The art and science of cloning QTLs in plants

SILVIO SALVI, MASSIMO BELLOTTI, SERGIO CONTI, ELISABETTA FRASCAROLI, SILVIA GIULIANI, PIERANGELO LANDI, MARCO MACCAFERRI, VINCENZO NATOLI, MARIA CORINNA SANGUINETI, GIORGIO SPONZA, VALENTINA TALAMÈ AND ROBERTO TUBEROSA*

Department of Agroenvironmental Sciences and Technologies, Viale Fanin 44, 40127 Bologna, Italy

*Corresponding author: roberto.tuberosa@unibo.it

Abstract

Recent technical progress in the area of molecular biology and genomics have made possible the molecular dissection of major loci (Quantitative Trait Loci: QTLs) responsible for the genetic control of quantitative traits. Until now, most plant QTLs have been cloned through a positional cloning approach after their identification in experimental crosses. In some cases it has been possible to establish an association between sequence variation at a candidate gene and a phenotype by the analysis of existing genetic accessions. Several refinements of these strategies are made possible by deploying appropriate genetic materials and the latest developments in genomics platforms (e.g. transcriptomics) which allow us to identify eQTLs (expression QTLs), i.e. functional polymorphisms influencing the level of expression of a particular gene. A strategy based on reverse genetics could also be deployed to clone QTLs. We foresee that while QTL analysis and cloning addressing naturally occurring genetic variation will shed light on mechanisms of plant adaptation, more emphasis on approaches relying on candidate gene identification and reverse genetics will accelerate the pace of discovery of the genes underlining QTLs. Although QTL cloning is still in its infancy, further refinement of genomics tools and platforms will make QTL cloning a more routine procedure.

Introduction

The inheritance of quantitative traits is classically thought to be controlled by a large number of genes with small effect scattered throughout different chromosomes. Such genes have been defined as polygenes (Mather 1941). However, early work showed that genetic loci with a major effect on quantitative traits do exist and can be experimentally mapped on chromosomes by evaluating the correlation between the quantitative trait value and the allelic states at linked genetic markers.
This led to define a quantitative trait locus (QTL; Geldermann 1975) as a genetic locus where functionally different alleles segregate and cause significant effects on a quantitative trait. A real constraint of this approach, i.e. the number of genetic markers available in order to adequately scan the genome for QTLs, was overcome during the 1980s with the introduction of molecular genetic markers, which are potentially unlimited and relatively cheap.

Currently, QTL mapping is a well-established procedure in quantitative genetics (Lynch and Walsh 1998; Hackett 2002). QTL mapping usually begins with the collection of genotypic (based on molecular markers) and phenotypic data from a segregating population, followed by a statistical analysis to reveal all possible marker loci where the allelic state correlates with the phenotype. Since this procedure only allows for an approximate mapping of the QTL, it is usually referred to as primary (or coarse) QTL mapping.

More recent technical progress in the area of molecular biology and genomics have made possible the cloning of QTLs, i.e. the identification of the DNA sequences (coding or non-coding) responsible for QTLs. Herein, we present a critical review of the results obtained in QTL cloning in plants and discuss its perspectives, with emphasis on a number of major limitations and promising novel approaches. The comparison of the number of research papers reporting new QTLs mapped in plants (Figure 1) with the number of reports dealing with QTL cloning in plants (Table 1) clearly indicates that QTL cloning is still in its infancy.

![Figure 1](image-url)
Table 1. Summary of the main characteristics of the QTLs cloned in plants.

<table>
<thead>
<tr>
<th>Species</th>
<th>Trait</th>
<th>QTL/gene</th>
<th>Function</th>
<th>Molecular identification</th>
<th>Candidate gene</th>
<th>R² (%)</th>
<th>Plants (no.)</th>
<th>ORF (no.)</th>
<th>Resolution (kb)</th>
<th>Identification of QTN</th>
<th>Functional proof</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td>Flowering time</td>
<td>ED1/CRY2</td>
<td>Cryptochrome</td>
<td>Pos. cloning</td>
<td>Yes (L)</td>
<td>28-56</td>
<td>1,822</td>
<td>15</td>
<td>45</td>
<td>Amino acid substitution</td>
<td>Transformation</td>
<td>El-Din El-Assal et al. 2001</td>
</tr>
<tr>
<td></td>
<td>Flowering time</td>
<td>FLOWFLM</td>
<td>Transcription factor</td>
<td>Pos. cloning</td>
<td>Yes (E)</td>
<td>27</td>
<td>NA</td>
<td>38</td>
<td>138</td>
<td>Deletion of whole gene</td>
<td>Transformation</td>
<td>Werner et al. 2005</td>
</tr>
<tr>
<td>Glucosinolates</td>
<td>Glucosinolates</td>
<td>GS-elong/MAM</td>
<td>MAM synthase</td>
<td>Pos. cloning</td>
<td>Yes (E)</td>
<td>NA</td>
<td>4,600</td>
<td>NA</td>
<td>NA</td>
<td>Nucleotide and gene indels</td>
<td>NA</td>
<td>Kroymann et al. 2003</td>
</tr>
<tr>
<td>Root morphology</td>
<td>Root morphology</td>
<td>BRK</td>
<td>Transcription factor</td>
<td>Pos. cloning</td>
<td>No</td>
<td>80</td>
<td>860</td>
<td>10</td>
<td>45</td>
<td>Premature stop codon</td>
<td>Transformation</td>
<td>Meachel et al. 2004</td>
</tr>
<tr>
<td>Maize</td>
<td>Plant architecture</td>
<td>Tb1</td>
<td>Transcription factor</td>
<td>Pos. cloning</td>
<td>Yes (E)</td>
<td>17-31</td>
<td>NA</td>
<td>NA</td>
<td>No, possibly regulatory</td>
<td>Complementation</td>
<td>Dobbley et al. 1995, 1997</td>
<td></td>
</tr>
<tr>
<td>Glume architecture</td>
<td>Tgt1</td>
<td>Tgt1</td>
<td>Transcription factor</td>
<td>Pos. cloning</td>
<td>No</td>
<td>3,106</td>
<td>1</td>
<td>1</td>
<td>Amino acid substitution</td>
<td>Recovery of mutant</td>
<td>Wang et al. 2005</td>
<td></td>
</tr>
<tr>
<td>Rice</td>
<td>Heading time</td>
<td>Hd1/Se1</td>
<td>Transcription factor</td>
<td>Pos. cloning</td>
<td>Yes (L)</td>
<td>67</td>
<td>1,505</td>
<td>2</td>
<td>12</td>
<td>No</td>
<td>Transformation</td>
<td>Yano et al. 2000</td>
</tr>
<tr>
<td>Heading time</td>
<td>Hsd60/CK2</td>
<td>Hsd60</td>
<td>Protein kinase</td>
<td>Pos. cloning</td>
<td>No</td>
<td>NA</td>
<td>2,807</td>
<td>1</td>
<td>26</td>
<td>Premature stop codon</td>
<td>Transformation</td>
<td>Takahashi et al. 2001</td>
</tr>
<tr>
<td>Heading time</td>
<td>Hsd3a</td>
<td>Hsd3a</td>
<td>Unknown</td>
<td>Pos. cloning</td>
<td>Yes (L)</td>
<td>NA</td>
<td>2,207</td>
<td>4</td>
<td>20</td>
<td>No</td>
<td>Transformation</td>
<td>Kojima et al. 2002</td>
</tr>
<tr>
<td>Heading time</td>
<td>Ehd1</td>
<td>Ehd1</td>
<td>B-type response regulator</td>
<td>Pos. cloning</td>
<td>No</td>
<td>NA</td>
<td>&gt;2,500</td>
<td>3</td>
<td>16</td>
<td>Amino acid substitution</td>
<td>Transformation</td>
<td>Doi et al. 2004</td>
</tr>
<tr>
<td>Grain number</td>
<td>Gst1/CKX2</td>
<td>Gst1</td>
<td>Cytokinin oxidase/ dehydrogenase</td>
<td>No</td>
<td>44</td>
<td>13,000</td>
<td>1</td>
<td>6.3</td>
<td>Several</td>
<td>Transformation</td>
<td>Ashikari et al. 2005</td>
<td></td>
</tr>
<tr>
<td>Tomato</td>
<td>Fruit sugar content</td>
<td>Bnx0-2-5/In5</td>
<td>Inverse</td>
<td>Pos. Cloning</td>
<td>No</td>
<td>7,000</td>
<td>1</td>
<td>0.5</td>
<td>Amino acid substitution</td>
<td>Complementation</td>
<td>Fridman et al. 2000, 2004</td>
<td></td>
</tr>
<tr>
<td>Fruit shape</td>
<td>Oute</td>
<td>Oute</td>
<td>Unknown</td>
<td>Pos. cloning</td>
<td>No</td>
<td>48-67</td>
<td>3,000</td>
<td>8</td>
<td>55</td>
<td>Premature stop codon</td>
<td>Transformation</td>
<td>Liu et al. 2002</td>
</tr>
<tr>
<td>Fruit weight</td>
<td>fcu2.2</td>
<td>fcu2.2</td>
<td>Unknown</td>
<td>Pos. cloning</td>
<td>No</td>
<td>30</td>
<td>3,472</td>
<td>4</td>
<td>92</td>
<td>Unknown regulatory variant</td>
<td>Transformation</td>
<td>Frary et al. 2000; Cong et al. 2002</td>
</tr>
</tbody>
</table>

a Evidence for candidate gene. (E) indicates early evidence, after primary QTL analysis; (L) indicates late evidence, after physical mapping and/or sequencing.

b Proportion of phenotypic variance explained by the QTL in the primary cross.

c Dimension of the population utilized for fine mapping.

d DNA physical interval completely linked with the QTL.

NA: not applicable or not available.
Positional cloning of QTLs

Among the handful of QTLs isolated so far in plants, the majority have been cloned via positional cloning. Positional cloning implies a number of steps (Figure 2) that enable us, by creating a large mapping population and by use of molecular markers, to assign a QTL to the shortest possible genetic interval (QTL fine genetic mapping) and to identify the corresponding interval on the DNA sequence (QTL physical mapping). Candidate genes genetically and physically co-segregating with the QTL are then identified and/or selected for evaluation. The increase in mapping resolution required to accomplish QTL positional cloning is substantial, since after primary mapping a QTL is positioned within a chromosome interval of ca. 10-30 cM which usually includes several hundred genes. Eventually, independent proof is required to validate the role of the identified allelic polymorphism on the observed phenotypic effect.

Figure 2. Flow chart depicting the major experimental steps involved in cloning a QTL using a map-based (positional cloning) approach.
**Genetic materials**

A widely-adopted strategy to more accurately estimate the position and effect of a coarsely mapped QTL is to create a new experimental population by crossing nearly isogenic lines differing only for the allelic constitution at the short chromosome segment (usually varying from ca. 10 to 30 cM in length) harbouring the QTL (QTL-NILs). The population should provide enough crossover events in order to genetically distinguish the gene responsible for the target QTL from other nearby genes. Factors like the physical vs. genetic distance ratio and gene density have to be duly considered when planning the size of the population. Figure 3 shows the number of chromosomes to be screened depending on the resolution needed in genetic terms. Such estimates should be considered with caution due to the variability of the kb/cM ratio on any plant chromosome.

In a nearly isogenic population, due to the absence of other segregating QTLs, the target QTL becomes the only genetic source of variation and, accordingly to the heritability of the trait, a major source of the total phenotypic variation, thus enabling the phenotypic means of the QTL genotypic classes (+/+,-/- and, when present and in case of prevailing additive effects, -/+ ) to be statistically differentiated. Appropriate replication and/or progeny testing are generally implemented based upon the heritability of the trait considered. Under such conditions, the QTL is considered Mendelized (Alonso-Blanco and Koornneef 2000) and cM distances between a QTL and the nearby molecular markers can be more precisely estimated.

![Figure 3](image-url)  
**Figure 3.** Number of chromosomes required to retrieve (with \( P = 0.9 \)) at least one crossing-over within a defined interval. Equation: number of chromosomes = \( \ln 0.1/\ln (1 – \text{interval length in Morgan}) \).
Starting from the same population where primary mapping was carried out, QTL-NILs can be produced by (i) marker-assisted backcross introgression (i.e. substitution) of one QTL allele into one or both parental genetic backgrounds or (ii) iteratively identifying and selfing individuals which are heterozygous at the QTL region (Tuinstra et al. 1997). QTL-NILs can also be efficiently identified within introgression libraries (ILs), i.e. collections of lines where each line is isogenic to a background parental line with the exception of a single short chromosome segment introgressed from a donor (Zamir 2001). Within an IL, the donor genome should be completely represented among the different IL lines so that a QTL-NIL will potentially exist for any segregating QTL. Remarkably, the same IL of the wild tomato *Lycopersicon pennellii* within the cultivated tomato genetic background (Eshed and Zamir 1994) provided the source of the QTL-NILs utilized for the cloning of three tomato QTLs.

NILs suitable for positional cloning are also produced by the advanced backcross QTL analysis (ABQA) method, which combines backcrossing chromosome segments from a wild accession within an elite line with some level of phenotypic selection against extreme phenotypes ( Tanksley and Nelson 1996). Critical aspects to be considered are the time and effort required for the development of collections of IL or ABQA lines as well as the limitation of the genetic variability provided by the use of only two parental lines: no matter how carefully the parental lines are chosen, the collections will only segregate for a fraction of the many more QTLs segregating for the same trait in other populations, which should be recognized. This important limitation can be partially overcome through the use of multiparental intercrossed populations as proposed by (Mott et al. 2000). These populations are generated by crossing a carefully chosen set of parental lines capturing much of the genetic variation of the species, and then by performing several cycles of intermating. However, while this approach promises to increase the efficiency of QTL mapping both in terms of detection (segregation is expected at many loci) and genetic resolution (many rounds of meiosis), a recent report from the same authors highlighted unexpected analytical complexities (Yalcin et al. 2004). It should be noted that a substantial increase in genetic resolution can also been obtained by intercrossing standard biparental populations (Lee et al. 2002).

During the fine mapping step, the resolution of the target QTL in two or more linked loci may bring positional cloning projects to an end when the proportion of phenotypic variability explained by each QTL is too small to be revealed with a realistically manageable number of replications. QTL clusters have indeed been observed in plants (Khavkin and Coe 1997; Tuberosa et al. 2002; Chen and Tanksley 2004). On the other hand, cloning was accomplished when one of the linked QTLs retained most of the effect (Fridman et al. 2002; Kojima et al. 2002).
Genotyping

The recruitment of polymorphic markers required for the fine mapping of a QTL is rather simple for Arabidopsis and rice, where the whole genome has been sequenced, and for species such as maize or tomato for which genomic sequencing is under way or information is available in terms of sequenced ESTs. However, in species for which detailed sequence information is not available or cannot be deduced from syntenic relatives, a large number of molecular markers (e.g. AFLPs) need to be screened in genotypes contrasted at the target region (e.g. pair of QTL-NILs). Figure 4 provides an estimate of the number of markers to be screened in positional cloning when new random markers need to be added. Synteny with Arabidopsis, Medicago/Lotus and rice will assist in the identification of additional markers for cloning projects within Brassicaceae, legumes and cereals, respectively (Schmidt 2000). As to genotyping techniques, microarray-based platforms appear

![Figure 4](image_url)

**Figure 4.** Number of random molecular markers to be screened in order to identify at least one polymorphic marker within a predetermined genomic interval. Five levels of polymorphisms are considered. Computations are for a probability of $P = 0.9$. The equation is:

$$\text{Number of markers} = \frac{\ln 0.1}{\ln \left[1 - \frac{\text{base pairs of interval}}{\text{base pairs of genome}}\right]}.$$
particularly promising for a high-throughput identification of polymorphisms (SNPs and insertion-deletions) at thousands of loci (Borevitz et al. 2003; Hardenbol et al. 2003).

**Physical mapping and candidate sequences**
When the genetic resolution approaches the cM level, the markers closest to the QTL are used for anchoring the genetic map to the physical map, i.e. the genomic sequence or a BAC (bacterial artificial chromosome) contig covering the QTL region. An early transfer of the information to the physical map allows for the efficient generation of new single-copy markers useful for refining the genetic mapping and searching candidate genes. Even if only a BAC contig is available, sequenced BAC ends can often be transformed in genetic markers and low-pass, shot-gun sequencing can provide a glimpse of local gene content. In this phase, bioinformatics provides an important contribution in terms of gene prediction and annotation, and exploitation of syntenic relationships.

Among the studies herein considered, four (targeting Brix-9-2-5, Gn1, Hd6 and Tga1; Table 1) managed to reduce the number of genes co-segregating with the target QTL to one. In one case, Hd6, the QTN (i.e. the nucleotide polymorphism responsible for the QTL effect) was a nucleotide substitution in one of the two alleles that caused a premature stop codon (Takahashi et al. 2001). When the physical region co-segregating with the QTL includes more than one gene (up to 38; Werner et al. 2005), candidates can either be identified via function prediction and selected for further testing (see Cry2, FLM, Hd1 and Hd3a; Table 1), or may not be very evident. When multiple coding sequences with no obvious candidate gene are identified, two possible options are to increase the mapping resolution or to functionally test each open reading frame (ORF).

**Validation of a candidate sequence**
The functional testing of a candidate gene/s can be performed by overexpressing or down-regulating the target gene through genetic engineering or RNAi (Waterhouse and Helliwell 2003), by genetic complementation of a known mutant (Doebley et al. 1997) or by rescuing and phenotypically and molecularly characterizing mutants at the candidate gene (Wang et al. 2005). If available within the species under investigation, reverse genetics tools such as T-DNA or transposon-tagged populations (Maes et al. 1999) and/or TILLING (McCallum et al. 2000) can also be exploited. As compared to transposon tagging, TILLING and RNAi are appealing alternatives for their almost universal applicability and for providing subtle changes of gene functionality comparable to those observed naturally. Gene replacement, still in its infancy but already reported in rice (Iida and Terada 2004) can be considered the ultimate tool for validating candidate genes.

The validation of QTNs in non-coding regions is one of the current major challenges. Regulatory regions close to (e.g. promoters) or far (e.g. enhancers/silencers)
from the regulated gene can be expected to host sequence polymorphisms causing variation in quantitative phenotypes. It can also be predicted that QTLs will be found at microRNA loci and at regions controlling chromatin methylation and/or organization (e.g. folding). Furthermore, transposon insertions have already been shown to be responsible for changes in gene expression (Kashkush et al. 2003). However, for most of these genomic features a structural characterization is still lacking, thus hindering the recognition of their role in the control of quantitative traits. Regulatory elements at several tens of kb from coding sequences have already been shown to act as QTLs (Salvi et al. 2002; Salvi et al. 2003; Clark et al. 2004) although the causal QTNs have not yet been identified.

After considering all of the above-mentioned aspects, it is clear that positional cloning of QTLs in plants remains a demanding and daunting undertaking. Additionally, positional cloning has been limited exclusively to major QTLs, since all the cloned QTLs showed, in the primary genetic analysis, an $R^2$ value higher than 15% (Table 1). It should be noted that $R^2$ values based on primary mapping can be grossly under- or overestimated (Beavis 1994) due to statistical artefacts; furthermore, epistasis can modify the genetic effect of the target QTL when the genetic background changes (Doebly et al. 1995), for instance during QTL-NIL production. Therefore, an independent evaluation of the QTL effect, (e.g. by developing and testing QTL-NILs. Landi et al. 2005) is recommended before embarking on QTL positional cloning.

Cloning QTLs by association mapping

As an alternative to positional cloning, QTLs can be molecularly resolved through association mapping (Cardon and Bell 2001), i.e. by identifying, within a set of genotypes (e.g. germplasm accessions, cultivated varieties, etc.), a statistical association between allelic variants at marker or candidate loci and the mean of the analyzed trait. The analysis evaluates the trait mean change caused by the substitution of one allele with another. For QTL cloning in plants, the interest lies in (i) the possibility of finding chromosome regions important for controlling quantitative traits without the costly and time-consuming production of large experimental populations (Morgante and Salamini 2003), (ii) the potentially high genetic resolution provided by the many meiotic events which occurred during past generations and (iii) the possibility of surveying a large number of functionally diverse alleles per locus.

A major factor to be considered in association mapping is the level of linkage disequilibrium (LD; level of non-random assortment of alleles at different loci) among the tested accessions. In plants, the extensive LD analyses conducted in Arabidopsis and maize (Flint-Garcia et al. 2003) have indicated that while LD persists over hundreds of kb in Arabidopsis, in maize LD decays after a few kb, although it can extend significantly farther in collections of elite germplasm (Flint-Garcia et al.
With high LD values, marker-trait association can theoretically be revealed with a manageable number of molecular markers. In this case, the expected mapping resolution will only be sufficient for the discovery and coarse mapping of the QTL. On the other hand, when testing germplasm panels with low LD, the diagnostic power of a single marker will only extend shortly and thus a prohibitively high number of markers would be required for a whole genome scan. In the latter situation, association mapping can still be used to fine map the QTL at the gene level after the QTL is positioned using standard mapping procedures. Based on this, it is conceivable that different sets of genotypes, characterized by high or low LD, can be assembled and used for QTL discovery or candidate gene validation, respectively, as suggested for human genetics (Reich et al. 2001). Notably, the presence of population structure, namely the possible presence of hidden subgroups (e.g. due to relatedness, selection, etc.) with an unequal distribution of alleles may influence the efficacy of this approach by causing spurious trait-marker associations (Pritchard et al. 2000).

A powerful approach for identifying different haplotypes (combinations of allelic variants) at target loci and making them available for association mapping is provided by EcoTILLING (Comai et al. 2004), which allows for the identification of virtually all SNPs and small insertion/deletions within a ca. 1-kb window in a set of genotypes at a fraction of the sequencing cost. This notwithstanding, the necessity to also screen regulatory regions often quite distant from the effector genes indicates that the selection of candidate sequences to be tested for association mapping is not a trivial task if the genomic scan aims to be comprehensive. Examples of the identification of association between haplotype variation at a candidate gene and a quantitative trait were reported in *Arabidopsis* (Olsen et al. 2004), *Brassica* (Osterberg et al. 2002; Gupta et al. 2004), potato (Simko et al. 2004) and in maize (Thornsberry et al. 2001; Guillet-Claude et al. 2004; Wilson et al. 2004).

It should be emphasized that the identification of a statistically significant association between haplotype variation at a candidate gene or sequence and a quantitative phenotype should be followed by validation experiments similar to those deployed within the positional cloning approach previously described.

**Functional genomics and QTL cloning**

Functional genomics is contributing to many aspects of QTL analysis and cloning. Transcriptional profiling between contrasting QTL genotypes can quickly provide a list of genes differentially expressed; subsequently, those genes functionally related to the target trait and mapping at the QTL region can be selected as candidates (Wayne and McIntire 2002). Unfortunately, the number of QTLs cloned so far in plants is too small to test the validity of this approach. Indeed, when the QTL was shown to involve a difference in gene expression level between alleles, those differ-
ences were either too low (ca. two-fold; Doebley et al. 1997) or showed too strong of a spatial and/or temporal pattern (Cong et al. 2002) to allow for their identification with a standard microarray-based transcriptome analysis. Other profiling platforms, such as SAGE (serial analysis of gene expression; Gowda et al. 2004) and MPSS (massively parallel signature sequencing; Brenner et al. 2000) are better suited to detect subtle differences in gene expression. Transcript profiling can reach the sub-tissue level of resolution if carried out in combination with laser-capture microscopy (Schnable et al. 2004).

The expression profiling of a mapping population at the mRNA or protein level allows us to treat the level of expression of a single gene as a quantitative trait and to dissect its genetic control by QTL analysis (Jansen and Nap 2001; Tuberosa et al. 2003; Brem and Kruglyak 2005; Figure 5). The loci controlling the level of gene expression have alternatively been named transcript quantity loci (TQLs), expres-

Figure 5. Expression profiling of a mapping population at the mRNA level via microarray analysis to identify expression QTLs (eQTLs) for specific cDNAs. Correspondence between an eQTL peak for a specific cDNA (e.g. cDNA-2) and a QTL peak for a trait causally linked to the function of the protein encoded by the cDNA provides circumstantial evidence supporting the role of the cDNA as a candidate gene for the target trait.
sion QTLs (eQTLs) or protein quantity loci (PQLs; Damerval et al. 1994; Schadt et al. 2003). Correspondences between eQTLs and/or PQLs for candidate genes with QTLs for morpho-physiological traits have already been observed in small- or medium-scale experiments (de Vienne et al. 1999; Francia et al. 2003; Guillaumie et al. 2004). Microarray-based studies have mapped eQTLs both at the same location of the gene whose expression was measured, thus indicating a role for cis-regulatory allelic variation, and also at distant chromosome positions (Brem et al. 2002; Schadt et al. 2003). The same studies highlighted the presence of eQTL “hot spots”, i.e. chromosome regions apparently responsible for controlling the simultaneous expression of many genes.

QTL tagging

Robertson (1985) suggested that qualitative mutant alleles and wild-type alleles at loci affecting quantitative traits are the extremes of a possible range of effects, with QTLs resulting from the segregation of naturally-available alleles with milder effects. Robertson’s hypothesis has been confirmed when a mutant was available for the gene subtending the target QTL (Doebley et al. 1995; El-Din El-Assal et al. 2001). On this line, it was recently argued that mutagenesis could be more efficient for dissecting the genetic basis of quantitative traits than QTL analysis, which only provides “accidents of history” allelic variants (Nadeau and Frankel 2000). A direct method for the identification of such genes would be to utilize a tagging (insertional) approach. Such framework (Figure 6) would require the phenotypic screening of an insertionally-mutagenized population for the target quantitative trait in order to identify those lines with a phenotypic mean value outside a predicted range due to environmental effects (Robertson 1985; Soller and Beckmann 1987). The complete screening experiment would involve a manageable number of plants (e.g. from a few thousands to a few tens of thousands) if multiple insertion systems are employed and several quantitative traits are concurrently evaluated (Soller and Beckmann 1987). The gene functionally modified or inactivated by the insertional agent can be rescued using standard molecular procedures. Following a similar approach, QTL tagging has already been successfully accomplished in Magnaporthe (Fujimoto et al. 2002), the causal agent of rice blast, and in Drosophila (Norga et al. 2003). In plants, QTL tagging could be carried out with a number of different approaches based on T-DNA as well as DNA-transposons and retrotransposons. However, systems relying on callus cultures (e.g. activation of rice TOS-17 retrotransposon; Hirochika et al. 1996) should be considered with caution due to the occurrence of somaclonal variation, i.e. the de novo variation, observed in plants regenerated from tissue culture and caused by changes in DNA-methylation, transposon activity and others molecular events (Kaeppler and Phillips 1993). Somaclonal variation can potentially alter any quantitative trait and therefore hin-
der the identification of the tagged QTL. Instead, interesting resources are the Ac-Ds-based insertional populations developed in rice (Jeon and An 2001); in these cases, following the introduction of the heterologous transposons, the majority of the mutational events were created by new transposition activity. In maize, a Mu-based insertional population has been developed in a non-segregating genetic background (Settles et al. 2004), where most of the quantitative variability can be attributed to the segregation of the tagged QTLs.

The role of candidate genes

Classically, a link between a gene and a quantitative trait can be hypothesized based on linkage information (all genes co-segregating with a QTL are positional candidate genes) or communality between the quantitative trait physiology and the biochemical function of the gene (functional candidate gene) or both (Pflieger et al. 2001). For instance, completion of genome sequences and improved bioinformatics will facilitate in silico cross-matching of candidate sequences with QTLs in pro-
ograms of positional cloning or association mapping. Additionally, a better understanding of the mechanisms behind the regulation of gene expression will extend the concept of candidate gene to include also cis-acting regulatory sequences. Therefore, it is conceivable that in the future QTL cloning will increasingly rely on candidate gene information and this will be made possible by exploiting the tools of reverse genetics.

Concluding remarks

The cloning of the first set of plant QTLs showed us that the type of mutations, genes, and cellular and physiological pathways involved in determining quantitative traits are not distinct from those underlining Mendelian traits. Among the QTLs cloned so far, the apparent abundance of regulatory genes or transcription factors, which potentially act on many downstream functions, was not unexpected due to the complexity of the traits that are usually investigated in QTL analysis.

Almost invariably, the QTLs cloned had shown the largest phenotypic effect in the original experimental populations, often produced by wide crosses between subspecies. Of course, targeting major QTLs simplifies the cloning process especially if based on positional cloning. However, the so-called minor QTLs (those showing a smaller effect on the trait in the original population) should be targeted with the same emphasis since they can represent potentially important genes and their appearance as “minor” could simply be due to the segregation of alleles of similar effect in the experimental cross. For this reason, minor but extremely important QTLs (and genes) are to be expected for yield and yield-related traits in crosses involving elite germplasm, where plant breeding has already culled alleles with undesirable effects. Hopefully, the constant improvement of the molecular platforms, new types of genetic materials, progress in bioinformatics and the increasing availability of tools for functionally testing candidate genes will offer the opportunity of targeting QTLs other than those with a major effect.

Finally, we emphasize the need to continue investigating natural variation to unlock the allelic richness present in germplasm collections, thus enabling us to more effectively deploy marker-assisted selection and genetic engineering to introduce or modify valuable alleles in crops. At the same time, quantitative approaches to genetically dissect mutants will also have to be extended if more subtle regulators of complex phenotypes are to be uncovered.
References


Eshed Y, Zamir D (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for
fine-mapping of genes. Euphytica 79:175-179
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. Proc Natl Acad Sci USA
Natl Acad Sci USA 90:8773-8776
Mather K (1941) Variation and selection of polygenic characters. J Genetics 41:159-193


Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics 8:552-560


Tuinstra MR, Ejeta G, Goldsborough PB (1997) Heterogeneous inbred family (HIF) analysis: a
method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet 95:1005-1011