The challenges in Marker Assisted Breeding

Johan D. Peleman and Jeroen Rouppe van der Voort
Keygene N.V. business unit Genetics, P.O. Box 216, 6700 AE, Wageningen, The Netherlands
Contact: johan.peleman@keygene.com, (+31) 317 466866.

Abstract: This paper describes our opinion on the future of Marker Assisted Breeding. Concomitantly, it provides a comprehensive overview of the different types of application in Marker Assisted Breeding.

Introduction

The potential value of genetic markers, linkage maps and indirect selection in plant breeding has been known for over 80 years. However, it was not until the advent of DNA marker technology in the 1980s, that a large enough number of environmentally insensitive genetic markers could be generated to adequately follow the inheritance of important agronomic traits. Since then, DNA marker technology has dramatically enhanced the efficiency of plant breeding. In the past decade, a number of breeding companies have, to varying degrees, started using markers to increase the effectiveness of selection in breeding and to significantly shorten the development time of varieties. Now, advances in automated technology enable a new approach in marker assisted breeding which we have called 'Breeding by Design'. The advances in applied genomics and the possibility to generate large scale marker data sets provide us with the tools to determine the genetic basis for all traits of agronomical importance. Also, methods for assessing the allelic variation at these agronomically important loci are now available. This combined knowledge will eventually allow the breeder to combine favorable alleles at all these loci in a controlled manner, leading to superior varieties.

Changing concepts and molecular approaches provide opportunities to develop rational and refined breeding strategies. Here, it is argued that knowledge about map position and allelic variation at agronomically important loci in concert with available, easy-to-assay molecular markers allow the design of superior varieties. Depending on the crop specific generation time, controlled marker assisted selection strategies can lead to the production of superior varieties within five to ten years.

In our view, different stages of impact can be distinguished in Marker Assisted Breeding. These stages are described below in more detail.

Stage 1: Marker Assisted Selection (MAS)

DNA markers are highly reliable selection tools as they are stable, not influenced by environmental conditions and relatively easy to score in an experienced laboratory. Compared to phenotypic assays, DNA markers offer great advantages to accelerate the variety development time as a result of:

1. increased reliability: the outcome of phenotypic assays is affected, among others, by environmental factors, the heritability of the trait, the number of genes involved, the magnitude of their effects and the way these loci interact. Hence, error margins on the measurement of phenotypes tend to be significantly larger than those of genotyping scores based on DNA markers.
2. increased efficiency: DNA markers can be scored at the seedling stage. This is especially advantageous when selecting for traits which are expressed only at later stages of

development, such as flower, fruit and seed characteristics. By selecting at the seedling stage, considerable amounts of time and space can be saved.

3. reducing cost: there are ample traits where the determination of the phenotype costs more than the performance of a PCR assay. In high throughput setting, the material and consumable cost for a PCR assay will typically not exceed 2 Euro. In comparison, the growth of a tomato or pepper plant to full maturity in a heated greenhouse will cost approximately 20 US$. Every plant that can be rejected before planting will in such settings save a considerable amount of money.

Before deciding to follow DNA marker-assisted approaches, practical concerns and cost-benefit analysis need to be addressed. Leaders of breeding programs must address a multifaceted evaluation of DNA marker-assisted approaches before committing to such new endeavours. Where none of the above arguments apply, then the utilization of markers does not make any sense, at least for the above described purposes.

The use of DNA markers for indirect selection offers greatest benefits for quantitative traits with low heritability as these are the most difficult characters to assess in field experiments. Obviously, the development of marker assisted assays for such traits is extra difficult and most costly due to the extensive phenotypic assays. However, once the knowledge exists to estimate the parameters which determine the trait of interest, a well designed experimental set up will result in the availability of marker assisted breeding tools which can reduce to a major extent future application of phenotypical assays.

Stage 2: The creation of novel varieties

The application of DNA markers in the breeding process can create substantially more added value than just improving the quality or cost of existing selection programs. By applying markers in a creative manner, new traits can be introduced which either could not or only be obtained with great difficulty by classical breeding. Therefore, the application of markers in breeding can create a major competitive advantage to those breeders / companies which have integrated markers as part of their working tools. Where breeding goals can not be achieved through traditional approaches, there is considerable scope for the use of molecular markers at almost any stage of development of new varieties. Here the limitation is not the facilitating technology but the imagination and motivation of the (Marker Assisted) Breeder. Below, a number of examples are provided where creative applications of markers clearly provide a major benefit over classical breeding.

Removal of linkage drag

In the mid nineties, Keygene was involved in a marker assisted breeding approach that led to the development of novel lettuce variety resistant to the aphid *Nasonovia ribisnigri* (Jansen, 1996). This aphid is a major problem in field grown lettuce areas in Europe and California causing reduced and abnormal growth in addition to spread of viral diseases. Resistance to this aphid could be introgressed from a wild relative of lettuce, *Lactuca virosa*, by repeated backcrossing. However, despite many rounds of backcrossing the new product was of extremely poor quality, bearing yellow leaves and a greatly reduced head. This could either have been caused by a pleiotropic effect of the resistance gene or by so called ‘linkage drag’, a negative trait closely linked to the positive trait of interest. Marker analysis eventually demonstrated that the reduced quality was caused by so called ‘linkage drag’. In this case, the linkage drag was recessive, only visible in the homozygous state, thereby seriously increasing the difficulty to select for recombinations based on the phenotype. It was decided to use DNA markers flanking the introgression to pre-select for individuals that are recombinant in the vicinity of the gene. More than thousand F2 plants were screened this way, leading to the

selection of some 100 individuals bearing a recombination or even double recombinations in the vicinity of the gene. Only those individuals needed to be phenotyped for both the resistance and, at the F3 level, for the absence of the negative characteristics. This approach eventually led to the selection of an individual bearing recombination events very close to each side of the gene thereby removing the linkage drag. The results demonstrated that the (recessive) linkage drag was located on both sides of the resistance gene on top of being tightly linked. This result would have been very hard to obtain by classical selection methods.

Pyramiding resistance genes

Another example using markers for the creation of novel varieties is by pyramiding disease resistance genes. This approach can offer great financial rewards through extending the life span of new varieties.

Such an approach has been used for the backcross transfer of QTL for downy mildew resistance in pearl millet (Witcombe and Hash 2000). Here a limited number of RFLP markers has been used for marker-assisted selection to improve disease resistance in both parent lines of a popular hybrid variety. Despite the labour intensive nature of this approach and the resulting limitation in population size in a given generation, good progress has been made and field evaluation of the finished projects is underway just four years after initiation of the project.

As resistance genes tend to reside in resistance gene clusters, interesting alleles of different resistance genes may be located in tandem but present in different accessions. In such case it is of paramount importance to precisely fine map the alleles of the different genes with respect to one another. This goal can be most easily achieved using DNA markers. Subsequently, the linked markers can be utilized to select for the rare recombinants that combine the favourable alleles in tandem.

Marker Assisted Breeding of polygenic traits

Keeping track of all genes involved in complex traits during a breeding program is an enormous challenge if not impossible task. In our experience, we have several times observed the loss of minor QTL when a round of marker assisted selection was replaced by a round of phenotypic selection. The utilization of markers can obviously prevent such loss of QTL.

Simulation studies, although based on some over-simplified scenarios, provide some interesting insight into optimum number of locations, replications and population size in molecular breeding programs (Moreau et al. 2000). In these simulation studies, marker-assisted approaches remained efficient for QTL with even very low heritabilities (0.15).

DNA markers allow us to unravel the genetic basis of traits expressing continuous phenotypic variations as they are abundant, and scattered throughout the genome. By using dense genetic marker maps, the contributions of separate regions of the genome on the trait values can be estimated once the mapping population is sufficiently large. In addition, agronomically important traits like nutritional quality, yield, flowering time and ‘durable’ resistance are all traits which appear to follow complex, polygenic inheritance patterns with multiple genes having small effects on the trait value. Nevertheless, various lines of evidence obtained from various crops indicate that even such complex traits appear to be determined by only a few major factors (Young et al. 1996; Frary et al., 2000; Thornsberry et al., 2001; Rouppe van der Voort et al. 2000).

We therefore argue that simplifications of these complex analyses can offer an important key to success in mapping the loci involved in these traits. Such simplifications can be obtained at several levels:

1. Simplification of the phenotype: division of a complex phenotype into its separate genetic components. For example, an extremely important phenotype like yield, is determined by

a vast array of component characters, such as root size, plant size, number of fruit, size of fruit, fruit contents, etc. Mapping the genes involved in these separate components provides a better understanding of the complex trait and a higher chance of success.

2. Simplification of the mapping: separating the effect of each QTL by generating Near Isogenic Lines (NIL's), using the technique of Introgression Line Libraries (Eshed and Zamir, 1995) and Reverse QTL Mapping (RQM; Wye et al., 2000, Peleman et al., submitted), enables the more precise measurement of the effect of the QTL and thereby the precise/fine mapping of the QTL. Fine mapping of a QTL is an essential step in exploiting the QTL by marker assisted selection.

These approaches significantly aid in unraveling the complexity of agronomically important traits. It is with such traits that, in the long term, the biggest benefits of MAS can be obtained.

Introduction of novel characteristics

An innovative approach for introducing novel polygenic characteristics from wild germplasm in a controlled manner is the application of Advanced Backcross Breeding (AB-Breeding). This approach involves the simultaneous discovery and transfer of important QTL from unadapted germplasm into elite breeding lines (Tanksley and Nelson 1996). The AB breeding strategy postpones QTL mapping until the BC2 or BC3 generation. During the generation of BC2 or BC3 populations, negative selection is being exercised to minimize the occurrence of unfavorable donor alleles. The advantage of focussing on BC2 or BC3 populations is that they offer sufficient statistical power for QTL identification at the one hand and on the other hand provide sufficient similarity to the recurrent parent to select for QTL-NILs in a short time span (within 1-2 years). By use of QTL-NILs, the QTLs discovered can be verified and the NILs may serve directly either as improved varieties or as a parent variety in case of hybrid crops.

The advanced backcross approach has been successfully used to identify markers for QTL contributing to fruit size, shape, colour and firmness together with soluble solids and total yield in tomato. On this basis, QTL marker associations were identified in one backcross generation and immediately applied in the subsequent backcross generation some six months later (Tanksley et al. 1996).

Effective exploitation of (exotic) germplasm

Breeders have traditionally been reluctant about the use of wild germplasm in their breeding programs due to complex, long-term and unpredictable outcomes, particularly in crops where quality traits are important market criteria. This is a pity because in most crops, the cultured germplasm only represents a small section of the vast diversity available in the species. Tanksley and coworkers have clearly demonstrated that wild relatives of tomato contain genes contributing to interesting culture characteristics which are generally not expected to reside in those species (Tanksley and McCouch, 1997; Frary et al. 2000). Marker assisted backcrossing now enables the breeders to precisely introgress small sectors of wild/exotic accessions thereby providing breeders with the tools to effectively unleash the vast resources held in germplasm collections.

DNA marker-based diversity analysis enables gene banks to define core collections, which will provide a user friendly entry point for breeders to access large and varied germplasm collections. A large scale genetic distance analysis of the complete CGN genebank of lettuce in The Netherlands has been done using DNA markers. The analysis involved more than 6.800 samples and a enormous data set of more than 1,35 million datapoints was produced in this study (Van Hintum, submitted).

This type of analyses will also greatly aid selection of genotypes for broadening the genetic base of breeding populations and for the development of heterotic populations for breeding F_1 hybrid varieties.

Using markers tightly linked to a gene of interest, so called locus haplotyping can be performed on accessions of germplasm to identify those samples that bear different alleles at the locus of interest. It enables identification of accessions/lines bearing different alleles at a single locus, which can then be evaluated into further detail with respect to performance. This enables the breeders to efficiently identify new traits or better versions of existing traits which then can be quickly introgressed into their breeding lines. This approach allows the effective exploitation of germplasm without the impossible task to have to phenotype all accessions.

Stage 3: Breeding By Design

In the previous paragraphs we have demonstrated that the application of markers in breeding can not only improve existing selection processes but can aid in creating novel varieties bearing new characteristics of agronomical importance. Extending on these capabilities, the understanding of the genetics basis of *all* agronomically important characters and the allelic variation at those loci, would enable the breeder to design superior breeding lines 'in silico'. This concept, to which we refer to as 'Breeding by Design' uses two types of mapping methods to generate the knowledge required:

- Mapping traits by using populations which simplify complex characters, such as Introgression Libraries (see also above in: Marker Assisted Breeding of polygenic traits);
- Mapping of traits and allelic variation by linkage disequilibrium (LD) mapping on germplasm;

These mapping approaches will be explained in more detail in a separate paper (Peleman and Rouppe van der Voort, in preparation). Eventually, the knowledge of the map positions of all loci of agronomic interest, the allelic variation at those loci, and their contribution to the phenotype will enable the breeder to design superior genotypes consisting of a combination of favorable alleles at all loci. Since the positions of all loci of importance are mapped precisely, recombination events can be accurately selected using flanking markers to collate the different favorable alleles next to each other. Software tools will be built that determine the optimal route to generate those mosaic genotypes by crossing lines and using markers to select for the specific recombinations that will eventually combine all those alleles. Since this is a precise process, selection by phenotyping can be reduced or even omitted during the process. Only the eventually obtained superior varieties will need to be evaluated for field performance in different environments.

In our view, the application of Breeding by Design in the coming decade(s) will represent a major breakthrough in breeding by enabling the creation of varieties with considerably improved characteristics such as yield, nutritional content, quality and disease resistances.

Acknowledgements

The authors gratefully acknowledge Henk Verbakel, Rolf Mank, Dr. Hans Dons, Dr. Jonathan Crouch (ICRISAT) and Dr. Tom Gerats (University of Nijmegen) for their contributions and discussions during the development of this paper.

References

Eshed, Y., Zamir, D. 1995. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141: 1147 – 1162.

Frary A., Clint Nesbitt, T., Frary, A., Grandillo, S., van der Knaap E., Cong B., Liu J., Meller J., Elber R., Alpert K. B., Tanksley S.D. 2000 *fw2.2*: A quanititative trait locus key to the evolution of tomato fruit size. Science 289: 85 – 88.

Jansen, J.P.A. 1996. Aphid resistance in composites. International application published under the patent cooperation treaty (PCT) No. WO 97/46080.

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

Peleman, J.D. and Rouppe van der Voort, J., Breeding by Design, in preparation.

Rouppe van der Voort, J., van der Vossen, E., Bakker, E., Overmars, H., van Zandvoort, P., Hutten, R., Klein Lankhorst, R. and Bakker J. 2000. Two additive QTLs conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters Theor. Appl. Genet 101 :1122-1130

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

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

Tanksley, S.D. and McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277: 1063 – 1066.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E.S. 2001. Dwarf8 polymorphisms associate with variation in flowering. Nat. Genet. 28: 286-289.

Hintum, Th.J.L. van, H. Verbakel, I.W. Boukema, E.C. de Groot and J. Peleman, (submitted) Wild germplasm in genebank collections: Molecular genetic diversity of wild *Lactuca* in the CGN collection. Submitted; see also: http://www.cgn.wageningen-ur.nl/pgr/research/lettuce/.

Wye, C., Rouppe van der Voort, J. and Peleman, J. 2000. Reverse QTL mapping (RQM), an efficient approach for high resolution mapping of QTL. Poster #.. Plant & Animal Genome Conference VII, San Diego.

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