Plant Systems Biology
Edited by Sacha Baginsky and Alisdair R. Fernie
© 2007 Birkhäuser Verlag/Switzerland

Natural and artificially induced genetic variability in crop and model plant species for plant systems biology

Christophe Rothan[1] and Mathilde Causse[2]

[1] *INRA-UMR 619 Biologie des Fruits, IBVI-INRA Bordeaux, BP 81, 71 Av. Edouard Bourlaux, 33883 Villenave d'Ornon cedex, France*
[2] *INRA-UR 1052, Unité de Génétique et Amélioration des Fruits et Légumes, BP 94, 84143 Montfavet cedex, France*

Abstract

The sequencing of plant genomes which was completed a few years ago for *Arabidopsis thaliana* and *Oryza sativa* is currently underway for numerous crop plants of commercial value such as maize, poplar, tomato grape or tobacco. In addition, hundreds of thousands of expressed sequence tags (ESTs) are publicly available that may well represent 40–60% of the genes present in plant genomes. Despite its importance for life sciences, genome information is only an initial step towards understanding gene function (functional genomics) and deciphering the complex relationships between individual genes in the framework of gene networks. In this chapter we introduce and discuss means of generating and identifying genetic diversity, i.e., means to genetically perturb a biological system and to subsequently analyse the systems response, e.g., the changes in plant morphology and chemical composition. Generating and identifying genetic diversity is in its own right a highly powerful resource of information and is established as an invaluable tool for systems biology.

Introduction

In the plant genomic era, huge amounts of sequence data have been obtained, mostly for model plants but also for an ever increasing number of non model plant species. Genome sequencing, which was completed a few years ago for *Arabidopsis* and rice, is currently underway for numerous crop plants of high commercial value such as maize, poplar, tomato, grape or tobacco. In addition, hundreds of thousands of EST sequences are publicly available for many plant species (e.g., at TIGR, http://www.tigr.org/tdb/tgi/plant.shtml) and may represent between 40 and 60% of the genes present in plant genomes. However, the identification of very large sets of gene sequences in any plant species is only an initial step towards (i) understanding gene function in the plant (functional genomics) and (ii) deciphering and representing the complex relationships between gene sequence and protein expression varia-

tion, corresponding pathways and networks, and changes in plant morphology and chemical composition (plant systems biology).

The recent development of high throughput methods for transcriptional profiling of genes using microarrays (Chapters by Foyer et al. and Hennig and Köhler) and for metabolite profiling using various separation and analytical techniques (metabolome) (Chapters by Steinhauser and Kopka, and Sumner et al.), as well as the current progress in large scale protein analysis (proteomics, Chapters by Brunner et al. and Schuchardt and Sickmann) and morphological phenotyping of plants, has revolutionised the way we now envisage plant systems biology. By studying plants to find out where and when, and under what conditions, whole sets of genes and proteins are expressed, and by analysing the correlations with corresponding changes in plant phenotype (development, morphology and chemical composition), we are now able to infer the putative functions of genes and to deduce the possible relationships between pathways, regulatory networks and phenotypes.

Linking phenotype to genotype: Strategies

Basically, two strategies, usually named forward and reverse genetics, will help bridge the gap between genotypic variations and associated phenotypic changes. Both are based on the use of natural or artificially induced allelic gene variation to gain insights into the relationship between genes, their function and their influence on phenotypic traits. The forward (traditional) genetic approach aims at discovering the gene(s) responsible for variations of known single Mendelian traits or of quantitative traits (Quantitative Trait Loci or QTL) previously identified through phenotypic screening of natural populations. In contrast, the main objective of reverse genetics is to unravel the physiological role of a target gene and to establish its effect on the plant phenotype.

Forward genetic approaches

Forward genetic approaches have been hampered until recently in many crop plants by the lack of detailed genetic maps, genomic resources (BACs, bacterial artificial chromosome) and genomic sequences. Due to the remarkable development of genetic marker technology over the last 15 years, genetic linkage maps are now available for most crop species, allowing the comparative mapping of crop species and model plants, the location of loci controlling Mendelian traits or QTL on linkage groups and finally the isolation by map-based cloning of the gene responsible for the phenotype. Today, the availability and use of high throughput and precise analytical tools for metabolic profiling (Chapters by Steinhauser and Kopka, and Sumner et al.) has considerably increased the number of compounds that can be identified and quantified in plants. This will enable the decomposition of previously identified complex quantitative traits into multiple single quantitative traits, potentially unravelling loci controlling whole metabolic pathways. The use of transcriptome or proteome profiling and genome sequence information will provide new candidate genes for characterising the sequences responsible for natural genetic variation.

Reverse genetic approaches

Genome and EST sequencing, and large scale analyses of transcript, protein and metabolite profiles, can give rise to a large number of candidate genes whose function needs to be evaluated in the context of the plant. Very efficient reverse genetic tools, mostly based on insertional mutagenesis and targeted silencing of specific genes by RNAi-based technology (Chapter by Johnson and Sundaresan), have therefore been developed in model plants. However, a comparable strategy is clearly impossible for most crop plants, due to cost or technical limitations such as a large genome size or the unfeasibility of large scale genetic transformation. One might consider that the information gained from model plants can easily be transferred to plant species. Currently, recent advances in plant studies indicate that results obtained from a model plant are not always applicable to other plant species, not only because many crop plants have specialised organs not present in the model plants *Arabidopsis* and rice (e.g., tubers in potato, root in sugar beet or fruit in tomato) but also because a considerable fraction of the genes are probably unique to the different taxa or even to the particular species to which they belong [1]. In addition, for certain categories of genes, e.g., those involved in signalling pathways or in regulatory processes such as transcription factors or kinases, knockout mutations can be lethal for the plant, induce phenotypic variations only distantly related to the real function of the target gene or, in some cases, give weaker phenotypes than those observed with missense mutations that produce dominant-negative mutants [2]. In these circumstances, natural or artificially induced allelic variants appear as the most appropriate strategy.

Forward genetics: Gene and QTL characterisation

The possibility of saturating the genome with molecular markers has allowed Mendelian mutations and QTL to be systematically mapped. Since the early 1990s, hundreds of studies have been conducted to map Mendelian mutations and QTL in plants. Several genes have been cloned through map-based cloning [3–5], but only a few QTL have been cloned and characterised. QTL are not different in nature from loci responsible for discrete variations, but, rather than a ‘mutant-wild-type’ opposition, there are moderate differences (of effects) between ‘wild-type’ (or active) alleles, which are responsible for the variation of quantitative characters. One can believe that systems biology and high-throughput genomic approaches will lead to a rapid increase in the number of gene/QTL cloned and of our understanding of the genetic basis of natural variation.

Principles and methods of QTL mapping

QTL mapping is based on a systematic search for association between the genotype at marker loci and the average value of a trait. It requires:

- a segregating population derived from the cross of two individuals contrasted for the character of interest.
- that the genotype of marker loci distributed over the entire genome is determined for each individual of the population (and thus a saturated genetic map is constructed).
- the measurement of the value of the quantitative character for each individual of the population.
- the use of biometric methods to find marker loci whose genotype is correlated with the character, and estimation of the genetic parameters of the QTL detected.

Several biometric techniques to find QTL have been proposed, from the most simple, based on analysis of variance or Student's test, applied marker by marker, to those that take into account simultaneously two or more markers [6]. The QTL are characterised by three parameters (a, d, R^2). The additive effect a is equal to $(m_{22} - m_{11})/2$, where m_{22} and m_{11} are the mean values of homozygous genotypes *A1A1* and *A2A2*, respectively. The degree of dominance is the difference between the mean of the heterozygotes *A1A2*, and half the sum of the homozygotes: $d = m_{12} - (m_{11} + m_{22})/2$ (Fig. 1). Each segregating QTL contributes to a certain fraction of the total phenotypic variation, which is quantified by the R^2, which is the ratio of the sum of squares of the differences linked to the marker locus genotype to the sum of squares of the total differences. Epistasis (interaction between QTL) may also be searched for by screening for interaction between every pair of markers, but due to the number of tests, very stringent thresholds must be applied and thus only very highly significant interactions are detected, unless a specific design is used. The advantage of QTL detection on individual markers is its simplicity. Other more powerful methods have been developed that allow us to precisely position QTL in the interval between the markers and to estimate their effects at this position. The most widespread method for testing for the presence of a QTL in an interval between two markers is based on the calculation of a LOD score. At each position on a chromosome (with a step of 2 cM for example), the decimal logarithm of the probability ratio below is calculated:

$$\mathrm{LOD} = \log_{10} \frac{V(a_1, d_1)}{V(a_0, d_0)}$$

where $V(a_1, d_1)$ is the value of the probability function for the hypothesis of QTL presence, in which the estimations of parameters are a_1 and d_1, and where $V(a_0, d_0)$ is the value of the probability function for the hypothesis of QTL absence, that is, when $a_0 = 0$ and $d_0 = 0$ [7]. A LOD of 2 thus signifies that the presence of a QTL at a given point is 100 times more probable than its absence; a LOD of 3 means 1,000 times more probable, etc. A curve of LOD can thus be traced as a function of the position on a linkage group. The maximum of the curve, if it goes beyond a certain

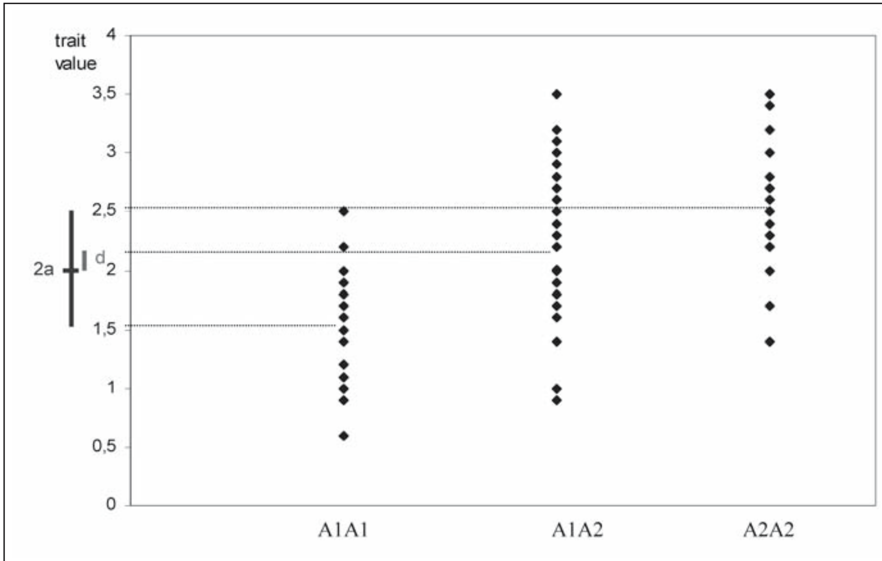

Figure 1. Genetic parameters related to a QTL. The plot shows average values of the three genotypic classes at the marker B (of Fig. 1) for the quantitative character studied. A significant difference between the means signifies that the effects of two alleles at the QTL are sufficiently different to have detectable consequences. The parameters *a* and *d* are then estimated. R^2 is related to the intraclass variance s^2 and to the sample size.

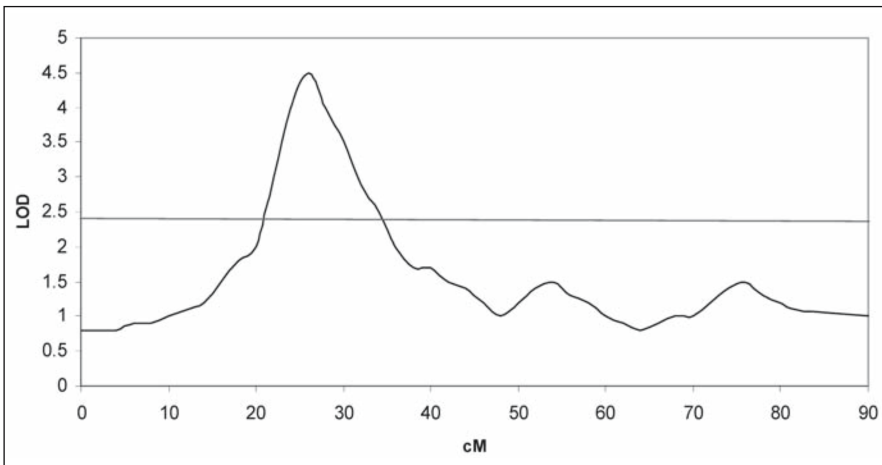

Figure 2. Example of Lod plot along a 90 cM chromosome.The most likely position of the QTL is shown with the confidence interval associated.

threshold, indicates the most probable position of the QTL (Fig. 2). The confidence interval of the QTL position is thus conventionally defined as the chromosomal fragment corresponding to a reduction in LOD of 1 unit in relation to the maximum LOD, which indicates that the probability ratio has fallen by a factor of 10. This method was first implemented in the Mapmaker/QTL software [8], which is coupled with the Mapmaker software for the construction of genetic maps. Several related methods have then been proposed including the composite interval mapping that takes the other QTL present in the genome, represented by markers that are close to them, as co-factors in the model. This reduces the residual variation induced by their segregation [9–10] and then substantially improves the precision of estimation of QTL effects and positions. These methods are implemented in several software. Access to most of these software is free and the addresses of sources can be found in databases including http://www.stat.wisc.edu/~yandell/qtl/software.

Factors influencing QTL detection

Although the principle of QTL detection is relatively simple, several parameters influence the results and must be taken into account to optimise the experimental setup. For a given sample size, the efficiency of QTL detection depends partly on the additive effect of QTL (a very small difference of effects between alleles will not be found significant) and partly on the variance within the genotypic classes. This variance depends on environmental effects (the environmental control of variations increases the efficiency of the test) on other segregating QTL in the genome, on the presence of epistasis and on the distance between markers and QTL (this is particularly important if the density of markers is low). Because of the large number of analyses carried out, low values of α must be chosen. For interval mapping, a global risk of $\alpha = 0.05$ for the entire genome imposes a fairly high LOD threshold per interval, which depends on the density of markers and the genetic length of the genome [7]. Thresholds are now usually estimated following permutation tests, based on a random resampling of data [11].

Efficiency of QTL detection and precision of QTL location depends more on population size than on marker density [12]. Once a mean marker density of 20 or 25 cM is attained, any supplementary means must be invested in analysing additional individuals rather than in increasing the number of markers. A QTL with a strong effect will be detected with a high probability whatever the population size, but for detection of a QTL with moderate effect (R^2 about 5%), it is necessary to use a larger number of individuals. It must also be noted that it is better to increase the number of genotypes in the population rather than the number of replications per genotype.

The populations in which QTL mapping is most efficient are those derived from crosses between two homozygous lines, such as F2, recombinant inbred lines (RIL), doubled haploid (DH) and backcross (BC). F2 are the only populations allowing the dominance effect to be estimated, while a mixture of *a* and *d* is estimated with BC. Highly recombinant inbred lines (HRIL) obtained after several cycles of intercrossing

individuals were proposed to increase the precision of marker ordering and subsequently also to increase the precision of QTL mapping [13]. When no homozygous parental lines are available (in allogamous species and species with a long generation time, such as trees), QTL detection is complicated because the parents may differ by more than two alleles, and because the phase (coupling or repulsion) of the marker-QTL linkage may change from one family to another. Various populations may nevertheless be used, from F1, BC or populations using information from two generations in families of full siblings [14]. Knowledge of the grandparent genotypes at marker loci can improve detection by allowing phases of associations between adjacent markers to be identified [15].

Tanksley and Nelson [16] proposed to search for QTL in populations of advanced backcross (BC2, BC3, BC4). Although the power of QTL detection is reduced, this strategy is interesting when screening positive alleles from a wild species, as it will allow the identification of mostly additive effects and will reduce linkage with unfavourable alleles and thus simultaneously advance the production of commercially desirable lines.

The efficiency of detecting a particular QTL in a segregating population is low because other QTL are segregating and major QTL mask minor ones. For this reason, Eshed and Zamir [17] proposed the use of introgression lines in which each line possesses a unique segment from a wild progenitor introgressed in the same genetic background. The whole genome has been covered with 75 lines and has created a sort of 'genome bank' of a wild species in the genome of a cultivated tomato. These lines can then be compared with the parental cultivated line to search for QTL carried by the introgressed fragments. The detection is more efficient than in a classical progeny because of the fixation of the rest of the genome. Greater test efficiency and a significant economy in terms of time and effort can also be achieved by molecular genotyping exclusively individuals showing the extreme values of the character studied (through selective genotyping) [18]. Nevertheless this approach is only useful for detecting QTL with major effects and can be applied only if one character is studied.

What have we learnt from QTL studies?

Ever since the mapping of QTL became possible, several studies have showed that even with populations of moderate size (sometimes less than 100 individuals), some QTL are almost always found, for all types of characters and plants [19–20]. Data compiled from maize and tomato, where many QTL have been mapped, indicate that the effects of QTL measured by their R^2 are distributed according to a marked L curve, with a few QTL having a strong or very strong effect, and most QTL having a weak or very weak effect. With populations of normal size (60 to 400 individuals), R^2 are usually overestimated [21] and depending on the characters, one to ten QTL are usually detected with an average of 4 QTL detected per study [22]. These numbers constitute a minimum estimate of the number of segregating QTL in the populations studied for several reasons: (i) Some QTL have an effect below the detection threshold, (ii) some chromosomal segments may contain several linked QTL when

only one is apparent and (iii) if two QTL of comparable effect are closely linked, but in repulsion phase, i.e., if the positive alleles at the two loci do not come from the same parent, no QTL will be detected, until fine mapping is attempted [23]. Moreover, the monomorphic QTL in a given population cannot be detected. For species and traits where a large number of studies have been performed with several progenies, it is frequent to compile more than 30 QTL [24, 25]. Using meta-analysis, Chardon and colleagues [26] summarised 22 studies and identified at least 62 QTL controlling flowering time in maize.

Transgressive QTL are frequently discovered. Even when highly contrasted individuals have been chosen as parents of a population, it is not rare to find a QTL showing an effect opposite to that expected from the value of the parents. Results from advanced backcross experiments in tomato showed for example unexpected positive transgressions from wild relatives, for various fruit traits [27].

When comparative mapping data are available, some QTL of a given character are frequently found at homologous positions on the genomes of species that are more or less related. This is the case for grain weight in several legume species [28–30], for domestication traits in cereals [31, 32] and for fruit-related traits in Solanaceae species [33].

Epistasis between QTL is rarely detected with classical populations [34], but this is mostly due to statistical limits of the populations studied. A way of increasing the reliability of epistasis analysis is to eliminate the ‘background noise’ due to other QTL by using near isogenic lines (differing only by a chromosome fragment) for a particular QTL as parents of the populations studied [35]. On the other hand, it is not because a QTL does not show epistatic interactions with other QTL taken individually that its effect is independent of the genetic background. For instance, the effects of two maize domestication QTL are much weaker when they are segregating in a ‘teosinte’ genetic background than in an F2 maize x teosinte background [36]. Similarly, significant QTL by genetic background interaction was shown in tomato by transferring the same QTL regions into three different lines [37].

QTL mapping is particularly interesting in attempting to analyse the determinism of complex characters, by focusing on components of these characters [38–40]. QTL mapping thus provides access to the genetic basis of correlations between characters. When characters are correlated, at least some of their QTL will be common (or at least genetically linked). In the case of apparent co-location of QTL controlling different characters, there is no direct method to highlight the existence of a single QTL with a pleiotropic effect or of two linked QTL. Korol and colleagues [41] proposed a statistical test to use the information of correlated traits to locate QTL simultaneously controlling several traits. They showed that this approach increased the power of QTL detection when compared to a trait by trait search. Nevertheless the best way to distinguish pleiotropy from linkage is through fine mapping experiments. Many fine mapping experiments have separated QTL that were initially thought to control two related traits [42–44].

The environment may have a significant impact on the effect of QTL: a QTL detected in one environment may no longer be detected in another, or its effect may vary. This has been frequently observed, even though the environmental influence

differs according to the characters and the range of environments studied. Certain QTL are detected in all or almost all the environments tested, while others are specific to a single environment. Several statistical methods for the estimation of QTL x environment interactions have been proposed [45–48]. Certain studies look directly at QTL involved in the response to environmental changes such as soil nitrogen [49] or drought [50]. Ecophysiological modelling may also be used to identify the biological processes underlying QTL and to distinguish loci affected by the environment [51–53].

Characterisation of QTL: Still a difficult task

Today, in plants, several Mendelian mutations have been characterised by positional cloning in plants, but still very few QTL have been definitively characterised at the molecular level ([54, 55], Tab. 1). Direct cloning of a QTL is more difficult than cloning a major gene because the QTL only partially influences character variation and its effect can only be appreciated by statistical methods. For this reason, the resources required are more considerable and the first QTL cloned by map-based cloning correspond to QTL with strong effects that are independent of the environment. Figure 3 illustrates the general strategy used to characterise a QTL. If nothing is known about the physiological and molecular determinism of the character, positional cloning is the most straightforward method to characterise a QTL. If on the other hand some genes involved in the expression of the character are known, it is possible to test whether the polymorphism of one of them (the 'candidate' gene) could explain the variation of the character. In both cases it is necessary to reduce the interval around the QTL through fine mapping.

The population sizes conventionally used do not allow for precise location of QTL with moderate effects (confidence intervals usually range from 10–30 cM). Such segments may comprise several hundreds of genes, so any attempt at characterising or positional cloning of QTL is impracticable. To fine map a QTL it is necessary to compare several near-isogenic lines differing only for a region containing the QTL that has to be located precisely. The QTL can be located more precisely by comparing these new lines to the initial recurrent line [42]. Such lines can be derived through backcrosses or using residual heterozygosity of RILs [56]. The QTL is 'mendelised' when it is the only source of variation for the trait. Introgression lines constitute another point of departure for fine mapping and cloning a QTL. By deriving an F_2 population from a cross between an introgression line and a cultivated line, then self-fertilising the individuals carrying a recombination in the fragment of interest, fixed lines for different subgroups of the initial fragment can be created [57].

Positional cloning can only really be considered when the QTL is precisely located in an interval much smaller than one centimorgan, in which case large insert libraries (YAC or BAC) can be screened. Ideally the distance between marker and QTL should be around the size of a BAC clone. This is obtained by studying a population of several thousands plants [58] and obtaining polymorphic markers closely linked to the QTL. To confirm that the isolated gene corresponds to the QTL

Table 1. Summary of the QTL cloned in plants. The gene function is indicated. When a candidate gene was proposed, it is indicated if it was early (E) or late (L) in the cloning process. Adapted from Salvi and Tuberosa [55]

Species	Trait	QTL	Function	Method	Candidate Gene*	QTN	Functional proof	Ref
Arabidopsis	Flowering time	ED1	CRY2 Cryptochrome	Pos. cloning	Yes (L)	a.a. substitution	Transformation	[69]
	Flowering time	FLW	Transcription Factor	Pos. cloning	Yes (E)	Gene deletion	Transformation	[76]
	Insect resistance	GS-elong	MAM synthase	Pos. cloning	Yes (E)	Nucleotide and indels	No	[78]
	Root morphology	BRX	Transcription Factor	Pos. cloning	No	Premature stop codon	Transformation	[73]
Rice	Heading time	Hd1	Transcription Factor	Pos. cloning	Yes (L)	Unidentified	Transformation	[79]
	Heading time	Hd3a	Unknown	Pos. cloning	Yes (L)	Unidentified	Transformation	[80]
	Heading time	Hd6	Protein kinase	Pos. cloning	No	Premature stop codon	Transformation	[72]
	Heading time	Ehd1	B-type response regulator	Pos. cloning	No	a.a. substitution	Transformation	[70]
	Salt tolerance	SKC1	HKT-type transporter	Pos. cloning	Yes (L)	a.a. substitution	Complementation	[71]
Tomato	Fruit sugar content	Brix9-2-5	Invertase	Pos. cloning	Yes (L)	a.a. substitution	Complementation	[59, 60]
	Fruit shape	Ovate	Unknown	Pos. cloning	No	Premature stop codon	Transformation	[74]
	Fruit size	fw2.2	Unknown	Pos. cloning	No	Unknown	Transformation	[61, 75]

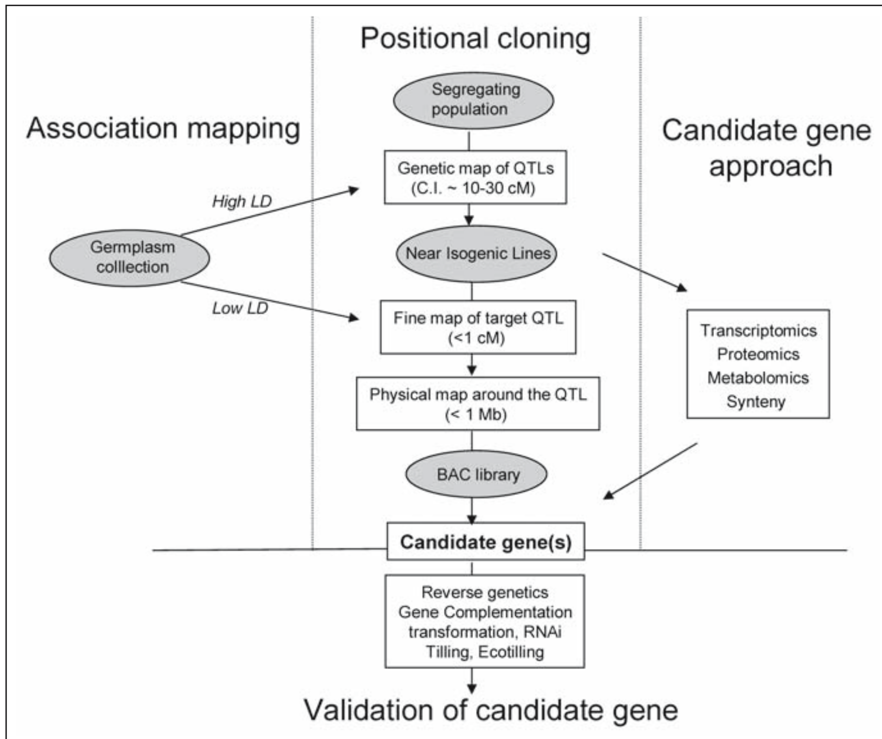

Figure 3. General strategy used to characterise a QTL.

of interest, the ideal situation is to obtain a recombinant within the candidate gene that leads to different values of the trait. For example the cloning of a QTL controlling the variation in sugar content of tomato fruit followed fine mapping [59] and benefited from the existence of recombinations within the gene to localise the QTL in a region of 484bp covering the sequence of a cell-wall invertase. The functional polymorphism was then delimited to an amino acid near the catalytic site which affects enzyme kinetics and fruit sink strength [60]. Transformation with contrasted alleles may allow us to definitively prove that the candidate gene is the QTL. A fruit weight QTL in tomato responsible for about 30% of the variation of this character has been isolated using the classical strategy of high resolution mapping by screening 3472 F2 plants, identifying 53 recombinants (between two markers 4.2 cM apart) and screening a YAC library. From a YAC likely to contain the required gene, a cosmid library was screened and three clones used to transform a tomato variety. The cosmid leading to differences in fruit size after transformation was sequenced and the two sequences corresponding to ORFs were used in a second round of transformation. This allowed the definitive identification of the clone corresponding to the QTL [61]. Certain problems may arise from validation by transformation, as generally we aim to modify the value of a trait by introducing a favourable allele, no easy task when the effect of the environment, the genetic background, and the transformation

(dose effects, gene silencing) may interfere. Constructions to overexpress the gene can be used but carry a risk of seeing artefactual positive effects on the trait.

For certain quantitative characters, the physiology of the plant indicates what the functions in question might be. For others, mutants with phenotypes resembling extreme variations of the character are available. If the corresponding genes are available, whether they are responsible for the QTL of the character studied depends on whether they are polymorphic and whether this polymorphism has repercussions on the variation of the character considered [40, 62, 63].

The confirmation of the role of a candidate gene in the variation of a character is not direct and must proceed via:

- fine mapping of the QTL; testing for co-segregation of the candidate gene and the QTL with thousands of plants may allow the rejection of several candidate genes
- the search for correlations between polymorphisms of the candidate gene and variation of the character in populations in which linkage disequilibrium is minimal (in such populations, only a cause-effect relationship ensures the durability of the correlation throughout the generations). This association mapping approach has already been useful to characterise several QTLs [64–68]
- analysis of the variation at biochemical and metabolic levels. A necessary but not sufficient condition for a gene coding for an enzyme to be a QTL is that the activity of the enzyme must be variable. This has allowed elucidation of the origin of variation at the Lin5 QTL [60]
- molecular analysis of alleles to find the molecular basis of variation; the identification of the polymorphism responsible for the QTL is not straightforward, as it can be either a nucleotide substitution (or indel) causing an amino acid modification [59, 60, 69–71], a stop codon [72, 73], a gene deletion [74] or a mutation in a regulatory sequence that may be very distant from the gene [75–77]. The exact nature substitutions or indels are detected [78]
- transformation, even though this poses specific problems in the case of QTL [77–80]
- complementation of a known mutation corresponding to the same gene [59, 60, 71, 81].

How can systems biology help QTL characterisation?

Functional genomics facilitates gene or QTL cloning at different levels. Due to high throughput technologies, the number of ESTs sequenced and mapped is rapidly increasing for many species, providing new candidate genes [82]. Apart to the access to all the ORFs carried by a genome fragment, this will provide a non limited number of molecular markers useful for map-based cloning. In *Arabidopsis*, the access to the whole genome sequenced has considerably reduced the time to positionally clone a gene [4]. Although the number of genomes fully sequenced is still limited, their number is rapidly growing, now covering a range of botanical families. Synteny with model species should then assist in identifying molecular markers and

candidate genes in related crop species [83]. Even distantly related species exhibit microsynteny (see for example tomato and *Arabidopsis* genomes [84]), thus markers and candidate genes can be transferable across species.

Microarray-based techniques may be helpful for high throughput identification of polymorphisms (SNP or Indels) at thousands of loci simultaneously [85]. Screening for candidate genes is also much more efficient when utilising high throughput tools for genome expression studies. Transcriptional profiling between near isogenic lines may provide a list of differentially expressed genes. Those which map in the QTL region are strong good candidates [86]. Expression profiling may also be used on a mapping population considering the level of expression of a gene as a trait (the QTL are thus expression QTL, called eQTL). These analyses provide important information about the organisation of regulatory networks [87], as eQTL are either located in the region of the corresponding gene (*cis*-regulation) or in a distant region (*trans*-regulation). A review of the first eQTL mapping experiments shows that (i) major effect eQTL are often detected, (ii) up to one-third of eQTL are *cis*-acting, and (iii) eQTL hot spots that explain variation for multiple transcripts are frequent [88]. Correspondence between eQTL and morpho-physiological QTL can then be researched [89]. It almost goes without saying however that this approach is limited by the fact that all the QTL are governed by alterations in RNA amounts.

An alternative approach consists of identifying loci affecting the quantities of protein (*Protein Quantity Loci* or PQLs) or loci responsible for the charge or molecular mass of protein isoforms (*Position Shift Loci* or PSLs) as detected by two-dimensional gel electrophoresis [90]. When a PQL cosegregates with a PSL, the variation of protein quantity can be due to a polymorphism within the protein itself. On the other hand, if PSL and PQL are mapped to distinct regions of the genome, the variation in protein quantity can be due to a trans-acting regulatory factor/gene [91]. In maize, this approach has been useful in discovering genes involved in water-stress tolerance [92]. Proteomic approaches, by revealing polymorphisms within genes as well as differences in protein expression are therefore complementary to DNA marker and mapping approaches. Metabolomic profiling combined to genetic studies may also provide insight on the physiological bases of quantitative trait and give clues on the candidate genes to screen [93]. At last, all the tools available for reverse genetics, collections of mutants, TILLING (Targeting Induced Local Lesions IN Genomes), RNAi (presented below) may be used to validate a candidate.

To recapitulate, forward genetics approaches are thus powerful tools for deciphering natural genotypic variability. They have also been applied to artificially induced mutants in crop and model plant species. In *Arabidopsis* for example, this strategy is yielding remarkable results by allowing the isolation of unknown genes involved in the control of specific phenotypes [94].

Reverse genetics strategies in plants

Several genome-wide gene targeting techniques have been widely developed in plants. In the absence of efficient and routine methods for homologous recombina-

tion in plants, insertional mutatagenesis using transferred DNA (T-DNA) from *Agrobacterium* or transposable elements has been the method of choice for genome size reverse genetics approaches in the model plants *Arabidopsis* and rice. Several populations of tens of thousand of mutagenised plants have been created with the objective to reach near saturation of the collections (e.g., *Arabidopsis* genetic resources at http://www.arabidopsis.org/portals/mutants/worldwide.jsp). Knockout mutants in a given gene can be screened by PCR-search of *Arabidopsis* insertion collections or even by BLAST search of the insertion flanking sequences. Since the probability to hit the gene is lower for small genes than for large genes, loss-of-function mutants for the target gene are not always identified and very large numbers of mutagenised plants are needed to reach near saturation of the collection [95]. Nonetheless, insertion collections have proved to be powerful reverse genetics tools for studying gene function in the context of the plant (as reviewed in [94]). In much the same way, collections of activation tagging lines resulting in gain-of-function phenotypes have been created. Target genes are activated by random insertion in the genome of T-DNA or transposable elements carrying strong promoters [96]. More recently, downregulation of specific genes by using RNAi-based technology [97] has been scaled up to genome-wide level in *Arabidopsis* (e.g., the AGRIKOLA project, http://www.agrikola.org/objectives.html). Genome-scale RNAi approaches take advantage of the easiness of *Agrobacterium* transformation of *Arabidopsis* using the floral dipping technique and of the recent development of site-specific recombination-based cloning vectors allowing efficient and high throughput insertion of inverted repeats of a gene sequence in plant transformation vectors [97, 98]. Though silencing efficiency may vary according to the gene studied, which often results in the observation of a range of more or less severe phenotypic effects in the RNAi silenced plants, this approach is particularly useful when analysing large gene families or classes of genes. In addition to the detailed functional analysis of individual genes, it also allows the study of detectable phenotypes by targeting the regions conserved among several genes in a multigene family, which is very useful when loss-of-function phenotypes are difficult to observe due to the high functional redundancy of plant genes [99]. This strategy may alleviate the need for multiple knockout mutants in order to detect phenotypic changes linked with the mutations in target genes belonging to the same family.

However, these strategies are mostly used for *Arabidopsis* [94] and, to a lesser extent, for rice [100, 101]. Most crop plants still await the development of similar high throughput methods for functional genomics. Considering the case of tomato is instructive. Tomato is the model plant for fleshy fruit development and for *Solanaceae* (among others: potato, tobacco, pepper), and at the same time, a commercial crop of prime importance. Tomato genome size is 950 Mb, i.e., several fold larger than the 125 Mb of *Arabidopsis* but much smaller than the 2,700 Mb of pepper and the 17,000 Mb of wheat, for example. Transposon-based insertional mutagenesis using the non-autonomous mobile elements Activator(Ac)/Dissociation(Ds) from maize have been developed in tomato and shown to be very effective for creating knockout mutants and for promoter-trap studies [102–104]. Activation-tagging lines using T-DNA insertions have also been developed, yielding very interesting

gain-of-function phenotypes (Mathews et al., 2003). However, given the genome size of tomato, near to 200,000 to 300,000 transposon-tagged lines are necessary to obtain 95% saturation of the genome, according to some estimates [106]. Since tomato genetic transformation is based on the low throughput *in vitro* somatic embryogenesis, this goal is still out of reach for most groups, including large consortiums, even when using the miniature tomato cultivar MicroTom suitable for high throughput reverse genetics approaches [102]. Insertional mutagenesis with T-DNA in tomato, which necessitates a plant transformation step to obtain each insertion line, would require even more efforts.

The two rate-limiting steps pointed out for tomato, i.e., large genome size and lack of high throughput transformation methods are common features to most crop plants. Ideally, mutagenesis methods for genome-wide reverse genetics should be applicable to any plant whatever the genome size, remain independent of the availability of high throughput transformation methods for that plant (if such method exists) and give a range of mutations prone to be detected by easy, robust, automated and cheap techniques. With the overwhelming increase in sequence data for model and most field-grown crop plants, such alternatives have been developed in recent years. These methods, based on the use of chemical or physical mutagenesis techniques and previously employed for decades for creating genetic variability, have been mostly exploited until recently in plant breeding programs and in forward genetics approaches aimed at identifying the genes behind the phenotypes.

Chemical mutagens and ionising radiations usually create high density of irreversible mutations ranging from point mutations to very large deletions, depending on the mutagenic agent used. As a consequence, saturated mutant collections can be obtained with only a few thousand mutagenised lines, which should be compared to the hundreds of thousand of lines necessary for reaching near saturation collections of insertional mutants [95]. Unknown mutations in target genes can be screened using low throughput classical methods, including DNA sequencing, which may eventually become the method of choice due to the large decrease in DNA sequencing prices over the last years. The recent development of PCR-based technologies allowing the detection of unknown mutations triggered the rapid development of mutant collections in crop and model plants and of high throughput mutation screening methods aimed at discovering the phenotypes behind the genes. An additional advantage of mutant plants in many countries, especially in some European countries opposed to GMO plants, is that they are not genetically modified organisms and, as such, not subjected to regulatory or public acceptance barriers. Mutant alleles can thus be used for crop improvement using traditional and marker assisted breeding programs.

The following section will describe two of the major reverse genetic techniques recently developed for functional genomics approaches in model and crop species: (i) fast neutron mutagenesis and detection [107] and (ii) TILLING (Targeting Induced Local Lesions IN Genomes) [108, 109].

Fast neutron mutagenesis and mutation detection

Fast neutron bombardment is a highly efficient mutagenic method that creates DNA deletions with size distribution ranging from a few bases to more than 30 kb. As a consequence, knockout mutants are obtained. Since the large deletions generated may encompass several genes, this general reverse-genetics strategy can be particularly useful in plant species where duplicated genes, which often show functional redundancy, are arranged in tandem repeats. Availability of tandem repeat knockouts may overcome the very difficult (or even impossible) task of obtaining double mutants. In addition, similar mutation frequencies are observed whatever the size of the genome of the plant [110], which renders this method very attractive for many crop species. One of its disadvantages is that the occurrence of large deletions may be problematic for subsequent genetic analyses. The construction of a deletion mutant collection is straightforward [102, 107, 111]. Basically, after conducting pilot studies aimed at determining the optimal dose necessary to achieve the rate of mutations desired (typically, half of the mutagenised M1 plants should be fertile enough; [112]), M0 seeds are mutagenised, giving M1 seeds which are sown. The M2 seeds are individually collected from the resulting M1 plants and a fraction of them are sown for collecting plant material for DNA extraction. The remaining M2 seeds can be sown for performing phenotypic and segregation analyses on the M2 families and/or stored until further use.

Screening the collection for mutations is a simple PCR-based technique (named Deleteagene for Delete-a-gene) described for rice and *Arabidopsis* [107, 112]. A region of the target gene is PCR-amplified from DNA samples collected from M2 plants using gene-specific primers. The primers and the length of the PCR extension time are carefully chosen so that deletions in target gene can be detectable by PCR in deletion mutants (typically, 1 kb deletions) but not in wild-type plants (wild-type DNA fragment with larger size is not amplified since extension time is too short). In addition, since PCR methods are highly sensitive, pools of up to 2,500 lines can be screened. Once a positive pool is detected, individual mutants can be detected using the same strategy by deconvolution of the pools and of the subpools, and further confirmed by DNA sequencing of the mutated target gene. Based on screenings performed in *Arabidopis*, about 50,000 mutagenised lines would be necessary to achieve an objective of deletion mutants in about 85% of the targeted loci. While possibly realistic in crop plants bearing dry fruits that are easy to collect (e.g., seeds), this objective is probably very difficult to achieve in some other species where seed harvesting is the limiting step, e.g., in the fleshy fruits such as tomato, melon or grape or in species with long reproductive cycles, e.g., the perennial trees. In tomato for example, the largest fast neutron mutagenesis collection includes several thousand M2 families in cv. M82 [102, 111] (http://zamir.sgn.cornell.edu/mutants/), which is already a huge task to produce. In addition, preliminary knowledge of genomic sequence is preferably needed for efficient PCR screening of deletion mutants thereby reducing the range of species for which this method can be used at the present time. For many crop species, forward genetics will probably remain the best adapted approach for using deletion mutant collections in the few coming years.

TILLING

TILLING is a general reverse-genetics strategy first described by McCallum et al. [108] who used this method for allele discovery for chromomethylase gene in *Arabidopsis* [113]. This method combines random chemical mutagenesis by EMS (ethylmethanesulfonate) with PCR-based methods for detecting unknown point mutations in regions of interest in target genes. Since the early description of the method, which was then performed by using heteroduplex analysis with dHPLC [108], the method has been refined and adapted to high throughput screening by using enzymatic mismatch cleavage with CEL1 endonuclease, a member of the S1 nuclease family [109, 114]. TILLING technology is quite simple, robust, cost-effective and thus affordable for many laboratories. In addition, it allows the identification of allelic series including knockout and missense mutations. For these reasons, this genome-wide reverse-genetics strategy has been applied very rapidly to a growing number of plants, including model plants and field-grown crops of diverse genome size and ploidy levels, and even to insects (*Drosophila* [115]). A number of TILLING efforts in plants have been reported for *Arabidopsis* [109, 116], *Lotus japonicus* [117], barley [118], maize [119] and wheat [120]. Recent reviews give excellent insights on the TILLING methods, from the production of the mutagenised population to the current technologies for mutation detection, and on the future prospects for TILLING [121–124]. In addition, a number of TILLING facilities have been created for various plants including facilities for *Arabidopsis* which already delivered >6,000 EMS-induced mutations in *Arabidopsis* and is also opened to other species [124] (ATP, http://tilling.fhcrc.org:9366/), maize at Purdue University (http://genome.purdue.edu/ maizetilling/), *Lotus* in Norwich (USA) (http://www.lotusjaponicus.org/tillingpages/ Homepage.htm), barley in Dundee (UK) (http://germinate.scri.sari.ac.uk/barley/mutants/), sugar beet in Kiel (Germany) (http://www.plantbreeding.uni-kiel.de/project_tilling.shtml), pea at INRA (Evry, France; http://www.evry.inra.fr/public/projects/tilling/tilling.html) and ecotilling at CanTILL (Vancouver, Canada) (http://www.botany.ubc.ca/can-till/).

Mutagenesis

EMS (ethylmethanesulfonate) is the mutagenic agent used for most of the plant TILLING projects cited above. As a result of EMS alkylation of guanine, more than 99% of mutations are G/C-to-A/T transitions, as experimentally shown by analysing (EMS)-induced mutations in *Arabidopsis* [116]. Other mutagens with genotoxic effects inducing point mutations, frameshifts or small insertion/deletions (InDel) are also likely to be applicable to a TILLING project using CEL1 endonuclease. Indeed, CEL1 technology allows the efficient detection of a broad range of mutations, i.e., the natural allelic variants found in different plant genotypes or ecotypes or the artificially-induced mutations in zebrafish induced by the *N*-ethylnitro-*N*-nitrosourea (ENU) mutagen [125]. With EMS, similar mutation frequencies are expected whatever the plant genome size [110], rendering this approach applicable to most crop species. However, considering the results from the diverse TILLING

projects in different species, the mutation density detected by TILLING may actually range from 1 mutation/Mb in barley [118] and 1 mutation/500 kb in maize [119] to 1 mutation/40 kb in tetraploid wheat and even 1 mutation/25 kb in hexaploid wheat [120]. By comparison, mutation densities are 1 mutation/170 kb in *Arabidopsis* (ATP project [116]) and 1 mutation/125 kb in MicroTom tomato (our own unpublished results). Polyploidy may confer tolerance to EMS mutations, thus explaining the high density of mutations found in wheat [124].

EMS treatment is usually done by soaking the seeds (referred to as M0 seeds) in EMS solution for several hours (usually 12–16 h overnight); mutagenised seeds are then referred to as M1 seeds (Fig. 4). Pollen can also be mutagenised, as done in maize [119, 124]. At this step, a delicate balance has to be found between (i) the primary objective of mutagenesis for TILLING, which is to obtain saturated mutagenesis (i.e., the highest density of mutations possible in the plant genome) in order to analyse a reduced number of lines, and (ii) the amount of mutagenesis that a plant can withstand without overwhelming problems of seed lethality or plant lethality and sterility. In tomato, we obtained high density mutations using EMS doses giving 50–70% of seed lethality after EMS treatment (M1 seeds) and 40–50% of sterile plants in the M1 plants. Since the necessary EMS concentrations may vary considerably according to the species, the physiological state of the seeds and even from batch to batch, pilot studies with different EMS concentrations (from 0.2–1.5%) should be carried out before large scale mutagenesis. The M1 plants obtained by sowing the mutagenised seeds are chimeric and cannot be further used for mutation detection. Indeed, in the embryo, each cell is independently mutagenised. Only a few cells in the apical meristem (e.g., two to three cells in tomato, A. Levy, personal communication) will give rise to reproductive organs and thus to gametes. In contrast, mutations in other embryonic cells are not inherited by the next generation (somatic mutations) and will give rise to chimeric tissues in M1 plants (e.g., the variegated plants with dark green and light green or white sectors often observed in M1 plants).

The M2 seeds, obtained after selfing (or crossing when necessary) the M1 plants, are individually collected from each plant and stored. One or a few M2 plants are usually grown in order to provide plant material for DNA extraction (Fig. 4). Another strategy that we use in tomato, though it involves a time-consuming step, is to grow 12 individual plants per M2 family and to collect M3 seeds and tissue samples from these plants. In addition to enabling the multiplication of the seeds, this strategy allows the description of the plant phenotypes and the segregation analyses of visible mutations in the M2 families. These data are collected and further compiled in a phenotypic description database. The rationale is that once a mutation in a target gene is detected in an individual M2 family, the information on the phenotypic and segregation data can give a first hint on the severity of the mutation and the functional role in the plant of the target gene without having to wait for the observations made on M3 plants. This approach can be particularly useful when dealing with crop species that have a long developmental cycle and/or with specific plant tissues (e.g., fruits or seeds).

In addition to the artificially-induced mutants obtained by using various physical or chemical mutagens in species such as rice [126] or tomato [111], natural allelic

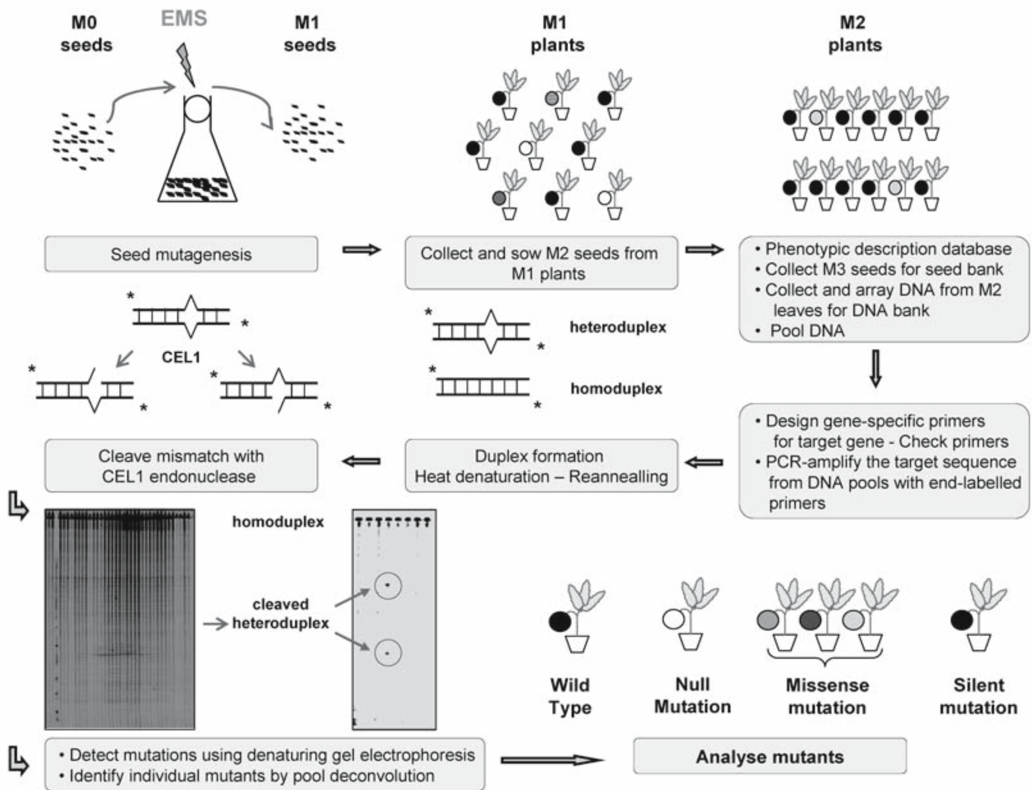

Figure 4. Schematic description of the TILLING procedure. Tomato TILLING strategy is shown. Seeds (M0) are mutagenised with ethylmethanesulfonate (EMS) giving M1 seeds, which are sown. M2 seeds from the resulting M1 plants are collected and sown. For each M2 family, 12 plants are grown and used for: (i) description of plant phenotype (data stored in a tomato mutant database); (ii), extraction of DNA from leaf tissue, later used for mutation detection; and (iii), collection of M3 seeds stored in a seed bank. For mutation detection, eightfold DNA pools are generated from M2 family DNA and gene-specific primers are designed to PCR-amplify the target gene from these pools. The resulting amplicon is heat denatured and reannealed, producing both homoduplexes and heteroduplexes (presence of a mismatch in the duplex). Heteroduplexes are cleaved at the 3' side of the mismatch by the CEL1 endonuclease and further detected by denaturing gel electrophoresis. Identification of the individual M2 family harbouring the mutation is done by deconvolution of the DNA pools using the same technology. Screening tomato mutant collection for a target gene (e.g., a gene involved in fruit colour) yields a series of mutant alleles. Some mutations (~5%) will create knockout mutants (null mutations, ~5%) or affect the biological function of the encoded protein (missense mutations, ~50%) while many mutations (~45%) will remain silent.

variants are already present in germplasm resources, which represent a large source of genetic variability for most crop and model species [57, 127]. Core collections may include related species, various accessions with high genetic diversity often collected near the centre of origin of the species, and cultivated lines and mutants obtained by breeders worldwide (e.g., the Tomato Genomic Resource Center at Davis: http://tgrc.ucdavis.edu/). In addition to the populations of artificially-induced mutants, these collections provide very useful resources for identifying natural alleles for a target gene using Ecotilling. This approach refers to the detection, using high throughput TILLING technology with CEL1 type endonuclease, of allelic variants in the species germplasm (e.g., ecotypes in *Arabidopsis*, hence the name of Ecotilling) [128]. This can be particularly useful in association genetics approaches, for example for the confirmation of the role of a candidate gene previously shown to be co-localised with a QTL.

Mutation detection

A recent review [124] describes in detail the current technologies for mutation and polymorphism detection while Yeung et al. [114] analyses and compares the diverse enzymatic mutation detection technologies available. Basically, three different technologies are used for high throughput mutation discovery in TILLING: (i), the denaturing high performance liquid chromatography (dHPLC), originally used in the first plant TILLING project described [108] and further improved since [118]. The dHPLC is a duplex DNA melting temperature-based system that allows the detection of duplex DNA fragments destabilised by mismatches using temperature-controlled hydrophobic columns. The system is automated and can be used for screening four family DNA pools. However, this technology displays best results with DNA fragment ranging from 300–600 bp and does not allow the precise location of the point mutation; (ii), the single-strand conformational polymorphism (SSCP), which detects conformational changes caused by point mutations and has been improved and automated for capillary DNA sequencers. However, it shows the same limitations as dHPLC, i.e., the limitation to pools of four DNA samples, the detection of fragments <500 bp, and the unknown location of the mismatch; and (iii) enzymatic mismatch cleavage using endonuclease enzymes, members of the S1 nuclease family, followed by electrophoresis separation of the cleaved fragments [109]. This technology has become the method of choice for high throughput TILLING [122].

Originally extracted from celery and later from other plant species, the CEL1 endonuclease is a mismatch cleavage enzyme showing very little sequence bias [114]. In addition, CEL1 has an exonuclease activity that cleaves the 5’ end of DNA fragments, thus releasing the labelled end used for detecting DNA fragments (Fig. 4), which can decrease the sensitivity of the detection. Efficient CEL1 enzyme preparations can be purified from many plant sources [124]. In addition, enzymes performing similar functions have been cloned and are commercially available such as the Surveyor mutation detection kit (http://www.transgenomic.com/flash/surveyor/Surveyor.asp [129]) or the ENDO1 enzyme (http://www.evry.inra.fr/public/projects/tilling/tilling.html).

The technology used for high throughput TILLING with CEL1 is very simple (Fig. 4) and affordable in main research centres. First, a DNA fragment of 0.5–2 kb is amplified from DNA pools (usually eight-fold pools when detecting heterozygous mutations, i.e., 1 genome in 16) with differentially labelled primers. The design of the primer will depend on the previous knowledge of the protein (the most interesting region to target for functional analysis according to the user, e.g., the interacting domain in a transcription factor or the catalytic site in an enzyme), the probability of finding knockout or missense mutations in the region, which can be estimated using the CODDLE (Codons Optimised to detect Deleterious Lesions) software developed by the Seattle group (http://www.proweb.org/coddle) or, more simply for many crop plants, the availability of EST or genomic sequences. Amplification of the DNA fragment with unlabelled primers is usually done in a first round to check the primers, especially when amplifying DNA fragments with no previous knowledge of genomic sequence of the target gene, e.g., EST sequences. In order to reduce the costs of labelled primers specifically designed for a target gene, a two-step strategy can also be followed for amplifying labelled DNA fragments [115]. The labelling of the primers will depend on the electrophoresis equipment used: infrared-based sequencers such as LI-COR, which is commonly used for TILLING due to its robustness and sensitivity [109, 121, 124], or fluorescence-based sequencers such as ABI sequencers [114]. Once the labelled DNA fragment has been amplified, the amplicon is subjected to a high temperature-denaturation/low temperature-reannealing cycle, in order to allow the formation of DNA homoduplexes and heteroduplexes. By using CEL1 endonuclease, which cuts at the 3' side of the mismatch, the heteroduplexes are then cleaved while homoduplexes are left intact by the enzyme (Fig. 4). The cleaved end-labelled DNA fragments can be readily separated from non-cleaved DNA fragments by electrophoresis on denaturing gel. Furthermore, the use of differentially labelled primers allows the precise location on the gel of the two cleaved fragments and thus the detection of the region in the DNA sequence where mutation occurs. In addition to the use of Photoshop software for gel image analysis and band detection, newly developed free software called GelBuddy (www.proweb.org/gelbuddy/index.html) facilitates image analysis of TILLING gels [124].

Once a mutant is detected in a pool of families, the deconvolution of the pool and the detection of the mutated family or plant can be done using the same technology (PCR amplification of target gene, CEL1 cleavage and denaturing gel detection). The mutation in the target gene can thus be confirmed, usually by using DNA sequencing or alternative Single Nucleotide Polymorphism (SNP) detection technologies [124].

Linking mutation to phenotype

EMS induces point mutations, mostly G/C-to-A/T transitions. Single-base change in protein-coding genes may be classified as silent, missense or truncation. Silent mutations do not affect the protein. Missense mutations arise when single base change in a given codon induces changes in the amino acid encoded. Amino acid substitutions can be conservative (similar function is expected) or non conservative (e.g., the substitution of the neutral amino acid glycine by the basic amino acid ar-

ginine, which is expected to modify the function of the protein). The SIFT (Sorting Intolerant From Tolerant) program can be used to predict the damage to protein function caused by missense mutation (http://blocks.fhcrc.org/sift/SIFT.html). Truncations of the protein resulting in knockout mutants are expected from single-base changes converting an amino acid codon to a stop codon or from mutations in splice junctions. From the TILLING experimental results obtained in *Arabidopsis* [116] the proportion of nonsilent mutations that may affect the biological function of the protein and hence the phenotype of the plant, was estimated to be 55%, including 5% of truncations and 50% of missense mutations. Interestingly, there was a considerable bias in favour of heterozygotes for the detection of the most severe mutations (truncations), suggesting that corresponding knockouts mutations in homozygotes were lethal. These overall results highlight the potential of TILLING for discovering allelic series, including knockouts and hypomorphic mutations that are highly informative for functional studies of target genes.

Once a mutation is discovered in a target gene and the corresponding family identified, the effect on the plant resulting from a possible lesion on the protein must be screened phenotypically, usually on the M3 plants. At this point, a major issue is how to differentiate the mutation in the target gene detected by TILLING from the other background mutations in the plant introduced by EMS mutagenesis. Actually, the strategy will depend on the objective of TILLING, i.e., for mutation breeding purposes or for functional study of a target gene. For crop improvement, a number of cycles of backcrossing are necessary before agronomic use. In the highly mutagenised wheat for example, Slade et al. [120] estimated that four backcrosses should be sufficient to derive lines very similar to the parents but did not exclude the need for additional backcrosses. For functional studies, it is generally considered that the fastest method for demonstrating that the mutant phenotype results from a mutation in the target gene is to isolate additional mutant alleles [94].

The optimum number of mutated alleles necessary for functional studies of a gene of interest will mostly depend on the target gene studied. Based on the results obtained in *Arabidopsis* [116], an allelic series including one knockout mutation and ~10 missense mutations that can possibly affect the biological function of the protein should roughly comprise 20 mutated alleles. Depending on the species and the density of mutations in the collection of mutants, this objective usually involves the screening of 3,000–6,000 mutant lines. According to calculations made with *Arabidopsis* TILLING collections [121], the frequency of misattributing a phenotype observed in M3 plants in these collections to a mutation in the target gene can be estimated to ~0.05% when the parent M2 plant is heterozygous. When the M2 plant is homozygous, a backcross is necessary before selfing and analysing the plants. Another possibility is to cross two independent lines mutated in the same target gene. Background mutations are heterozygous in the resulting plants carrying the two non-complementing mutations that can therefore be considered as responsible for the phenotype observed.

Plant systems biology and reverse genetics approaches

During the last few years, tremendous efforts have been made in developing genome-size reverse genetics tools and genetic resources in model and crop plants for studying gene function in the context of the plant. At the same time, the development of high-throughput approaches for global analyses of transcripts, proteins and metabolites paved the way for a comprehensive description of complex networks involved in signal transduction cascades, in regulation and activity of primary or secondary metabolism pathways, and in many other aspects of plant development. These studies have major consequences on our present way of studying plants. First, they allow the discovery of new candidate genes putatively involved in the operation of plant functional networks [94]. Other candidate genes are being generated in both model and crop plants by the forward genetic approaches aimed at identifying the genes underlying the QTLs controlling traits of interest, as previously described. Second, beyond the mere functional study of a single gene, genomic-scale approaches now allow the study of plant biology from the systems level. Visualisation of metabolic pathways and cell functions is already facilitated in some model and crop plants by tools such as MAPMAN which uses transcriptome and metabolome data [130, 131], and models describing complex networks begin to be constructed in plants [132].

Plant mutants have already proved valuable tools for plant functional genomic studies, e.g., for the discovery of the function of new candidate genes and the analysis of their possible contribution to functional complexes or metabolic pathways [94, 133]. Given the very large collections of insertional mutants available in *Arabidopsis*, most of the studies have been focused on knockout mutants. Indeed, null mutants can be very helpful genetic tools for systems biology approaches, as demonstrated in yeast [134], for example. In this genome-scale study, knockout mutants with functions in central metabolism used in combination with computational analyses, flux data and phenotypic analyses gave access to the relative contribution of network redundancy and of alternative pathways to genetic network robustness in yeast. Although comparable studies are still difficult to carry out in plants, integrated analyses of plant primary and secondary metabolic networks using null mutants or overexpressing lines have been attempted [132, 133] and should progress with the availability of new mutant collections and analytical technologies.

In that context, the recent development of large-scale RNAi in *Arabidopsis* and, especially, of the TILLING and Ecotilling approaches in model and crop plants is very promising. The RNAi approach is already used in some model organisms such as *C. elegans* for inducing systematic perturbations of networks in order to study the functional relationships between the components of interacting complexes involved in a signalling pathway [135]. Systems biology approaches can also make use of TILLING and Ecotilling, which reveal allelic series corresponding to several independent point mutations or other small mutations in target genes. Point mutations are more prone than null mutants to cause a range of discrete variations close to those observed in natural populations, where most traits are controlled by Quantitative Trait Loci (QTLs). One advantage of the artificially-induced mutants for

systems biology studies is that they share exactly the same genetic background and can thus be directly compared, while the lines containing the natural allelic variants usually differ by several tens or hundreds of genes, even in Nearly Isogenic Lines. Nonsilent point mutation usually results in protein lesion, the severity of which will cause a more or less profound effect on the biological function of the protein. Point mutations may also produce dominant-negative mutants, which are very useful tools for revealing functional interactions between the components of a complex or a signalling pathway [2], or even gain-of-function mutants such as the tomato LIN5 invertase variant with altered kinetic properties [60], originally cloned as a QTL controlling solid soluble solids content in tomato fruit [59]. The wide collection of mutants available for a gene of interest identified through TILLING should be particularly amenable for systems biology approaches since a range of quantitative effects, and not only of qualitative effects as in null mutants, can be obtained.

How to use these mutants? One of the most immediate applications in network analysis for mutants detected by TILLING is probably the study of the regulation of metabolic pathways. Although few TILLING results have been published to date, two of the target genes analysed were involved in sugar metabolism, either in starch synthesis [120] or in the synthesis of callose, a beta-1,3-glucan [136]. Metabolite profiling is a high throughput technology with limited cost per sample that allows the initial screening of the allelic mutants identified, even those showing no visual phenotype. Furthermore, since the establishment of network regulation needs large-scale studies involving as many different mutants in several target genes as possible [132, 134], metabolic profiling can be reduced in a first step to rapid metabolic fingerprinting of the mutants, as already experimented with mutants displaying a silent phenotype [137, 138]. In this approach, the most interesting mutants showing significant perturbations in metabolite profiles can be subsequently subjected to more detailed analyses, including transcriptome, proteome and metabolome profiling. The global set of data obtained can be further combined and analysed with the array of tools already available ([130, 139] and Chapters by Dieuaide-Noubhani et al., Nikiforova and Willmitzer, and Ahrens et al.), in order to validate the underlying hypotheses on the functional role of the target gene studied and/or to give a comprehensive view of the metabolic network [140]. One delicate step for fully understanding the changes in the metabolic network induced by the mutation in the target genes remains the analysis of the metabolic fluxes ([141] and Chapter by Dieuaide-Noubhani et al.), which can hardly be carried out in a high throughput manner in plants, and, therefore, will probably remain restricted to a limited number of mutants previously selected through global analyses.

Summary

The first experiments on gene and QTL mapping date from the late 1980s. Since that time, hundreds of mapping experiments have been performed, providing information on the genetic basis of individual traits or allowing complex traits to be dissected into their component parts. The number of Mendelian mutations characterised by a candi-

date gene approach or positional cloning has rapidly increased, but very few QTL have been characterised to date. Accumulated data from several species suggest a continuum between discrete variations (mutant genes) and continuous variations (QTL), and the identification of QTL will improve our understanding of the molecular and physiological basis to complex character variation. In this context, gene maps and large EST data sets will prove useful as sources of candidates. The access to a growing number of sequenced genomes, and to transcriptomic and proteomic approaches, should increase the efficiency of QTL characterisation. Furthermore ecophysiological modelling and metabolomic profiling will give clues to the physiological processes underlying QTL and the potential candidate genes. In this context, fine mapping of the QTL and validation of the candidate genes will become the most restrictive steps.

The development of large scale DNA sequencing facilities and of high throughput gene and protein expression and metabolite profiling technologies in model and crop plants has triggered the development of genome-wide reverse genetics tools aimed at identifying and characterising the function of candidate genes in the context of the plant. Insertional mutagenesis using T-DNA or transposons that creates knockout or activation-tagged mutants and, more recently, large scale gene targeting by RNAi have been the methods of choice for functional genomics in the model plants *Arabidopsis* and rice. However, most of the above mentioned tools are unavailable in crop plants due to limitations (low throughput genetic transformation technologies, size of the genome) inherent to the species. For these reasons, new technologies for detecting unknown mutations created by chemical mutagens or ionising radiations have emerged in the recent years. Among them, the TILLING (Targeting Induced Local Lesions In Genomes) technology, which is mostly based on the generation by a chemical mutagen (EMS) of high density point mutations evenly distributed in the genome and on the subsequent screening of the mutant collection by a PCR-based enzymatic assay, has become very popular and is currently applied to a wide variety of model and crop plants. Chemical mutagenesis used in the TILLING procedure generates a range of mutated alleles for a target gene, including knockouts and missense mutations, thereby affecting more or less severely the biological function of the corresponding protein and the phenotype of the plant. These allelic series should prove valuable tools for plant systems biology studies by enabling the comparative analysis of metabolic or other complex networks in plants showing genetic variability for a target gene with the help of genomics (transcriptome, proteome, metabolome) and data analysis/modelling tools.

References

1. Rensink WA, Lee Y, Liu J, Iobst S, Ouyang S, Buell RC (2005) Comparative analyses of six solanaceous transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. *BMC Genomics* 6: 124
2. Diévart A, Clark SE (2003) Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. *Curr Opin Plant Biol* 6: 507–516
3. Tanksley SD, Ganal MW, Martin GB (1995) Chromosome landing – a paradigm for map-based gene cloning in plants with large genomes. *Trends Genet* 11: 63–68

4. Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol* 129: 440–450
5. Xu YB, McCouch SR, Zhang QF (2005) How can we use genomics to improve cereals with rice as a reference genome? *Plant Mol Biol* 59: 7–26
6. Hackett CA (2002) Statistical methods for QTL mapping in cereals. *Plant Mol Biol* 48: 585–599
7. Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199
8. Lincoln SE, Daly MJ, Lander ES (1992) Constructing genetic maps with MAPMAKER/EXP version 3.0. A tutorial and reference manual http://linkage.rockeffeller.edu/soft/mapmaker
9. Jansen RC, Stam P (1994) High-resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136: 1447–1455
10. Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468
11. Churchill GA, Doerge RW (1996) Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971
12. Darvasi A, Soller M (1997) Simple method to calculate resolving power and confidence interval of QTL map location. *Behav Genet* 27: 125–132
13. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Mol Biol* 48: 453–461
14. Grattapaglia D, Bertolucci FL, Sederoff RR (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Theor Appl Genet* 90: 933–947
15. Bradshaw HD, Stettler RF (1995) Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. *Genetics* 139: 963–973
16. Tanksley SD, Nelson JC (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
17. 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 QTLs. *Genetics* 141: 1147–1162
18. Ronin Y, Korol A, Shtemberg M, Nevo E, Soller M (2003) High-resolution mapping of quantitative trait loci by selective recombinant genotyping. *Genetics* 164: 1657–1666
19. Lynch M, Walsh B (eds) (1998) Genetics and analysis of quantitative traits. Sinauer Associates
20. De Vienne D, Causse M (2002) Mapping and characterizing quantitative trait loci. In: D de Vienne (ed): *Molecular markers in Plant genetics and Biotechnology*. Sci Publisher Inc, 89–125
21. Xu SZ (2003) Theoretical basis of the Beavis effect. *Genetics* 165: 2259–2268
22. Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; where are we now? *Heredity* 80: 137–142
23. Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. *Nature* 435: 95–98
24. Grandillo S, Ku HM, Tanksley SD (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor Appl Genet* 99: 978–987

25. Fulton TM, Bucheli P, Voirol E, López J, Pétiard V, Tanksley SD (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. *Euphytica* 127: 163–177
26. Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* 168: 2169–2185
27. Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Sayama H, Uhlig J, Zamir D et al. (1998) Advanced back-cross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium*. *Theor Appl Genet* 97: 1191–1196
28. Fatokun CA, Menanciohautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mung bean based on rflp mapping. *Genetics* 132: 841–846
29. Timmerman-Vaughan GM, Mccallum JA, Frew TJ, Weeden NF, Russell AC (1996) Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). *Theor Appl Genet* 93: 431–439
30. Maughan PJ, Maroof MAS, Buss GR (1996) Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. *Theor Appl Genet* 93: 574–579
31. Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SR, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269: 1714–1718
32. Devos M, Gale D (1997) Comparative genetics in the grasses. *Plant Mol Biol* 35: 3–15
33. Frary A, Doganlar S, Daunay MC, Tanksley S D (2003) QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. *Theor Appl Genet* 107: 359–370
34. Tanksley SD (1993) Mapping Polygenes. *Annu Rev Genet* 27: 205–233
35. Eshed Y, Zamir D (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* 143: 1807–1817
36. Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* 386: 485–488
37. Chaib J, Lecomte L, Buret M, Causse M (2006) Stability over genetic backgrounds, generations and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. *Theor Appl Genet* 112: 934–944
38. Prioul JL, Quarrie SA, Causse M, de Vienne D (1997) Dissecting complex physiological functions through the use of molecular quantitative genetics. *J Exp Bot* 48: 1151–1163
39. Lefebvre V, Palloix A (1996). Both epistatic and additive effects of QTLs are involved in polygenic induced resistance to disease: a case study, the interaction pepper – *Phytophthora capsici* Leonian. *Theor Appl Genet* 93: 503–511
40. Sergeeva LI, Keurentjes JJB, Bentsink L, Vonk J, van der Plas LHW, Koornneef M, Vreugdenhil D (2006) Vacuolar invertase regulates elongation of *Arabidopsis thaliana* roots as revealed by QTL and mutant analysis. *Proc Natl Acad Sci USA* 103: 2994–2999
41. Korol AB, Ronin YI, Itskovich AM, Peng J, Nevo E (2001) Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics* 157: 1789–1803

42. Paterson AH, de Verna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. *Genetics* 124: 735–742
43. Monforte AJ, Tanksley SD (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. *Theor Appl Genet* 100: 471–479
44. Lecomte L, Saliba-Colombani V, Gautier A, Gomez-Jimenez MC, Duffé P, Buret M, Causse M (2004) Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato. *Mol Breeding* 13: 1–14
45. Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci USA* 90: 10972–10976
46. Jansen RC, van Ooijen JW, Stam P, Lister C, Dean C (1995) Genotype-by environment interaction in genetic mapping of multiple quantitative trait loci. *Theor Appl Genet* 91: 33–37
47. Romagosa I, Ullrich SE, Han F, Hayes PM (1996) Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theor Appl Genet* 93: 30–37
48. Moreau L, Charcosset A, Gallais A (2004) Use of trial clustering to study QTL x environment effects for grain yield and related traits in maize. *Theor Appl Genet* 110: 92–105
49. Loudet O, Chaillou S, Merigout P, Talbotec J, Daniel-Vedele F (2003) Quantitative trait loci analysis of nitrogen use efficency in *Arabidopsis*. *Plant Physiol* 131 : 345–358
50. Juenger TE, Mckay JK, Hausmann N, Keurentjes JJB, Sen S, Stowe KA, Dawson TE, Simms EL, Richards JH (2005) Identification and characterization of QTL underlying whole-plant physiology in *Arabidopsis thaliana*: delta C-13, stomatal conductance and transpiration efficiency. *Plant Cell Environ* 28: 697–708
51. Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the response of maize leaf growth to temperature and water deficit. *Plant Physiol* 131: 664–675
52. Quilot B, Kervella J, Genard M, Lescourret F (2005) Analysing the genetic control of peach fruit quality through an ecophysiological model combined with a QTL approach. *J Exp Bot* 56: 3083–3092
53. Yin XY, Struik PC, Kropff MJ (2004) Role of crop physiology in predicting gene-to-phenotype relationships. *Trends Plant Sci* 9: 426–432
54. Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. *Trends Genet* 19: 303–306
55. Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci* 10: 298–304
56. Tuinstra MR, Ejeta G, Goldbrough PB (1997) Heterogeneous inbred families (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theor Appl Genet* 95: 1005–1011
57. Zamir D (2001) Improving plant breeding with exotic genetic libraries. *Nat Rev Genet* 2: 983–988
58. Durrett RT, Chen KY, Tanksley SD (2002) A simple formula useful for positional cloning. *Genetics* 160: 353–355
59. 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* 97: 4718–4723

60. Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 305: 1786–1789
61. Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289: 85–88
62. Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach. A Review. *Molecular Breeding* 7: 275–291
63. Byrne PF, McMullen MD, Snooks ME, Musket TA, Theuri JM, Widstrom NW, Wiseman BR, Coe EH (1996) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc Natl Acad Sci USA* 93: 8820–8825
64. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* 28: 286–289
65. Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weinig C, Schmitt J, Purugganan MD (2004) Linkage disequilibrium mapping of *Arabidopsis* CRY2 flowering time alleles. *Genetics* 167: 1361–1369
66. Osterberg MK, Shavorskaya O, Lascoux M, Lagercrantz U (2002) Naturally occurring indel variation in the *Brassica nigra* COL1 gene is associated with variation in flowering time. *Genetics* 161: 299–306
67. Gupta V, Mukhopadhyay A, Arumugam N, Sodhi YS, Pental D, Pradhan AK (2004) Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. *Theor Appl Genet* 108: 743–749
68. Guillet-Claude C, Birolleau-Touchard C, Manicacci D, Rogowsky PM, Rigau J, Murigneux A, Martinant JP, Barriere Y (2004) Nucleotide diversity of the ZmPox3 maize peroxidase gene: Relationships between a MITE insertion in exon 2 and variation in forage maize digestibility. *BMC Genetics* 5: Art. No. 19
69. El-Assal SED, Alonso-Blanco C, Peeters AJM, Raz V, Koornneef M (2001) A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. *Nat Genet* 29: 435–440
70. Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, Yano M, Yoshimura A (2004) Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-Iike gene expression independently of Hd1l. *Genes Dev* 18: 926–936
71. Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nat Genet* 37: 1141–1146
72. Takahashi Y, Shomura A, Sasaki T, Yano M (2001) Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proc Natl Acad Sci USA* 98: 7922–7927
73. Mouchel CF, Briggs GC, Hardtke CS (2004) Natural genetic variation in *Arabidopsis* identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. *Genes Dev* 18: 700–714
74. Liu JP, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc Natl Acad Sci USA* 99: 13302–13306
75. Cong B, Liu JP, Tanksley SD (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc Natl Acad Sci USA* 99: 13606–13611
76. Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J, Weigel D (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc Natl Acad Sci USA* 102: 2460–2465

77. Clark RM, Linton E, Messing J, Doebley JF (2004) Pattern of diversity in the genomic region near the maize domestication gene tb1. *Proc Natl Acad Sci USA* 101: 700–707
78. Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc Natl Acad Sci USA* 100: 14587–14592
79. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y et al. (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS *Plant Cell* 12: 2473–2483
80. Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) Hd3a, a rice ortholog of the *Arabidopsis* FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol* 43: 1096–1105
81. Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* 386: 485–488
82. Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S (2002) Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell* 14: 1441–1456
83. Schmidt R (2000) Synteny: recent advances and future prospects. *Current Opin Plant Biol* 3: 97–102
84. Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 14: 1457–1467
85. Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res* 13: 513–523
86. Wayne ML, McIntyre LM (2002) Combining mapping and arraying: An approach to candidate gene identification. *Proc Natl Acad Sci USA* 99: 14903–14906
87. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297–302
88. Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends Genet* 21: 616–623
89. Guillaumie S, Charmet G, Linossier L, Torney V, Robert N, Ravel C (2004) Colocation between a gene encoding the bZip factor SPA and an eQTL for a high-molecular-weight glutenin subunit in wheat (*Triticum aestivum*). *Genome* 47: 705–713
90. Damerval C, Maurice A, Josse JM, de Vienne D (1994) Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* 137: 289–301
91. de Vienne D, Leonardi A, Damerval C, Zivy M (1999) Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. *J Exp Bot* 50: 303–309
92. Consoli L, Lefevre A, Zivy M, de Vienne D, Damerval C (2002) QTL analysis of proteome and transcriptome variations for dissecting the genetic architecture of complex traits in maize. *Plant Mol Biol* 48: 575–581
93. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Brudigam C, Kopka J et al. (2006) Genetics of metabolite in fruits of interspecific introgressions of tomato. *Nat Biotechnol* 24: 447–454

94. Ostergaard L, Yanofsky MF (2004) Establishing gene function by mutagenesis in *Arabidopsis thaliana*. *Plant J* 39: 682–696
95. Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 11: 2283–2290
96. Schneider A, Kirch T, Gigoshvili T, Mock H-P, Sonnewald U, Simon R, Flügge U-I, Werr W (2005) A transposon-based activation-tagging population in *Arabidopsis thaliana* (TAMARA) and its application in the identification of dominant developmental and metabolic mutations. *FEBS Lett* 579: 4622–4628
97. Waterhouse PM, Heliwell CA (2003) Exploring plant genomes by RNAi-induced gene silencing. *Nat Rev Genet* 4: 29–38
98. Karimi M, Inzé D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* 7: 193–195
99. Meinke DW, Meinke LK, Showalter TC, Schissel AM, Mueller LA, Tzafrir I (2003) A sequence-based map of *Arabidopsis* genes with mutant phenotypes. *Plant Physiology* 131: 409–418
100. An G, Jeong D-H, Jung K-H, Lee S (2005) Reverse genetics approaches for functional genomics of rice. *Plant Mol Biol* 59: 111–123
101. Hirochika H, Guiderdoni E, An G, Hsing Y-I, Eun MY, Han C-D, Upadhyaya N, Ramachandran R, Zhang Q, Pereira A et al. (2004) Rice mutant resources for gene discovery. *Plant Mol Biol* 54: 325–334
102. Meissner R, Jacobson Y, Melamed S, Levyatuv S, Shalev G, Ashri A, Elkind Y, Levy AA (1997) A new model system for tomato genetics. *Plant J* 12: 1465–1472
103. Meissner R, Chague V, Zhu Q, Emmanuel E, Elkind Y, Levy AA (2000) A high throughput system for transposon tagging and promoter trapping in tomato. *Plant J* 38: 861–872
104. Gidoni D, Fuss E, Burbidge A, Speckmann GJ, James S, Nijkamp D, Mett A, Feiler J, Smoker M, de Vroomen MJ et al. (2003) Multi-functional T-DNA/Ds tomato lines designed for gene cloning and molecular and physical dissection of the tomato genome. *Plant Mol Biol* 51: 83–98
105. Mathews H, Clendennen SK, Caldwell CG, Connors K, Matheis N, Schuster DK, Menasco DJ, Wagoner W, Lightner J, Wagner DR (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *Plant Cell* 15: 1689–1703
106. Emmanuel E, Levy AA (2002) Tomato mutants as tools for functional genomics. *Curr Opin Plant Biol* 5: 112–117
107. Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y (2001) A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J* 27: 235–242
108. McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. *Nat Biotechnol* 18: 455–457
109. Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S (2001) High-throughput screening for induced point mutations. *Plant Physiol* 126: 480–484
110. Koornneef M, Dellaert LWM, van den Veen JH (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidospis thaliana* (L.) Heynh. *Mutat Res* 93: 109–123
111. Menda N, Semel Y, Peled D, Eshed Y, Zamir D (2004) *In silico* screening of a saturated mutation library of tomato. *Plant J* 38: 861–872
112. Li X, Zhang Y (2002) Reverse genetics by fast neutron mutagenesis in higher plants. *Funct Integr Genomics* 2: 254–258

113. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* 292: 2077–2080
114. Yeung AT, Hattangadi D, Blakesley L, Nicolas E (2005) Enzymatic mutation detection technologies. *Biotechniques* 38: 749–758
115. Winkler S, Schwabedissen A, Backasch D, Bökel C, Seidel C, Bönisch S, Fürthauer M, Kuhrs A, Cobreros L, Bran M, Gonzalez-Gaitan M (2005) Target-selected mutant screen by TILLING in *Drosophila*. *Genome Res* 15: 718–723
116. Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR et al. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* 164: 731–740
117. Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, Parniske M (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol* 131: 866–871
118. Caldwell DG, McCallum N, Shaw P, Muehlbauer GJ, Marshall DF, Waugh R (2004) A structured mutant population for forward and reverse genetics in Barley (*Hordeum vulgare* L.). *Plant J* 40: 143–150
119. Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR et al. (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biology* 4: 12
120. Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, non transgenic approach to wheat crop improvement by TILLING. *Nat Biotechnol* 23: 75–81
121. Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. *Annu Rev Plant Biol* 54: 375–401
122. Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* 13: 524–530
123. Gilchrist EJ, Haughn GW (2005) TILLING without a plough: a new method with applications for reverse genetics. *Curr Opin Plant Biol* 8: 211–215
124. Comai L, Henikoff S (2006) TILLING: practical single-nucleotide mutation discovery. *Plant J* 45: 684–694
125. Wienholds E, van Eeden F, Kosters M, Mudde J, Plasterk RHA, Cuppen E (2003) Efficient target-selected mutagenesis in zebrafish. *Genome Res* 13: 2700–2707
126. Wu JL, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba MRS, Ramos-Pamplona R, Mauleon R, Portugal A, Ulat J et al. (2005) Chemical- and irradiation-induced mutants of Indica Rice IR64 for forward and reverse genetics. *Plant Mol Biol* 59: 85–97
127. Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277: 1063–1066
128. Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR et al. (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J* 37: 778–786
129. Qiu P, Shandilya H, D'Alessio JM, O'Connor K, Durocher J, Gerard GF (2004) Mutation detection using Surveyor nuclease. *Biotechniques* 36: 702–707
130. Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Wilmitzer L, Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO rep* 4: 1–5
131. Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37: 914–939

132. Sweetlove LJ, Fernie AR (2005) Regulation of metabolic networks: understanding metabolic complexity in the system biology era. *New Phytol* 168: 9–24
133. Fridman E, Pichersky E (2005) Metabolomics, genomics, proteomics and the identification of enzymes and their substrates and products. *Curr Opin Plant Biol* 8: 242–248
134. Blank LM, Kuepfer L, Sauer U (2005) Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutation in yeast. *Genome Biol*. 6: R49
135. Tewari M, Hu PJ, Ahn JS, Ayivi-Guedelhoussou N, Vidalain P-O, Li S, Milstein S, Armstrong CM, Boxem M, Butler MD et al. (2004) Systematic interactome mapping and genetic perturbation analysis of a *C. elegans* TGF-β signalling network. *Mol Cell* 13: 469–482
136. Enns LC, Kanaoka MM, Torii KU, Comai L, Okada K, Cleland RE (2005) Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant Mol Biol* 58: 333–349
137. Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic pathways unravel the effect of silent plant phenotypes. *Proc Natl Acad Sci USA* 101: 7809–7811
138. Scholtz M, Gatzek S, Sterling A, Fiehn O, Selbig J (2004) Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* 20: 2447–2454
139. Fernie AR, Trethevey RN, Krotzky AJ, Wilmitzer L (2004) Metabolite profiling: from diagnostic to system biology. *Nat Rev* 5: 1–7
140. Sweetlove LJ, Last RL, Fernie AR (2003) Predictive metabolic engineering: a goal for systems biology. *Plant Physiol* 132: 420–425
141. Rontein D, Dieuaide-Noubhani M, Dufourc EJ, Raymond P, Rolin D (2002) The metabolic architecture of plant cells: Stability of central metabolism and flexibility of anabolic pathways during the growth cycle of tomato cells. *J Biol Chem* 277: 43948–43960