REVIEW / SYNTHÈSE

Inexpensive fine mapping and positional cloning in plants using visible, mapped transgenes

Stephen J. Dinka and Manish N. Raizada

Abstract: Vast numbers of crop, fungal, and animal accessions as well as insect vectors and evolving eukaryotic pathogens await molecular analysis. Inexpensive methods are required to make map-based gene isolation accessible to more of the world's researchers. Today, positional cloning relies on genotyping and phenotyping large numbers of progeny to detect chromosome recombination events that break linkage between the trait of interest and flanking molecular markers following meiosis. In the postgenome era, positional cloning will no longer be limited by the availability of high-density molecular markers but rather by the skilled labour and the expense of genotyping and phenotyping 10^3–10^4 progeny to detect rare recombination events in a narrow chromosome block flanking the target gene of interest. Here, we review how linked, mapped transgenes that encode dominant, visible traits such as green fluorescent protein can be used to preselect meiotic recombinants inexpensively, thus reducing progeny genotyping and phenotyping requirements by >95% during positional cloning. Because transgene markers such as green fluorescent protein are genotype independent, transgenes created in one inbred line may be used to fine-map genetic variation in large numbers of genotypes.

Key words: positional cloning, map-based cloning, genetic mapping, plant genomics, genetic markers TMARS/FMARS.

Résumé : Il existe un grand nombre d'accessions fongiques et animales, ainsi que d'insectes vecteurs et de pathogènes eucaryotes en évolution, en attente d'analyse moléculaire. Un besoin existe également pour des méthodes peu dispendieuses pour rendre l'isolation de gène basée sur des cartes, accessible à un plus grand nombre de chercheurs du monde. Aujourd'hui, le clonage positionnel est basé sur le génotypage et le phénotypage d'une nombreuse progéniture pour détecter les évènements chromosomiques qui brisent le lien entre le caractère d'intérêt et les marqueurs moléculaires correspondants, suite à la méiose. À l'ère post-génomique, le clonage positionnel ne sera plus limité par la disponibilité de marqueurs à haute densité, mais plutôt par l'expertise, et ceci aux dépens du génotypage et du phénotypage de 10^3–10^4 progénitures, pour détecter les rares évènements de recombinaison, dans un bloc étroit de chromosomes reliés au gène cible d'intérêt. Les auteurs révisent comment les transgènes liés et cartographiés, qui codent pour des caractères dominants et visibles, comme la protéine fluorescente verte, peuvent être utilisés à peu de frais pour la pré-sélection de recombinants méiotiques, réduisant ainsi les besoins en génotypage et phénotypage de >95 %, au cours du clonage positionnel. Parce que les marqueurs transgéniques, comme la protéine fluorescente verte, sont indépendants du génotype, les transgènes générés dans une lignée consanguine, peuvent être utilisés pour produire des cartes fines de la variation génétique chez un grand nombre de génotypes.

Mots clés : clonage positionnel, clonage basé sur une carte, carte génétique, génomique végétale, marqueurs génétiques, TMARS/FMARS.

[Traduit par la Rédaction]

Introduction

The vast majority of the world's genetic variation remains to be characterized. A key breakthrough occurred 25 years ago when the co-inheritance of parent-specific DNA polymorphisms with parent-specific traits as linkage blocks following meiosis was demonstrated (Botstein et al. 1980). As a result of this observation, map-based gene isolation became possible by employing meiotic recombination to break co-inheritance between the trait of interest and parent-specific linked DNA markers that serve to anchor the trait to a narrowing physical block on a parental chromosome (Paterson et al. 1988). Unfortunately, >95% of F_2 test-crossed progeny do not carry a recombination event within the 5 cM interval flanking a target gene of interest. As a result, almost all postmeiotic progeny analyzed are noninformative. Because organisms such as plants, fungi, insects, fish, worms, and amphibians are not limited by the numbers of meioses per generation, researchers simply genotype and phenotype large numbers of F_2 progeny during positional

Received 18 May 2005. Published on the NRC Research Press Web site at http://canjbot.nrc.ca on 31 March 2006.

S.J. Dinka and M.N. Raizada.[1] Canadian Laboratory of Inexpensive Maize Biotechnologies, Department of Plant Agriculture, University of Guelph, Guelph, ON N1G 2W1, Canada.

[1]Corresponding author (e-mail: Raizada@uoguelph.ca).

 doi:10.1139/B05-170 © 2006 NRC Canada

cloning to detect only a few informative recombinants; this requires significant labour and money (Lukowitz et al. 2000; Jander et al. 2002). Large-scale genotyping and phenotyping are beyond the reach of many of the world's researchers. In this review, we (*i*) analyze the current effort required for positional cloning in plants, (*ii*) review how mapped transgenes encoding visible traits may be used to accelerate positional cloning, and (*iii*) summarize the feasibility of employing existing transgene resources to accelerate positional cloning in plant species today.

A review of current positional cloning efforts in plants

How many postmeiotic progeny do plant researchers currently genotype to find rare recombination events during positional cloning? In Table 1, we summarize the mean number of progeny that researchers genotyped during 84 positional cloning experiments in 11 plant species (S.J. Dinka and M.N. Raizada, unpublished analysis). The mean number of progeny genotyped varied between species, ranging from approx. 1100 in *Arabidopsis* to approx. 2400 in rice and wheat to approx. 3100 in barley. There was tremendous variation in the number of progeny that were genotyped between different studies within a species. For example, in *Arabidopsis*, the number of progeny genotyped ranged from 174 to 4563 plants. In some instances, including the positional cloning of the *Arabidopsis CRR4* gene, researchers had prior sequence knowledge of the gene of interest and were able to accurately select candidate genes from a large interval of 304 kb containing many potential candidates, thereby reducing the number of progeny genotyped to just 243 F_2 plants (Kotera et al. 2005). In contrast, low levels of recombination in the region surrounding the gene of interest can increase the number of progeny that need to be genotyped to detect rare recombination events. During the positional cloning of the *VTC2* locus in *Arabidopsis*, Jander et al. (2002) encountered suppressed recombination frequency around *VTC2*. They genotyped 3700 plants and discussed the wisdom of genotyping larger mapping populations than the theoretical minimum number needed to ensure that enough recombinants are identified. There is an element of chance that plays a part in determining how many progeny need to be genotyped to achieve a desired resolution (Durrett et al. 2002), as recombination frequencies vary within the genome, but this is especially true when the target locus is located in a region of a chromosome with suppressed meiotic recombination, in which case, screening for interval recombinants is especially daunting.

Although the mean number of progeny that were genotyped in *Arabidopsis* is up to threefold lower than in larger genomes, in sequenced genomes, it may be misleading to conclude that the effort required to narrow the search to a single gene is more efficient in species with small genomes. For example, in an analysis of positional cloning studies in *Arabidopsis* shown in Table 1, the mean final map resolution reported is 16 genes (in a mean resolution of 86 kb) out of 20 studies, which may be a reflection of the high mean genes to centiMorgans ratio in *Arabidopsis* (56.3 genes/cM, calculated by dividing the genome-wide gene number by total genome centiMorgans). In contrast, although three studies in wheat reported genotyping approx. 2.4-fold more progeny than the mean number in *Arabidopsis*, a recent positional cloning study demonstrated that only two genes were present within a 324-kb target interval in wheat (Yan et al. 2003). In genomic regions that have not been sequenced, clearly, a low kilobase to centiMorgan ratio is advantageous to be able to map a gene to a bacterial artificial chromosome clone, but in sequenced, well-annotated genomes, the rate-limiting step is not kilobase distances between breakpoints (kilobases/centiMorgan) but between genes; hence, the critical ratio of interest is the number of genes per centiMorgan in such situations. Both of these ratios vary throughout any genome, as genes in species with large genomes are sometimes clustered into gene-rich islands (Barakat et al. 1997).

Second, differences in the number of progeny genotyped between organisms (lower in *Arabidopsis* compared with cereal genomes) may simply reflect differences in what is considered an acceptable number of candidate genes. In *Arabidopsis*, a 16-gene map resolution is acceptable because complementation is easy. In sharp contrast, in maize, wheat, and many of the world's most important crops, transformation is time consuming (6 months to 1 year) and genotype dependent. Because alleles of interest are typically not in transformable genotypes (e.g., maize inbred A188, wheat 'Bobwhite'), complementation would require several generations of backcrossing to introgress a transgene into the mutant background. Therefore, researchers must achieve a better candidate gene resolution (by obtaining more recombinants) in species that are difficult to transform.

To achieve single-gene map resolution with 95% probability of success, we calculate that 6×10^3 to $>10^4$ F_2 testcross progeny would need to be genotyped in rice, maize, and *Arabidopsis* using crude, mean-genome indicators in an F_2 test-cross population. This is considered a conservative estimate using the equation $N = (4.744 \times 100 \times R)/T$, where N is the number of F_2 testcross progeny required, R is the genome mean kilobase to centiMorgan ratio, and T is the gene block resolution desired (adapted from Durrett et al. 2002). The actual numbers will vary throughout the genome, but the point is that the number of progeny required to be genotyped is very large, and researchers who do not employ other selection strategies must either genotype large numbers, rely on chance as recently calculated by Durrett et al. (2002), or rely on complementation, which as noted is not feasible for many crop species. Alternatively, researchers must have multiple alleles of a mutant that can be sequenced to correlate genotype with phenotype, but this approach becomes less feasible as the candidate gene number increases or when examining natural variation based phenotypes. Note that the studies summarized in Table 1 are not a random sampling of positional cloning attempts but only those that were successful. Clearly, a novel approach is needed to achieve fine gene resolution mapping in diverse species.

In addition to genotyping, positional cloning is limited by the need to correlate inheritance of a polymorphic phenotype with a molecular marker (Botstein et al. 1980). In theory, positional cloning only requires that a few progeny be phenotyped, those known to carry an interval recombinant; in

© 2006 NRC Canada

Table 1. Mean number of progeny sampled during positional cloning gene discovery using current random genotyping methods.

Species	No. of studies analyzed	Mean no. of progeny genotyped	Range for no. of progeny genotyped	Mean target interval (kb)	Target interval range (kb)
Arabidopsis thaliana	57	1082	174–4563	86	10–305
Oryza sativa	9	2402	339–3305	56	8–150
Lycopersicon esculentum	5	1153	747–1815	149	17–400
Triticum monococcum and *Triticum durum*	3	2428	1340–3095	300	260–324
Lotus japonicus	3	778	448–1279	113	72–150
Hordeum vulgare	2	3141	2022–4259	70	30–110
Medicago sativa	1	2576	na	160	na
Raphanus sativus	1	6907	na	22	na
Capsicum annuum	1	1600	na	100	na
Solanum tuberosum × *Solanum spegazzinii*	1	1046	na	200	na
Cucumis melo	1	662	na	75	na

Note: This table represents a summary of the analyses of over 130 studies, 93 of which reported both the size and the nature of the mapping population used and the smallest physical interval containing the gene of interest as defined by recombination breakpoints The details of the analyses will be published elsewhere. Nine studies, which used phenotypic markers to preselect informative progeny, a population of recombinant inbred lines, or a mapping population that was smaller than 150 F_2 plants, were excluded from the calculation of means. na, not applicable.

practice, all progeny may need to be phenotyped. For example, if a species is short-lived (e.g., *Caenorhabditis elegans*), the trait persists only in the juvenile period (e.g., embryo), tissue sampling is destructive, or genotyping is slow, then 10^2–10^4 progeny may need to be phenotyped prior to being genotyped. Large-scale phenotyping is labour intensive when the trait to be scored is biochemical, physiological, behavioural, or quantitative (e.g., Arondel et al. 1992; Hirel et al. 2001). As a result, in many species, the number of progeny required to be genotyped on a large-scale and the need to associate phenotype with genotype have become the rate-limiting steps in exploring induced or natural genetic variation, especially as molecular markers become increasingly nonlimiting in the postgenome era.

Transgene marker assisted selection using one flanking transgene (TMARS1)

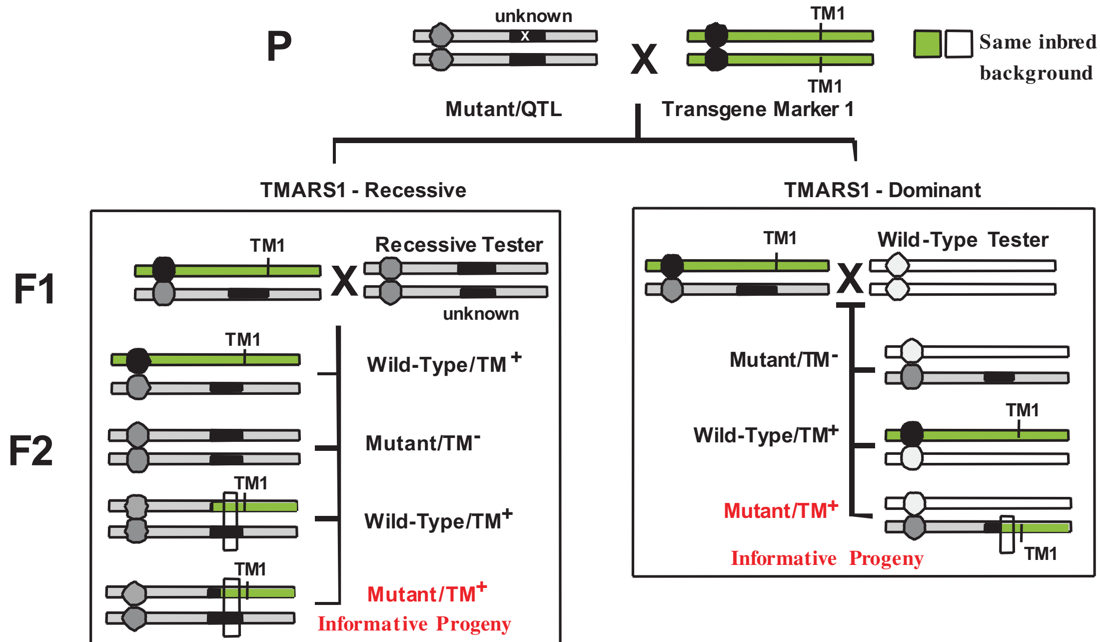
Prior to the molecular biology revolution, genes were mapped by employing flanking phenotypic markers to detect meiotic recombination visually (Sturtevant 1913). Even recently, positional cloning studies in *Arabidopsis* have reported the use of conventional phenotypic markers (e.g., *tt2*, light-coloured seed) to preselect progeny harbouring recombination events in the region of interest (Ziegelhoffer et al. 2000). What is noteworthy is that by preselecting meiotic recombinants prior to molecular analysis, the number of progeny genotyped in three of these types of experiments was only 34, 59, and 268, respectively (Sherson et al. 1999; Ziegelhoffer et al. 2000; Lukowitz et al. 2004), an average of 120 progeny. This represents an almost 10-fold reduction in the number of progeny genotyped compared with the studies shown in Table 1 and a similar reduction in the size of the average target interval containing the gene (from 86 to 52 kb). However, phenotypic markers are rare and genotype dependent, and the research community has thus largely used DNA markers for the past two decades (Botstein et al. 1980; Paterson et al. 1988). During this time, however, efficient gene transformation protocols have been established, and transgene reporters that encode chemical resistance, pigmentation, or fluorescence have been developed (e.g., green fluorescent protein, GFP) (Shaner et al. 2004). We propose a much wider use of mapped transgenes that encode dominant, visible, genotype-independent reporters such as GFP to permit phenotype-based preselection for rare interval recombinants to dramatically reduce progeny genotyping and phenotyping needs during positional cloning.

In Fig. 1, we illustrate two genetic strategies that use mapped transgenes to isolate linked alleles encoding recessive or dominant traits. For a recessive trait, the first step is to map the trait to within a 5 to 10 cM interval. The second step is to cross this allele to a transgene mapping line, which should carry the wild-type or dominant allele and a mapped transgene marker located at either the distal or proximal edge of the 5 to 10 cM interval. The interval region of the transgene mapping line should contain sufficient nucleotide polymorphisms, with respect to the parent encoding the allele of interest, to employ biallelic molecular markers to distinguish linkage blocks from the two parents following meiosis. The third step is to testcross the F_1 hybrid to the recessive parent. The fourth step is to select F_2 progeny that express the transgene reporter (approx. 50% of progeny) and then score for expression of the recessive trait within this population. By design, these rare progeny will contain the desired recombination events in the interval between the transgene and the candidate locus. The fifth step is to genotype only these rare recombinant progeny using biallelic molecular markers. The region of the interval that is heterozygous for molecular markers from both parents, but has inherited both copies of the recessive allele, can be excluded from possessing the gene of interest. In the last step of this strategy, a second mapped transgene is employed to identify recombination events on the opposite side of the locus. Therefore, in this strategy, cosegregation of the recessive trait and the transgene reporter is used to identify interval recombinants instead of random genotyping, similar to classic interval mapping using visible genetic markers.

For a dominant trait (Fig. 1), the strategy to detect meiotic recombination is similar, except that after the parent containing the trait of interest is crossed to the transgene map-

© 2006 NRC Canada

Fig. 1. TMARS1 methodologies to reduce progeny genotyping requirements during map-based cloning by using a single, linked transgene marker to detect recombination events flanking a candidate map region of interest. Strategies are shown to detect recombinants flanking an allele encoding a recessive trait and a dominant trait. A possible disadvantage of TMARS1 is that 50% of progeny at the F_2 generation, those that are positive for the transgene marker, require trait phenotyping. TM, transgene marker.

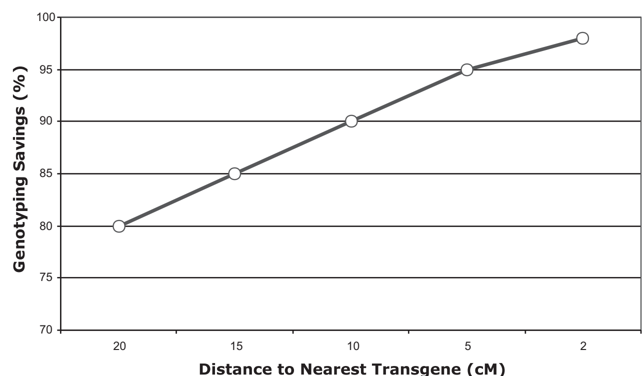


ping line, the F_1 progeny are crossed to a wild-type tester (either parental genotype) to generate F_2 progeny. Desired recombination events in the interval between the transgene and the candidate locus will be detected by co-inheritance of the transgene reporter and dominant trait phenotype.

For simplicity, we refer to the strategies shown in Fig. 1 as transgene marker assisted recombinant selection using one transgene (TMARS1). We refer to the parent containing a mapped transgene as the TMARS line.

Figure 2 illustrates the reduction in F_2 progeny genotyping requirements when employing linked transgenes as genetic markers to detect recombination flanking a target locus compared with current strategies in which recombinants are detected by random genotyping. The key advantage of TMARS1 is that there is a reduction in genotyping requirements by 95% using a 5-cM-linked transgene, which is reduced by 99% with a 1-cM-linked transgene (for an F_2 testcross). TMARS does not reduce the number of F_2 progeny that must be generated, only the number that must be genotyped. TMARS1 is most feasible when an individual parent is able to produce $>10^3$ progeny and is therefore most amenable to organisms such as plants, fungi, insects, fish, and invertebrates, which form the majority of eukaryotes. TMARS1 is also most amenable to species that are relatively easy to outcross. The principal disadvantage of TMARS1 is that 50% of F_2 progeny, those that express the transgene, must be phenotyped for the trait of interest. Therefore, TMARS1 is most useful when the trait of interest is simple to assay and more feasible to score than large-scale genotyping. Although we have outlined a diploid scheme in Fig. 1, TMARS1 is perfectly suited for use in haploid organisms (e.g., fungi) or polyploid species (e.g., wheat) because transgene reporters such as GFP are dominant.

Fig. 2. Savings in the number of F_2 testcross progeny that must be genotyped to detect meiotic recombination flanking a target recessive allele by employing mapped, transgene reporters and the TMARS1 genetic strategy. The graph demonstrates that as the distance between the target locus being fine-mapped and the transgene is reduced, there is a linear decrease in the number of progeny that must be genotyped to detect recombination. The graph was derived from the simple equation percent genotype savings = frequency $(1 - R) \times 100\%$, where R is the frequency of recombinants, which equals the recombination frequency between the allele and the transgene and corresponds to the homozygous recessive allele/transgene class of TMARS1 F_2 testcross progeny. It was assumed that double crossovers would not significantly contribute to recombination (the Kosambi map function) as demonstrated by Perkins (1962) for intervals <20 cM. When calculating the map distance between the transgene and the recessive allele of interest, because only half of the recombinants will be scored in TMARS1 (those containing the transgene), the apparent recombination frequency within this class must be doubled.



© 2006 NRC Canada

Transgene marker assisted selection using two flanking transgenes (TMARS2)

For traits that are difficult to assay, we have proposed a two-transgene marker strategy as shown in Fig. 3 that preselects interval recombinants without the need to genotype or phenotype large numbers of progeny. The first step is to map a trait to within a 5 to 10 cM interval. The second step is cross this allele to a TMARS line containing a transgene marker (TM1). The third step is to cross the F_1 hybrid to a second TMARS line containing a different transgene reporter (TM2) phenotypically distinguishable from TM1 (e.g., red fluorescent protein (RFP)). The two transgene markers should flank the candidate locus interval. The fourth step is to employ molecular marker genotyping to find a recombination event between TM1 and the locus of interest in the F_2 generation near the edge of the genetic interval of the unknown locus. If TM1 is located 5 cM from the locus, then only approx. 20 progeny would need to be genotyped to find such a recombinant. Alternatively, the progeny may be selfed or test-crossed and phenotyped to confirm preservation of the trait of the interest and TM1. In the next step, the chosen F_2 progeny would then be crossed to the appropriate (nontransgenic) testers, and meiotic recombination would be scored phenotypically between TM1 and TM2 in the F_3 generation by screening for cosegregation of the two transgene phenotypes following meiosis. Genotyping and phenotyping of these rare F_3 recombinants would then be used to detect persistence of linkage between a molecular marker from the parent containing the allele of interest and the trait of interest. TMARS2 preselects the rare interval recombinant progeny, avoiding genotyping and phenotyping of large numbers of progeny. TMARS2 should identify recombination events located at both the proximal and the distal sides of the locus of interest to narrow the position of the gene of interest.

TMARS2 is most useful for species that are capable of producing large numbers of embryos from a single parent (male or female) in the F_2 generation. Alternatively, the F_2 genotype may be preserved and then used to generate F_3 progeny sequentially, although care must be taken to ensure that the desired genotype persists. Both TM1 and TM2 must be in a genotype distinguishable from the trait-containing parent; TM1 and TM2 may be in the same or different genetic backgrounds. It is important to note that the interval distance between transgenes does not affect the number of progeny that must be initially tested to search for recombinants but rather, the number of subsequent recombinants that must be genotyped. However, a key advantage of TMARS2 is that recombinants can be detected rapidly, for example, by screening for coexpression of GFP and a second fluorescent protein at a juvenile stage of development (e.g., seeds, seedlings). This should permit inexpensive high-throughput screening with considerable savings in growth space because 95% of progeny can be discarded. In the case of quantitative trait locus (QTL) fine-mapping, only the rare recombinants would need to be propagated to screen for the extent of phenotypic variation in the next generation. Because of the many challenges involved in QTL mapping, only time will tell whether or not TMARS2 will be an effective approach in isolating alleles underlying quantitative traits. To conclude, TMARS2 may be used to accelerate the positional cloning of dominant, recessive, and possibly QTLs. TMARS1 and TMARS2 have the potential to make map-based cloning more feasible in many species, regardless of genome size, gene density, local frequencies of recombination, or effort required for phenotypic trait analysis. In genomes where the fear is a high kilobase to centiMorgan ratio (i.e., large genomes), TMARS2 may permit near-saturation of breakpoints in the target interval, allowing a researcher to assign a target allele to a bacterial artificial chromosome clone.

Transgene marker assisted mapping

Mapped transgenes may also be used in the initial stages of mapping to assign a mutation to a chromosome arm. A suite of arm-specific transgenes may be crossed to an unmapped mutant and then selfed or test-crossed. Linkage between a transgene marker and the locus of interest would be assayed as F_2 progeny that fail to independently segregate the visible transgene reporter from the mutant phenotype. We refer to this mapping strategy as transgene marker assisted mapping (TMAM).

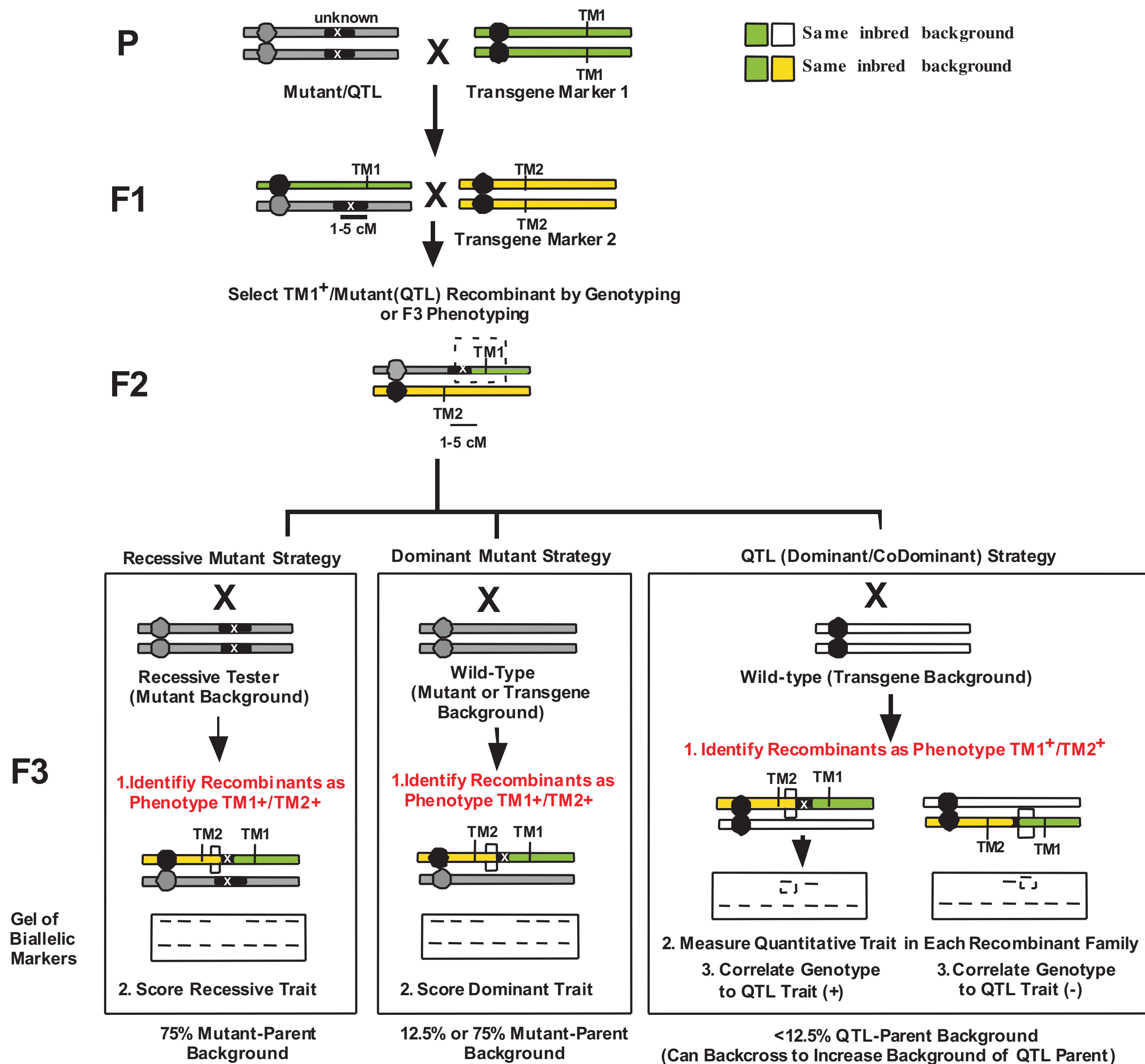
Feasibility and limitations

The current mapping resolution in published *Arabidopsis* positional cloning studies is 16 genes (see the section above: A review of current positional cloning efforts in plants). This is an acceptable number because complementation is straightforward and transformation is genotype independent in *Arabidopsis*. Unfortunately, this is not the case in many plant species. To achieve single-gene resolution in many crop species, we calculated that >6000 progeny must be genotyped in diverse angiosperms, an unacceptable number. TMARS2, in particular, makes high-resolution mapping more feasible across species by dramatically reducing the genotyping requirements. TMARS2 also has the potential to dramatically reduce phenotyping requirements that can limit fine-mapping.

Before a researcher chooses to employ TMARS, it is critical to note the drawbacks of the strategy. First, TMARS does not reduce the costs or markers required to initially map a gene to 5–10 cM (unless using TMAM, see above), nor does it reduce the need to employ high-density interval markers to fine-map a gene once the recombinants have been preselected. It simply reduces the costs in one step of the process and accordingly makes fine/single-gene resolution mapping feasible in organisms in which complementation would otherwise be costly and time consuming. In terms of time, TMARS requires two or three additional generations after a researcher has already created an F_2 population. The extra time investment required for TMARS may not be worthwhile if candidate gene resolution is not critical; this situation occurs when complementation is not time consuming (e.g., *Arabidopsis*) or there are multiple mutant alleles. TMARS may also not be the method of choice when the trait of interest is simple to phenotype (e.g., developmental), there is no reason to suspect a low local frequency of recombination, and reagent cost or genotyping labour is not a critical issue compared with time. However, in species where complementation is challenging and geno-

© 2006 NRC Canada

Fig. 3. TMARS2 methodologies. Strategies to dramatically reduce the need to genotype and phenotype progeny during map-based cloning are shown. Two linked transgene markers (e.g., GFP and RFP) are used to detect recombination events flanking a candidate map region of interest visually. TMARS2 requires that each of the transgene markers (TM1 and TM2) encodes a distinguishable visible trait. The asterisk indicates the step in which interval recombinants are identified. TM, transgene marker; NIL, near-isogenic line; QTL, quantitative trait locus.



type dependent and where there is a danger that multiple candidate genes cannot be distinguished (e.g., lack of multiple mutant alleles), then the two or three generations required for TMARS may be equal to or faster than the time required for transformation and transgene backcrossing with much less effort required and more confidence in the outcome. An additional disadvantage of TMARS is that the method requires outcrossing in contrast with current methods that accept selfing. In organisms such as maize, outcrossing is straightforward, but in certain species, outcrossing can be challenging. In both TMARS1 and TMARS2, a sufficient number of crosses would need to be performed to generate 10^3–10^4 outcross progeny; in *Arabidopsis*, about 50 crosses should be sufficient to generate such a population. It is important to note that an advantage of an outcross (e.g., a testcross) is that recessive alleles are not obscured, allowing all potentially informative breakpoints to be associated with a phenotype. In situations where phenotyping is more challenging than genotyping and would not otherwise be necessary (e.g., in the case of an embryonic phenotype that must be scored prior to genotyping), the large-scale phenotyping requirements of TMARS1 make this approach less feasible than the current genotyping method; in this circumstance, however, TMARS2 is ideal because it dramatically reduces both genotyping and phenotyping requirements and requires only one additional generation. Finally, we will not know unless there is a success whether or not TMARS2 will accelerate the isolation of al-

© 2006 NRC Canada

leles underlying QTLs; tissue culture mutagenicity and the importance of the inbred background genotype are two concerns.

TMARS/TMAM may already be feasible in certain species today, as up to 10^5 T-DNA and transposon insertion resources have been generated in the last 10 years in model plant species and crop plants. A sample of these resources is summarized in Table 2. The most comprehensive insertion resources are available in *Arabidopsis*, rice, tomato, poplar, *Lotus japonicus*, and the moss *Physcomitrella patens*. Note that TMARS/TMAM does not require the use of fluorescent reporters but any means of visible selection including herbicide resistance and antibiotic resistance that are available now. For example, Chen et al. (1998) used kanamycin resistance in distal and proximal mapped *Ds* transposons in a TMARS1-like strategy to positionally clone the *AGRAVITROPIC1* (*AG1*) gene in *Arabidopsis*. A key advantage of fluorescent reporters, however, is that these reporters are genotype independent, and theoretically, they can be detected at the seed stage (Stuitje et al. 2003). In fact, during the revision of this manuscript, a mapped population of multiple fluorescent reporters, which permits the detection of meiotic recombination in *Arabidopsis* seeds, was reported (Melamed-Bessudo et al. 2005). This is a significant step forward in making TMARS1 and TMARS2 more feasible, at least in *Arabidopsis*, by scoring for interval recombination simply by scoring fluorescence at the seed stage.

In species where mapped transgene resources do not exist, how many transgenes will be required and is this a feasible number? We calculate the number of transgenes required assuming random integration using the equation $N = \log(1 - P)/\log[(X - T)/X]$, where N is the number of independent transgenic events required, P is the probability of successfully achieving the transgene interval, X is the total genome recombination in centiMorgans, and T is the target transgene interval in centiMorgans. To achieve a 95% probability that no target locus is more than 10 cM from a transgene (90% genotyping savings), 226, 200, and 460 stable mapped transgenes would be required in the cereals *Oryza sativa*, *Zea mays*, and *Triticum aestivum*, respectively. These are achievable numbers. The transgene number calculation depends on the transgene density required, but this correlates with the total map distance of the genome, not the physical genome size. Owing to transgene silencing and multiple integration events, the initial transgene number must exceed the numbers given. As *Agrobacterium* provides simpler integration events than particle bombardment, this must be the transformation method of choice. Initial transformants must be analyzed for at least two or three generations to check for transgene stability and expected segregation. Thermal asymmetric interlaced polymerase reaction might provide the most efficient means to map the insertions. Developing a population of 200–400 mapped insertions may require the cooperation of multiple laboratories to benefit the entire community, and more cooperation between laboratories in the developed and developing world in the case of “orphan crops” that are not widely studied in wealthier nations. Will the cost and effort required to generate TMARS resources not outweigh the benefits of brute-force genotyping? If only a few alleles will be a target for positional cloning within a species, then TMARS is not a cost-effective strategy. However, if many alleles may be of interest globally, then it is important to note that mapped transgenes in any inbred have the potential to help fine-map alleles in hundreds or thousands of other accessions of that species, outweighing the initial infrastructure costs.

As noted above, transgenes suffer from gene silencing. The TMARS strategies were designed, however, to prevent false positives caused by reporter silencing. Transgene integration events must be confirmed to be at a single locus and then tested to ensure that there is no associated mutant phenotype and that transgene expression is stable across multiple generations. Because of dominant–recessive relationships of different accessions or hybrid background effects on specific traits, it may be useful to construct TMARS/TMAM in multiple inbreds. To move forward, libraries of randomly spaced, stably expressing, single-locus mapped transgenes encoding dominant, genotype-independent visible markers are now needed. A variety of fluorescent reporters have recently become available (Shaner et al. 2004) that may be more useful in plants than GFP, the detection of which suffers from chlorophyll autofluorescence. Many fluorescent reporters, however, should be detectable in nonfluorescent tissue such as roots and seeds, as already noted above. It must be noted, however, that even a *single* mapped transgene is useful to a subset of researchers fortunate enough to have a linked allele. The Raizada Laboratory is in the process of constructing a public TMARS database (online in 2006) where plant researchers can share stably expressing mapped transgene reporter resources in different plant species as these transgenes become available worldwide.

TMARS2 may expedite the exploration of genetic variation in plants and plant pathogens

In addition to accelerating the positional cloning of simple, mutant alleles, TMARS, and TMAM should facilitate the molecular exploration of natural and human-selected genetic variation in wild and domesticated plants, accelerating the discovery of valuable alleles for agriculture and medicine. Because transgene markers such as GFP are genotype independent, TMARS resources constructed in only one genotype have the potential to detect recombination between any transgenic line and all other accessions. TMARS may prove valuable in identifying complex crop pathogenesis-causing genes and pesticide-resistance alleles in evolving fungal, nematode, insect, and weed parasites that affect agriculture, including *Fusarium* and *Striga*. For example, there are >100 000 fungal species (Hawksworth 1991) awaiting molecular characterization, many with large numbers of accessions, that are either beneficial or detrimental to plants. Hundreds of thousands of crop accessions have been deposited into seedbanks, and using these accessions, plant breeders have mapped significant numbers of QTLs of agronomic importance, representing >10 000 years of breeding selection (Tanksley and McCouch 1997). The loci underlying these QTLs await discovery, and only empirical evidence can demonstrate whether or not transgene markers introduced into these crops will accelerate this discovery.

© 2006 NRC Canada

Table 2. Summary of available T-DNA and transposon insertion resources in plants.

Species	Ecotype/cultivar	Selectable marker/ reporter[a]	No. of insertions[b]	Reference(s) for resource	Web site
Arabidopsis thaliana	Col-0	KAN	145 589	Alonso et al. 2003	http://signal.salk.edu/tabout.html
Arabidopsis thaliana	Ws	GUS, BAR, KAN	31 560	Samson et al. 2002	http://193.51.165.9/projects/FLAGdb++/HTML/data.shtml
Arabidopsis thaliana	Col-0	SUL	60 651	Rosso et al. 2003	http://www.mpiz-koeln.mpg.de/GABI-Kat/
Arabidopsis thaliana	Col-0	BAR	48 434	Sessions et al. 2002	http://www.tmri.org/en/partnership/sail_collection.aspx
Arabidopsis thaliana	Nossen	GUS, HYG	18 551	Kuromori et al. 2004	http://rarge.gsc.riken.go.jp/dsmutant/index.pl
Arabidopsis thaliana	Col-0	GUS, KAN	23 411		http://www.jic.bbsrc.ac.uk/science/cdb/exotic/index.htm
Arabidopsis thaliana	Ler	GUS, KAN	5 169		http://genetrap.cshl.org/
Arabidopsis thaliana	Col-0	BAR, HYG	10 459		http://www.hort.wisc.edu/krysan/DS-lox/
Arabidopsis thaliana	Ler	GUS, KAN	878	Parinov et al. 1999	http://nasc.nott.ac.uk/ima.html
Arabidopsis thaliana	Col-0	HYG	975	Mathur et al. 1998	http://www.szbk.u-szeged.hu/~arabidop/T-DNA.html#mapped
Oryza sativa	Nipponbare	HYG, GUS, GFP	7 480 (40 000)	Sallaud et al. 2004	http://genoplante-info.infobiogen.fr/OryzaTagLine/
Oryza sativa	Dongjin or Hwayoung	GUS; GUS, GFP	24 299 (75 000)	Jeon et al. 2000; Jeong et al. 2002	http://www.postech.ac.kr/life/pfg
Oryza sativa	Nipponbare or Zhonghua No. 11	Not reported	1 009	Chen et al. 2003	http://www.genomics.zju.edu.cn/ricetdna
Lycopersicon esculentum	Moneymaker	KAN	140 (265)	Gidoni et al. 2003	
Lycopersicon esculentum	Mini-Tom	KAN	(10 427)	Mathews et al. 2003	
Populus tremula × *Populus alba*		BAR or KAN	(627)	Busov et al. 2003	
Populus tremula × *Populus alba*		KAN, GUS	38 (1 344)	Groover et al. 2004	
Lotus japonicus	GIFU B-129-S9	KAN	(1 112)	Schauser et al. 1998	
Physcomitrella patens		GUS, KAN	(5 264)	Nishiyama et al. 2000	

[a]Plant selectable marker genes/reporter genes; KAN, kanamycin resistant; GUS, beta glucuronidase reporter gene; BAR, phosphinothricin resistant; SUL, sulfadiazine resistant; HYG, hygromycin resistant; GFP, green flourescent protein reporter gene.

[b]Number of mapped insertions as of 1 May 2005 as reported on the SIGnAL "T-DNA Express" Arabidopsis Gene Mapping Tool; available from http://signal.salk.edu/cgi-bin/tdnaexpress. Numbers in parentheses indicate unmapped insertions.

© 2006 NRC Canada

Conclusion

The vast majority of human-selected and natural genetic variation in plants awaits characterization at the molecular level. Today, the majority of the world's laboratories cannot afford to conduct routine fine-scale molecular mapping because of the large number of postmeiotic progeny that must be genotyped and phenotyped to uncover rare meiotic recombination events that break linkage between molecular markers and the trait of interest. Here, we have proposed that the plant research community should consider returning back to the era of employing mapped, visual markers to detect recombination. Today, mapped transgenes that encode visible traits such as fluorescence or herbicide resistance may be used as genetic markers. Unlike traditional genetic markers, transgene reporters may be randomly inserted in large numbers throughout the genome, and they encode proteins that are genotype independent. As a result, transgenes mapped in only one inbred may be used to detect recombinants in large numbers of accessions. The transformation requirements of TMARS/TMAM can be met by many poor countries for their crops or pathogen species of interest because cell culture and transformation are often used in training programs; however, TMARS would still require a laboratory to use molecular markers and only reduces part of the reagent cost required. Nevertheless, TMARS/TMAM may make fine-scale interval mapping for breeding and positional cloning economically more feasible for many researchers and in many species.

Acknowledgments

We acknowledge grants from the Natural Sciences and Engineering Research Council of Canada, the Ontario Ministry of Agriculture and Food, and the University of Guelph Food System Biotechnology Centre and from an Ontario Premier's Research Excellence Award to M.N.R. We thank Steve Chatfield for his critical reading of the manuscript and two thoughtful anonymous reviewers.

References

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana.* Science (Washington, D.C.), **301**: 653–657.

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H.M., and Somerville, C.R. 1992. Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis.* Science (Washington, D.C.), **258**: 1353–1355.

Barakat, A., Carels, N., and Bernardi, G. 1997. The distribution of genes in the genomes of Graminae. Proc. Natl. Acad. Sci. USA, **94**: 6857–6861. doi:10.1073/pnas.94.13.6857.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32**: 314–331.

Busov, V.B., Meilan, R., Pearce, D.W., Ma, C., Rood, S.B., and Strauss, S.H. 2003. Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature. Plant Physiol. **132**: 1283–1291. doi:10.1104/pp.103.020354.

Chen, R., Hilson, P., Sedbrook, J., Rosen, E., Caspar, T., and Masson, P.H. 1998. The *Arabidopsis thaliana* AGRAVITROPIC 1 gene encodes a component of the polar-auxin-transport efflux carrier. Proc. Natl. Acad. Sci. USA, **95**: 15112–15117. doi:10.1073/pnas.95.25.15112.

Chen, S., Jin, W., Wang, M., Zhang, F., Zhou, J., Jia, Q., Wu, Y., Liu, F., and Wu, P. 2003. Distribution and characterization of over 1000 T-DNA tags in the rice genome. Plant J. **36**: 105–113. doi:10.1046/j.1365-313X.2003.01860.x.

Durrett, R.T., Chen, K.Y., and Tanksley, S.D. 2002. A simple formula useful for positional cloning. Genetics, **160**: 353–355.

Gidoni, D., Fuss, E., Burbidge, A., Speckmann, G.J., James, S., Nijkamp, D., Mett, A., Feiler, J., Smoker, M., de Vroomen, M.J., Leader, D., Liharska, T., Groenendijk, J., Coppoolse, E., Smit, J.J., Levin, I., de Both, M., Schuch, W., Jones, J.D., Taylor, I.B., Theres, K., and van Haaren, M.J. 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. doi:10.1023/A:1020718520618.

Groover, A., Fontana, J.R., Dupper, G., Ma, C., Martienssen, R., Strauss, S., and Meilan, R. 2004. Gene and enhancer trap tagging of vascular-expressed genes in poplar trees. Plant Physiol. **134**: 1742–1751. doi:10.1104/pp.103.034330.

Hawksworth, D.L. 1991. The fungal dimension of biodiversity: magnitude, significance and conservation. Mycol. Res. **95**: 641–655.

Hirel, B., Bertin, P., Quillere, I., Bourdoncle, W., Attagnant, C., Dellay, C., Gouy, A., Cadiou, S., Retailliau, C., Falque, M., and Gallais, A. 2001. Towards a better understanding of the genetic and physiological basis for nitrogen use efficiency in maize. Plant Physiol. **125**: 1258–1270. doi:10.1104/pp.125.3.1258.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M., and Last, R.L. 2002. *Arabidopsis* map-based cloning in the post-genome era. Plant Physiol. **129**: 440–450. doi:10.1104/pp.003533.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., Kim, C., Jang, S., Yang, K., Nam, J., An, K., Han, M.J., Sung, R.J., Choi, H.S., Yu, J.H., Choi, J.H., Cho, S.Y., Cha, S.S., Kim, S.I., and An, G. 2000. T-DNA insertional mutagenesis for functional genomics in rice. Plant J. **22**: 561–570. doi:10.1046/j.1365-313x.2000.00767.x.

Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K., and An, G. 2002. T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. **130**: 1636–1644. doi:10.1104/pp.014357.

Kotera, E., Tasaka, M., and Shikanai, T. 2005. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature, **433**: 326–330. doi:10.1038/nature03229.

Kuromori, T., Hirayama, T., Kiyosue, Y., Takabe, H., Mizukado, S., Sakurai, T., Akiyama, K., Kamiya, A., Ito, T., and Shinozaki, K. 2004. A collection of 11800 single-copy Ds transposon insertion lines in *Arabidopsis.* Plant J. **37**: 897–905. doi:10.1111/j.1365.313X.2004.02009.x.

Lukowitz, W., Gillmor, C.S., and Scheible, W.R. 2000. Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. Plant Physiol. **123**: 795–805. doi:10.1104/pp.123.3.795.

Lukowitz, W., Roeder, A., Parmenter, D., and Somerville, C. 2004. A MAPKK kinase gene regulates extra-embryonic cell fate in *Arabidopsis.* Cell, **116**: 109–119. doi:10.1016/S0092-8674(03)01067-5.

© 2006 NRC Canada

Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., Schuster, D.K., Menasco, D.J., Wagoner, W., Lightner, J., and Wagner, D.R. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell, **15**: 1689–1703. doi:10.1105/tpc.012963.

Mathur, J., Szabados, L., Schaefer, S., Grunenberg, B., Lossow, A., Jonas-Straube, E., Schell, J., Koncz, C., and Koncz-Kalman, Z. 1998. Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspension. Plant J. **13**: 707–716. doi:10.1046/j.1365-313X.1998.00059.x.

Melamed-Bessudo, C., Yehuda, E., Stuitje, A.R., and Levy, A.A. 2005. A new seed-based assay for meiotic recombination in *Arabidopsis thaliana.* Plant J. **43**: 458–466. doi:10.1111/j.1365-313X.2005.02466.x.

Nishiyama, T., Hiwatashi, Y., Sakakibara, I., Kato, M., and Hasebe, M. 2000. Tagged mutagenesis and gene-trap in the mass, *Physcomitrella patens* by shuttle mutagenesis. DNA Res. **7**: 9–17. doi:10.1093/dnares/7.1.9.

Parinov, S., Sevugan, M., Ye, D., Yang, W.C., Kumaran, M., and Sundaresan, V. 1999. Analysis of flanking sequences from *Dissociation* insertion lines: a database for reverse genetics in *Arabidopsis.* Plant Cell, **11**: 2263–2270. doi:10.1105/tpc.11.12.2263.

Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E., and Tanksley, S.D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature, **335**: 721–726. doi:10.1038/335721a0.

Perkins, D.D. 1962. Crossing-over and interference in a multiply marked chromosome arm of *Neurospora.* Genetics, **47**: 1253–1274.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. 2003. An *Arabidopsis thaliana* T-DNA mutagenized population (gabi-kat) for flanking sequence tag-based reverse genetics. Plant Mol. Biol. **53**: 247–259. doi:10.1023/B:PLAN.0000009297.37235.4a.

Sallaud, C., Gay, C., Larmande, P., Bes, M., Piffanelli, P., Piegu, B., Droc, G., Regad, F., Bourgeois, E., Meynard, D., Perin, C., Sabau, X., Ghesquiere, A., Glaszmann, J.C., Delseny, M., and Guiderdoni, E. 2004. High throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. Plant J. **39**: 450–464. doi:10.1111/j.1365-313X.2004.02145.x.

Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M., and Lecharny, A. 2002. Flagdb/fst: a database of mapped flanking insertion sites (fsts) of *Arabidopsis thaliana* T-DNA transformants. Nucleic Acids Res. **30**: 94–97. doi:10.1093/nar/30.1.94.

Schauser, L., Handberg, K., Sandal, N., Stiller, J., Thykjaer, T., Pajuelo, E., Nielsen, A., and Stougaard, J. 1998. Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus.* Mol. Gen. Genet. **259**: 414–423.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J.D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitzel, T., Katagiri, F., Glazebrook, J., Law, M., and Goff, S.A. 2002. A high-throughput *Arabidopsis* reverse genetics system. Plant Cell, **14**: 2985–2994. doi:10.1105/tpc.004630.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E., and Tsien, R.Y. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* spp. red fluorescent protein. Nat. Biotechnol. **22**: 1567–1572. doi:10.1038/nbt1037.

Sherson, S., Gy, I., Medd, J., Schmidt, R., Dean, C., Kreis, M., Lecharny, A., and Cobbett, C. 1999. The arabinose kinase, *ARA1*, gene of *Arabidopsis* is a novel member of the galactose kinase gene family. Plant Mol. Biol. **39**: 1003–1012. doi:10.1023/A:1006181908753.

Stuitje, A.R., Verbree, E.C., van der Linden, K.H., Mietkiewska, E.M., Nap, J.-P., and Kneppers, T.J.A. 2003. Seed-expressed fluorescent proteins as versatile tools for easy (co)transformation and high-throughput functional genomics in *Arabidopsis.* Plant Biotechnol. J. **1**: 301–309.

Sturtevant, A.H. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. J. Exp. Zool. **14**: 43–59. doi:10.1002/jez.1400140104.

Tanksley, S.D., and McCouch, S.R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science (Washington, D.C.), **277**: 1063–1066. doi:10.1126/science.277.5329.1063.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., and Dubcovsky, J. 2003. Positional cloning of the wheat vernalization gene *VRN1*. Proc. Natl. Acad. Sci. USA, **100**: 6263–6268. doi:10.1073/pnas.0937399100.

Ziegelhoffer, E.C., Medrano, L.J., and Meyerowitz, E.M. 2000. Cloning of the *Arabidopsis Wiggum* gene identifies a role for farnesylation in meristem development. Proc. Natl. Acad. Sci. USA, **97**: 7633–7638. doi:10.1073/pnas.130189397.

© 2006 NRC Canada