*Review*

# Map-based cloning of quantitative trait loci: progress and prospects

D. L. REMINGTON*, M. C. UNGERER AND M. D. PURUGGANAN

*Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614*

**Summary**

Map-based cloning has been considered problematic for isolating quantitative trait loci (QTLs) due to the confounding phenotypic effects of environment and other QTLs. However, five recent studies, all in plants, have succeeded in cloning QTLs using map-based methods. We review the important features of these studies and evaluate the prospects for broader application of the techniques. Successful map-based cloning requires that QTLs represent single genes that can be isolated in near-isogenic lines, and that genotypes can be unambiguously inferred by progeny testing. In plants or animals for which map-based cloning of genes with discrete phenotypes is feasible, the modified procedures required for QTLs should not be limiting in most cases. The choice between map-based cloning and alternative methods will depend on details of the species and traits being studied.

**Introduction**

Map-based positional cloning has been used to isolate a large number of genes that are responsible for Mendelian trait differences. The availability of numerous molecular markers, large-insert genomic libraries, and suitable mapping populations that segregate for traits of interest has made map-based cloning of genes with discrete effects theoretically straightforward, if laborious in practice, for a number of organisms. However, map-based cloning has been considered problematic for quantitative trait loci, or QTLs (Falconer & Mackay, 1996).

In map-based cloning, the marker interval containing the gene is refined by genetic mapping to a region small enough to be physically mapped, by scoring the phenotypes of individuals that are recombinant between flanking markers. Additional markers in this interval are derived from the physical map and used to delimit the recombinational breakpoints. Physical contigs or even individual large-insert clones spanning the smallest marker interval containing the gene are identified, and known or predicted candidate genes within the region are tested using complementation or other methods. The process is

tractable even in organisms with relatively large genomes if the requisite resources are available (Tanksley *et al.*, 1995). (In this review, we will refer to this approach as "map-based" rather than "positional" cloning to avoid confusion with association mapping methods, which are often described as positional cloning in the human genetics literature.)

Quantitative trait variation, however, is controlled by an unknown number of loci as well as environmental factors, so the phenotypes of individual recombinants do not provide reliable genotypic information (Falconer & Mackay, 1996; Mackay, 2001). Moreover, even if this problem can be overcome, there is no guarantee that an identified QTL represents a single gene (Graham *et al.*, 1997; Pasyukova *et al.*, 2000). These complications have led to doubts about the utility of map-based cloning for QTLs, and even to suggestions that QTL analysis in general is an inferior tool for understanding the basis of genetic variation (Nadeau & Frankel, 2000).

Nevertheless, success in map-based cloning of QTLs has been reported in five separate studies in plants in the past two years; namely *Fw2.2* (Alpert & Tanksley, 1996; Frary *et al.*, 2000) and *Brix9-2-5* (Fridman *et al.*, 2000) in tomato, *FRIGIDA* in *Arabidopsis* (Johanson *et al.*, 2000), and *Hd1* and *Hd6* in rice (Yano *et al.*, 2000; Takahashi *et al.*, 2001). These recent successes point out a need to reevaluate the

* Corresponding author. Tel: 919-515-5738. Fax: 919-515-3355.
e-mail: dlreming@unity.ncsu.edu

techniques for QTL gene discovery. Do these studies presage widespread success for an approach that until recently has been dismissed as generally unworkable? And, why are the reported successes all in plants? Is map-based cloning of QTLs likely to work in animal systems as well, or do plants possess unique characteristics that make them particularly tractable to this methodology? Our purpose in this brief review is to address these questions and suggest conditions under which map-based cloning of QTLs is likely to be a useful option for gene discovery.

### Features of successful plant QTL cloning experiments

The five reported instances of map-based cloning of QTLs are summarized in Table 1. In each case, the entire QTL effect appears to be resolvable to a single gene. Of particular interest are the effects of the isolated QTLs. The *FRIGIDA* (*FRI*) locus in *Arabidopsis thaliana* has such a large effect on the number of days to flowering in the mapping population that individual plants could be scored as discrete early and late flowering phenotypes. Three of the other four QTLs also had major effects on trait values, with the two homozygous classes differing by 20–30% of the trait means. On the other hand, *Hd6* in rice had a relatively modest effect on heading date, and was not even detected in the original QTL mapping experiment (Yano *et al.*, 1997; Yamamoto *et al.*, 2000). Thus, while map-based cloning has been used primarily to isolate major QTLs so far, this does not appear to be an essential limitation.

Two modifications to the typical map-based cloning procedure summarized above were used in most of the studies to overcome the special difficulties of identifying QTL phenotypes. In effect, these modifications converted QTLs into single Mendelian factors by minimizing both residual genetic and environmental variation. First, near-isogenic lines (NILs) were developed by repeated backcrossing of an $F_1$ plant to one of the parents. This allowed evaluation of the QTL in a nearly uniform genetic background. Fine-scale QTL mapping was then performed with first or second generation offspring ($F_2$ or $F_3$ progeny, respectively) of NILs. Second, to reduce the effects of environmental variation on trait evaluation, progeny tests were used to infer the QTL genotypes of recombinant individuals. Progeny sets had to be large enough that the QTL genotypes of the recombinants could be inferred reliably from the means and distributions of progeny phenotypic values.

QTLs were localized by fine-scale mapping to regions ranging from ∼ 150 kb for *FW2.2* and *FRI*, down to only 484 bp for *Brix9-2-5*. The sizes of these regions appear not to be closely related to genome size, which ranges from 125 Mb in *Arabidopsis* (The

Arabidopsis Genome Initiative, 2000) to about 1000 Mb in tomato (Tanksley *et al.*, 1995), but rather to the size of the fine-scale mapping population, local variation in recombination frequency, and in at least one case (*Brix9-2-5*) to serendipity. An essential requirement for map-based cloning is the ability to generate thousands of progeny in which to screen for recombinants in small genetic intervals, which is true for map-based cloning of genes with discrete phenotypes as well.

Strategies used to identify the responsible gene within the recombinationally delimited region depended on the size of the region and the available genomic resources. *Agrobacterium*-mediated transformation with one or more cosmids and/or subclones within the candidate region was used either to localize the QTL effect to a single gene or verify the candidate gene. In the case of *Brix9-2-5*, the chance isolation of QTL effects between adjacent recombination events only 484 bp apart identified a subgenic region without further characterization.

### Are plants uniquely suited to map-based QTL cloning?

The techniques used in these studies were, for the most part, based on map-based cloning strategies previously proposed for plants (Tanksley *et al.*, 1995; Yano & Sasaki, 1997). The particular challenges posed by QTLs, however, seem unlikely by themselves to prevent application of these techniques in animals. The most daunting limitation is the need to generate a large number of offspring from crosses between individual parents, clones or breeding lines, in order to generate multiple recombinants within sub-cM genetic intervals. This is equally limiting, however, for all map-based cloning experiments, not just those involving QTLs. The additional difficulties of producing NILs and progeny testing should be relatively minor in most cases compared to those of generating large linkage mapping populations. Map-based cloning has been extensively used in mice and rats, and strategies to extend it to QTLs have been initiated in both systems (Galli *et al.*, 1999; Talbot *et al.*, 1999; Nabika *et al.*, 2000; Liu *et al.*, 2001). We see no reason to believe these strategies would be any less successful in these species than in plants.

Several classes of organisms will present special difficulties for map-based cloning. It will be difficult if not impossible to produce the $F_2$ and $F_3$ families required for fine mapping in self-incompatible plants, as well as NILs with some self-incompatibility systems. Plants and animals exhibiting high levels of inbreeding depression may present similar difficulties, especially when repeated inbreeding to produce NILs is required (Remington & O'Malley, 2000). The time required to

Table 1. *Summary of features of studies reporting positional cloning of QTLs*

| Feature | Quantitative trait locus[a] | | | | |
| --- | --- | --- | --- | --- | --- |
| | Tomato *FW2.2* | Tomato *Brix9-2-5* | *Arabidopsis FRI* | Rice *Hd1* | Rice *Hd6* |
| *Trait* | fruit size | fruit sugar content | flowering time | heading date | heading date |
| *Trait variation explained* | $\leqslant 30\%$ of mean value | 30% of mean of cultivated species | Discrete classes | 20–30 days (20–25% of trait mean) | $\sim 11$ days ($\sim 10\%$ of trait mean) |
| *Fine mapping population* | $F_2$ offspring of NIL | $F_2$ offspring of NIL | $F_2$ from early × late cross | $F_3$ offspring of $BC_3$ | $F_3$ offspring of $BC_3$ |
| *Number of plants evaluated for recombination* | 3472 | 7000 | 330 homozygous recessive ($\sim 1/4$ of total plants) | $\sim 9000$ | 2807 |
| *Size of region delineated by recombination* | 150 kb | 484 bp | $\sim 100$–200 kb (?) | 12 kb | 26·4 kb |
| *Identification of specific gene within region* | Complementation testing of 4 cosmids | Subgenic region identified by recombination alone | Complementation testing of 33 cosmids, plus additional subclones | Gene prediction identified 2 genes including candidate | BLAST against rice genome EST database found single cDNA |
| *Number of progeny evaluated for w/i class trait value* | 5 | 48 | 1 (?) | $\geqslant 20$ (?) | $\geqslant 20$ |
| *Responsible gene* | Previously-unknown gene family | Apoplastic invertase-encoding | Previously-unknown gene family | *CONSTANS* homologue | Similar to *CK2α* |
| *Probable basis of trait differences* | Not identified, but probably in promoter | Regulatory, in intron | Indels creating null alleles | Frameshift deletions and intron indels | Premature stop codon |
| *Verification* | Transformation | None additional | Transformation, plus population genetic study | Transformation, plus analysis of natural variants | Transformation, plus analysis of natural variants |

[a] See text for references.

generate NILs and additional generations for progeny tests may be prohibitive in plants or animals with long generation times, in which each additional generation adds years rather than months to the study. This would generally not be an obstacle for cloning of genes with discrete Mendelian effects. Fine-scale mapping to BAC-sized regions may be nearly impossible in organisms with very large genomes. This may depend in part on the degree to which recombination is elevated in gene-containing regions. The fact that one QTL could be localized by recombination to a region less than 1 kb in the relatively large tomato genome (Fridman et al., 2000) may be encouraging in this regard.

Finally, it may be difficult to verify isolated QTLs in organisms that lack a system for transformation of genomic fragments encompassing entire genes along with their regulatory regions, or even to unambiguously identify the responsible gene in the first place if multiple genes are located in the recombinationally delimited region. The utility of Agrobacterium-mediated transformation makes verification relatively straightforward in model plants relative to animal systems.

### Alternative and emerging strategies

The availability of alternative strategies has superseded the use of conventional map-based cloning techniques for QTLs in some systems. For example, quantitative complementation mapping with chromosomal deficiency stocks allows localization of QTL effects to small chromosomal regions in Drosophila without the need to generate large numbers of offspring to find recombinants (Mackay, 2001). Interpretation of results from quantitative complementation tests is not as straightforward as with conventional QTL mapping under some circumstances, but the technique has been applied successfully in fine-scale mapping of loci affecting longevity (Pasyukova et al., 2000). If deficiency lines were to become widely available in model plants, complementation mapping could become a valuable supplement to conventional map-based cloning.

Association mapping based on linkage disequilibrium (LD) in more-or-less randomly mated populations is another alternative to fine-scale mapping that can be used to refine locations of detected QTLs (Falconer & Mackay, 1996), or prospectively to scan entire genomes (Risch & Merikangas, 1996). Association mapping has been applied extensively in human disease studies, but has several potential limitations including confounding effects of population admixture (Weir, 1996; Pritchard et al., 2000a; Pritchard et al., 2000b), inconsistent relationships between LD and physical distance (Hill & Weir, 1988;

Hill & Weir, 1994), and LD decay rates that are either too slow (Farnir et al., 2000) or too rapid (Remington et al., 2001) to be useful for systematic scanning of QTL regions.

If candidate genes reside within a QTL interval, it is possible to proceed directly to association or complementation testing of individual genes without fine structure mapping. This approach has been used to verify genes responsible for several QTLs in Drosophila (Lai et al., 1994; Long et al., 1998; Lyman & Mackay, 1998; Long et al., 2000) and maize (Doebley et al., 1995; Thornsberry et al., 2001). Its success, however, relies on a well-characterized genome, the presence of candidate genes in QTL regions, and actual involvement of the candidate gene in phenotypic variation. A QTL region may harbor hundreds of genes, any of which could affect the trait in question. The candidate gene approach is biased in favor of genes with known functions related to the trait, and will entirely miss genes with no previously-assigned functions. Of the five cases of map-based QTL cloning discussed here, only the Hd1 CONSTANS orthologue would have been an obvious a priori candidate gene, and two of the QTLs turned out to belong to completely novel gene families.

The advent of new genomic technologies holds the promise of simplifying the process of QTL cloning. The availability of increasing numbers of genetic markers in many species reduces the amount of physical mapping required for chromosome walking; this may be important in large genomes (Tanksley et al., 1995). The completion of physical mapping and genomic sequencing in a number of model organisms will greatly simplify several aspects of map-based cloning (Lukowitz et al., 2000; Yano, 2001). Single-gene knockout lines, which are being developed on a large scale in several species (Ramachandran & Sundaresan, 2001; Thorneycroft et al., 2001), can be used for quantitative complementation tests of individual genes and as genetic backgrounds for transformation experiments to test candidate QTLs more directly (Mackay, 2001). Steady improvement of transformation techniques in a growing number of species, including the development of homologous recombination methodologies, promises even more direct methods to introduce specific polymorphisms and evaluate their effects of on quantitative traits (Rong & Golic, 2000).

### Choosing the right tools

In summary, it is evident that map-based cloning can indeed be used effectively for QTLs, provided they can be crossed into an isogenic background and progeny testing can be used to determine the QTL genotypes of recombinants. One of the key limitations will be the

nature of the QTLs themselves; QTL regions consisting of multiple genes affecting the trait will present difficulties if the individual effects cannot be effectively resolved. Limitations imposed by the organism include difficulties of producing large numbers of offspring to identify recombinants, long generation times, self-incompatibility, or high levels of inbreeding depression. These limitations are no more likely overall to apply to animals than plants, and will generally affect map-based cloning of genes with discrete phenotypes as well as QTLs. The relative ease of producing advanced generations and large progeny sets from individual crosses in plants and the availability of relatively simple transformation techniques have undoubtedly sped the process of map-based QTL cloning in plants. However, suitable fine-mapping populations have been developed in mice and rats (Galli *et al.*, 1999; Talbot *et al.*, 1999; Liu *et al.*, 2001), demonstrating that map based approaches are feasible in some of the most important animal models as well.

The availability of chromosomal deficiency lines for quantitative complementation testing provides an alternative to fine-scale recombinational mapping in *Drosophila* (Mackay, 2001). Linkage disequilibrium mapping and direct testing of promising candidate genes within larger QTL intervals are also valuable options, but these approaches have drawbacks that may limit their application or the reliability of the results. The effectiveness of all methods will increase as better genomic tools become available in more species. On the balance, investigators should give full consideration to all the available options including map-based cloning. The appropriate choice of methods will ultimately depend on details of the particular species, traits, and QTLs being studied.

## References

Alpert, K. B. & Tanksley, S. D. (1996). High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: A major fruit weight quantitative trait locus in tomato. *Proceedings of the National Academy of Sciences of the USA* **93**, 15503–15507.

Doebley, J., Stec, A. & Gustus, C. (1995). *Teosinte branched 1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**, 333–346.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics.*, 4th edn. Harlow, Essex: Longman.

Farnir, F., Coppieters, W., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. & Georges, M. (2000). Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**, 220–227.

Frary, A., Nesbitt, T. C., 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 quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88.

Fridman, E., Pleban, T. & Zamir, D. (2000). A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proceedings of the National Academy of Sciences of the USA* **97**, 4718–4723.

Galli, J., Fakhrai-Rad, H., Kamel, A., Marcus, C., Norgren, S. & Luthman, H. (1999). Pathophysiological and genetic characterization of the major diabetes locus in GK rats. *Diabetes* **48**, 2463–2470.

Graham, J. I., Wolff, D. W. & Stuber, C. W. (1997). Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. *Crop Science* **37**, 1601–1610.

Hill, W. G. & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology* **33**, 54–78.

Hill, W. G. & Weir, B. S. (1994). Maximum-likelihood estimation of gene location by linkage disequilibrium. *American Journal of Human Genetics* **54**, 705–714.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R. & Dean, C. (2000). Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347.

Lai, C., Lyman, R. F., Long, A. D., Langley, C. H. & Mackay, T. F. C. (1994). Naturally occurring variation in bristle number and DNA polymorphisms at the *scabrous* locus of *Drosophila melanogaster*. *Science* **266**, 1697–1702.

Liu, X., Oliver, F., Brown, S. D. M., Kenny, P. & Keightley, P. D. (2001). High-resolution quantitative trait locus mapping for body weight in mice by recombinant progeny testing. *Genetical Research* **77**, 191–197.

Long, A. D., Lyman, R. F., Langley, C. H. & Mackay, T. F. C. (1998). Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**, 999–1017.

Long, A. D., Lyman, R. F., Morgan, A. H., Langley, C. H. & Mackay, T. F. C. (2000). Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. *Genetics* **154**, 1255–1269.

Lukowitz, W., Gillmor, C. S. & Scheible, W.-R. (2000). Positional cloning in Arabidopsis: why it feels good to have a genome initiative working for you. *Plant Physiology* **123**, 795–805.

Lyman, R. F. & Mackay, T. F. C. (1998). Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in *Drosophila melanogaster*: the *Delta-Hairless* gene region. *Genetics* **149**, 983–998.

Mackay, T. F. C. (2001). Quantitative trait loci in *Drosophila*. *Nature Reviews Genetics* **2**, 11–20.

Nabika, T., Kobayashi, Y. & Yamori, Y. (2000). Congenic rats for hypertension: how useful are they for the hunting of hypertension genes? *Clinical and Experimental Pharmacology and Physiology* **27**, 251–256.

Nadeau, J. H. & Frankel, D. (2000). The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nature Genetics* **25**, 381–384.

Pasyukova, E. G., Vieira, C. & Mackay, T. F. C. (2000). Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Genetics* **156**, 1129–1146.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000a). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Pritchard, J. K., Stephens, M., Rosenberg, N. A. & Donnelly, P. (2000b). Association mapping in structured populations. *American Journal of Human Genetics* **67**, 170–181.

Ramachandran, S. & Sundaresan, V. (2001). Transposons as tools for functional genomics. *Plant Physiology and Biochemistry* **39**, 243–252.

Remington, D. L. & O'Malley, D. M. (2000). Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family. *Genetics* **155**, 337–348.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M. & Buckler, E. S., IV (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the USA* **98**, 11479–11484.

Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.

Rong, Y. S. & Golic, K. G. (2000). Gene targeting by homologous recombination in *Drosophila*. *Science* **288**, 2013–2018.

Takahashi, Y., Shomura, A., Sasaki, T. & Yano, M. (2001). *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the α subunit of protein kinase CK2. *Proceedings of the National Academy of Sciences of the USA* **98**, 7922–7927.

Talbot, C. J., Nicod, A., Cherny, S. S., Fulker, D. W. & Collins, A. C. (1999). High-resolution mapping of quantitative trait loci in outbred mice. *Nature Genetics* **21**, 305–308.

Tanksley, S. D., Ganal, M. W. & Martin, G. B. (1995). Chromosome landing: a paradigm for map-based cloning in plants with large genomes. *Trends in Genetics* **11**, 63–68.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.

Thorneycroft, D., Sherson, S. M. & Smith, S. M. (2001). Using gene knockouts to investigate plant metabolism. *Journal of Experimental Botany* **52**, 1593–1601.

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D. & Buckler, E. S., IV (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* **28**, 286–289.

Weir, B. S. (1996). *Genetic Data Analysis II*. Sunderland, MA: Sinauer.

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.

Yano, M. (2001). Genetic and molecular dissection of naturally occurring variation. *Current Opinion in Plant Biology* **4**, 130–135.

Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y. & Sasaki, T. (1997). Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *Theoretical and Applied Genetics* **95**, 1025–1032.

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y. & Sasaki, T. (2000). *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *The Plant Cell* **12**, 2473–2483.

Yano, M. & Sasaki, T. (1997). Genetic and molecular dissection of quantitative traits in rice. *Plant Molecular Biology* **35**, 145–153.