0.25^{10}), or one in just 17. A rational strategy for MAS might therefore be to fix a small number of loci at each generation, thereby keeping population size and MAS assay number per year to a manageable level. The danger of this strategy is that unless sufficient material is retained for each subsequent selection generation, non-selected loci might become fixed, by chance, in the 'wrong' allelic state, so this strategy still implies a significant genotyping effort at each generation.

This analysis raises an important major question: how many breeders' traits are (even partially) determined by allelic variation at a small number of loci? Traits with simple inheritance, unless their expression is particularly prone to environmental interference, are, by definition, those that are the least likely to require indirect selection methods. In any case, genetic variation for many of the simply inherited traits might already be largely exhausted as a result of the decades of scientifically based breeding that have been applied to an intensively bred species such as wheat. Fully unravelling complex inheritance is not an activity that any breeder is in a position to undertake because it requires genome-wide genetic mapping as well as the phenotyping of populations, often over locations and years – a so-called QTL analysis. Such analyses are beguilingly simple in concept, but achieving an acceptable level of accuracy requires control over several complex factors, some of which are hardly controllable [10]. In mitigation, however, the common situation is that for a trait controlled by multiple QTL, not all the loci contribute equally, so that there remains an advantage in being able to indirectly select just for one, or a small number of components, rather than necessarily for all.

The practicalities of early generation MAS

Whether or not to deploy markers to aid selection early in the pedigree is rapidly becoming a question of resource rather than feasibility. In this context, the value of the end product becomes highly relevant. An illustrative contrast

can be made between a commodity crop such as wheat, and a high-value crop such as tomato. What becomes crucial 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 that is deficient in amylose [11], which might be 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 likelihood that MAS will improve the efficiency of conventional breeding systems is largely dependent on the predictable development of affordable 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 should encourage more wide crossing than is presently undertaken [12].

MAS in wheat – the future

MAS is a capital-intensive endeavour. Looking to the future, the expectation is of a continued centralization 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'. The rapid development of DNA assays, particularly single nucleotide polymorphisms (SNPs), can be anticipated, along with an acceleration in the design and deployment of automation platforms, aimed at both the assays themselves and at the pre-assay (particularly DNA acquisition) and post-assay (data point collection) stages. Overall, these should serve to reduce the unit cost per assay and thus allow an increase in the number of assays possible. Similar considerations of course also apply to the widely heralded exploitation of human DNA polymorphisms to predict differential drug response. A recent report states that the average cost of an SNP assay has fallen from US$1.00 to US$0.10 over a 12-month period, but that a further order of magnitude reduction to US$0.01 per assay will be required before wide-scale usage of the technology becomes feasible [13]. The contrast between the high-value end product of pharmacogenetics (medicinal drugs) and the lower-value end product of plant breeding (finished varieties) only underlines the dramatic reduction in technology cost necessary to allow the widespread use of MAS in crop improvement.

Conclusion

MAS will be increasingly applied in four main areas. First, for the accelerated selection of a small number of traits that are difficult to manage via phenotype, owing to low penetrance and/or complex inheritance. Second, for the maintenance of recessive alleles in backcrossing pedigrees. Third, for the pyramiding of disease-resistance genes, and fourth, for aiding in the choice of parents in crossing programmes, to ensure minimal levels of duplication of alleles across sets of genes targeted for selection, and to promote fixation. Overall, it will remain vital that 'laboratory-based breeding' should remain the servant of the field breeder and not its master – because if large-scale MAS deployment results in a 'magic bullet' approach to breeding, in which major breeding targets are resolved by a single gene approach, then the holistic advances that have been achieved by the phenotypic selection of minor genes will be lost and the varieties that emerge might become vulnerable to future shifts in production system/climate/end use and so on. Our thesis is that the breeding paradigm that has served the industry well over many decades will be touched, but not overturned, by genomics-driven MAS. Wheat breeding will continue to be driven primarily by selection in breeders' plots, rather than by detection in microtitre plates.

References

1 Young, N.D. (1999) A cautiously optimistic vision for marker-assisted selection. *Mol. Breed.* 5, 505–510
2 Eagles, H.A. *et al.* (2001) Implementation of markers in Australian wheat breeding. *Aust. J. Agric. Res.* 52, 1349–1356
3 Payne, P.I. *et al.* (1983) The high molecular weight subunits of glutenin: classical genetics, molecular genetics and the relationship of breadmaking quality. In *Proc. 6th Int. Wheat Genetics Symp.* (Sakamoto, S., ed.), pp. 827–834, Kyoto, Japan
4 Kolster, P. *et al.* (1991) 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, 277–285
5 Summers, R. *et al.* (1988) The use of an isozyme marker in breeding wheat (*Triticum aestivum*) resistant to the eyespot pathogen (*Pseudocercosporella herpotrichoides*). *Proc. 7th Int. Wheat Genetics Symp.* (Miller, T.E., Koebner, R.M.D. eds), pp. 1195–1197, Cambridge, UK
6 Huang, X.Q. *et al.* (2000) Molecular mapping of the wheat powdery mildew resistance gene *Pm24* and marker validation for molecular breeding. *Theor. Appl. Genet.* 101, 407–414
7 Raupp, W.J. *et al.* (2001) Cytogenetic and molecular mapping of the leaf rust resistance gene *Lr39* in wheat. *Theor. Appl. Genet.* 102, 347–352
8 Bariana, H.S. *et al.* (2001) Mapping of durable adult plant and seedling resistances to stripe rust and stem rust diseases in wheat. *Aust. J. Agric. Res.* 52, 1247–1255
9 Anderson, J.A. *et al.* (2001) DNA markers for Fusarium head blight resistance QTLs in two wheat populations. *Theor. Appl. Genet.* 102, 1164–1168
10 Asins, M.J. (2002) Present and future of quantitative trait locus analysis in plant breeding. *Plant Breed.* 121, 281–291
11 McLauchlan, A. *et al.* (2001) 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, 1409–1416
12 Tanksley, S.D. and McCouch, S.R. (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066
13 Roses, A.D. (2002) Pharmacogenetics place in modern medical science and practice. *Life Sci.* 70, 1471–1480