PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2005) **360**, 1503–1511
doi:10.1098/rstb.2005.1670
Published online 7 July 2005

Selection in backcross programmes

Frédéric Hospital*

INRA, UMR de Génétique Végétale, Ferme du Moulon, 91190 Gif-sur-Yvette, France

Backcrossing is a well-known and long established breeding scheme where a characteristic is introgressed from a donor parent into the genomic background of a recurrent parent. The various uses of backcrossing in modern genetics, particularly with the help of molecular markers, are reviewed here. Selection in backcross programmes is used to either improve the genetic value of plant and animal populations or fine map quantitative trait loci. Both cases are helpful in our understanding of the genetic bases of quantitative traits variation.

Keywords: backcross breeding; marker-assisted selection; quantitative trait loci; isogenic lines; congenic strains

1. INTRODUCTION

Backcrossing is a well-known and long established breeding scheme where a characteristic is introgressed from a donor parent into the genomic background of a recurrent parent. The characteristic could be a trait, a gene or even an anonymous locus or chromosome segment. In successive generations, progeny are selected for the characteristic of interest and then backcrossed to the recurrent parent. This ensures that the proportion of genome from the donor parent tends to zero as generations accumulate, except for the part hosting the characteristic of interest. The objective is to reduce the latter to the smallest size necessary. If selection is applied for the desired characteristic only, then the proportion of donor genome is expected to be reduced by one-half (50%) at each generation, except on the chromosome holding the characteristic. On this chromosome, the rate of decrease is slower (Hanson 1959; Stam & Zeven 1981; Naveira & Barbadilla 1992) resulting in *linkage drag*. Obviously, if selection can also be applied against the donor genome proportion, then its rate of decrease can become faster. Selection on phenotypic resemblance to the recurrent parent (or against the donor) has long been used by breeders. Selection can also be based on molecular marker alleles typical of either parent. Historically, this was among the first suggested uses of molecular markers to assist breeding programmes (Tanksley & Rick 1980; Beckmann & Soller 1983; Burr *et al.* 1983). Reduction of linkage drag is the most difficult goal to achieve because of the selection for the target. Hence, this is where use of marker-assisted selection (MAS) is most rewarding (Young & Tanksley 1989; Hospital 2001).

Backcross and introgression are useful for genetic improvement in breeding programmes. Backcrossing is also useful to dissect the genetic architecture of quantitative traits because it isolates a gene, or chromosomal region, in a different genetic background (the genetic background of the recurrent parent). In fact, it is one of the few reliable methods to validate the additive effect of a quantitative trait locus (QTL) or a candidate gene. In addition, backcrossing could be used prior to, or in conjunction with, QTL detection to increase the precision of QTL mapping. Here, I will review the various ways in which the backcross breeding scheme can still be useful in the field of modern genetics and highlight W. G. Hill's contributions to the field.

As will be seen, the various methods described in the literature have different names, although the same principles hold true for all variants of the method. Names not only vary between organisms and species but also within species, which makes it sometimes difficult to search and link articles otherwise addressing the same questions. More seriously, different names for the same idea in different areas of genetics have led to an unfortunate compartmentalization of the corresponding bodies of literature. In particular, some methodological advances in plant and animal breeding seem to remain unknown to human and animal model geneticists, although I am not the first to point this out.

Hence, in preparing this review I have tried my best to exhaustively research the various terms that could hide recent developments, particularly those linked to the use of molecular markers. I have tried to gather similar methods around a few basic ideas. Given that some of these results might save time in the search for the genetic bases of human diseases, I would be happy if this effort could contribute to bridging the gap between geneticists.

2. BACKCROSSING AND GENETIC IMPROVEMENT

(a) *Optimization of marker-assisted backcrossing*

In addition to conventional breeding methods, various aspects of the use of molecular markers (for controlling the target genes, accelerating the recovery of recurrent genome or reducing linkage drag) to improve the efficiency of introgression in backcross breeding

* (fred@moulon.inra.fr).

One contribution of 16 to a Theme Issue 'Population genetics, quantitative genetics and animal improvement: papers in honour of William (Bill) Hill'

programmes have been investigated from a theoretical standpoint in recent years. These were reviewed recently (Visscher *et al.* 1996; Whittaker 2001; Dekkers & Hospital 2002; Hospital 2003) and will not be detailed here. Whether it is called marker-assisted introgression or marker-assisted backcross, use of markers in backcross breeding programmes is efficient. As far as theory is concerned, relatively simple calculations based on classical Mendelian genetics (probability of recombination) and simulations show that the use of markers may improve backcross breeding at all levels. Several studies have shown that selection on markers for the recovery of genetic background provides a gain in time equivalent to about two backcross generations. Selection against genetic drag can save tens of generations (Young & Tanksley 1989), not necessarily at high cost (Hospital 2001). This is true even if the target gene is in fact a QTL located with a given error on the genetic map (Visscher *et al.* 1996; Hospital & Charcosset 1997). However, the number of targets is then limited. It is generally not possible to introgress more than four or five QTL, even with the largest population sizes. However, this assumes that the QTL is a 'true' QTL, that is, that it has an effect on the trait of interest (not a false positive) and that this effect is sustained over the breeding programme; specifically, the effect will be unmodified by changes of genetic background, environment or epistatic relationships with other genes. This assumption is not always true for all QTL, as will be discussed below when reviewing real-world QTL introgression experiments.

Methodological optimization of backcross programmes is still needed for two main purposes: (i) the scattered theoretical results from particular, sometimes antagonistic optimization questions should be integrated into comprehensive breeding scenarios that breeders could readily apply to fulfill their breeding objectives (which are possibly constrained by factors such as time, cost, organism biology and molecular techniques). Some attempts have already been made in this direction (Frisch *et al.* 1999; Ribaut *et al.* 2002*a*,*b*; Servin 2003; Stam 2003) but this should be pursued further. (ii) Optimization efforts should concentrate not only on the average value of backcross progeny but also on their variance. The variance in genomic composition of backcross progeny sharing the same genotype at selected molecular markers at the end of the programme is an important criterion. It controls the number of individuals that should be genotyped (Visscher 1996; Servin 2005). Both studies are based on an elegant analytical derivation by Hill (1993) which is of wider interest than just for backcross breeding schemes (e.g. McElroy 1999; Perez-Enciso & Varona 2000).

(b) *The animal view*

'What if Mendel had studied sheep?' asks Hill (2001) in an amusing book review. Indeed, working on animals instead of plants sometimes makes a big difference for the geneticist. For example, although one of the nicest animals (pig) is a result of introgression (Giuffra *et al.* 2000), marker-assisted introgression is more difficult in livestock than in plants for several reasons. The number of offspring is much smaller in animals, inbreeding is a problematic issue and animal breeding is more costly. Hence, it is generally not possible to screen hundreds of individuals and select only one or a few for backcrossing, as is usual in plants. In addition, the generation interval is much longer in animals (quantitative geneticists treat trees as animals rather than plants). Animal breeders have to take into account the increasing discrepancy of genetic value for the non-introgressed traits between the introgressed and the non-introgressed populations. Such 'genetic lag' happens when introgression populations cannot be selected for traits other than the target of introgression because of a small population size. This raises specific questions (Gama *et al.* 1992; Groen & Smith 1995; Koudande *et al.* 1999; Visscher & Haley 1999; Van der Waaij & Van Arendonk 2000; Wall *et al.*, submitted).

(c) *Marker-assisted introgression: experimental results*

Published results of marker-assisted introgression or any other type of MAS are still rare but have recently begun to accumulate at an increasing rate. It might be time to try drawing tentative conclusions, at least for plant breeding. Animal results are still very few (e.g. introgression of the *naked-neck* gene in chickens (Yancovich *et al.* 1996)). Of course, more results are available for animal models such as the mouse but these will be reviewed in §2d. Results reported here concern the use of markers for genetic improvement. As the topic of this review is backcrossing, it concentrates only on introgression experiments. Experimental results for other MAS strategies (population screening, recurrent index selection etc.) are also available, though the general conclusions are the same.

There are different ways to organize the published results available. Hospital (2003) goes from 'simple' to 'complex' schemes or traits, while Bernardo (2002) goes from 'successful' experiments to those with limited success and even 'failure'. However, the references rank in approximately the same order in both classifications and this is probably part of the answer: the rate of 'failed' or better said 'unexpected' (see below) results increases with the complexity of the studied traits.

Relatively 'simple' and certainly highly successful results start with the integration of the *Bt* transgene into different maize genetic backgrounds (Ragot *et al.* 1995). In this case, there is a single target, which is a well known transgenic construction, so the 'marker' equals the target without recombination. This confirmed the theoretical prediction that the use of markers to speed up the recovery of recipient genome background provides a gain in time equivalent to two generations of backcrossing. Although few other respective results have been published, the technique is now largely used, particularly by private plant breeding companies.

Other successful experiments report the manipulation of known genes (not transgenes) with indirect (linked) markers. This includes 'pyramiding' of several major resistance genes in rice, from near-isogenic lines (NIL, see §3a) that each carry only one gene, into

a common background (Huang *et al.* 1997; Hittalmani *et al.* 2000).

Finally, a few successful reports concerned unknown, QTL genes. Toojinda *et al.* (1998) successfully introgressed two QTL for stripe rust resistance in barley into a genetic background different from the one used to map the QTL. The effects of both QTL were confirmed and additional QTL were detected in the new background, including some resistance alleles brought in by the susceptible parent. Note that it is unclear how to assess the latter observation. It should be considered as more than successful from the breeder's perspective. However, would someone concerned with a more fundamental understanding of the genetic bases of quantitative traits consider such unexpected results 'successful'? Probably the answer relies on which genes were polymorphic in both populations. Chee *et al.* (2001) also report the successful transfer of a QTL for grain protein concentration in wheat into a different genetic background. Ahmadi *et al.* (2001) successfully introgressed two QTL for resistance to yellow mottle virus in rice. Yousef & Juvik (2002) successfully selected on three markers linked to QTL that enhanced seedling emergence in sweet corn.

The rate of success starts to decrease for introgression of larger numbers of target QTL. If Stuber (1994) claims success in increasing grain yield in maize lines by introgression of six favourable chromosome segments, it must be noted that none of the improved lines had all six segments together. Starting with the introgression lines of Eshed & Zamir (1995), Lawson *et al.* (1997) introgressed four target chromosomal regions containing five QTL for pest resistance (acylsugar accumulation) from wild tomato into cultivated tomato. The introgression of the four regions was successful at the genomic level. However, the level of acylsugar accumulation in the progeny introgressed for the five QTL was lower than expected, and, in particular, lower than that of the interspecific F_1 hybrid. Sebolt *et al.* (2000) performed marker-assisted backcrossing of two QTL for seed protein concentration in soybean. Only one QTL was confirmed in $BC_3F_{4:5}$ progeny (these are common notations in plant breeding: BC_iF_j refers to progeny obtained after i backcrosses followed by j selfing generations; $F_{j:k}$ means that the value of F_j individuals are estimated from the average value of their F_k progeny with $k=j+1$). When that QTL was introgressed in three different genetic backgrounds it had no effect in one background.

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 doubled haploid lines, they produced a number of BC_3F_3 lines, each introgressed for one or two QTL at most. Among the four QTL, one exhibited the expected effect in the progeny, one was finally revealed as a false positive, one segment was shown to contain two QTL in repulsion phase (+/−) that reduced its expression and one segment did not exhibit the expected effect.

Ribaut *et al.* (2002*a*,*b*) introgressed five target regions containing QTL for drought tolerance (reduction of anthesis-silking interval (ASI)) in maize. The results depended on the condition of the phenotypic assay of the progeny: the introgressed progeny exhibited a reduced ASI under stress conditions (drought) but the introgression had no visible effect in the absence of stress.

Bouchez *et al.* (2002) performed the introgression of favourable alleles at three QTL for two traits (earliness and yield) between maize elite lines with marker-assisted backcrossing. They showed that the use of markers to improve background selection is efficient, even with few markers, especially on non-carrier chromosomes. Foreground selection on markers to control the three target regions without the help of phenotypic assay was also efficient. However, results of the phenotypic evaluation of introgressed progeny, as well as the redetection of QTL among those progeny, depended upon the complexity of the trait under control. For the simple trait (earliness), QTL effects in the progeny were in general accordance with those expected from the original detection in the parental lines. For the more complex trait (yield), results were generally not as good as expected and one high-yielding allele putatively detected from the low-yielding parent finally exhibited an effect opposite to the expectation (i.e. reducing yield).

Lecomte *et al.* (2004) introgressed five QTL strongly involved in tomato fruit quality into three different recipient lines through MAS. The breeding efficiency varied strongly with the recipient parent and significant interactions between QTL and genetic backgrounds were shown for all the studied traits. About 50% of the QTL were confirmed in each new background and new QTL were detected. The QTL with the largest effects were the most stable.

Thabuis *et al.* (2004) transferred resistance to *Phytophthora capsici* alleles at four QTL from a small-fruited pepper into a bell pepper recipient by three cycles of marker-assisted backcrossing. Introgression was successful but a decrease of the effect from the moderate-effect QTL and of the epistatic interaction between QTL was observed.

Finally, in some cases none of the introgressed QTL had any effect, for example, three QTL for grain yield in barley (Kandemir *et al.* 2000) and three QTL for high yield in soybean (Reyna & Sneller 2001).

(d) *Why QTL introgression may produce unexpected results.*

More examples could be added but the literature cited above contains most of the information. Marker-assisted introgression is not always successful. One major limitation is not the ability of marker-based selection to produce the selection objective at the molecular level—all experiments report that genomic composition of the produced genotype is close to that predicted by theory. Rather, success of introgression depends on the ability of the target genes to exhibit the expected effects once introgressed in a new genetic background (the genetic background of the recurrent parent). Introgression alters the epistatic interactions between the target and the donor background. Then, it is essentially the additive effect of the target that can show up in the new background, unless new epistatic interactions are built up. Hence, the 'success' or

'failure' of introgression experiments may help discover whether QTL effects are mostly additive. Conversely, such experiments could be ideal for determining the extent to which epistasis affects the genetic architecture of complex traits. However, when the target QTL fails to exhibit the expected additive effect, it can be that its effect was epistatic (i.e. non-additive) or that the QTL had no effect at all (i.e. it was a false positive). Hence, we are faced with two extreme interpretations (among others that will be detailed below) of the results. It might then be more appropriate to think not in terms of 'success' versus 'failure' but rather in terms of 'expected' versus 'unexpected results' (I thank one of the reviewers of the first draft for helping me point this out more clearly).

Moreover, the rate of unexpected results seems to increase when moving from known genes to QTL, when increasing the number of targets and when dealing with more 'complex' traits (complexity comes from the large number of genes controlling the trait, interactions between genes due to linkage and epistasis, low heritability and interaction between genes and environment). Actually, many of the unexpected results refer to cases where one tried to introgress multiple QTL for yield, which is generally considered by plant breeders as one of the most 'complex' traits because it integrates most of the plant's physiological functions.

From a breeder's point of view, it is clear that it is risky to embark in a selection programme based only on markers unless the target genes are few and have large effects. In other cases, it is probably wiser to support selection decisions with phenotypic evaluation. However, note that the relative merits of marker-based and phenotypic selection must incorporate an economic perspective, so the choice depends on the species and the breeding context (for example, some traits are very difficult, very time consuming or very costly to improve by conventional selection, in which case marker-based selection is greatly valuable even though not greatly efficient).

It is more puzzling, and maybe more interesting for the fundamental quantitative geneticist who wants to elucidate the genetic bases of quantitative trait variation, to understand the reasons why some introgression experiments for quantitative trait loci failed to produce the expected effects.

The first obvious reason is that the putative QTL may in fact be a false positive. As wisely stated by Bernardo (2004) among others, '… the false discovery rate (FDR) should be kept low so that resources are not wasted in introgressing false QTL. Perhaps the success or failure in attempts to introgress QTL may be partly due to the α_C (significance) level used to identify QTL'. In addition, it is known that estimated QTL effects are generally biased for several reasons (Beavis 1994; Bost *et al.* 2001). Some groups have engaged in an extensive and valuable effort to empirically estimate the repeatability of QTL detection and correct biases in estimated effects (e.g. Melchinger *et al.* 1998; Schon *et al.* 2004).

It might be interesting to perform a quantitative survey of the published results but it does not seem that putative QTL displaying no effect after introgression is the most frequent case. Many times, the QTL is still detected after introgression but its effect is reduced. In the worst cases, it is even opposite to the expected effect, as in Bouchez *et al.* (2002). Such cases may not be explained completely by statistical error or imprecision. Another potential cause of unexpected results is the possibility of QTL by environment interactions, which have definitely been shown to exist in some cases (e.g. Ribaut *et al.* 2002*a*,*b*). Generally, genotype by environment interactions are frequent in plants but less so in animals.

Another explanation worthy of greater consideration is the possibility that the chromosomal segments detected as QTL hold not just one but several genes. Recombination between those genes would then simply modify the effect of the introgressed segments. Such an observation is actually not infrequent after fine mapping of QTL segments (e.g. Eshed & Zamir 1995; Monna *et al.* 2002; Steinmetz *et al.* 2002; Christians & Keightley 2004).

Finally, the last, though probably not the least, cause of unexpected introgression results is epistasis, either between QTL or between QTL and the genetic background. As W. G. Hill (personal communication) said, he would sooner join the club of marker assisted introgression experiments with unexpected results, as we shall certainly learn more about this cause in the near future.

Note that epistasis can be beneficial and MAS very rewarding in some cases. In a very nice experiment, Ahmadi *et al.* (2001) confirmed the epistatic relationship between two QTL that were detected in a population. Introgression lines hosting one of the QTL, but not the other, displayed no effect. Conversely, the line hosting both QTL exhibited the expected effect. In such cases, use of marker-based selection is clearly valuable because manipulating epistatic relationships by phenotypic selection only is generally very difficult.

3. BACKCROSS SELECTION AND QTL DETECTION

Because backcrossing isolates a gene or chromosomal region in a different genetic background (the genetic background of the recurrent parent), it helps to dissect the genetic architecture of quantitative traits. In fact, it is one of the few reliable methods to validate the additive effect of a QTL or candidate gene after it is putatively detected. In addition, backcrossing could be used prior to detection to increase the precision of QTL mapping or exploit wild genetic resources. Introgressing one gene into a different genetic background removes or modifies the possible epistatic interactions between that gene and the rest of the genome; it is then useful to study the additive (non-epistatic) effect of the gene. Combinations of introgression lines with different genes and different backgrounds can thus be used to study epistatic interactions.

(a) *Isogenics or congenics?*

Isogenic lines is the term used in plant genetics and congenic strains is the term used in genetics of animal models (mouse, *Drosophila*). Both refer to the same type of material. It consists of lines (plants) or strains

(mouse) that are fixed for identical genomes, except for a small part of genome that differs between lines/strains. The part of the genome that differs can be a single gene or a chromosome segment, in which case we speak of NIL. Of course, this chromosome segment can be a segment found to be hosting a putative QTL, in which case the 'QTL–NIL' will serve to confirm the QTL (Van Berloo *et al.* 2001). In addition, backcrossing can be applied further to cut a chromosomal segment hosting a putative QTL into pieces for finer mapping of the QTL. Snell (1948) pioneered this approach in his Nobel Prize winning work on the major histocompatibility complex. This procedure is now widely used in both plants and animals (Gurganus *et al.* 1999; Mackay 2001; Brouwer & St Clair 2004). In addition to validating a QTL, it can serve to investigate a QTL's effects in different genetic backgrounds (Christians *et al.* 2004), resolve linkage between QTL (Monna *et al.* 2002; Takeuchi *et al.* 2003; Christians & Keightley 2004) or check for dominance and/or epistatic relationships between QTL (Eshed & Zamir 1995, 1996; Lin *et al.* 2000, 2003; Yamamoto *et al.* 2000). Such studies should accumulate for more species and more traits in the near future and, hence, provide us with a better understanding of the genetic bases of quantitative trait variation, which may still be more complex than one thought at the beginning of the QTL revolution (Flint & Mott 2001; Mackay 2001; Barton & Keightley 2002; Christians & Keightley 2002; Stylianou *et al.* 2004).

Derivation of isogenic or congenic material can be also performed in a systematic way, without any prior knowledge of putative QTL location. In this case, a collection of isogenic/congenic material is produced, representing, if possible, most of the donor genome split into small fragments and introgressed in the recipient genome. These lines are then evaluated for quantitative traits in order to directly detect QTL by, for example, comparing the value of a line with a donor segment to the value of the recurrent parent. As stated by Stuber *et al.* (1999), from the breeder's point of view: 'a major advantage of this NIL approach is that once a favourable QTL has been identified, it is already fixed in the elite recipient line and the breeding work is essentially complete. In addition, because only a small segment of the genome of the recipient line has been modified, the enhanced line is nearly identical to the original line and the amount of field testing required is minimal. In addition, lines with favourable QTL alleles can be easily maintained and then used for pyramiding several favourable QTL alleles into a single line'. Gene pyramiding will be addressed below.

Derivation of isogenic/congenic material can take different forms and different names, depending on the breeding scheme used and on the proportion of donor genome introgressed into the recurrent parent genome. Chromosome substitution series derived from monosomic and nullisomic series have been used in plants such as wheat (Sears 1953) where polyploidy made it possible. Thanks to molecular markers, it is now possible in other species. Chromosome substitution strains, where the fraction of the genome from the donor is an entire chromosome, are becoming a central tool in mouse genetics (Nadeau *et al.* 2000; Belknap 2003; Singer *et al.* 2004). This is a valuable intermediate, where QTL segments could be first assigned to one chromosome, then possibly dissected further by recombination in additional backcrosses. The work is then based on recombinant congenics (*e.g.* Santos *et al.* 2002), interval-specific congenic strain (Darvasi 1997), genome-tagged mice (Iakoubova *et al.* 2001) and interval-specific congenic recombinant lines (Bennett *et al.* 2002). Note that *Drosophila* geneticists have already used these techniques for a long time.

In plants, in addition to NIL, the various material derived by marker-assisted introgression includes backcross inbred lines (which gives my favourite acronym here: BIL!), which are BC_1F_5 progeny obtained after self-pollinating BC_1F_1 plants for five generations by single-seed descent (Sato *et al.* 2003), chromosome segment substitution lines (Kubo *et al.* 2002) and an 'intervarietal set of part chromosome substitution lines' (Burns *et al.* 2003).

(b) *Using markers to speed congenics*

'Speed congenics' is the name used by mouse geneticists to refer to the use of molecular markers for speeding up the derivation of congenic material. Again, several strategies have been proposed: 'interval-specific congenic strain' (Darvasi 1997), 'QTL-marker-assisted counter selection' (Bennett & Johnson 1998), marker-assisted congenic screening (Collins *et al.* 2003), MAS protocol (Estill & Garcia 2000), chromosome elimination strategy (Weil *et al.* 1997) and marker-assisted congenics (Deng *et al.* 2001).

It seems quite obvious that, except for the biology of the species studied, 'speed congenics' has a lot in common with the 'marker-assisted introgression' breeding schemes outlined in §1 and that at least the theoretical work on the optimization of both schemes should benefit each other. However, that has not been the case. Although Visscher (1999) has pointed out that one of the most cited theoretical works on speed congenics is based on an incorrect treatment of recombination, the literature in mouse genetics still largely ignores the results derived in the context of plant and animal breeding, the paper by Wakeland *et al.* (1997) being one of the rare exceptions.

Conversely, the derivation of congenics could be improved by applying the methodologies developed for plants and animals. More generally, it should be possible to derive methods to produce either congenic strains or isogenic lines in a more efficient way (i.e. faster) by an optimal use of marker-based selection. Given that such populations are likely to be developed as central tools for the dissection of complex traits in numerous species (see also below), such effort would certainly be helpful.

(c) *Recurrent selection backcross schemes*

Sewall Wright (1952) had suggested using a breeding scheme with repeated backcrossing and selection on a quantitative trait to isolate genes of large effects on that trait in the recurrent genetic background. This method attracted renewed interest after Hill (1998) revisited it and is now sometimes called recurrent selection backcross (RSB). With many markers surrounding the fixed QTL segment, the NIL obtained could then be used for

fine mapping the QTL. Hill (1998) computed the probability that a QTL of specified effect remains segregating as a function of its effect on the trait, the intensity of selection and the number of generations of backcrossing. This method works best for QTL of large effects. However, as suggested by Hill (1998), interspersing a generation of *inter se* mating between each generation of backcrossing (RSBI for RSB intercross) makes it possible to apply stronger selection and, hence, to fix QTL of smaller effects. The theoretical basis for the RSB/RSBI-based QTL mapping method was further investigated by Luo *et al*. (2002) who studied its optimization and efficacy, and compared the latter to conventional interval mapping of QTL. It was concluded that RSB/RSBI does not use the same information as interval mapping and may be relevant to quantitative traits of a different genetic architecture. However, RSB still has some advantages over interval mapping, particularly in exploiting very dense marker coverage around the QTL. The authors argue that 'Given that many years of considerable research efforts to isolate genes affecting complex traits have resulted in slow progress, we would not consider the long duration of the RSB breeding program to be an expensive investment for significant improvement in mapping precision and resolution in the QTL locations that may lead directly to cloning of QTL'. In addition, they show that the precision of RSB-based mapping can only increase when increasing the duration of the scheme (number of generations), which is not always the case for other 'highly recombinant' schemes using multiple generations of intercross, for example, advanced intercross lines (Darvasi & Soller 1995). Recently, Luo & Ma (2004) gave a theoretical formulation for predicting heterozygosity of a putative marker locus linked to two QTL in an RSB scheme.

(d) *Backcross and genetic resources*

The advanced backcross QTL analysis proposed by Tanksley & Nelson (1996) is similar to the RSB scheme described above in that a QTL from an exotic resource is introgressed into 'elite' genetic background by selection. However, selection here is not necessarily for the trait of interest, but rather against the unwanted phenotypic characteristics of the donor. It could also be 'natural' selection (i.e. no conscious selection). This approach has been used with some success (Tanksley & Nelson 1996; Bernacchi et al. 1998).

More generally, exotic genetic resources may hide genes for agriculturally important traits of larger effects than those already segregating in the commercial population, which was indeed shown in some plants (Eshed *et al*. 1996). Based on this fact, there is a growing interest in using molecular markers to 'unlock the genetic potential from the wild' (Tanksley & McCouch 1997). It is then advocated to 'strengthen the resources of the research community, which is positioned between the seed banks and the commercial plant breeders, so that this community can bring about germplasm enhancement through the development of exotic introgression lines' (Zamir 2001). It can be argued that it may be more urgent to develop such libraries for inbred plants where the degree of polymorphism in commercial varieties, like tomato, is low.

In any case, this is another reason to expect the development of numerous 'introgression lines libraries' in the near future. This will require expensive community resources, large databases (Gur *et al*. 2004) and new selection methods to be able to 'pyramid' all detected QTL into new improved genetic material (Servin *et al*. 2004).

4. CONCLUSION

It is likely that the old backcross-breeding scheme still has a lot to offer the genetics community. Improved selection methods based on marker information are still sought to make it more efficient, leading to improved agricultural populations of plants and animals, finer description of the genetic architecture of quantitative traits, better understanding of epistasis between QTL and wider knowledge of genotype by environment interactions. Finally, it may help us tap into the genetic resources present in the wild relatives of most crops.

The author warmly thanks Peter Keightley, Olivier C. Martin and two anonymous reviewers for careful reading and helpful comments on earlier versions of the manuscript.

REFERENCES

Ahmadi, N., Albar, L., Pressoir, G., Pinel, A., Fargette, D. & Ghesquiere, A. 2001 Genetic basis and mapping of the resistance to rice yellow mottle virus. III. Analysis of QTL efficiency in introgressed progeny confirmed the hypothesis of complementary epistasis between two resistance QTLs. *Theor. Appl. Genet.* **103**, 1084–1092.

Barton, N. H. & Keightley, P. D. 2002 Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3**, 11–21.

Beavis, W. D. 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies. In *Proc. 49th Annual Corn and Sorghum Research Conference*, pp. 250–266. Washington, DC: American Seed Trade Association.

Beckmann, J. S. & Soller, M. 1983 Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor. Appl. Genet.* **67**, 35–43.

Belknap, J. K. 2003 Chromosome substitution strains: some quantitative considerations for genome scans and fine mapping. *Mamm. Genome* **14**, 723–732.

Bennett, B. & Johnson, T. E. 1998 Development of congenics for hypnotic sensitivity to ethanol by QTL-marker-assisted counter selection. *Mamm. Genome* **9**, 969–974.

Bennett, B., Beeson, M., Gordon, L., Carosone-Link, P. & Johnson, T. E. 2002 Genetic dissection of quantitative trait loci specifying sedative/hypnotic sensitivity to ethanol: mapping with interval-specific congenic recombinant lines. *Alcohol Clin. Exp. Res.* **26**, 1615–1624.

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D. & Tanksley, S. 1998 Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theor. Appl. Genet.* **97**, 381–397.

Bernardo, R. 2002 *Breeding for quantitative traits in plants*. Woodbury, MN: Stemma Press.

Bernardo, R. 2004 What proportion of declared QTL in plants are false? *Theor. Appl. Genet.* **109**, 419–424.

Bost, B., de Vienne, D., Hospital, F., Moreau, L. & Dillmann, C. 2001 Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. *Genetics* **157**, 1773–1787.

Bouchez, A., Hospital, F., Causse, M., Gallais, A. & Charcosset, A. 2002 Marker assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics* **162**, 1945–1959.

Brouwer, D. J. & St Clair, D. A. 2004 Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. *Theor. Appl. Genet.* **108**, 628–638.

Burns, M. J., Barnes, S. R., Bowman, J. G., Clarke, M. H., Werner, C. P. & Kearsey, M. J. 2003 QTL analysis of an intervarietal set of substitution lines in *Brassica napus*. I. Seed oil content and fatty acid composition. *Heredity* **90**, 39–48.

Burr, B., Evola, S. V., Burr, F. A. & Beckmann, J. S. 1983 Application of restriction fragment length polymorphism to plant breeding. *Genet. Eng.* **5**, 45–59.

Chee, P. W., Elias, E. M., Anderson, J. A. & Kianian, S. F. 2001 Evaluation of a high grain protein QTL from *Triticum turgidum* L. var. dicoccoides in an adapted durum wheat background. *Crop Sci.* **41**, 295–301.

Christians, J. K. & Keightley, P. D. 2002 Genetic architecture: dissecting the genetic basis of phenotypic variation. *Curr. Biol.* **12**, R415–R416.

Christians, J. K. & Keightley, P. D. 2004 Fine mapping of a murine growth locus to a 1.4-cM region and resolution of linked QTL. *Mamm. Genome* **15**, 482–491.

Christians, J. K., Rance, K. A., Knott, S. A., Pignatelli, P. M., Oliver, F. & Bunger, L. 2004 Identification and reciprocal introgression of a QTL affecting body mass in mice. *Genet. Sel. Evol.* **36**, 577–591.

Collins, S. C., Wallis, R. H., Wallace, K., Bihoreau, M. T. & Gauguier, D. 2003 Marker-assisted congenic screening (MACS): a database tool for the efficient production and characterization of congenic lines. *Mamm. Genome* **14**, 350–356.

Darvasi, A. 1997 Interval-specific congenic strain (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. *Mamm. Genome* **8**, 163–167.

Darvasi, A. & Soller, M. 1995 Advanced intercross lines, an experimental population for fine genetic-mapping. *Genetics* **141**, 1199–1207.

Dekkers, J. C. M. & Hospital, F. 2002 The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genetics* **3**, 22–32.

Deng, A. Y., Dutil, J. & Sivo, Z. 2001 Utilization of marker-assisted congenics to map two blood pressure quantitative trait loci in Dahl rats. *Mamm. Genome* **12**, 612–616.

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.

Eshed, Y. & Zamir, D. 1996 Less than additive epistatic interactions of QTL in tomato. *Genetics* **143**, 1807–1817.

Eshed, Y., Gera, G. & Zamir, D. 1996 A genome-wide search for wild-species alleles that increase horticultural yield of processing tomatoes. *Theor. Appl. Genet.* **93**, 877–886.

Estill, S. J. & Garcia, J. A. 2000 A marker assisted selection protocol (MASP) to generate C57BL/6J or 129S6/SvEv-Tac speed congenic or consomic strains. *Genesis* **28**, 164–166.

Flint, J. & Mott, R. 2001 Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat. Rev. Genet.* **2**, 437–445.

Frisch, M., Bohn, M. & Melchinger, A. E. 1999 Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci.* **39**, 1295–1301.

Gama, L. T., Smith, C. & Gibson, J. P. 1992 Transgene effects, introgression strategies and testing schemes in pigs. *Anim. Prod.* **54**, 427–440.

Giuffra, E., Kijas, J. M. H., Amarger, V., Carlborg, Ö. & Andersson, L. 2000 The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**, 1785–1791.

Groen, A. F. & Smith, C. 1995 A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *J. Anim. Breed. Genet.* **112**, 161–170.

Gur, A., Semel, Y., Cahaner, A. & Zamir, D. 2004 Real ime QTL of complex phenotypes in tomato interspecific introgression lines. *Trends Plant Sci.* **9**, 107–109.

Gurganus, M. C., Nuzhdin, S. V., Leips, J. W. & Mackay, T. F. 1999 High-resolution mapping of quantitative trait loci for sternopleural bristle number in *Drosophila melanogaster*. *Genetics* **152**, 1585–1604.

Hanson, W. D. 1959 Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. *Genetics* **44**, 833–837.

Hill, W. G. 1993 Variation in genetic composition in backcrossing programs. *J. Hered.* **84**, 212–213.

Hill, W. G. 1998 Selection with recurrent backcrossing to develop congenic lines for quantitative trait loci analysis. *Genetics* **148**, 1341–1352.

Hill, W. G. 2001 What if Mendel had studied sheep? *Nature* **414**, 586.

Hittalmani, S., Parco, A., Mew, T. V., Zeigler, R. S. & Huang, N. 2000 Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. *Theor. Appl. Genet.* **100**, 1121–1128.

Hospital, F. 2001 Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics* **158**, 1363–1379.

Hospital, F. 2003 Marker-assisted breeding. In *Plant molecular breeding* (ed. H. J. Newbury), pp. 30–56. London: Blackwell Scientific.

Hospital, F. & Charcosset, A. 1997 Marker-assisted introgression of quantitative trait loci. *Genetics* **147**, 1469–1485.

Huang, N., Angeles, E. R., Domingo, J., Magpantay, G., Singh, S., Zhang, G., Kumaravadivel, N., Bennett, J. & Khush, G. S. 1997 Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theor. Appl. Genet.* **95**, 313–320.

Iakoubova, O. A., *et al.* 2001 Genome-tagged mice (GTM): two sets of genome-wide congenic strains. *Genomics* **74**, 89–104.

Kandemir, N., Jones, B. L., Wesenberg, D. M., Ullrich, S. E. & Kleinhofs, A. 2000 Marker-assisted analysis of three grain yield QTL in barley (*Hordeum vulgare* L.) using near isogenic lines. *Mol. Breed.* **6**, 157–167.

Koudande, O. D., Thomson, P. C. & Van Arendonk, J. A. M. 1999 A model for population growth of laboratory animals subjected to marker-assisted introgression: how many animals do we need? *Heredity* **82**, 16–24.

Kubo, T., Aida, Y., Nakamura, K., Tsunematsu, H., Doi, K. & Yoshimura, A. 2002 Reciprocal chromosome segment substitution series derived from Japonica and Indica cross of rice (*Oryza sativa* L.). *Breed. Sci.* **52**, 319–325.

Lawson, D. M., Lunde, C. F. & Mutschler, M. A. 1997 Marker-assisted transfer of acylsugar-mediated pest resistance from the wild tomato, *Lycopersicon pennellii*, to the cultivated tomato, *Lycopersicon esculentum*. *Mol. Breed.* **3**, 307–317.

Lecomte, L., Duffé, P., Buret, M., Servin, B., Hospital, F. & Causse, M. 2004 Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. *Theor. Appl Genet.* **109**, 658–668.

Lin, H. X., Yamamoto, T., Sasaki, T. & Yano, M. 2000 Characterization and detection of epistatic interactions of 3 QTLs, Hd1, Hd2, and Hd3, controlling heading date in rice using nearly isogenic lines. *Theor. Appl. Genet.* **101**, 1021–1028.

Lin, H. X., Liang, Z. W., Sasaki, T. & Yano, M. 2003 Fine mapping and characterization of quantitative trait loci Hd4 and Hd5 controlling heading date in rice. *Breed. Sci.* **53**, 51–59.

Luo, Z. W. & Ma, L. 2004 An improved formulation of marker heterozygosity in recurrent selection and backcross schemes. *Genet. Res.* **83**, 49–53.

Luo, Z. W., Wu, C. I. & Kearsey, M. J. 2002 Precision and high-resolution mapping of quantitative trait loci by use of recurrent selection, backcross or intercross schemes. *Genetics* **161**, 915–929.

Mackay, T. F. C. 2001 The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339.

McElroy, D. 1999 Moving agbiotech downstream—companies must turn to effective process development as a competitive weapon in agricultural biotechnology. *Nat. Biotechnol.* **17**, 1071–1074.

Melchinger, A. E., Utz, H. F. & Schon, C. C. 1998 Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* **149**, 383–403.

Monna, L., Lin, H. X., Kojima, S. & Sasaki, T. 2002 Genetic dissection of a genomic region for a quantitative trait locus, *Hd3*, into two loci, *Hd3a* and *Hd3b*, controlling heading date in rice. *Theor. Appl. Genet.* **104**, 772–778.

Nadeau, J. H., Singer, J. B., Matin, A. & Lander, E. S. 2000 Analysing complex genetic traits with chromosome substitution strains. *Nat. Genet.* **24**, 221–225.

Naveira, H. & Barbadilla, A. 1992 The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species. *Genetics* **130**, 205–209.

Perez-Enciso, M. & Varona, L. 2000 Quantitative trait loci mapping in F-2 crosses between outbred lines. *Genetics* **155**, 391–405.

Ragot, M., *et al.* 1995 *Marker-assisted backcrossing: a practical example. Techniques et utilisations des marqueurs moléculaires,* Les Colloques, no. 72. Paris: INRA.

Reyna, N. & Sneller, C. H. 2001 Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. *Crop Sci.* **41**, 1317–1321.

Ribaut, J. M., Banziger, M., Betran, J., Jiang, C., Edmeades, G. O., Dreher, K. & Hoisington, D. 2002 Use of molecular markers in plant breeding: drought tolerance improvement in tropical maize. In *Quantitative genetics, genomics, and plant breeding* (ed. M. S. Kang), pp. 85–99. Wallingford: CABI.

Ribaut, J.-M., Jiang, C. & Hoisington, D. 2002 Simulation experiments on efficiencies of gene introgression by backcrossing. *Crop Sci.* **42**, 557–565.

Santos, J., *et al.* 2002 A new locus for resistance to gamma-radiation-induced thymic lymphoma identified using inter-specific consomic and inter-specific recombinant congenic strains of mice. *Oncogene* **21**, 6680–6683.

Sato, T., Ueda, T., Fukuta, Y., Kumagai, T. & Yano, M. 2003 Mapping of quantitative trait loci associated with ultraviolet-B resistance in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **107**, 1003–1008.

Schon, C. C., Utz, H. F., Groh, S., Truberg, B., Openshaw, S. & Melchinger, A. E. 2004 Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* **167**, 485–498.

Sears, E. R. 1953 Nullisomic analysis in common wheat. *Am. Nat.* **87**, 245–252.

Sebolt, A. M., Shoemaker, R. C. & Diers, B. W. 2000 Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.* **40**, 1438–1444.

Servin, B. 2003 Méthodes de construction de génotypes assistée par marqueurs. Ph.D. thesis, Université Paris XI, Orsay, France.

Servin, B. 2005 Using markers to reduce the variation in the genomic composition in marker assisted backcrossing. *Genet. Res.*

Servin, B., Martin, O. C., Mezard, M. & Hospital, F. 2004 Toward a theory of marker-assisted gene pyramiding. *Genetics* **168**, 513–523.

Shen, L., Courtois, B., McNally, K. L., Robin, S. & Li, Z. 2001 Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. *Theor. Appl. Genet.* **103**, 75–83.

Singer, J. B., *et al.* 2004 Genetic dissection of complex traits with chromosome substitution strains of mice. *Science* **304**, 445–448.

Snell, G. D. 1948 Methods for the study of histocompatibility genes. *J. Genet.* **49**, 87–108.

Stam, P. 2003 Marker-assisted introgression: speed at any cost? In *Proc. Eucarpia Leafy Vegetables Section Meeting* (ed. Th. J. L. van Hitum, A. Lebeda, D. Pink & J. W. Schut), pp. 117–124. Wageningen: Centre for Genetic Resources.

Stam, P. & Zeven, A. C. 1981 The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. *Euphytica* **30**, 227–238.

Steinmetz, L. M., Sinha, H., Richards, D. R., Spiegelman, J. I., Oefner, P. J., McCusker, J. H. & Davis, R. W. 2002 Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330.

Stuber, C. W. 1994 Successes in the use of molecular markers for yield enhancement in corn. In *Proc. Corn Sorghum Industrial Research Conf.* **49**, 232–238.

Stuber, C. W., Polacco, M. & Senior, M. L. 1999 Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Sci.* **39**, 1571–1583.

Stylianou, I. M., Christians, J. K., Keightley, P. D., Bunger, L., Clinton, M., Bulfield, G. & Horvat, S. 2004 Genetic complexity of an obesity QTL (*Fob3*) revealed by detailed genetic mapping. *Mamm. Genome* **15**, 472–481.

Takeuchi, Y., Lin, S. Y., Sasaki, T. & Yano, M. 2003 Fine linkage mapping enables dissection of closely linked quantitative trait loci for seed dormancy and heading in rice. *Theor. Appl. Genet.* **107**, 1174–1180.

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

Tanksley, S. D. & 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. & Rick, C. M. 1980 Isozymic gene linkage map of the tomato: applications in genetics and breeding. *Theor. Appl. Genet.* **57**, 161–170.

Thabuis, A., Palloix, A., Servin, B., Daubèze, A. M., Signoret, P., Hospital, F. & Lefebvre, V. 2004 Marker-assisted introgression of 4 *Phytophthora capsici* resistance QTL alleles into a bell pepper line: validation of additive and epistatic effects. *Mol. Breed.* **14**, 9–20.

Toojinda, T., Baird, E., Booth, A., Broers, L., Hayes, P., Powell, W., Thomas, W., Vivar, H. & Young, G. 1998 Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker-assisted line development. *Theor. Appl. Genet.* **96**, 123–131.

Van Berloo, R., Aalbers, H., Werkman, A. & Niks, R. E. 2001 Resistance QTL confirmed through development of QTL-NILs for barley leaf rust resistance. *Mol. Breed.* **8**, 187–195.

Van der Waaij, E. H. & Van Arendonk, J. A. M. 2000 Introgression of genes responsible for disease resistance in a cattle population selected for production: genetic and economic consequences. *Anim. Sci.* **70**, 207–220.

Visscher, P. M. 1996 Proportion of the variation in genomic composition in backcrossing programs explained by genetic markers. *J. Hered.* **87**, 136–138.

Visscher, P. M. 1999 Speed congenics: accelerated genome recovery using genetic markers. *Genet. Res.* **74**, 81–85.

Visscher, P. M. & Haley, C. S. 1999 On the efficiency of marker-assisted introgression. *Anim. Sci.* **68**, 59–68.

Visscher, P. M., Haley, C. S. & Thompson, R. 1996 Marker-assisted introgression in backcross breeding programs. *Genetics* **144**, 1923–1932.

Wakeland, E., Morel, L., Achey, K., Yui, M. & Longmate, J. 1997 Speed congenics: a classic technique in the fast lane (relatively speaking). *Immunol. Today* **18**, 472–477.

Wall, E., Visscher, P. M., Hospital, F. & Woolliams, J. A. 2005 Genomic contributions in livestock gene introgression programmes. *Genet. Sel. Evol.* **37**, 291–313. (doi:10.1051/gse:2005003.)

Weil, M. M., Brown, B. W. & Serachitopol, D. M. 1997 Genotype selection to rapidly breed congenic strains. *Genetics* **146**, 1061–1069.

Whittaker, J. C. 2001 Marker assisted selection and introgression. In *Handbook of statistical genetics* (ed. D. J. Balding, M. Bishop & C. Cannings), pp. 673–695. New York: Wiley.

Wright, S. 1952 *Quantitative inheritance*. London: Her Majesty's Stationery Office.

Yamamoto, T., Lin, H., Sasaki, T. & Yano, M. 2000 Identification of heading date quantitative trait locus *Hd6* and characterization of its epistatic interactions with *Hd2* in rice using advanced backcross progeny. *Genetics* **154**, 885–891.

Yancovich, A., Levin, I., Cahaner, A. & Hillel, J. 1996 Introgression of the avian naked neck gene assisted by DNA fingerprints. *Anim. Genet.* **27**, 149–155.

Young, N. D. & Tanksley, S. D. 1989 RFLP analysis of the size of chromosomal segments retained around the *tm-2* locus of tomato during backcross breeding. *Theor. Appl. Genet.* 77, 353–359.

Yousef, G. G. & Juvik, J. A. 2002 Enhancement of seedling emergence in sweet corn by marker-assisted backcrossing of beneficial QTL. *Crop Sci.* **42**, 96–104.

Zamir, D. 2001 Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* **2**, 983–989.