GENOMICS ARTICLE

Function Search in a Large Transcription Factor Gene Family in Arabidopsis: Assessing the Potential of Reverse Genetics to Identify Insertional Mutations in R2R3 *MYB* Genes

Ruth C. Meissner,[a,1] Hailing Jin,[b,1] Eleonora Cominelli,[c] Marten Denekamp,[d] Antonio Fuertes,[e] Raffaella Greco,[c,2] Harald D. Kranz,[f] Steven Penfield,[a] Katia Petroni,[c] Ana Urzainqui,[e] Cathie Martin,[b] Javier Paz-Ares,[e] Sjef Smeekens,[d] Chiara Tonelli,[c] Bernd Weisshaar,[f] Elvira Baumann,[g] Victor Klimyuk,[h] Sylvestre Marillonnet,[h] Kanu Patel,[h] Elly Speulman,[i] Alain F. Tissier,[h,3] David Bouchez,[j] Jonathan J. D. Jones,[h] Andy Pereira,[i] Ellen Wisman,[g] and Michael Bevan[a,4]

[a] John Innes Centre, Department of Molecular Genetics, Colney Lane, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[b] John Innes Centre, Genetics Department, Colney Lane, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[c] Dipartimento di Genetica e di Biologia dei Microrganismi, Universita degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy

[d] Department of Molecular Cell Biology, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[e] Centro Nacional de Biotecnologia–Consejo Superior de Investigaciones, Campus de Cantoblanco, 28049 Madrid, Spain

[f] Max-Planck Institut für Züchtungsforschung, Abteilung Biochemie, Carl-von-Linne-Weg 10, 50829 Köln, Germany

[g] Max-Planck Institut für Züchtungsforschung, Abteilung Saedler, Carl-von-Linne-Weg 10, 50829 Köln, Germany

[h] John Innes Centre, Sainsbury Laboratory, Colney Lane, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[i] CPRO-DLO, Department of Molecular Biology, Droevendaalsesteeg 1, 6700 AA Wageningen, The Netherlands

[j] Institut National de la Recherche Agronomique, Centre de Versailles, Laboratoire de Biologie Cellulaire, Route de Saint-Cyr, 78026 Versailles Cedex, France

More than 92 genes encoding MYB transcription factors of the R2R3 class have been described in Arabidopsis. The functions of a few members of this large gene family have been described, indicating important roles for R2R3 MYB transcription factors in the regulation of secondary metabolism, cell shape, and disease resistance, and in responses to growth regulators and stresses. For the majority of the genes in this family, however, little functional information is available. As the first step to characterizing these genes functionally, the sequences of >90 family members, and the map positions and expression profiles of >60 members, have been determined previously. An important second step in the functional analysis of the *MYB* family, through a process of reverse genetics that entails the isolation of insertion mutants, is described here. For this purpose, a variety of gene disruption resources has been used, including T-DNA–insertion populations and three distinct populations that harbor transposon insertions. We report the isolation of 47 insertions into 36 distinct *MYB* genes by screening a total of 73 genes. These defined insertion lines will provide the foundation for subsequent detailed functional analyses for the assignment of specific functions to individual members of the R2R3 *MYB* gene family.

INTRODUCTION

The assignment of biological function to the large number of genes that have now been sequenced, with new se-

quence data being compiled rapidly, is currently one of the most challenging goals in biology. Genetic analysis, particularly the effects of loss-of-function mutations, is of central importance to achieving this goal. In plants, unlike yeast, targeted gene disruption is laborious and inefficient (Kempin et al., 1997). Gene silencing by antisense or sense suppression is also a common approach to studying plant gene function (Kooter and Mol, 1993; Baulcombe, 1996), but the specificity and extent of gene disruption through such methods have not been extensively tested,

so the analysis and interpretation of suppression sometimes have proved difficult (van der Krol et al., 1990; Höfgen et al., 1994).

T-DNA and transposable elements can alter gene function upon insertion into coding or regulatory sequences (Feldmann, 1991; Martienssen, 1998). Many lines of Arabidopsis harboring either T-DNA or transposon insertions have been generated (Koncz et al., 1992; Azpiroz-Leehan and Feldmann, 1997; Bouchez and Höfte, 1998; Martienssen, 1998; Wisman et al., 1998a, 1998b). Individual lines carrying insertions within a gene of interest can be identified by polymerase chain reaction (PCR) by using a gene-specific primer in combination with a primer complementary to border sequences of the insertion element. This method was first described for Drosophila (Ballinger and Benzer, 1989; Kaiser and Goodwin, 1990) but subsequently has been applied to several other organisms, including *Caenorhabditis elegans*, petunia, and maize (Zwaal et al., 1993; Das and Martienssen, 1995; Koes et al., 1995). In Arabidopsis, this process of "reverse" genetics (McKinney et al., 1995) was used to screen for T-DNA insertions in actin gene sequences. Subsequently, an increasing number of genes, such as those involved in signal transduction (Krysan et al., 1996) and ion transport (Gaymard et al., 1998; Hirsch et al., 1998), has been similarly characterized using T-DNA–mutagenized populations. Recently, the use of populations containing transposons for reverse genetics also has been reported. Examples include insertion mutant lines carrying the *Enhancer/Suppressor-mutator* (*En/Spm*) transposon in genes involved in flavonoid biosynthesis and gravitropism (Müller et al., 1998; Wisman et al., 1998b). Reverse genetics is particularly attractive for assigning functions to individual members of gene families. Winkler et al. (1998) demonstrated the systematic isolation of Arabidopsis lines containing T-DNA insertional mutations in various members of the P450 gene family and were able to ascribe functions to one member in particular.

MYB transcription factors contain a common DNA binding domain that consists of one to three imperfect helix-turn-helix repeats that are denoted R1, R2, and R3. The first *MYB* gene identified was *v-MYB*, the oncogenic component of avian myoblastoma virus, that has a cellular protooncogenic counterpart in animals designated *c-MYB* (Lüscher and Eisenman, 1990). Subsequently identified members of the *MYB* gene family represent all major eukaryotic groups (Rosinski and Atchley, 1998). In animals and yeast, the number of identified *MYB* genes is small (Thompson and Ramsay, 1995; Rosinski and Atchley, 1998), but a much larger number of *MYB* genes has been identified in plants (Martin and Paz-Ares, 1997). Although some of these plant *MYB* genes contain only a single repeat (Baranowskij et al., 1994; Kirik and Bäumlein, 1996; Feldbrügge et al., 1997), the largest group is the R2R3 class, containing two imperfect MYB-like repeats in their DNA binding domains (Jackson et al., 1991; Avila et al., 1993; Lin et al., 1996; Romero et al., 1998). To date, >92

R2R3-type *MYB* genes have been characterized in Arabidopsis (Kranz et al., 1998).

Whereas many animal *MYB* genes have been found to function in the control of cell proliferation and differentiation (Thompson and Ramsay, 1995), the functions of *MYB* genes in plants appear to be far more diverse and, with the exception of a few intensely studied examples, remain much less clear. Some plant R2R3 *MYB* genes are known to regulate secondary metabolism, particularly in the phenylpropanoid pathway (Paz-Ares, 1987; Grotewold et al., 1994; Sablowski et al., 1994; Moyano et al., 1996; Tamagnone et al., 1998) and tryptophan biosynthesis (Bender and Fink, 1998). Others contribute to processes of cellular morphogenesis (Oppenheimer et al., 1991; Noda et al., 1994; Waites et al., 1998), signal transduction in plant growth (Gubler et al., 1995; Iturriaga et al., 1996), abiotic stress (Urao et al., 1993; Magaraggia, 1997; Hoeren et al., 1998), and pathogen defense (Yang and Klessig, 1996). Nevertheless, very little is known about the functions of the majority of R2R3 MYB transcription factors in Arabidopsis.

To characterize the large R2R3 *MYB* family in Arabidopsis, we have described previously the sequences of >90 genes and determined the map positions and expression patterns of >60 members (Kranz et al., 1998). Here, we describe our strategies and results in isolating insertional mutations in a considerable number of these genes, an essential step toward investigating their functions in a variety of cellular processes. Specifically, distinct populations containing T-DNA or transposon insertions (Table 1) were used to screen for mutations within sequenced *MYB* genes, and the subsequent comparison of these populations has allowed us to evaluate their potential utility for more general reverse genetic analyses in Arabidopsis.

The populations used in our experiments (Table 1) include the Versailles and CSIC (for C̲onsejo S̲uperior de I̲nvestigaciones C̲ientíficas) T-DNA lines, carrying an average of 1.5 insertions per line, which were constructed upon vacuum infiltration of the *Agrobacterium tumefaciens* strain containing the pGKB5 vector (Bouchez et al., 1993). The AMAZE collection (Baumann et al., 1998; Wisman et al., 1998a, 1998b) contains several copies per line of the autonomous transposable element *En/Spm.* The Wageningen lines (see Speulman et al., 1999, in this issue) comprise a two-element system (Aarts et al., 1995) consisting of a stable homozygous *En/Spm*-transposase source along with a nonautonomous *Inhibitor*/defective *Spm* (*I/dSpm*) transposon that is represented at a copy number of 20 to 30. The SLAT lines (for S̲ainsbury L̲aboratory A̲rabidopsis t̲haliana; see Tissier et al., 1999, in this issue) were generated via a two-element system consisting of a transposase *En/Spm* source and contain one to three *I/dSpm* insertions per line. The transposase source was segregated from the mutagenic transposon insertions after transposition (before screening), so the insertions in the SLAT lines are generally stable.

Table 1. Summary of the Insertion Populations

	SLAT Population	Wageningen Population	AMAZE Population	Versailles and CSIC Population
Origin	J. Jones, Norwich, UK	A. Pereira, Wageningen, The Netherlands	E. Wisman, Cologne, Germany	D. Bouchez and CSIC, Versailles, France and Madrid, Spain
Ecotype	Columbia	Landsberg *erecta*	Columbia	Wassilewskija
Lines screened	28,800	2592	3000 + 5000	9264 Versailles + 4000 CSIC
Insertion	*En*/d*Spm* transposon	*En*/d*Spm* transposon	*En*/*Spm* transposon	T-DNA
System	Two-element system	Two-element system	Autonomous transposon	
State	Stable (reactivation possible)	Mobile	Mobile	Stable
Copy number	1 to 3	20 to 30	2 to 20	1 to 3
Pools	Superpools plus matrix of two dimensions	Matrix of three dimensions	Superpools plus matrix of three dimensions	Hyperpools, superpools, and single pools
Strategy	Identifies positive pool of 50 plants	Identifies positive plant	Identifies positive plant	Identifies positive pool of 50 plants

RESULTS

General Strategy

The R2R3 *MYB* genes encode proteins of between 220 and 394 amino acids (Kranz et al., 1998). Most of the genes contain two introns, one of which occurs at a highly conserved site that intervenes within the coding sequence for the R3 *MYB* repeat. Both introns are generally small, resulting in a total gene size of ~1.6 kb, including potential 5′ and 3′ regulatory sequences. Because of their small size, we were able to screen for transposon insertions over the entire gene sequence and promoter regions with just one gene-specific primer. This primer was designed to anneal to the 3′ end of each *MYB* gene, allowing amplification toward the 5′ end of each gene. The consensus N terminus of MYB proteins is homologous to many other gene products, and therefore the 5′ ends of *MYB* genes are not particularly amenable to gene-specific probes. Because T-DNA insertions often have only one intact border (Krysan et al., 1996; Nacry et al., 1998), on the other hand, the screening for T-DNA insertions within a given gene requires both a 3′ and a 5′ end primer to anneal to the insertion. Our PCR analyses generally relied on primers between 21 and 32 nucleotides in length. Those primers characterized by higher T_m values (75 to 82°C) allowed for higher annealing temperatures during the PCR and increased specificity. All primers used in screening were tested on wild-type DNA to assess specificity. The 3′ region of each *MYB* gene under investigation was amplified by PCR for subsequent use as a gene-specific hybridization probe. Gene-specific probes for each *MYB* family member were used in screening to avoid cross-hybridization with other family members. To screen the T-DNA–containing lines with the 5′-end primer, we used a full-length cDNA of the gene as a hybridization probe to identify PCR bands generated by insertions in both ends of the gene.

If a gene carries an insertion, a PCR band should be amplified only from the DNA of pools containing the insertion mutant specified by the combination of a gene-specific and a transposon or T-DNA primer. However, the pools contain DNA from many plants and therefore contain only a very small amount of DNA from each individual plant, making hybridization necessary to detect the specifically amplified bands. Nevertheless, bands hybridizing to gene-specific probes often proved to be irreproducible artifacts rather than insertion junctions. The frequent occurrence of false positives was the most significant problem encountered in screening all lines. Consequently, at least two dimensions of the DNA pools were screened simultaneously, if possible. This strategy distinguished genuine bands from false positives at early stages of the screens, and only bands amplified in two different dimensions were followed up in further analysis.

Semi-nested or nested PCR helped to reduce PCR artifacts and also enabled visualization of the bands on agarose gels stained with ethidium bromide before hybridization. The semi-nested approach used a given gene-specific primer in combination with a nested primer from the insertion and was followed by sequence analysis as detailed in Methods. The three transposon insertion populations and two T-DNA insertion populations (Table 1) proved to contain 47 insertions among 36 different *MYB* genes.

PCR Screening of the AMAZE Lines

The transposase encoded by the autonomous *En* element of the AMAZE lines produces insertions that are both somatically and germinally unstable (Table 1). Germinal excisions, which result in the complete loss of an insertion, are rare,

Table 2. Performance of the Insertion Populations

| | Populations Screened | | | | | |
	SLAT	Wageningen	AMAZE 1st	AMAZE 2nd	T-DNA	Total
Genes screened	33	18	13	37	39	73
Genes with insertions	9	2	5	15	7	36
Total number of insertions	9	4	7	20	7	47
Insertions/genes screened (%)	27	22	54	54	18	64
Insertions in ORF	5	3	3	11	0	22
Genes with altered RNA	5+ not tested	Not tested	2	6+ not tested	2	18+ not tested

whereas excisions that lead to sectors of somatic reversion are frequent. Two AMAZE populations were screened (Table 1). In the first (3000 line) population, 13 *MYB* genes were screened, resulting in the identification of seven insertions among five different genes. These data are summarized in Table 2. Only three of these insertions, two of which occurred in *AtMYB51*, disrupted the open reading frame (ORF). Both insertions in *AtMYB51* resulted in undetectable transcript levels. The effects on RNA levels were not determined for the third insertion. The identified insertions are characterized in Table 3. An insertion in the poly(A)$^+$ signal of the 3' untranslated region of *AtMYB46* reduced transcript abundance.

The second (5000 line) AMAZE population was screened for insertions in 37 *MYB* genes. Twenty insertions, among 15 genes, were isolated (Table 2). Eight of the 15 genes were disrupted by insertions in the ORF (Table 3), and all of those examined showed reduced RNA transcript levels. Although an insertion in the middle of the 5' end of the ORF might be predicted to abolish detectable levels of transcripts, reverse transcription–PCR (RT-PCR) analyses revealed substantially reduced but detectable levels of transcripts in those genes tested. Detectable transcript levels in such instances could be explained by somatic reversions giving rise to small sectors of tissue with wild-type levels of *MYB* gene expression in an otherwise null background. Nevertheless, an insertion 4 bp before the initial ATG codon of *AtMYB03*, and one insertion in an intron of *AtMYB67*, did not affect transcript levels (Table 3). It is possible in these instances that transcripts from very closely related genes may mask reduced transcript levels actually caused by the insertions.

Screening of the SLAT Lines

PCR and hybridization analyses (see Methods) of the 28,800 SLAT lines revealed seven insertions in the 30 *MYB* genes screened (i.e., a mutagenesis rate of 23%; Table 2). Four of these insertions were located in ORFs and caused reduction

of transcript levels to below the detection limit. Another insertion, in the second intron (*MYB45*), resulted in significantly reduced levels of full-length transcripts. Fifteen of the genes that gave a negative result in the PCR screenings were hybridized onto the filters of the 576 pools, but no additional insertions were identified with this strategy. These data are presented in Table 3.

Recently, we have begun to screen a second collection of 864 50-line pools (43,200 lines) by filter hybridization. In this hybridization screen, an insertion in a previously targeted gene was found (*AtMYB16*; Table 3). The insertion targeting of three additional *MYB* genes was detected exclusively on the filters representing the collection of 43,200 lines; one of these insertions was identified in an ORF.

PCR Screening of the Wageningen Lines

A stable, homozygous, T-DNA–derived transposase source is present in the screened lines, and the d*Spm* elements are therefore highly active. The outcrossing of lines bearing an identified insertion along with subsequent selfing of the F$_1$, however, stabilizes the line of interest after selection against the presence of the transposase in the F$_2$. In addition, outcrossing significantly reduces the number of transposon copies per line. Heritable insertions were confirmed by sequencing analysis to circumvent possibilities of germinal excision or the initial identification of insertions confined to large somatic sectors. Given the high transposition frequency of the Wageningen lines, it was necessary to consider both of these possibilities. Ultimately, the screening of 18 *MYB* genes resulted in the isolation of four insertions, three of which occurred in the ORFs of two distinct genes (see Tables 2 and 3).

PCR Screening of the Versailles and CSIC (T-DNA) Lines

Of 39 *MYB* genes examined for T-DNA insertion, seven different genes were identified (Table 2). All of these insertions were found in the 5' promoter regions (Table 3). One

Table 3. Identified Insertions in R2R3-*MYB* Genes from Arabidopsis

| *Myb* Genes | Population | | Position of the Insertions | | | | RNA Level |
	Transposon	T-DNA	Upstream of ATG	In the ORF	Intron	Downstream of the ORF	
atmyb2	Wageningen			Myb domain			
atmyb2	Wageningen				2nd		
atmyb3	AMAZE2		4 bp				Normal
atmyb4	SLAT			Myb domain			Null
atmyb5		Versailles	720 bp				
atmyb6	SLAT			Myb domain			Null
atmyb10	SLAT (filter)			Myb domain			
atmyb11	Wageningen			Myb domain			
atmyb11	Wageningen			3′ ORF			
atmyb12	AMAZE2			3′ ORF			
atmyb15	SLAT			Myb domain			Null
atmyb16	SLAT (filter)					250 bp	
atmyb16	AMAZE1		150 bp				Normal
atmyb17		CSIC	500 bp				
atmyb27	AMAZE2		250 bp				Reduced
atmyb30	AMAZE2				1st		
atmyb36	AMAZE2		800 bp				
atmyb40		CSIC	70 bp				Null
atmyb43	AMAZE2			3′ ORF			
atmyb44		Versailles	770 bp				
atmyb45	SLAT				2nd		Reduced
atmyb46	AMAZE1					In polyA signal	Reduced
atmyb47	AMAZE2			Myb domain			Reduced
atmyb47	AMAZE2			3′ ORF			Reduced
atmyb50	AMAZE2		250 bp				Normal
atmyb51	AMAZE1		48 bp				
atmyb51	AMAZE1			Myb domain			Null
atmyb51	AMAZE1			3′ ORF			Null
atmyb54	AMAZE2				2nd		
atmyb58	AMAZE1		450 bp				Normal
atmyb61	AMAZE2			Myb domain			Reduced
atmyb61	AMAZE2			3′ ORF			
atmyb67	AMAZE2				1st		Normal
atmyb67	AMAZE2			3′ ORF			Reduced
atmyb67	AMAZE2		155 bp				Normal
atmyb69	SLAT		102 bp				
atmyb70		Versailles	697 bp				
atmyb71	SLAT					150 bp	
atmyb76	AMAZE2			3′ ORF			
atmyb78	AMAZE2			3′ ORF			
atmyb84		Versailles	550 bp				
atmyb92	SLAT			3′ ORF			Null
atmyb94	AMAZE2		13 bp				Aberrant
atmyb101		CSIC	76 bp				Aberrant
atmyb101	AMAZE1			3′ ORF			
atmyb102	AMAZE2			Splicing site			Reduced
atmyb102	AMAZE2			Myb domain			Reduced
Total	40	7	17	22	5	3	

line containing an insertion 70 bp upstream of the start codon of *AtMYB40* did not produce detectable levels of *AtMYB40* RNA. Another insertion, 76 bp upstream of the start codon of *AtMYB101*, produced an aberrant transcript (see Discussion). All other insertions were >500 bp upstream of the start codon and, as far as they were tested, did not result in detectable effects on RNA levels of the *MYB* gene in question.

Identification and Preliminary Inspection of Homozygous Insertion Lines

The detailed analysis of the insertion mutants required the production of lines homozygous for each insertion. Although some of the isolated transposon lines were already homozygous for the insertion due to early transposition events, the majority of the lines isolated were heterozygous for their insertions. Twenty progeny plants of a heterozygous parent were tested by PCR or DNA gel blot analysis for the presence of the insertion and the wild-type alleles. Those plants displaying only the insertion allele were homozygous for the mutant gene. Sister plants homozygous for the wild-type allele were used, if available, as negative controls for all further analysis. The genetic background of these control plants was nearly identical to their siblings homozygous for the insertion. Consequently, detected differences should be caused solely by the insertion into the gene.

To date, we have isolated 32 lines homozygous for insertions in any of 26 genes. Many of these insertions occurred in coding regions, and most of these insertions affected mRNA levels. None of these disruption lines displayed an obvious phenotype when grown in normal conditions on soil. Disrupted lines that had reduced or undetectable transcript levels therefore were systematically screened for phenotypes by using a variety of greenhouse and plate-based assays. Controls were sister plants that did not contain the insertion. Although the systematic screening for phenotypes caused by disruptions in the 26 *MYB* genes has not been

Table 4. Phenotype Screens

Conditions	Phenotypes
Red light	
Far-red light	
Blue light	
White light	
UVB light	UVB-tolerant line
Short day (8 hr)	
Long day (18 hr)	
Cold adaptation (4°C, 7 days)	
Heat adaptation (30°C, 7 days)	
Pseudomonas treatment	
Compatible interaction	
Incompatible interaction	
Abscisic acid[a]	
Gibberellins[a]	
Epibrassinolide[a]	
Kinetin[a]	
Indoleacetic acid[a]	
6% glucose	Two glucose-insensitive germination lines (seedling phenotype)
7.5 mM mannose	
6.8% sucrose	
12% sucrose	
HPLC analysis	Two lines with altered flavonoid levels, one line with altered phenolic levels
Scanning electron microscopy analysis	
Shoot epidermis cell number and size	
Trichome number and branching	
Freeze-fracture leaf cell size and flower cell size	
UV autofluorescence	
Maule reagent	
Seedling phenotype	Ethylene biosynthesis
Root gravitropism and movement	

[a] 0.05, 0.1, 0.5 μM.

completed, most of those that so far have been inspected have no discernable phenotype under the conditions tested (Table 4).

DISCUSSION

The *MYB* Gene Family as a Model for Functional Genomics

PCR-based reverse genetic screens provide an attractive strategy for analyzing gene function, particularly in plants, where homologous recombination is inefficient and where the sequences and chromosomal locations of ~9000 genes already have been revealed. The presence of large gene families in the Arabidopsis genome indicates the likelihood of overlapping functions among multiple genes. This feature in particular requires functional analyses that are feasible through reverse genetics, whereby insertions in individual members of gene families can be isolated (Krysan et al., 1996; Winkler et al., 1998). Mutant lines generated in this manner can then be analyzed for gene-specific insertions and phenotypes, and individual lines containing multiple mutations can be produced by crossing to reveal the extent of potential functional redundancy and to reveal new phenotypes associated with defects in multiple genes.

The R2R3 *MYB* gene family consists of at least 92 closely related members (Kranz et al., 1998) that play important roles in the regulation of secondary metabolism, the control of cell shape, disease resistance, and hormone and stress responses (Martin and Paz-Ares, 1997). These diverse functions suggest that the many uncharacterized members of the family have roles in regulating diverse, and possibly plant-specific, aspects of development, metabolism, and environmental interactions. We have embarked on the systematic characterization of the functions of all members of the R2R3 *MYB* gene family.

Beyond our interests in the basic biology of the gene family, however, the R2R3 *MYB* genes are particularly well-suited for testing the utility and power of large, insertionally mutagenized Arabidopsis populations for identifying mutants by reverse genetics. First, the characterized members of the R2R3 *MYB* gene family are not significantly clustered within the Arabidopsis genome (Kranz et al., 1998), thereby permitting analysis of disruption frequencies in all areas of the genome. Second, an average gene size of 1.6 kb, including potential 5′ and 3′ regulatory components, poses a rigorous challenge to the targeting range of PCR, inasmuch as the average Arabidopsis gene size is ~4 to 5 kb (Bevan et al., 1998; Kaneko et al., 1998). Third, the relatively small size of *MYB* genes permits the use of single gene-specific primers to detect insertion within the 1.6-kb target range by PCR.

Our system has provided important new information about the relative efficiencies of two T-DNA–bearing populations. Moreover, three different transposon insertion populations, based on the maize *En/Spm* transposable element, have been assessed. The populations collectively have provided a relatively large number of useful insertion mutants, although there are specific practical advantages and disadvantages to each system tested.

Comparative Utility of Insertion Populations for Mutant Screening

T-DNA–Insertional Populations

The T-DNA–insertional populations consisted collectively of 13,264 lines with ~1.5 T-DNA loci per line. A 1.6-kb target locus gives a probability of 25% for finding an insertion in a given screened gene, a value with which our data are in fairly good agreement; insertions were found in seven of 39 (18%) of the screened genes. All seven insertions were upstream of the start ATG codon. Two insertions <100 bp upstream of the ATG codon either abolished the gene-specific detection of transcript or produced an aberrant transcript as the result of a read-through product from the strong 35S promotor driving the BASTA selection marker on the insert (data not shown). The frequency of insertions detected is consistent with the estimated copy number of T-DNA loci per plant, indicating that the screening process identified most insertions. It is clear, however, that the 35S promoter can direct complex transcription patterns across insertion sites, which may make insertion mutants difficult to interpret.

Although the isolation of T-DNA insertions in the genes of interest presented here has been successful, the use of T-DNAs as insertional mutagens can be problematic. In one population, only 25% of the lines carried an intact right border, and only 50% carried an intact left border junction due to truncations and rearrangements at the T-DNA ends (Krysan et al., 1996). Consequently, each section of a gene should be screened for insertions from both directions, thereby requiring two gene-specific primers in combination with each of the two T-DNA primers. T-DNA insertion lines also have a high frequency of loci consisting of several T-DNA copies linked together.

Multicopy-Transposon Populations

Two populations with multiple copies of transposable elements per line were screened. The AMAZE lines contain between two and 20 copies of an autonomous *En/Spm* element (i.e., a mean of five to seven copies). The Wageningen lines contain 20 to 30 copies of a nonautonomous d*Spm* element.

Two populations of the AMAZE lines were screened, the first containing ~15,000 independent insertions (Baumann et al., 1998) and the second containing ~35,000 independent insertions (Wisman et al., 1998b). Insertions in ~18% of the genes screened in the first AMAZE population and in ~40% of the genes screened in the second population would be predicted for a 1.6-kb target region. Nevertheless, the first population was found to contain seven insertions among five of the 13 target genes. Three insertions were found in *AtMYB51* alone, which maps to the top arm of chromosome 1. Such "hot spots" of unexpectedly high-frequency insertions also have been reported for insertions in *GL2* and *PIN1*, which lie on the bottom arm of chromosome 1 (Wisman et al., 1998b). Even if the "hot spot" is discounted, the insertion frequency remains a high ~30%. In the second AMAZE population, 20 insertions were found among 15 of the 37 target genes (54%), also higher than the predicted probability of 40%.

The screening strategy for the AMAZE lines culminates in the identification of a single line whose progeny can be analyzed directly for a (segregating) phenotype. Therefore, these lines provided the quickest strategy for identifying insertion mutants among the types tested. However, complete elimination of transcript accumulation was rare, probably because the unstable nature of the autonomous transposable element gave somatic sectors expressing wild-type transcript levels, thereby complicating the analysis of cell nonautonomous phenotypes and phenotypes based solely on biochemical assays. A possibility to overcome this problem is the isolation of stable footprint alleles (Wisman et al., 1998b), a method that requires at least one generation and additional screening.

At the time of screening, the Wageningen collection was estimated to contain 30,000 independent insertions, providing a theoretical 35% chance of finding an insertion in a 1.6-kb target region. Four insertions were found in two genes out of the 18 target genes investigated, an efficiency of 22%. The insertion frequency was thus slightly lower than predicted, and the number of genes with insertions recovered was additionally low due to multiple insertions in the same gene. Higher insertion frequencies have been achieved using the population from the sixth generation of single-seed descent containing an estimated 75,000 independent insertions (A. Pereira, unpublished data). The screening strategy for the Wageningen lines also might be more sensitive to small changes in the PCR protocol than those for the other populations because of the use of relatively short primers and the multiplex nature of the screens. An advantage of the Wageningen lines is that insertions can be stabilized relatively easily upon segregation of the active transposase source away from the *I/Spm* insertion, thereby making subsequent phenotypic screens easier.

The relatively small AMAZE and Wageningen populations are organized into arrays of multiple dimensions that can be systematically pooled for direct identification of individual positive lines. Such pooling strategies offer a significant practical advantage. Furthermore, it is possible to isolate insertions in more than one target gene or multiple insertions in one target gene. However, the large number of insertions in the genetic background can make subsequent phenotypic analysis complex and very time consuming.

Low-Copy-Transposon Populations

The SLAT lines contain one to three copies of stable insertions of a nonautonomous d*Spm* element. Given the low copy number of transposon insertions per line, a large population is necessary to saturate the genome with mutations. In comparison to the other low-copy populations containing T-DNA insertions, however, only half the number of PCRs was required because unidirectional priming was sufficient to cover the gene, a factor of some significance in analyzing large gene families. Using the PCR screening method, we found insertions in nine of the 33 genes searched (27%). This rate is consistent with the 18% expected from the estimated 15,000 independent insertions in the 28,800 lines. The 43,200 lines now available may therefore provide a 50% chance of obtaining an insertion in a 1.6-kb target gene. Indeed, preliminary data from screens of the total 43,200 lines reveal two insertions in the four *MYB* genes screened, but the mutants have not been recovered as homozygous lines. Given the stable nature of the SLAT line insertions, the high proportion of mutants with reduced or null transcript levels perhaps is not surprising. The SLAT lines are, in this respect, particularly amenable to phenotypic analysis. In addition, reactivation of the transposon in the SLAT lines is possible, although reactivation is time-consuming compared with lines already containing active elements. Reactivation, moreover, enables screens for revertant sectors and the possibility of tagging closely linked sites.

Technology of Insert Detection

Generally, the appearance of false positives in PCR screens was the most significant technical difficulty encountered. This problem was ameliorated through the simultaneous screening of at least two dimensions in each PCR screening approach. Identical bands appearing in all dimensions then were used as the criterion for further analysis to identify an insertion in the gene of interest. It was also possible to screen for insertions in the SLAT collection by hybridizing a gene-specific probe to filters carrying inverse PCR fragments from the pooled lines. The main advantage of the filters was the reduction of the amount of DNA used from each individual pool, which, in light of limitations on DNA pools, should permit the wider distribution of this material to the user community. Filters of inverse PCR products from the original 28,800 SLAT lines were used to screen for insertions in 15 genes that had not contained an insertion according to previous PCR screens and
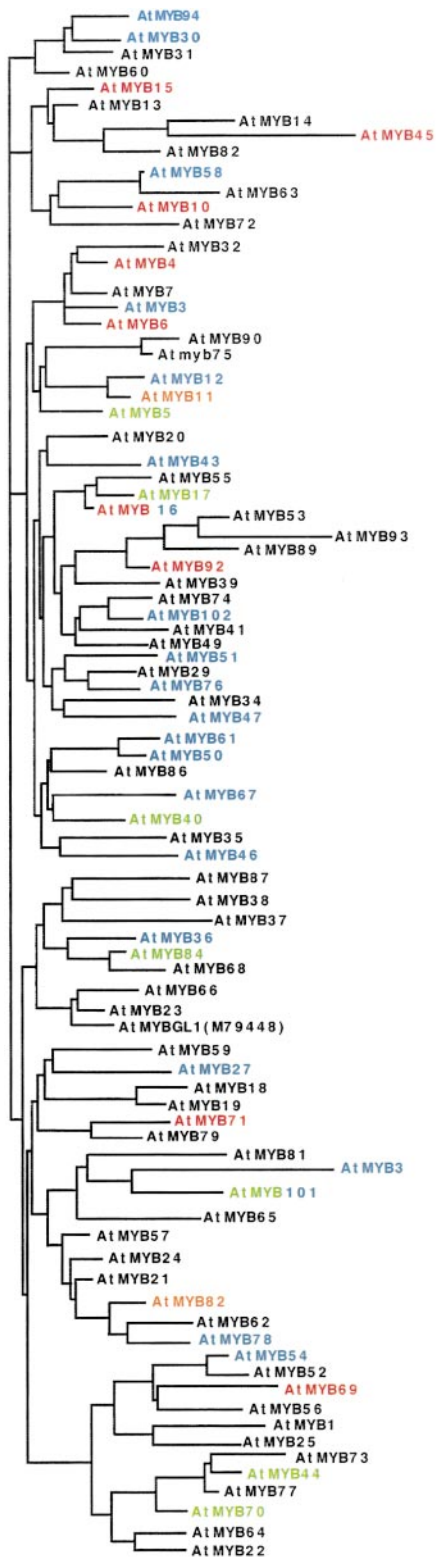
Figure 1. Phylogenetic Relationships of *AtMYB* Genes That Bear Insertion Mutations.

for insertions in two genes previously shown by PCR to contain an insert. No additional insertions were detected in the previously screened genes, and both genes with previously described insertions were detected on the filters (data not shown), demonstrating that filter screening was as effective as PCR-based screens. The very fast and easy first screening step, requiring a single hybridization with a gene-specific probe, is a significant advantage of the filter screen. Admittedly, the signal obtained from a filter does not indicate the position of the insertion in the gene, because insertions flanking the target gene of interest give the same signal on the filter as an insertion in the gene. Therefore, insertions identified by hybridization need to be confirmed by PCR to reveal the likely position of the insertion site, which also needs to be unequivocally demonstrated by sequencing the insertion site.

Toward the Functional Characterization of the R2R3 *MYB* Gene Family

The data presented in this study and elsewhere show that efficient and effective resources are available for isolating gene disruptions in Arabidopsis. The challenge now is to establish screens that provide insights into the functions of individual genes at the cellular and whole-plant level. The analysis of gene expression patterns within insertion populations is a further concern that must be addressed (Martienssen, 1998).

The isolation of 47 insertions among 36 members of the R2R3 *MYB* gene family, as depicted in the family tree in Figure 1, provides a useful starting point for the systematic determination of *MYB* gene function. Thirty insertions among 24 genes were located in transcribed regions, and an additional four were situated <110 bp upstream of the initiating ATG codon. Although transcript levels from these insertion alleles still are being assessed, most lines tested showed some degree of reduction in transcript levels; aberrant transcripts also were observed.

Some of the mutations identified in the R2R3 *MYB* gene family are predicted to represent loss-of-function alleles, although none of the insertions analyzed to date gives rise to morphological phenotypes visible in soil-grown plants. Nevertheless, preliminary evidence for phenotypes associated with *MYB* gene disruptions is being obtained by growing plants and seedlings in a variety of conditions, as shown in

The dendrogam displays the relationships within the Arabidopsis R2R3 *MYB* gene family. Genes with insertions are shown in color according to the insertional mutagenesis system used: red, SLAT population; blue, AMAZE population; green, T-DNA–harboring populations; orange, Wageningen population; black, no insertion. Genes with identified insertions in two systems appear in two colors.

Table 4. These screens are being refined, with an emphasis on stress responses and metabolite profiling. Additional assessment of phenotypes will include outcrossing of insertions and their subsequent homozygous recovery and either complementation or reversion of insertions. In addition, crosses are being made to generate double and triple mutants in closely related *MYB* genes and other genes that show expression patterns similar to *MYB* genes. The emerging data corroborate the involvement of *MYB* genes in a wide variety of cellular processes.

METHODS

Plant Growth Conditions

Arabidopsis thaliana plants were grown in soil in temperature-controlled greenhouses with additional light in winter to ensure long-day conditions (16-hr photoperiod). Where appropriate, BASTA (2.5 mL of Challenge [Agrevo, Frankfurt, Germany] in 1 liter of water) selection was performed by spraying the plants 2 and 3 weeks after sowing.

Plant Populations Screened

Three distinct populations of Arabidopsis bearing the maize transposable element *En/Spm* were analyzed (Table 1). The SLAT collection of 28,800 lines was generated in the ecotype Columbia. Due to clonal events, not all of the lines are the result of independent transposition events, so the estimated number of independent insertions in the screened lines is 15,000. The Wageningen collection contains 2592 lines (ecotype Landsberg *erecta*) harboring ~65,000 defective *Spm* (d*Spm*) elements. However, the estimated number of independent insertions in the screened population was 30,000. The first AMAZE subpopulation comprised 3000 lines containing ~15,000 autonomous transposons (Baumann et al., 1998; Wisman et al., 1998a), and the second AMAZE subpopulation comprised 5000 lines with ~35,000 transposons (Wisman et al., 1998b). Both of the AMAZE populations were in the ecotype Columbia.

Two T-DNA–bearing populations were analyzed (both ecotype Wassilewskija; Table 1): the Versailles population of 9264 lines collectively contains ~14,000 T-DNA insertion sites and is publicly available from the Nottingham Arabidopsis Stock Centre (Nottingham, UK); the second population of 4000 lines with 6000 insertions was produced by a consortium of Spanish laboratories (C̲onsejo S̲uperior de I̲nvestigaciones C̲ientificas [CSIC]). Several of the lines described here are under continual development, and the number of lines and independent insertions now available has increased significantly.

Calculations of Probability

The probability of an insertion into a 1.6-kb *MYB* sequence in the genome was calculated by the formula $P = 1 - \exp(n \times \ln[1 - f])$, where P is the probability of obtaining an insertion, *n* is the number of independent insertion events, and f is the inverse of the number of potential *MYB* targets. The value for f used in our calculations was 1/62,500, which assumes a genome size of 100,000 kb.

Molecular Techniques

Standard molecular techniques were used according to Sambrook et al. (1989). Genomic DNA was extracted for DNA gel blot analysis as described by Carroll et al. (1995), except that RNase (10 μg/mL) was added to the extraction buffer, which obviated the need for RNase treatment subsequent to DNA precipitation. Alternatively, a DNA extraction method for polymerase chain reaction (PCR) analysis, as described by Edwards et al. (1991), was used with minor modifications. Total RNA was isolated according to Chomczynski and Sacchi, (1987) or as described by Prescott and Martin (1987). Sequencing was performed using the ABIPRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer, Warrington, UK) according to manufacturer's protocols.

RNA Expression Analysis

The RNA levels of the insertion lines were tested by reverse transcription–PCR (RT-PCR) or RNA gel blot analysis, depending on the wild-type expression level of the gene being analyzed. RT using Superscript reverse transcriptase (Life Technologies, Gaithersburg, MD) was performed as instructed by the manufacturer. As controls for RT-PCR, parallel reactions were performed using primers for actin, ubiquitin, or other *MYB* genes.

DNA Gel Blots

Genomic DNA (2 to 5 μg) was digested with appropriate restriction enzymes. Digested genomic DNA or PCR products were resolved on agarose gels and blotted onto positively charged membranes (Boehringer Mannheim). These membranes then were incubated with probes that were labeled with digoxygenin (DIG; Boehringer Mannheim) according to manufacturer's protocols.

PCR Screening of the Insertion Populations

AMAZE Lines

The lines are organized in four dimensions so that each line can be designated according to its three-tray, single-tray, row, and column specifications (Baumann et al., 1998). The first population was contained among 28 three-tray arrays, and the second population was distributed among 44 three-tray arrays. We screened DNA from three-tray pools of lines by means of primers En205 (*En* 5′ end: 5′-AGAAGCACGACGGCTGTAGAATAGGA-3′) and En8130 (*En* 3′ end: 5′-GAGCGTCGGTCCCCACACTTCTATAC-3′), each in combination with a gene-specific primer (21 to 32 mer) corresponding to the 3′ end of a given *MYB* open reading frame (ORF). The following PCR regime was used: 85°C for 2 min, 40 cycles of 94°C for 40 sec, 65°C for 1 min, 72°C for 2 min, followed by 72°C for 5 min. Each 50-μL reaction contained 50 ng of DNA as amplification template, each primer at a final concentration of 0.4 μM, 75 μM dNTPs (Promega, Southhampton, UK), PCR buffer, and 2.5 units Taq polymerase (Amplitaq; Perkin-Elmer). The PCR products were separated on agarose gels, transferred to membranes, and hybridized with a DIG-labeled gene-specific probe.

Upon a positive PCR by using the DNA from a given three-tray pool of lines, subsequent rounds of semi-nested PCR were per-

formed individually with DNA originating from single-tray, column, and row pools, and amplification products then were visualized on ethidium bromide–stained gels. Depending on the orientation of the insert, semi-nested PCR also was performed on a dilution (1:100) of the three-tray PCR product, using either En91R (*En* 5′-end nested primer: 5′-TGCAGCAAAACCCACACTTTTACTTC-3′) or En8166 (*En* 3′-end nested primer: 5′-TGCAGCAAAACCCACACTTTTACTTC-3′) primers and the given gene-specific primer. The same PCR regime as described above then was implemented for 30 cycles. PCR products were gel purified and sequenced directly to confirm the transposon insertion and to determine the insertion site. The coordinates of positive PCRs in terms of the single-tray, row, and column pools (i.e., the intersection of positive PCRs among the three dimensions within the three-tray array) were taken to stipulate plants that deserved further analysis. Specifically, the progeny of such plants were tested for the presence of the insertion and phenotypically characterized.

SLAT Population

Mutant lines within the population are arranged in 576 pools of 50 plants in three dimensions. Each superpool (i.e., the first dimension) contains 48 50-plant pools arranged in eight columns (second dimension) and six rows (third dimension) of 50-plant pools. The screening of the entire population of lines took place in two stages, first by PCR screening of the 576 50-line pools, and second by hybridization to nylon filters spotted with DNA of the pooled inverse PCR fragments from the 576 pools.

In the PCR-based screen, 12 superpools were screened. Reaction volumes (25 μL) contained either superpool DNA (50 ng) or the DNA (25 ng) representative of a row or column of 50-plant pools and were otherwise prepared, in terms of primers and temperature cycling, as described for the AMAZE lines (above). For subsequent nested PCR, a combination of primers, including a primer specific for the 5′ end of the d*Spm* insert (i.e., dspm11, 5′-GGTGCAGCAAAACCCACACTT-TTACTTC-3′), the nested primer dspm5 (5′-CGGGATCCGACA-CTCTTTAATTAACTGACACTC-3′), a primer for the corresponding 3′ end of the insert (i.e., dspm1, 5′-CTTATTTCAGTAAGAGTGTGG-GGTTTTGG-3′), and nested primer dspm8 (5′-GTTTTGGCCGAC-ACTCCTTACC-3′), was used along with the *MYB*-specific primer. The reaction regime of this nested PCR protocol was as follows: 94°C for 2 min; then 35 cycles of 94°C for 15 sec, 65°C for 30 sec, and 68°C for 4 min; and finally 5 min at 68°C. Reaction mixtures (20 μL) contained either superpool DNA (50 ng) or DNA (25 ng) representative of a row or column of 50-plant pools with each primer at 0.3 μM, 0.25 mM dNTPs, PCR buffer, and 1.25 units of Taq polymerase. From the resulting PCR products, diluted 1:100, semi-nested or nested PCR was performed for 25 cycles, and the products were resolved on agarose gels, blotted, and hybridized with a gene-specific DIG-labeled probe. Positive PCRs that intersected in terms of superpool, pooled column, and pooled row were taken to designate a single 50-plant pool for subsequent analysis, which began with confirmation of the positive PCR and subsequent sequencing of the confirmed PCR product. Mixed seeds (200 to 500) from the insertion-positive 50-plant pool were sown and sprayed with BASTA to select against those individual plants lacking insertions, and the resulting progeny were tested by PCR (again in a pooling strategy) so that individual plants containing the insertion could be identified.

The SLAT filters were hybridized with a gene-specific probe labeled with DIG. Positive pools were tested by PCR by using a gene-specific primer and the *En*/d*Spm* primers, as described above in the semi-nested strategy. This method also identified single positive pools, which were confirmed by sequencing the PCR products. Single plant lines carrying the insertions were identified as described above.

Wageningen Lines

Lines were grown in three sets of nine trays, with each tray divided into eight rows and 12 columns (for a schematic depiction, see Speulman et al., 1999, in this issue). Eighty-seven three-primer PCR mixtures (representing 87 pooled lines from three sets of trays plus columns plus rows ($3 \times [9 + 8 + 12]$)) were performed. Primers specific for both ends of the d*Spm* transposon (itir2: 5′-CTTGAC-GTTTTCTTGTAGTG-3′; and itir3: 5′-CTTGCCTTTTTTCTTGTAGTG-3′) were used together with the *MYB*-specific primer in each PCR. Reaction mixtures (50 μL) contained PCR buffer, 1.5 mM MgCl$_2$, 1.0 μM each primer, 0.25 mM dNTPs, and 1 unit of Taq polymerase. The following PCR regime was used: 94°C for 4 min; 30 cycles of 94°C for 45 sec, 60°C for 45 sec, 72°C for 3 min; followed by a final incubation for 7 min at 72°C.

PCR products were resolved on agarose gels, transferred to filters, and hybridized with a gene-specific DIG-labeled probe. Individual lines containing insertions within *MYB* sequences were identified upon the appearance of a specific PCR product common to pools of lines from a given set that intersected in terms of tray, column, and row. To confirm the presence of insertions within a single line, we tested at least 12 progeny plants by PCR for the presence of the insertion, which we then verified by sequencing. To stabilize the insertion within the gene of interest and to limit the number of transposon insertions in the background, we crossed progeny plants carrying the insertion to wild-type plants. After selfing of the resulting F$_1$ plants, the F$_2$ generations were tested for the presence of the insertion and for the absence of the transposase source.

Versailles and CSIC (T-DNA) Lines

The T-DNA lines were organized in 285 pools of 48 lines. Both sets (i.e., Versailles and CSIC) of T-DNA lines were further pooled into 36 superpools (each containing roughly eight pools of 48 lines) or 12 hyperpools (each containing three superpools). Superpools and hyperpools were screened by PCR by using the 3′ and 5′ border-specific primers for the T-DNA (i.e., respectively, primer TAG3 [5′-CTGATA-CCAGACGTTGCCCGCATA A-3′] and primer TAG5 [5′-CTACAAATT-GCCTTTTCTTATCGAC-3′]). Because T-DNA insertions often have only one intact border (Krysan et al., 1996; Nacry et al., 1998), T-DNA–harboring lines were screened with two gene-specific primers, one specific for each of the ends of the *MYB* gene. A touch-down PCR program was used: 94°C for 2 min; 10 touch-down cycles: 94°C for 5 sec, 65°C for 30 sec (each cycle decreasing by 1°C), 72°C for 2 min; 35 cycles of 94°C for 15 sec, 55°C for 15 sec, 72°C for 1 min; 72°C for 2 min. PCR mixtures were prepared either according to the method described for the AMAZE lines (see above) or in reaction volumes of 25 μL containing 100 ng of DNA template and PCR buffer, 2.5 mM MgCl$_2$, 1 μM each primer, 0.2 mM dNTPs, and 1 unit of Taq polymerase. Because T-DNA borders often carry large deletions, nested PCR was not used to screen the T-DNA–containing lines. The products of the PCR were separated on agarose gels, transferred to filters, and hybridized with a gene-specific DIG-labeled probe. After identifying a positive superpool, its corresponding eight pools (48 lines per pool) were tested for the presence of the same

PCR band. Subsequently, positive PCR bands were reamplified, gel purified, and sequenced for confirmation and verification of the insertions. Twenty seeds from each of the 48 plants of a given positive pool were sown and tested in pools by PCR for the presence of the insertion to isolate a single positive line.

Phenotype Screens

Seeds from homozygous insertion lines, and from sister plants with no insertions as controls, were germinated on plates or grown in soil under various conditions as shown in Table 4.

ACKNOWLEDGMENTS

REFERENCES

Aarts, M.G.M., Corzaan, P., Stiekema, W.J., and Pereira, A. (1995). A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana.* Mol. Gen. Genet. **247,** 555–564.

Avila, J., Nieto, C., Canas, L., Benito, M.J., and Paz-Ares, J. (1993). *Petunia hybrida* genes related to the maize regulatory *C1* gene and to animal *myb* proto-oncogenes. Plant J. **3,** 553–562.

Azpiroz-Leehan, R., and Feldmann, K.A. (1997). T-DNA insertion mutagenesis in Arabidopsis: Going back and forth. Trends Genet. **13,** 152–156.

Ballinger, D.G., and Benzer, S. (1989). Targeted gene mutations in *Drosophila.* Proc. Natl. Acad. Sci. USA **86,** 9402–9406.

Baranowskij, N., Frohberg, C., Prat, S., and Willmitzer, L. (1994). A novel DNA binding protein with homology to Myb oncoproteins containing only one repeat can function as a transcriptional activator. EMBO J. **13,** 5383–5392.

Baulcombe, D.C. (1996). RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. Plant Mol. Biol. **32,** 79–88.

Baumann, E., Lewald, J., Saedler, H., Schulz, B., and Wisman, E. (1998). Successful PCR-based reverse genetic screens using an *En-1*–mutagenised *Arabidopsis thaliana* population generated via single-seed descent. Theor. Appl. Genet. **97,** 729–734.

Bender, J., and Fink, G.R. (1998). A MYB homologue, ATR1, activates tryptophan gene synthesis in *Arabidopsis.* Proc. Natl. Acad. Sci. USA **95,** 5655–5660.

Bevan, M., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana.* Nature **391,** 485–488.

Bouchez, D., and Höfte, H. (1998). Functional genomics in plants. Plant Physiol. **118,** 725–732.

Bouchez, D., Camilleri, C., and Caboche, M. (1993). A binary vector based on Basta resistance in planta transformation of *Arabidopsis thaliana.* C. R. Acad. Sci. Ser. III Sci. Vie **316,** 1188–1193.

Carroll, B.J., Klimyuk, V.I., Thomas, C.M., Bishop, G.J., Harrison, K., Scofield, S.R., and Jones, J.D.G. (1995). Germinal transpositions of the maize element *Dissociation* from T-DNA loci in tomato. Genetics **139,** 407–420.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocynate-phenol-chloroform extraction. Anal. Biochem. **162,** 156–159.

Das, L., and Martienssen, R. (1995). Site-selected transposon mutagenisis at the *hcf106* locus in maize. Plant Cell **7,** 287–294.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. **19,** 1349.

Feldbrügge, M., Sprenger, M., Hahlbrock, K., and Weisshaar, B. (1997). PcMYB1, a novel plant protein containing a DNA-binding domain with one MYB repeat, interacts in vivo with a light-regulatory promoter unit. Plant J. **11,** 1079–1093.

Feldmann, K.A. (1991). T-DNA insertion mutagenesis in Arabidopsis: Mutational spectrum. Plant J. **1,** 71–82.

Gaymard, F., Pilot, G., Lacombe, B., Bouchez, D., Bruneau, D., Boucherez, J., Michaux-Ferriere, N., Thilbaud, J.-B., and Sentenac, H. (1998). Identification and disruption of a plant shaker-like outward channel involved in K$^+$ release into the xylem sap. Cell **94,** 647–655.

Grotewold, E., Drummond, B.J., Bowen, B., and Peterson, T. (1994). The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. Cell **76,** 543–553.

Gubler, F., Kalla, R., Roberts, J.K., and Jacobsen, J.V. (1995). Gibberellin-regulated expression of a myb gene in barley aleurone cells: Evidence for Myb transactivation of a high-pI alpha-amylase gene promoter. Plant Cell **7,** 1879–1891.

Hirsch, R.E., Lewis, B.D., Spalding, E.P., and Sussman, M.R. (1998). A role for the AKT1 potassium channel in plant nutrition. Science **280,** 918–920.

Hoeren, F.U., Dolferus, R., Wu, Y., Peacock, W.J., and Dennis, E.S. (1998). Evidence for a role for AtMYB2 in the induction of the Arabidopsis alcohol dehydrogenase gene (*ADH1*) by low oxygen. Genetics **149,** 479–490.

Höfgen, R., Axelsen, K.B., Kannangara, C.G., Schuttke, I., Pohlenz, H.D., Willmitzer, L., Grimm, B., and Wettstein, D.v. (1994). A visible marker for antisense mRNA expression in plants: Inhibition of chlorophyll synthesis with a glutamate-1-semialdehyde aminotransferase antisense gene. Proc. Natl. Acad. Sci. USA **91,** 1726–1730.

Iturriaga, G., Leyns, L., Villegas, A., Gharaibeh, R., Salamini, F., and Bartels, D. (1996). A family of novel myb-related genes from the resurrection plant *Craterostigma plantagineum* are specifically expressed in callus and roots in response to ABA or desiccation. Mol. Gen. Genet. **32,** 707–716.

Jackson, D., Culianez Macia, F., Prescott, A.G., Roberts, K., and Martin, C. (1991). Expression patterns of myb genes from Antirrhinum flowers. Plant Cell **3,** 115–125.

Kaiser, K., and Goodwin, S.F. (1990). "Site-selected" transposon mutagenisis of *Drosophila.* Proc. Natl. Acad. Sci. USA **87,** 1686–1690.

Kaneko, T., Kotani, H., Nakamura, Y., Sato, S., Asamizu, E., Miyajima, N., and Tabata, S. (1998). Structural analysis of *Arabidopsis thaliana* chromosome 5. V. Sequence features of the regions of 1,381,565 bp covered by twenty-one physically assigned P1 and TAC clones. DNA Res. **5,** 131–145.

Kempin, S., Liljegren, S.J., Block, L.M., Rounsley, S.D., and Yanofsky, M.F. (1997). Targeted disruption in *Arabidopsis.* Nature **389,** 802–803.

Kirik, V., and Bäumlein, H. (1996). A novel leaf-specific *myb*-related protein with a single binding repeat. Gene **183,** 109–113.

Koes, R., et al. (1995). Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. Proc. Natl. Acad. Sci. USA **92,** 8149–8153.

Koncz, C., Nemeth, K., Redei, G.P., and Schell, J. (1992). T-DNA insertional mutagenesis in Arabidopsis. Plant Mol. Biol. **20,** 963–976.

Kooter, J.M., and Mol, J.N.M. (1993). *Trans*-inactivation of gene expression in plants. Curr. Opin. Biotechnol. **4,** 166–171.

Kranz, H.D., Denekamp, M., Greco, R., Jin, H., Levya, A., Meissner, R.C., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., Smeekens, S., Tonelli, C., Paz-Ares, J., and Weisshaar, B. (1998). Towards functional characterisation of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana.* Plant J. **16,** 263–276.

Krysan, P.J., Young, J.C., Tax, F., and Sussman, M.R. (1996). Identification of transferred DNA insertions within Arabidopsis genes involved in signal transduction and ion transport. Proc. Natl. Acad. Sci. USA **93,** 8145–8150.

Lin, Q., Hamilton, W.D.O., and Merryweather, A. (1996). Cloning and initial characterization of 14 myb-related cDNAs from tomato (*Lycopersicon esculentum* cv. Ailsa Craig). Plant Mol. Biol. **30,** 57–64.

Lüscher, B., and Eisenman, R.N. (1990). New light on Myc and Myb. Part II. Myb. Genes Dev. **4,** 2235–2241.

Magaraggia, F., Solinas, G., Valle, G., Giovinazzo, G., and Coraggio, I. (1997). Maturation and translation mechanisms involved in the expression of a *myb* gene of rice. Plant Mol. Biol. **35,** 1003–1008.

Martienssen, R.A. (1998). Functional genomics: Probing plant gene function and expression with transposons. Proc. Natl. Acad. Sci. USA **95,** 2021–2026.

Martin, C., and Paz-Ares, J. (1997). MYB transcription factors in plants. Trends Genet. **13,** 67–73.

McKinney, E.C., Ali, N., Traut, A., Feldman, K.A., Belostotsky, D.A., McDowell, J.M., and Meagher, R.B. (1995). Sequence-based identification of T-DNA insertion mutations in Arabidopsis: Actin mutants act2-1 and act4-1. Plant J. **8,** 613–622.

Moyano, E., Martinez Garcia, J.F., and Martin, C. (1996). Apparent redundancy in myb gene function provides gearing for the control of flavonoid biosynthesis in Antirrhinum flowers. Plant Cell **8,** 1519–1532.

Müller, A., Guan, C., Gälweiler, L., Tänzler, P., Huijser, P., Marchlant, A., Parry, G., Bennett, M., Wisman, E., and Palme, K. (1998). *AtPIN2* defines a locus of *Arabidopsis* for root gravitropism control. EMBO J. **17,** 6903–6911.

Nacry, P., Camilleri, C., Courtial, B., Caboche, M., and Bouchez, D. (1998). Major chromosomal rearrangments induced by T-DNA transformation in Arabidopsis. Genetics **149,** 641–650.

Noda, K.-I., Glover, B.J., Linstead, P., and Martin, C. (1994). Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. Nature **369,** 661–664.

Oppenheimer, D.G., Herman, P.L., Sivakumaran, S., Esch, J., and Marks, M.D. (1991). A *myb* gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. Cell **67,** 483–493.

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A., and Saedler, H. (1987). The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. EMBO J. **6,** 3553–3558.

Prescott, A.G., and Martin, C. (1987). Rapid method for the quantitative assessment of levels of specific mRNAs in plants. Plant Mol. Biol. Rep. **4,** 219–224.

Romero, I., Fuertes, A., Benito, M.J., Malpica, J.M., Leyva, A., and Paz-Ares, J. (1998). More than 80R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana.* Plant J. **14,** 273–284.

Rosinski, J.A., and Atchley, W.R. (1998). Molecular evolution of the Myb family of transcription factors: Evidence for polyphyletic origin. J. Mol. Evol. **46,** 74–83.

Sablowski, R.W.M., Moyano, E., Culianez Macia, F.A., Schuch, W., Martin, C., and Bevan, M. (1994). A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. EMBO J. **13,** 128–137.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Speulman, E., Metz, P.L.J., van Arkel, G., te Lintel Hekkert, B., Stiekema, W.J., and Pereira, A. (1999). A two-component *Enhancer-Inhibitor* transposon mutagenesis system for functional analysis of the Arabidopsis genome. Plant Cell **11,** 1853–1866.

Tamagnone, L., Merida, A., Parr, A., Mackay, S., Culianez-Macia, F.A., Roberts, K., and Martin, C. (1998). The AmMYB308 and AmMYB330 transcription factors from Antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. Plant Cell **10,** 135–154.

Thompson, M.A., and Ramsay, R.G. (1995). Myb: An old oncoprotein with new roles. BioEssays **17,** 341–350.

Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G., and Jones, J.D.G. (1999). Multiple independent defective *Suppressor-mutator* transposon insertions in Arabidopsis: A tool for functional genomics. Plant Cell **11,** 1841–1852.

Urao, T., Yamaguchi Shinozaki, K., Urao, S., and Shinozaki, K. (1993). An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell **5,** 1529–1539.

van der Krol, A.R., Mur, L.A., de Lange, P., Gerats, A.G.M., Mol, J.N.M., and Stuitje, A.R. (1990). Antisense chalcone synthase genes in petunia: Visualization of variable transgene expression. Mol. Gen. Genet. **220,** 204–212.

Waites, R., Selvadurai, H.R.N., Oliver, I.R., and Hudson, A. (1998). The *PHANTASTICA* gene encodes a MYB transcription factor

involved in growth and dorsoventrality of lateral organs in *Antirrhinum.* Cell **93**, 779–789.

Winkler, R.G., Frank, M.R., Galbraith, D.W., Feyereisen, R., and Feldmann, K.A. (1998). Systematic reverse genetics of transfer-DNA-tagged lines of Arabidopsis. Plant Physiol. **118**, 743–750.

Wisman, E., Cardon, G.H., Fransz, P., and Saedler, H. (1998a). The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis. Plant Mol. Biol. **37**, 989–999.

Wisman, E., Hartmann, U., Sagasser, M., Baumann, E., Palme, K., Hahlbrock, K., Saedler, H., and Weisshaar, B. (1998b). Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. Proc. Natl. Acad. Sci. USA **95**, 12432–12437.

Yang, Y., and Klessig, D.F. (1996). Isolation and characterization of a tobacco mosaic virus–inducible *myb* oncogene homolog from tobacco. Proc. Natl. Acad. Sci. USA **93**, 14972–14977.

Zwaal, R.R., Broeks, A., van Meurs, J., Groenen, J.T.M., and Plasterk, R.H.A. (1993). Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. Proc. Natl. Acad. Sci. USA **90**, 7431–7435.