USE OF MARKERS IN PLANT BREEDING: LESSONS FROM GENOTYPE BUILDING EXPERIMENTS

F. Hospital¹, A. Bouchez², L. Lecomte³, M. Causse³ and A. Charcosset¹

¹Station de Génétique Végétale, INRA-UPS-INAPG, Ferme du Moulon, 91190 Gif-sur-Yvette, France.
²Laboratoire de Génétique et d’Amélioration des Plantes, INRA, Domaine de Brunehaut, 80200 Estrées-Mons, France.
³Unité de Génétique et Amélioration des Fruits et Légumes, INRA, Domaine Saint-Maurice, BP 94, 84143 Montfavet cedex, France

INTRODUCTION

Molecular markers offer several possibilities to improve plant or animal breeding schemes in marker assisted selection (MAS) programs. Because of the specificities of most crop plants (mainly a shorter generation interval, easier use and derivation of pure inbred lines via selfing, and somehow lesser costs), it is easier in plant than animal breeding to plan MAS schemes based only on the individual genotypes at markers, and not backed up by classical phenotypic evaluation of the selection candidates, at least during the course of the program. Such breeding schemes are termed ‘Genotype Building’ (GB) schemes, here. A typical GB scheme involves three phases: i) phenotypic evaluation before the start of the MAS program, so that phenotypic effects are associated with some marker genotypes (e.g., through QTL detection or multiple regression); ii) one or more generations where selection is only for marker genotypes previously detected as favorable in phase i), and no phenotypic re-evaluation of the effects associated to markers is performed; and finally iii) at the end of the MAS program, phenotypic evaluation is performed in order to evaluate the agronomic value of the resulting progenies.

One case of GB widely used in plant breeding is marker-assisted introgression in backcross programs, where markers are used either to control one or more target genes (possibly QTL) and/or to speed up the recovery of recipient parent genome at non-target chromosomal locations. Other possible GB schemes include gene pyramiding over two or more generations, or screening of recombinant inbred lines (RIL) or doubled-haploid (DH) populations for their marker genotypes (for more details, see for example Dekkers and Hospital, 2002).

Obviously, phenotypic information may also be used in phase ii) in addition to, or in combination with, molecular information. But, compared to breeding schemes involving phenotypic evaluation at each selection step, GB schemes where selection in phase ii) is based solely on molecular information capitalize on a potentially important saving of time and/or experimental means, particularly in cases where phenotypic information is longer, more difficult, and/or more expensive to score than molecular information (e.g., testcross breeding for yield in maize involving progeny testing, malting quality in barley, breeding for disease resistance in most crops). However, such a breeding strategy assumes that the agronomic values realized in iii) will meet the expectations based upon the estimations performed in i), i.e., that effects associated to markers in phase i) are sufficiently well estimated, and
sustainable across agronomic conditions and genetic backgrounds. This assumption is likely to hold if the trait selected is rather simple, and controlled by a few major genes of large effects, hopefully well characterized by flanking markers. But, GB strategies are more risky if the trait is more complex, of medium to low heritability, and controlled by several genes (QTL) of not very large effects. In such cases, experimental validation is requested.

We wish here to review some published results of GB strategies, and focus in more details on two GB experiments performed in our labs: one involving marker-assisted introgression of QTL for two traits between maize elite lines; one involving marker-assisted introgression of QTL for fruit quality in tomato into different genetic backgrounds.

SURVEY OF PUBLISHED MAS EXPERIMENTS IN PLANTS

Few results of real MAS experiments have been published, most of them recently, and most of them in plants. Some are presented below by increasing level of complexity for the use of markers, the genes manipulated and/or the traits under control.

Using markers as simple marks to fasten the recovery of recipient genome background (background selection) in backcross (BC) introgression programs for the transfer of a single well-identified target region (direct marker) has been nicely proved efficient by the integration of the Bt transgene into different maize genetic backgrounds (Ragot et al. 1995). This confirmed the theoretical prediction that use of markers provides a gain in time of approximately 2 BC generations, compared to phenotypic selection (Hospital et al. 1992; Visscher et al. 1996). If few other results on this matter have been published, it is known that the technique is now largely used, in particular by private plant breeding companies. However, in the case of Ragot et al. (1995) it is likely that the same efficiency could have been reached with even fewer markers. More work is needed towards the optimization of the information provided by markers in this context, in order to reduce the costs (Visscher 1996; Servin and Hospital 2002).

Other experimental reports for the manipulation of known genes with indirect (linked) markers include ‘pyramiding’ of several major resistance genes in rice, from near-isogenic lines (NIL), each carrying only one gene, into a common background (Huang et al. 1997; Hittalmani et al. 2000). In all cases, control of the target genes by indirect linked markers was successful, as later checked by phenotypic assay of resistance. Huang et al. (1997) pyramidmed four genes for blight resistance into different combinations (2, 3 or 4 genes) that exhibited higher level of resistance and/or wider spectrum than the original parents. Moreover, some pyramidmed lines showed resistance to pathogen races to which all parents were susceptible. Results were also generally successful for Hittalmani et al. (2000), who pyramidmed 3 genes for blast resistance into different combinations. However, in this case some multiple-genes combinations did not perform any better than the single-gene one, indicating that a good knowledge of the spectrum of gene effects is necessary prior to performing the MAS program. In any case, pyramidming multiple resistance genes is a valuable step towards more durable and stable crop resistance, that could hardly be achieved without the use of marker-based selection, because epistasis and/or the masking effects of genes limit the efficiency of conventional (phenotypic) breeding.
methods. Moreover, use of markers not only alleviates this limit, but also provides a better understanding of these gene interactions.

Experimental results of MAS for the manipulation of QTL (not known major genes) are more contrasted. Toojinda et al. (1998) introgressed 2 QTL for stripe rust resistance in barley, through one backcross followed by one haplo-diploidisation with selection on marker genotype and phenotype, into a genetic background different from the one used to map QTL. Both QTL were confirmed, and additional QTL were detected in the new background, including some resistance alleles brought by the susceptible parent. Probably those alleles were fixed in the mapping population, but this illustrates the importance of the genetic background, both for QTL detection and MAS. Han et al. (1997) manipulated 2 QTL for a component of malting quality in six-row barley, a trait that is very difficult and costly to work phenotypically. They screened and selected DH lines with four different strategies: i) phenotype alone, ii) marker genotype alone, iii) genotype followed by phenotype in tandem selection, or iv) genotype and phenotype combined in an index. This either on a single-trait, or a multiple-trait basis. Results were successful for one QTL, but not for the other QTL, for which tandem and combined selection based on both marker genotype and phenotype did not perform any better than selection on phenotype alone, probably because the location of the QTL was inaccurate. However, the authors point out that, even not performing any better, tandem selection provides a valuable gain in time and efforts, compared to phenotypic selection. Lawson et al. (1997) introgressed four target chromosomal regions containing five QTL for pest resistance (acylsugar accumulation) from wild tomato into cultivated tomato. Starting with the introgression lines of Eshed and Zamir (1995), each carrying one target region, they performed three backcrosses followed by one intermating generation to obtain progenies homozygous for the resistance alleles at the five QTL. Selection was based on both marker genotype and phenotype. The introgression of the four regions was successful at the genomic level. However, the level of acylsugar accumulation in the progenies introgressed for the five QTL was lower than expected, and in particular lower than that of the interspecific F₁ hybrid, indicating that some genetic factors (QTL) of the accumulation were missing, either lost or not controlled in the program. Shen et al. (2001) manipulated four QTL for drought resistance (root depth) in rice, a trait that is very difficult to manage phenotypically. Starting from DH lines, they produced a number of BC₁F₃ lines, each introgressed for one or two QTL at most, using selection on marker genotype alone, not phenotype. They re-detected and fine mapped the QTL in the progenies. Among the four QTL, one exhibited the expected effect in the progenies, one was finally revealed as a false-positive, one segment was shown to contain in fact two QTL in repulsion phase (+/-) that reduced its expression, and one segment did not exhibit the expected effect, either because the QTL was lost in the program, or because its effect was masked by epistatic interactions. This again highlights the problems linked to the precision of the initial QTL detection with regard to the position and effect of the QTL, and the effect of possible epistatic interactions on the expression of the QTL in the progenies. Ribaut et al. (2002) introgressed five target regions containing QTL for drought tolerance (reduction of ASI) in maize. The results depended on the condition of the phenotypic assay of the progenies: under stress conditions (drought), the introgressed progenies exhibited a reduced ASI, while the introgression had no visible effect in the absence of stress. Zhu et al. (1999) screened DH lines of barley for the presence of several QTL for yield, a very complex composite trait, based on selection for marker genotype alone. They evaluated phenotypically the progenies in five
environments, including four locations and two years. The results indicate that the position of the QTL were confirmed as correct in the progenies. However, the effect of the QTL on the progenies were often different in magnitude and sign from the expectation. Moreover, the authors detected epistatic interactions between QTL, as well as numerous GxE interactions. They conclude that selection for complex traits should focus on allelic combinations (based on epistatic interactions) rather than on individual QTL effects.

INTROGRESSION OF QTLS FOR TWO TRAITS IN MAIZE
An experiment performed at Gif-sur-Yvette included: (i) detection of QTL for three quantitative traits in a cross between two elite maize inbred lines; (ii) marker-assisted introgression of the favorable alleles at three detected QTL from one inbred parent into the genomic background of the other parent; and (iii) agronomic evaluation of the effect of introgression and re-estimation of the individual QTL effects in the new genetic background (Bouchez et al. 2002).

Earliness and yield are the two major traits of interest for maize breeding in northern Europe. We focused on two earliness-related traits, silking date (SD) and grain moisture at harvest (GM), and one yield-related trait: dry grain yield (DGY). Two elite inbred lines were chosen for their complementarity with respect to these traits: one flint line (F2) for its earliness and one indent line (Io) for its high yield potential. Recombinant inbred lines were developed through successive self-fertilizing generations, in order to identify QTL for these traits. QTL detection was completed in 1992 on 96 F5 recombinant inbred lines (RIL) from the F2 x Io population. RIL were crossed to two inbred testers (F252 and Co255). Hybrid families were evaluated at two locations in France: Gif-sur-Yvette (North) and Clermont-Ferrand (Center).

QTL were detected for the three traits of interest. A molecular ideotype was designed subsequently in order to maximize yield while maintaining, or even improving, the earliness level. To restrain experimental cost, we chose to limit the introgression to three segments. The F2 type segment on chromosome 10 is expected to improve yield and earliness (i.e., decrease SD and GM) simultaneously. The F2 type segment on chromosome 8 is expected to confer a major decrease in SD and GM (respectively -1.8 days and -1.6% with tester F252). The F2 type segment on chromosome 5 is expected to improve yield but to increase SD and GM. Introgression of these three F2 segments in Io background should provide an elite line of Io type with increased yield (up to +12.4 qx/ha), decreased silking date (down to -1.6 days) and decreased grain moisture (down to -1.6%).

In the RIL population, we looked for the best genotype to start the introgression with, based on the ideotype. This selected RIL (#89) was then crossed with recipient line (Io). This progeny was considered equivalent to a BC1 because the RIL already contains an expected 50% of recipient genome. This BC1 progeny was backcrossed to the recipient line to produce a ‘BC2’ population. One selected BC2 individual was backcrossed again to produce a ‘BC3’ population. Finally, one selected BC3 individual was selfed to fix the QTL segments in homozygous donor state, producing the ‘BC3S1’ population. Within this population, we looked for the progeny closest to the defined ideotype. Note that the BC3S1 population was also used as a whole to re-detect QTL. At each cycle, marker-assisted selection was in two step. 1) The
presence of donor type alleles on QTL segments was checked by three markers selected to delimit a zone of 30 to 35 cM around each QTL to take into account the confidence interval around the estimated position. Three markers are sufficient to control such intervals, and insure a probability above 96% of ‘not losing’ the donor allele at the QTL for each controlled segment at each generation (Hospital and Charcosset 1997). 2) Background selection was achieved with one to three markers on each non-carrier chromosome, to control return to homozygous recipient type. For carrier chromosomes, no strong background selection was applied, because we had little confidence on QTL positions. Only a few markers were checked on carrier chromosomes and return to recipient type on QTL vicinity was never favored with regard to return to recipient type on parts of the genome unlinked to the QTL segments. For the BC$_2$ and BC$_3$ population, 175 individuals were genotyped. With this population size, the risk of not obtaining at least one individual carrying donor alleles at all foreground selection markers controlling the QTL segments is reduced below 1% (Hospital and Charcosset 1997). With selfing the expected frequency of the target genotype at each locus is only $\frac{1}{4}$ compared to $\frac{1}{2}$ with backcross. Also, background selection was proved to be more efficient in later generations (Hospital et al. 1992). Therefore, an enlarged population of 250 BC$_3$S$_1$ individuals was genotyped.

We estimated as precisely as possible the recipient genome content (RGC) of the individuals selected at each generation of the introgression scheme using the program MDM (Servin et al. 2002). This computation includes not only the marker genotypes at the generation considered (if available), but also the marker genotypes at previous and/or following generations (if informative). This permits to make the most of the genotyping information available for the complete breeding scheme and provides a better estimate of genome contents. For the chromosomes carrying the QTL segments, the RGC of the individuals selected at each generation of the introgression scheme are quite lower than expected without background selection (i.e., random selection among carriers). Conversely, return to recipient genome on non-carrier chromosomes is quite satisfactory. RGC in BC$_3$ (98%) is almost 5 RGC(%) above expected value. Note that a RGC of 98% would have been reached only in BC$_5$ if no background selection on markers had been applied. Hence, the gain is about two BC generations, which is again consistent with theoretical results. There is no gain in BC$_3$S$_1$ because no background selection could be performed at that generation. Overall, despite the low RGC of carrier chromosomes, the total recipient genome content of the BC$_3$S$_1$ individual selected is above expectation. Compared to the selected RIL, the marker assisted introgression scheme leads to a gain of +48.5 RGC(%), which is satisfactory.

The whole BC$_3$S$_1$ population was evaluated in order to estimate (i) the agronomic effect of introgression and (ii) the individual effect of each QTL introgressed in the final genetic background. Among BC$_3$S$_1$ plants, 217 were selfed in order to produce BC$_3$S$_1$;$_2$ families. These were crossed to tester line F252, yielding a total of 217 hybrid families. These were planted in 1997 and 1998 at three French locations: Clermont-Ferrand, Gif-sur-Yvette, and Mons (northern France). Mean of the BC$_3$S$_1$ population was highly significantly different from recipient parent Io for all traits. Introgressed families displayed on average earlier flowering than Io (-2.3 days on average) and slightly lower grain moisture at harvest (-0.9% on average). Concerning yield, introgression was unfavorable in most locations (-5.6 q/ha on average),
with the exception of one trial. Genetic effect within the BC$_3$S$_1$ population was highly significant for all traits and conditions. Genetic variance was significant but low when compared to those observed in the RIL population. This is consistent with fixation of a large fraction of the genome (more than 70%). Comparison of the introgressed family that was the closest to the ideotype (BC$_3$S$_1$ #194) with recipient line Io illustrates a major effect of introgression for the three traits of interest. For earliness traits, introgression effect is consistent in sign with that expected from initial QTL results. Magnitude of introgression effect is close to that expected for GM (-1.9% vs. -1.6%) and higher for SD (-4.1 days vs. -1.6 days). On the contrary, introgression had a negative effect on DGY (-12.9 qx/ha), opposite to what was expected (+12.4 qx/ha).

Precision of the initial QTL detection, pleiotropy and changes in environmental conditions are possible causes of the difference. Their possible role was investigated by reanalyzing the initial (RIL) data with a more refined statistical approach (Composite Interval Mapping, which was not possible in 1992) and comparing the results with those of QTL detection in the final BC$_3$S$_1$ population. No evidence for epistasis between QTL was found in either the RIL or the BC$_3$S$_1$ population. For SD, the unfavorable effect of F2 on chromosome 5 (+1.7 days in RIL, Table 3) was not confirmed within the BC$_3$S$_1$ population. For GM, all QTL detected in the RIL population were confirmed after introgression. More discrepancies were observed for DGY. For chromosomes 5 and 10, QTL effects expected to be positive according to RIL results turned out to be negative after introgression and therefore affected negatively the yield performance of individual BC$_3$S$_1$ #194. For chromosome 8, earliness QTL which showed no pleiotropic effect on yield within the RIL population finally displayed a negative pleiotropic effect on yield within the BC$_3$S$_1$ population and also affected negatively the yield performance of individual BC$_3$S$_1$ #194. This illustrates (i) that DGY is very sensitive to GxE effects and (ii) that the pleiotropic effect of earliness QTL on yield is highly dependent on environmental conditions.

TRANSFER OF QTL FOR TOMATO FRUIT QUALITY TRAITS INTO DIFFERENT GENETIC BACKGROUNDS
Tomato fruit quality is becoming of paramount importance for consumer acceptance. However, organoleptic quality is a complex characteristic, with several components, some of them being antagonistic with physical traits (e.g. fruit weight and sugar content). A program of QTL detection for fruit quality traits has been achieved in Montfavet. A range of 144 recombinant inbred lines of tomato derived from a cross between a cherry tomato line (C) chosen for its good taste and aromatic intensity and a large-fruited line (L) with common taste was analyzed with segregating molecular markers. An almost saturated map was constructed with RFLP, AFLP and RAPD marker (Saliba-Colombani et al. 2000). Each line was also evaluated for fruit physical (fruit size, color and firmness) and chemical (sugar, pigment and acid contents) characteristics. Taste and texture were determined by descriptive sensory profiling. Taste was analyzed through sweetness, sourness, the overall aroma intensity, together with candy, lemon, citrus fruit and pharmaceutical aroma. Firmness, meltiness, mealiness, juiciness and difficulty to swallow the skin characterized texture. A wide range of overall variation was shown for all the attributes and significant differences among lines were detected. The overall aroma intensity was positively correlated with sweetness and sourness, as well as with lemon, candy
and citrus fruit aromas. It was negatively correlated with mealiness. A negative correlation was detected between fruit weight and dry matter weight. Molecular markers were used to map QTL (Saliba-Colombani et al. 2001; Causse et al. 2001). One to five QTL were detected per trait. The proportion of phenotypic variation explained by each QTL ranged from 8% to 45%. QTL were mainly distributed on chromosomes 1, 2, 3, 4, 9, 11 and 12 and a few chromosome regions appeared to explain a major part of the variation for most of the traits.

Most of the favorable alleles for organoleptic quality came from the cherry tomato line, showing the potential usefulness of this line for tomato organoleptic quality improvement. Consecutively, a marker-assisted backcross has been set up in order to transfer five regions of the cherry tomato line C with the largest effects on fruit quality into 3 different recurrent lines: L, which is the other parent of the population where QTL were detected, and two other lines, B and D. The five regions were chosen according to the QTL effects and their involvement in complementary quality traits. When four of them were only controlling organoleptic quality traits, the last one (the chromosome 2 region) was controlling antagonistic traits. Indeed, selecting the cherry allele at the chromosome 2 region implied selecting a reduced fruit weight together with an improved organoleptic quality. Thus, for the three other fruit weight QTL detected on chromosomes 3, 11 and 12, recurrent parent alleles (providing bigger fruits) were selected. Introgression of the five QTL regions was performed through three backcross generations followed by two selfing generations, leading to BC3S1 progenies.

The size of the regions to transfer was estimated relatively to the size of the confidence intervals evaluated by Mapmaker/QTL. A LOD decrease of 1.5 was chosen, approximately corresponding to a 5% type I risk (Mangin et al. 1994). To optimize the population size as well as the number and location of markers to be used at each generation, probability calculations were performed, based on the analytical formulas proposed by Hospital and Charcosset (1997). Thus, one to three markers were chosen to control each region, leading to a foreground selection realized with 12 markers. The probability to control each QTL varied from 0.95 to 0.99 and the overall probability was estimated to 0.86, leading to study at least 267 individuals in order to have a 5% risk to lose one QTL. Actually, 300 individuals were genotyped per generation, only the 3 to 5 individuals having the donor allele at all the 12 markers on the five segments were selected, and only one was retained at each generation, based on its genetic background (four to five segregating markers on non-carrier chromosomes allowed this last selection). After selfing generations, QTL segments were fixed in different combinations in order to study interactions between QTLs (BC3S1 individuals will be tested in 2002).

Impact of genetic background on QTL detection was tested in 2001 through the evaluation of respectively 106 and 103 individuals of BC3S1(C×L) and BC3S1(C×B) populations, for which the five selected regions were segregating. Overall, QTL detection appeared strongly affected by genetic background. Pooled over four physical traits (weight, firmness, color and locule number) and five chemical traits (dry matter weight, soluble-solids content, sugar content, titratable acidity and pH), number of QTL detected were 22, 22, and 25 for populations RIL(C×L), BC3S1(C×L) and BC3S1(C×B), respectively. Ten QTL of these were common to the three populations, but the magnitude of QTL effects varied with the genetic background considered. Also, seven were common to only RIL(C×L) and BC3S1(C×L) populations, suggesting that these QTL did not express in B genetic background. At least, while only two
QTL were specific of BC$_3$S$_1$(C×L) population, eleven were specific of BC$_3$S$_1$(C×B) population, so new QTL have to be considered when the genetic background is different.

**CONCLUSION**

Taken as a whole, results of GB experiments available so far indicate that the efficiency of MAS is highly dependent on the complexity of the genetic architecture of the trait(s) of interest. For complex traits controlled by QTL with pleiotropic effects highly affected by environment, it appears necessary to have an accurate evaluation of QTL effects in varying environments before initiating an introgression program of favourable alleles. It might also be risky to perform selection based solely on markers, without confirming the estimated effects by phenotypic evaluation at some step during the introgression process. In such situations, marker-assisted recurrent selection schemes with periodic re-evaluation of marker-trait associations may be recommended (Hospital et al. 1997). Experimental validation of such strategies is still in progress (Moreau et al. 2001).

**REFERENCES**