

GENOMICS ARTICLE

# A Two-Component *Enhancer-Inhibitor* Transposon Mutagenesis System for Functional Analysis of the Arabidopsis Genome

Elly Speulman, Peter L. J. Metz, Gert van Arkel, Bas te Lintel Hekkert, Willem J. Stiekema, and Andy Pereira<sup>1</sup>

Centre for Plant Breeding and Reproduction Research, Department of Molecular Biology, P.O. Box 16, NL-6700 AA Wageningen, The Netherlands

A modified *Enhancer-Inhibitor* transposon system was used to generate a series of mutant lines by single-seed descent such that multiple *I* insertions occurred per plant. The distribution of original insertions in the population was assessed by isolating transposon-flanking DNA, and a database of insertion sites was created. Approximately three-quarters of the identified insertion sites show similarity to sequences stored in public databases, which demonstrates the power of this regimen of insertional mutagenesis. To isolate insertions in specific genes, we developed three-dimensional pooling and polymerase chain reaction strategies that we then validated by identifying mutants for the regulator genes *APETALA1* and *SHOOT MERISTEMLESS*. The system then was used to identify inserts in a class of uncharacterized genes involved in lipid biosynthesis; one such insertion conferred a *fiddlehead* mutant phenotype.

## INTRODUCTION

Over the past two decades, Arabidopsis has become a model organism for research in plant physiology, development, biochemistry, and pathogenesis. Recent developments in Arabidopsis genomics research include the construction of a nearly complete physical map of the genome based on yeast artificial chromosomes (Choi et al., 1995; Schmidt et al., 1996, 1997; Zachgo et al., 1996; Camilleri et al., 1998), bacterial artificial chromosomes (BACs; Mozo et al., 1998), P1 clones (Liu et al., 1995; Kaneko et al., 1998), the systematic sequencing of expressed sequence tags (ESTs; Höfte et al., 1993; Newman et al., 1994), and coordinated multinational sequencing efforts (Goodman et al., 1995; Bevan et al., 1998).

The large amount of sequence information generated by Arabidopsis genome and EST sequencing projects will provide the basis for systematic studies of gene function and eventually will allow for unlimited comparisons with the yeast (Mewes et al., 1997) and Caenorhabditis (*C. elegans* Sequencing Consortium, 1998) genomes. Such cross-kingdom analyses might reveal those genes that are necessary for the development and maintenance of eukaryotic cells and those that are associated with functions unique to either

plants or animals. An intriguing finding to have emerged already from these studies is that half of the Arabidopsis ESTs bear no significant similarity to sequences from the previously established databases (Rounsley et al., 1996).

Over the past decade, insertional mutagenesis has proven to be one of the most efficient ways of isolating and identifying genes via the traditional approaches of “forward” genetics (Feldmann, 1991; Koncz et al., 1992; Aarts et al., 1993; Bancroft et al., 1993; Jones et al., 1994; Okuley et al., 1994; Azpiroz-Leehan and Feldmann, 1997). In light of the ongoing accumulation of ever more DNA sequences of unknown function, insertional mutagenesis now offers the means for determining the function of newly elaborated sequences by a process of “reverse” genetics. The strategy begins with the broad insertional mutagenesis of the genome through the use of T-DNA or transposons followed by the identification of individuals carrying insertions in a gene of interest and phenotypic and functional analysis (Ballinger and Benzer, 1989; Kaiser and Goodwin, 1990).

In one well-described reverse genetics strategy, the identification of gene-specific insertions from among a large mutagenized population involves polymerase chain reactions (PCRs) in which the first primer is designed to anneal to the insertion sequence and the second anneals to the targeted gene. Thus, PCR products are specified by a combination of the target gene and the insertion element, thereby indicating the proximity of insertion to the gene of interest. Specific

<sup>1</sup>To whom correspondence should be addressed. E-mail A.Pereira@cpro.dlo.nl; fax 31-317-418094.

pooling strategies then can lead to identification of the desired mutant individual or individuals within the treated population. This target-selected mutational strategy has been applied successfully within studies of *Drosophila*, *Caenorhabditis*, maize, and petunia (Kaiser and Goodwin, 1990; Rushforth et al., 1993; Zwaal et al., 1993; Das and Martienssen, 1995; Koes et al., 1995; Mena et al., 1996).

In *Arabidopsis*, most reverse genetics screens have used T-DNA insertional mutagenesis (McKinney et al., 1995; Krysan et al., 1996; Azpiroz-Leehan and Feldmann, 1997; Bouchez and Höfte, 1998; Winkler et al., 1998) and have required the use of large populations so that only a few inserts per plant would result in genome saturation. To develop an effective reverse genetics strategy in *Arabidopsis* based on transposons, we have previously described a two-component maize *Enhancer-Inhibitor* system (*En-I*; *Enhancer* is sometimes designated as *Suppressor-mutator* [*Spm*], *Inhibitor* as defective *Spm* [*dSpm*]). Unlike the use of T-DNA as the insertional mutagen, transposons can be mobilized subsequent to their mutagenic insertion, thereby yielding revertants that can confirm the phenotypic consequences of the mutation. Another advantage is offered in that the maize *Activator-Dissociation* (*Ac-Ds*) and *En/Spm* transposable elements tend to transpose to linked sites, a phenomenon that can be used for local mutagenesis or for targeted insertion into specific domains of a gene (Peterson, 1970; Bancroft and Dean, 1993; Cardon et al., 1993; Aarts et al., 1995b; James et al., 1995). Indeed, the autonomous *En* element recently has been shown to be useful for reverse genetics in *Arabidopsis* (Baumann et al., 1998; Wisman et al., 1998).

Our two-component system consists of a mobile *I* transposon and an immobilized *trans*-active *En* transposase under control of the cauliflower mosaic virus 35S promoter. A stable transposase locus, T-En5, was shown to mediate frequent transposition of *I* elements throughout plant development (Aarts et al., 1993, 1995a). The continuous, frequent transposition of this transposon system was used for the generation of a population of 2592 lines containing multiple *I* elements per line. After six generations of selfing and propagation by single-seed descent, each line was estimated to harbor ~20 *I* inserts.

To demonstrate the use of this population in reverse genetics approaches, we use the well-studied developmental regulator genes *APETALA1* (*AP1*; Mandel et al., 1992) and *SHOOT MERISTEMLESS* (*STM*; Endrizzi et al., 1996; Long et al., 1996) here as examples. Additionally, to complement our previous forward genetics analysis of genes involved in cuticular wax biosynthesis (Aarts et al., 1995b), we selected specific ESTs, potentially related to key genes of very-long-chain fatty acid synthesis, from public databases and used them to find mutants present in our test population. We recovered an insertion in an *FAE1* (for *FATTY ACID ELONGATION 1*) homolog that displayed a *fiddlehead* (*fdh*; Lolle et al., 1992) mutant phenotype. This example again shows the use of our system for identifying inserts that yield biological information via a reverse genetics approach. Our transpo-

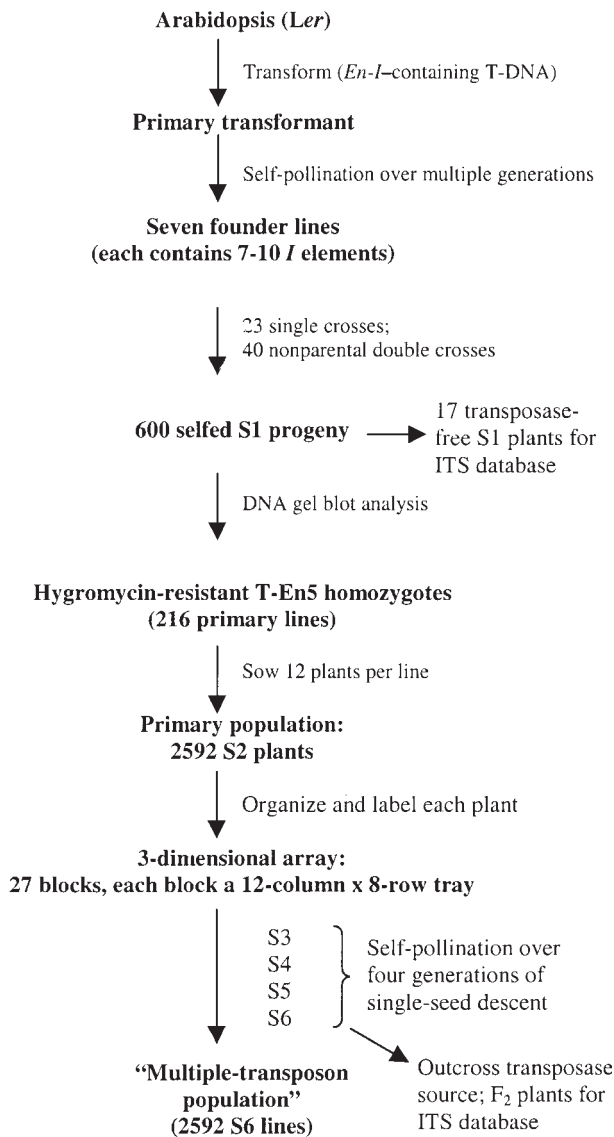
son population also has been used to generate an *Inhibitor*-tagged site (ITS) database by isolating transposon-flanking DNA, using PCR-based methods, followed by database analyses. Our results show that our *I* element insertions are present in genes as well as noncoding regions, distributed all over the sequenced chromosomes. Our methodologies will soon become increasingly attractive and cost effective as the *Arabidopsis* genome is fully elaborated.

## RESULTS

### An ITS Database from the Multiple-Transposon Population

Our population of 2592 *Arabidopsis* lines, each line containing multiple *I* elements, was generated as outlined in Figure 1. In each of the seven original "founder" lines used to start the population, approximately seven to 10 independent inserts were present (see Aarts et al., 1995a). Upon intercrossing these original seven lines, the genomic representation of *I* inserts accumulated to ~20, as estimated by genomic DNA gel blot hybridizations with an *I* element-specific probe. With further increases in copy number, the resolution of inserts became difficult due to high somatic transposition with variable intensities and overlapping bands (data not shown). In the "primary" ( $S_2$ ) population of 2592 plants, most plants had similar transposon insertion sites. After four additional generations of selfing by single-seed descent, and based on a calculated independent transposition frequency of 10 to 30% (Aarts et al., 1995a), we estimate that the resulting "multiple-transposon population" ( $S_6$ ; see Figure 1) contains an average of ~20 to 25 independent *I* elements per line and 50,000 to 65,000 *I* element insertions in total. This number of insertions would correspond to an insertion every 2 kb in the *Arabidopsis* genome.

To assess and predict the distribution of inserts in the population, we chose a set of 17  $S_1$  lines (see Figure 1) that were T-En5 transposase free and thus represented the base population of independent inserts within the seven original lines. Inverse PCR products were isolated from these 17 lines, cloned, and sequenced. An initial database of ~250 primary sequences was generated, after vector and transposon end removal. Table 1 displays a summary of the unique sequences (ITS 1 to 99) from these lines after analysis and database searches. We identified between 10 and 18 *I* element inserts per plant, but this range probably underestimates the actual number because not all inserts are equally accessible by the inverse PCR technique with the one enzyme (*Hinf*I) used. The transposase-free plants were selected by segregation of the transposase, and in this process, one-fourth of the inserts that are linked or randomly segregated were lost. We estimate, based on these extrapolations and genomic DNA gel blots, that the plants at the



**Figure 1.** Schematic Representation of the Generation of 2592 Arabidopsis Lines Containing Multiple *I* Insertions per Line.

onset of population development contained ~20 to 25 *I* elements per plant.

The ITS database was filtered for poor and redundant sequences, yielding a total of 99 nonredundant ITSs in the 17 S<sub>1</sub> plants. The overall efficiency for recovery was 40%. Sequences were given an ITS identity number for entry in public databases.

To generate a population of stable transposons from the advanced S<sub>6</sub> generation of plants containing multiple independent transposons, we outcrossed a set of plants to the wild-type Landsberg *erecta* (*Ler*) ecotype to segregate the

T-En5 transposase source. We established a simple hygromycin-sensitive screening procedure (see Methods) supplemented by a PCR screen to select T-En5 transposase-free (F<sub>2</sub>) segregants from the S<sub>6</sub> generation. Approximately 50% of the hygromycin-sensitive seedlings turned out to be free of T-En5 by the PCR test, the escapes probably due to a weak and nonuniform resistance reaction. In setting up the procedure for establishing the stable transposon population, seven plants were used for isolation and sequencing of ~25 different inverse PCR products. On this limited scale, 18 new ITSs (designated ITS 100 to 117) were obtained.

Perhaps the most remarkable observation of the ITS database comes from the compilation of target site duplications (TSDs). Specifically, there is a striking prevalence of A+T (73%) nucleotides and a complete absence of trinucleotide sequences consisting solely of G+C. This sample thus suggests a bias in insertional specificity toward A-T-rich sites in the genome.

The ITS sequences were compared with those of the public databases by using the BLAST suite and the nonredundant nucleotide, protein, or EST databases. The sequences with a significant BLAST score are indicated in Table 1. Remarkably, 87 of the presented 117 ITSs (74%) show significant similarity to some sequence in the databases. This high frequency sequence similarity is due to the extensive sequence information (average 320 bp) available from the clones derived by inverse PCR.

Figure 2 displays the positions of the *I* elements that were scattered over the sequenced portions of the chromosomes, except for chromosome III, for which little sequence information is available. The number of ITSs that could be chromosome-positioned (76, i.e. 65%) correlates to the extent that chromosomal sequences are defined in the databases. This correlation thus demonstrates the power of the sequence-based system described here for mapping and validates the approach as a method for identifying insertions in those genes annotated in the database entries. The concentration of insertions (15 of the chromosome-positioned 76) at the bottom half of chromosome II, the original source of the *I* elements, suggests local transposition from the T-DNA (T-En5). Indeed, ~10% of all transpositions occurred near the original T-DNA insert. Surprisingly, very few insertions are mapped to the end of chromosome II that is distal to the T-En5 locus. This absence probably reflects a bias in transposition or the direction of recombination events in generation of the transposase-free lines. An encouraging feature is that genes even very closely situated to T-En5, such as the *LTP* gene or the *Mlo*-hi gene, situated ~5 centimorgans (cM) away, can be tagged and recovered in a T-En5-free line, presumably after recombinational segregation.

Significantly, known or predicted genes could be exactly positioned by our approach. Among the gene families for which our approach proved effective were those encoding MYB-like proteins, homeodomain proteins, cytochrome P450-like proteins, AP2-domain containing proteins, actin, invertase, sucrose transporters, MLO-like proteins, 2S1

**Table 1.** ITSs Similar to Sequences Deposited in Public Databases

ITS	Plant	TSD <sup>a</sup>	Size <sup>b</sup>	Nucleotide Similarity <sup>c</sup>	E <sup>d</sup>	Ch <sup>e</sup>	Clone	Predicted Gene/Protein Similarity	E <sup>e</sup>
1	C1-6	GAA	804	AB018109 @64582	9 e-85	V	K17N15	Ribosomal protein L11	0.002
2	C1-6	AAA	329	AC006283 @64775	e-115	II	T1B3	GATA binding TF CAA74002	2 e-21
3	C1-6	ATA	224						
4	C1-6	TTT	430						
5	C1-6	TTG	154	AC005287 @7353	e-66	I	F20D21	Hypothetical protein	
6	C2-6	CAA	133					AC004473 MYB isolog (chrom 1)	e-5
7	C2-6		171	AC005896 @75493	e-75	II	F3G5	ATRZAT11	2 e-31
8	C2-6	AAA	253						
9	C2-6	TCT	272	AB009050 @63923	e-130	V	MDF20	AF002109 CHP-rich hypothetical	2 e-23
10	C2-6	ATG	372						
11	C2-6	CTG	246	ATF7H19 @60091	e-118	IV	F7H19		
12	C2-6		228	AB022220 @74125	e-105	III	MLN21		
13	C3-3	TTG	230						
14	C3-3	TCT	110	AL080282 @65311	2 e-30	IV	T13K14	Promoter TFIIIE-β homolog	
15	C3-3	TCT	163	AC002535 @72066	5 e-10	II	T30B22	T30B22.24 myb related	2 e-5
16	D3-3	TAT	726	AB017070 @33356	9 e-28	V	MNL12		
17	C4-6	TAA	230	AC002986 @3921	e-20	I	YUP8H12R	Rat liver specific transport protein	0.001
18	C4-6	TTT	440						
19	C4-6	GTA	334	A1080253 @99427	e-123	IV	F9F13	Receptor protein kinase cluster	2 e-7
20	C4-6	CAC	189	AB020747 @23401	7 e-84	V	MP110		
21	C4-6	TTT	326	AC003979 @46764	e-6	I	T22J18	Sucrose transporter SUC2 exon 2	0.049
22	C5-5	GAT	599						
23	C5-5	ATA	691	AC006919.4 @43068	e-102	II	F1011	Promoter CBF1 F1011.8	
24	C5-5	TTT	158	AC002409 @71376	e-63	II	T20B5		
25	C5-5	TTT	70	AC006234 @36917	4 e-27	II	F5H14	Vacuolar ATP synthase subunit B	4 e-6
26	C5-5		674	AC004697 @78462	0.00	II	T16B24	T16B24.16 Mlo-hi	6 e-9
27	C5-5	TAT	408	AF069299 @80416	e-133	IV	F6N15	F6N15.25 predicted protein	3 e-15
28	C5-5	GCT	334	AB008267 @53586	6 e-5	V	MMG4	Reverse transcriptase AC005311	7 e-27
29	C5-5	TAG	376	ATHAT2SI @1066	e-130	IV	T24A18	2S seed storage protein ARATH2SI	8 e-31
30	C6-6	TTT	89	AB028609 @20919	e-27	III	K7P8		
31	C6-6	AAG	263	AB012247 @63906	e-57	III	MSL1		
32	C6-6	AGA	110	AC002130 @28965	2 e-20	I	F1N21		
33	C6-6	CTT	487	AF058919 @89167	0.00	IV	F6N23	AHDP homeodomain protein intron	2 e-7
34	C6-6	AAG	200	AB016871 @40390	7 e-10	V	K16L22		
35	C7-4	ACT	165	ATF 2206 @3697	3 e-33	III	F2206	Invertase ATU011033	e-4
36	C7-4	AGA	411	AB005247 @24261	0.00	V	MXA21	AC009187-3 transposase	2 e-22
37	C7-4	CTT	101						
38	C7-4	AAT	180	AB010068 @46958	8 e-37	V	K18P6		
39	C7-4		361						
40	C7-4	GTA	514	AB020752 @49384	0.00	V	MTH16		
41	C7-4	AGT	457	AC005499 @87807	0.00	II	T6A23	LTP2 @1751 3' region	
42	C7-4	TAT	212	AC002339 @39352	e-88	II	T11A7	T11A07.10 / Castor protein	3 e-29
43	C7-4	TTT	152						
44	C7-4	TAA	420	AL031032 @24279	e-132	IV	F17I5	P-loop containing protein, exon 2	2 e-8
45	C7-4		199	ATTS0717/Z18212	e-9				
46	C7-4	TAG	410	AC004044 @69038	0.00	IV	T5J8	Hypothetical protein T5J8.19	
47	C7-4	AAT	466	AB023039 @56064	0.00	V	MIF21	Leucine-rich repeat gene <i>C. elegans</i>	
48	C8-1	CTG	136	AC002396 @15675	e-35	I	F3I6	Zn metalloproteinase	e-7
49	C8-1	AAC	232	AC007727 @41316	e-67	I	F8K7	Predicted gene	
50	C8-1	TTG	602	AB016819	2 e-18			UDP-glucose glycosyltransferase	e-49
51	C8-1	AAA	268	AB025638 @7353	e-108	V	MWF20		
52	D8-1	CAA	250						
53	D8-1	CTA	561	AC007260 @61703	1.6 e-87	I	T30F21		
54	D8-1	CAC	353	AC004260 @40555	7.8 e-43	I	T14N5		
55	D8-1	CTG	579						
56	D8-1	CGA	414						
57	C10-1	TGA	383						
58	C10-1	TAC	62	AB028622 @40558	5 e-25	III	MZN24		
59	D10-1	GCT	372	AF039380 C4-RGL	1.3 e-24	I	F24J19	AAC78631 <i>RPP8</i> gene	3 e-22
60	D10-1		150						
61	C10-4	CAC	435						
62	C10-4	TTC	485	Z97336 @131719	e-141	IV	ATFCAL	Promoter TEL1 homolog	
63	C10-4	GAG	279	AC004667 @11773	e-113	II	T4C15	T4C15.6 DNA binding protein	2 e-24

*continued*

Table 1. (continued)

ITS	Plant	TSD <sup>a</sup>	Size <sup>b</sup>	Nucleotide Similarity <sup>c</sup>	E <sup>d</sup>	Ch <sup>e</sup>	Clone	Predicted Gene/Protein Similarity	E <sup>e</sup>
64	D10-4	CAT	569	AF085279 @52380	0.00	II	T7M7	ABA regulated cluster intergenic	
65	D10-4	GCT	100	AC003114 @50389	3 e-51	I	T12M4	ABI1 homolog downstream	
66	C11-1	GCA	335	Z99707 @7794	e-167	IV	ATAP21	Cytochrome P450 oxidase	e-42
67	C11-1	GTT	436	181K21T7	e-15				
68	D11-1	AAT	664						
69	D11-1	ATA	689	ATU53501 @6309	5.5 e-72	I	G8261	Zinc finger protein promoter	
70	D11-1	CTT	352	AC007591 @107928	e-107	I	F9L1	3' of glyoxalase gene	
71	C12-3	TGT	502	AC002343 @50981	0.00	IV	T19F6	Predicted exon	
72	C12-3	GAT	468	173C20T7	e-119			Rubber latex abundant protein	2 e-4
73	C12-3	ATG	152	AF001308 @71322	e-31	IV	T10M13	T10M13.14 glycosyl transferase	9 e-9
74	C12-3	GGT	446						
75	C12-3	CAT	436	AC000103 @23613	0.00	I	F21J9	F21J9.6 AP2 domain protein	4 e-75
76	C12-3		395						
77	D12-2	TAT	202	AC000098 @59616	6 e-66	I	YUP8H12R	Meristematic homeobox ATML1	3 e-6
78	C13-3	AGT	105	ATF17A8 @84729	e-25	IV	F17A8	3' of myc-type F17A8.170	
79	C13-3	AGA	140	AB010693 @35046	5 e-34	V	K21C13	RPS2 homolog	
80	C13-3	ATA	493						
81	C13-3	TTT	204	Z25970/ATTS1247	4.9 e-7			UDP-glucose dehydrogenase	2 e-12
82	D13-3	CAT	406	AC006282 @68097	e-171	II	F13K3	Glucosyl transferase	
83	D13-3	AGT	174	AF108666 @634	e-6			AAD04627 CYT1 protein	2.5 e-9
84	D13-3	CAC	566	AC007504 @37746	e-4	I	F13F21		
85	D13-3	AAA	229						
86	C14-5	TTG	352	ATU11764 @358	3 e-59			Tomato GAST1 (GA inducible)	
87	C14-5	AGG	250	AL022605 @52536	e-125	IV	T19P19	Trehalose-6-phosphate phosphatase	2 e-7
88	C14-5	AAA	257	AL049640 @77345	9 e-75	IV	T20K18	Gamma interferon-induced IP30	
89	C14-5		211	ATU90439	e-111	II	T6D20	Actin isolog U90439	2 e-32
90	D14-5	AAC	227	B23048 @132	4.8 e-17	I	F24P8		
91	D14-5	TTT	268	AB015468 @56417	3 e-7	V	K15N18		
92	C15-7	TAT	505						
93	D15-7	TTG	233	AL035708 @63045	4 e-89	IV	T5J17	Cytochrome P450 like	
94	D15-7	AGA	108	AL049171 @70055	3 e-46	IV	T25K17		
95	D15-7	TAT	153						
96	C16-2	TAA	223	AC007369 @88968	7 e-89	I	F9H16	Promoter phosphate transporter	
97	H12-2	CTA	177	AB006700 @81572	8 e-46	V	MHF15	RAP2.6/VP1 homolog	e-8
98	H12-2	AAA	440	AA09252	5 e-58			e1351774 predicted <i>C. elegans</i>	8 e-4
99	H12-2	AAT	758	AC003028 @81333	0.0	II	F16M14	SRGI 2 <sup>nd</sup> intron /ACC oxidase	2 e-52
100	R9	TTA	236	AC003671 @27965	8 e-19	I	F1707	F1707.8 predicted protein	2 e-8
101	R9	TGC	300	AF085279 @27136	e-83	II	T7M7	Promoter protein phosphatase 2C	
102	R14	CTA	466	AB024029 @34359	4 e-10	V	K21L19	N resistance gene homolog	2 e-33
103	R14		215						
104	S61A	CAA	161	AB016819 @791	2 e-21			UDP-glucose glycosyl transferase	
105	S61A	TGA	126	ATT5T17 @89007		IV	T5T17	Serine threonine kinase intron 8	
106	S61A	AAT	108						
107	S61A	AAT	37	ATF4B14 @32566	3 e-10	IV	F4B14	Prolyl 4-hydroxylase alpha (II)	
108	S61A	AAT	16						
109	S61A	AAG	95						
110	S61A	ATT	98	AL035440.2 @58591	3 e-41	IV	F10M23	Promoter NPK1 related kinase	
111	S61B	CAT	597						
112	S61C	CGA	596	AC006223 @28631	3 e-15	II	F22D22		
113	S61D	TTT	150						
114	S61E	TAG	595	AB023035 @9783	0.00	V	K9H21		
115	S61G	TTT	290	X16077 @1636	6 e-70			ATRRN18/18s rRNA	
116	S61G	TTG	205	AC003287 @7353	2 e-60	I	F20D20		
117	S61H	TCG	415						

<sup>a</sup> Target site duplications. The average trinucleotide TSD contains 73% (A + T).

<sup>b</sup> Size of DNA sequence that flanks the / insertion in base pairs.

<sup>c</sup> Nucleotide similarity to the accession number reported by a BLAST search; @ is the exact nucleotide position of / insertion.

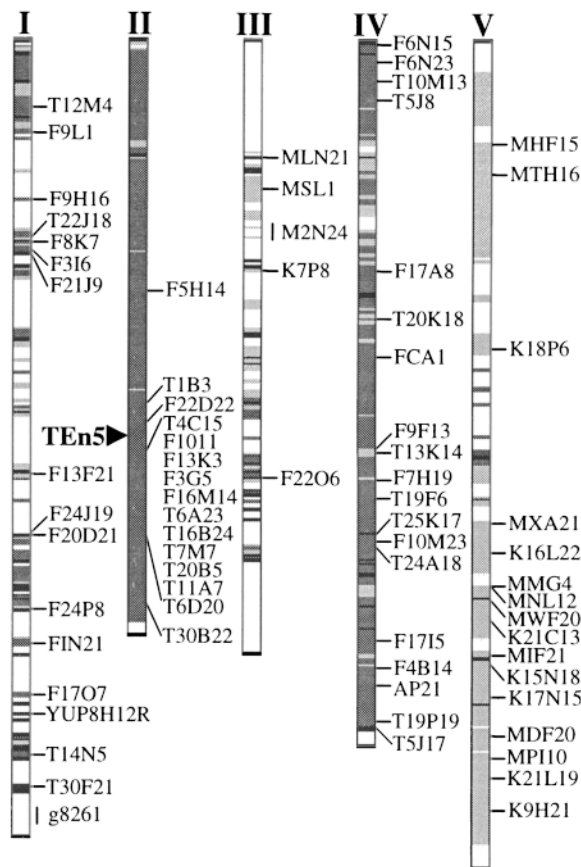
<sup>d</sup> Expect-value in BLAST search reporting significance of similarity.

<sup>e</sup> Ch is the chromosome position determined by sequence identity to a sequenced BAC.

albumin, metalloproteinase, ACC oxidase, and transferases. These results demonstrate that such an ITS database, generated from transposase-free plants, can be directly used for mutational analyses.

### Three-Dimensional Pooling and PCR Strategies

To identify insertions, we divided a multiple transposon population of 2592 *S<sub>6</sub>* lines into pools to minimize the amount of PCR. In our pooling strategy, outlined in Figure 3, the population of 2592 plants was divided into three sets of nine blocks, each block containing 96 plants. From the main inflorescence of each plant, two to four flower buds were harvested and appropriately documented with regard to block,



**Figure 2.** Chromosomal Positions of *I* Insertions Represented by ITSs.

The ITS positions were either determined by comparison to sequences compiled by the Arabidopsis Genome Initiative or mapped by BAC fingerprinting. The T-En5 locus is the original T-DNA source of the *I* inserts transposed into the genome. The chromosomes were imaged from a recent AtDB sequence map overview; the gray/black tones represent the sequenced portions of the chromosomes.

column, and row. A total of 87 “intersecting” DNA pools, 29 from each set, were obtained after DNA extraction of floral material. As seen in Figure 3, a pool of one block represents 96 plants, a pool from one row represents 108 plants, and a pool from one column represents 72 plants. Two *I* element primers, Itir2 and Itir3, complementary to the left and right terminal inverted repeat ends of the *I* element, respectively, were used in combination with a gene-specific primer for PCR.

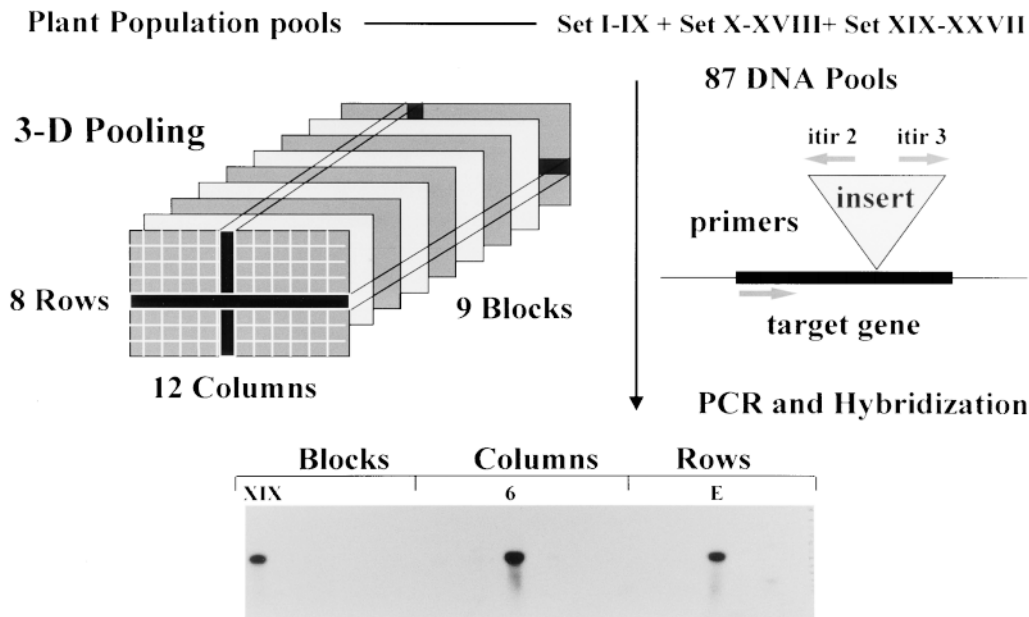
To standardize parameters for screening the population and to test the screening strategy for its use in reverse genetics, we used the genes *AP1* and *STM*, with well-characterized mutant phenotypes, to identify *ap1::I* and *stm::I* insertion alleles. Phenotypes, possibly related to insertions in *AP1* and *STM*, had been observed in the population at addresses V-11-G and XII-5-B, respectively. It therefore was expected, provided that these phenotypes were caused by *I* element insertions, that the three-dimensional pooling and PCR strategies would successfully identify the mutant plants.

### Identification and Analysis of an *stm::I* Mutant Allele

To identify insertions within the *STM* gene, we used two gene-specific primers, designated 1074 and 1075, complementary to the 5′ and 3′ ends of *STM*, in two separate sets of PCRs. Fragments of amplification products were visible in agarose gels from both screens and were confirmed by DNA gel blot hybridization to a gene-specific probe as shown in Figure 4A. In the second of the three sets of 864 (2592/3) plants, pools from block XII, column 5, and row B each displayed a very intense hybridization signal of identical size; both gene-specific primers, moreover, indicated a common “address” for the *stm::I* mutant, thereby indicating fidelity of the reaction. Offspring of plant XII-5-B were grown on Murashige and Skoog medium in a growth chamber, and the DNA from a pool of seedlings was extracted, which proved upon PCR analysis to contain the *stm::I* allele (Figure 4B). PCR products obtained with each gene-specific and both transposon primers were cloned and sequenced, revealing that the *I* element insertion is present in the second exon of *STM*, 24 nucleotides from the border between the second exon and the second intron, as displayed in Figure 5A. The insertion had generated a 3-bp TSD, characteristic for the *En/Spm* transposon family (Aarts et al., 1995a).

### Identification and Analysis of *ap1::I* Mutant Alleles

For the identification of insertions in the *AP1* gene, two gene-specific primers again were used for PCR analysis, as schematized in Figure 5B. Two individuals tested positive for the *ap1::I* allele. From set 3, positive hybridization signals for *AP1* arose from block XIX, column 6, and row E (shown in Figure 3). Similar results were found for individual V-11-G of set 1 (data not shown). The progeny produced upon self-



**Figure 3.** Schematic Representation of Three-Dimensional Pooling and PCR Strategies.

A population of 2592 *I*-containing Arabidopsis lines was divided into three sets, each containing nine blocks: set 1 (blocks I to IX), set 2 (blocks X to XVIII), and set 3 (blocks XIX to XXVII). Each block contained 96 plants (eight rows and 12 columns). Floral material was harvested from plants as needed and similarly arranged in sets according to the same three dimensions as shown. Plants or floral material was pooled in each of the three dimensions according to set: first by block, next by the row, and finally by column. This resulted in 29 DNA pools per set (12 columns + eight rows + nine blocks) and 87 DNA pools in total. For the identification of an *I* insertion in a gene of interest, one gene-specific and two *I*-specific primers were used in 87 PCRs, as described in the text. PCR products then were resolved by gel electrophoresis, blotted, and hybridized with a gene-specific probe. The autoradiogram shows the result of a PCR with an *AP1* gene-specific primer (1078) by using the 29 DNA pools of set 3. PCR products visible in the autoradiogram specify an insert common to the pools defined by block XIX, column 6, and row E. The physical intersection of these three "dimensions" (i.e., plant XIX-6-E) identifies a plant containing an insert in the *AP1* gene. See the text for details.

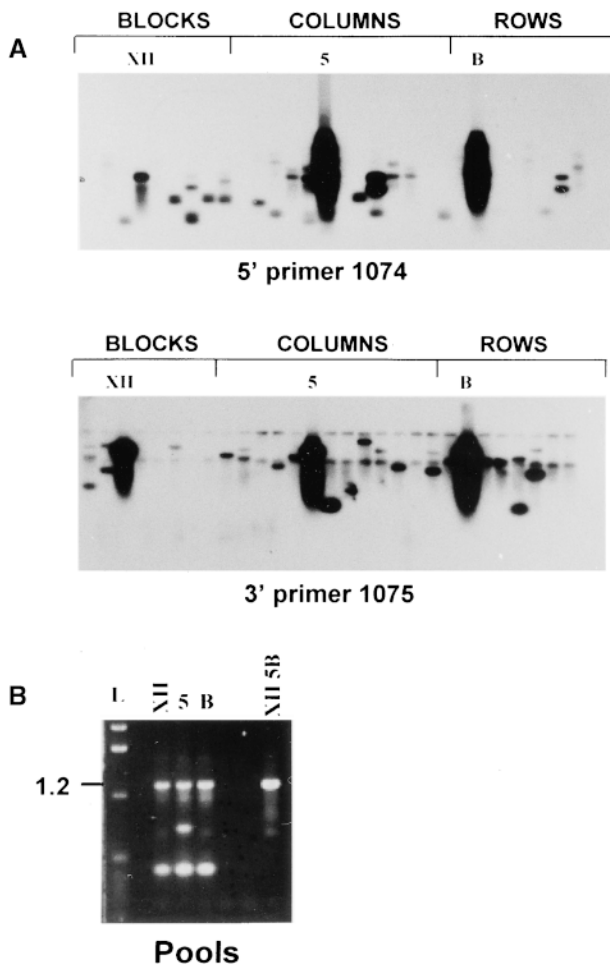
ing of V-11-G and XIX-6-E segregated for wild-type and *ap1* mutant plants with phenotypes as described by Mandel et al. (1992). A genomic DNA gel blot, displayed in Figure 6, showed that each plant contained a high degree of somatic reversions. Progeny of V-11-G expressing a mutant phenotype (Figure 6A, lanes 1, 3, 4, 6, 7, 9, 11, and 12) carried the *ap1::I(V-11-G)* allele at a frequency of 65%, whereas wild-type plants (Figure 6A, lanes 2, 5, 8, and 10) carried the *AP1* allele at a frequency of 70% (see Methods for estimation of frequencies). Among the progeny of individual XIX-6-E, mutant plants (Figure 6B, lanes 1, 2, 7, and 8) were homozygous for the *ap1::I(XIX-6-E)* insertion allele and displayed an average somatic reversion frequency of 12%, which is relatively low compared with that of the progeny of V-11-G.

The insertion sites in the *AP1* alleles (Figure 6B) were characterized by cloning and sequencing of PCR products, revealing that the insertion in *ap1::I(V-11-G)* was present in the seventh exon, seven nucleotides from the exon/intron boundary of exon 7 and intron 6, whereas the insertion in allele *ap1::I(V-11-G)* was present in the first intron, 17 nucle-

otides from the second exon. The *ap1::I(XIX-6-E)* allele, containing the *I* element in the intron, displayed a mutant phenotype. The *I* insert is oriented (considering coordinates of the *En/Spm* element transcription) in the same direction as that of the *AP1* gene transcription. To test if the mutant phenotype were transposase dependent (suppressor function; McClintock, 1954), we outcrossed the mutant allele to the wild type and screened the resultant  $F_2$  for mutant segregation. The recovery of the predictable 25% *ap1* mutants, including transposase free mutants (data not shown), confirms that this stable insertion in the intron displays a mutant phenotype.

#### Reverse Genetics of *FAE1* Homologs

To identify genes involved in epicuticular wax biosynthesis through our reverse genetics system, we identified genes in databases that had a potential role in very-long-chain lipid biosynthesis. The EST database was searched for homologs



**Figure 4.** Identification of an *I* Element Insertion into *STM* by Three-Dimensional Screening.

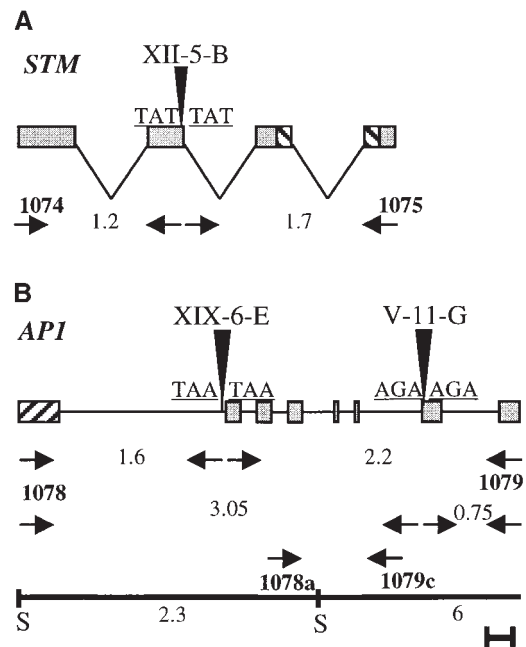
(A) Results of a three-dimensional screening with two gene-specific primers (1074 and 1075) for *STM*.

(B) Agarose gel displaying PCR products of DNA pools amplified with primer 1074. Lane L contains a DNA size marker (kilobase ladder) with the 1.2-kb band amplified from population DNA pools XII, 5, and B, as shown in (A). The amplification products from the progeny pool are in lane XII 5B.

of the *FAE1* gene (James et al., 1995) that has been shown to be involved in the production of C20 and C22 very-long-chain fatty acids in the seed. A great many partial EST clones with significant homology were found that had been sequenced completely, and a set of five independent homologs was chosen for further analysis. Gene-specific primers were used for PCR screens, and putative inserts were obtained for three genes that were monitored in the individual progeny. The insert in the gene homologous to the EST clone sequence ATTS3252 produced progeny with a striking *fdh* phenotype, characterized by epidermis fusion in the

leaves and inflorescences that tend to remain closed. The insertion site was sequenced and shown to be 950 bp from the primer, as shown in Figure 7A, in the putative second exon of the gene. Cosegregation analysis of 60  $F_2$  progeny showed that the *fdh* phenotype, as well as the presence of the PCR product, was linked to the transposase locus on chromosome II, thereby suggesting that it was the previously isolated *fdh* mutant (Lolle et al., 1992).

Genomic DNA gel blot analysis (Figure 7B) of the *fdh* mutant progeny (lanes 6 to 11) revealed a 5.5-kb insertion allele; a band representing the 3.3-kb excision allele also was visible at a scanned density that was 30% of that of the 5.5-kb band (see Methods). A plant showing a revertant phenotype (lane 12) displayed a majority (65%) of the 3.3-kb wild-type



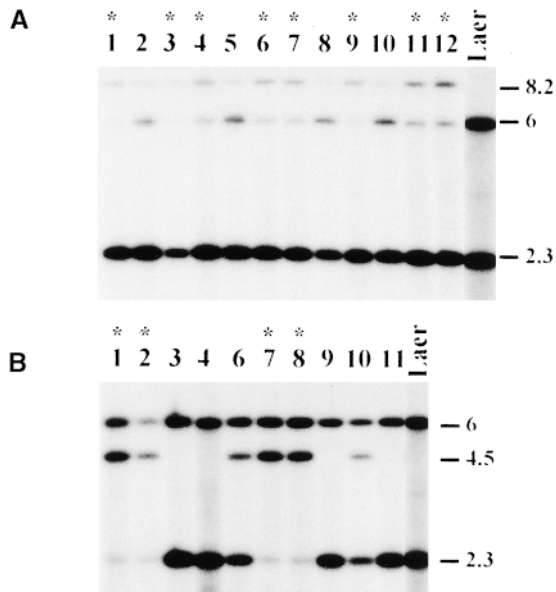
**Figure 5.** Schematic Representations of the *I*-Tagged Alleles of *STM* and *AP1*.

Exons are represented by gray boxes, the homeodomain and the MADS domain by hatched bars, and introns by a single line. The primers used for screening are indicated by boldface numbers. The large arrowheads indicate the position of the *I* insertions, with the TSD shown by underlined nucleotides. Horizontal arrows show primer positions for PCR fragments, with sizes entered in kilobases.

(A) Position of an *I* element insertion in the *stm::I* allele. Introns are not drawn to scale because their size is not known.

(B) Position of both *I* element insertions in the *ap1::I* alleles. Sequence data of the *AP1* gene were used to draw both exons and introns to scale (M. Yanofsky, personal communication). The position of *SacI* (S) restriction sites in *AP1*, with sizes (in kb) of both *SacI* fragments, is indicated. Primers 1078a and 1079c were used to generate a gene-specific probe in genomic DNA gel blot analysis. The bar in the bottom right-hand corner represents 200 nucleotides.





**Figure 6.** Analysis of *I* Element Insertion Alleles of *AP1*.

DNA gel blot of *SacI*-digested DNA of the offspring of *ap1* mutant alleles identified by reverse genetics screens. Plants with an *ap1* phenotype are indicated by an asterisk. The difference in intensity between mutant and wild-type bands in mutant plants was used to establish the frequency of somatic reversions in these lines. The bandwidths are indicated to the right in kilobases. Lane Laer contains wild-type Landsberg *erecta* DNA.

**(A)** In *ap1::(V-11-G)* progeny, the wild-type *AP1* allele is represented by the 6-kb band, whereas the mutant *ap1::(V-11-G)* allele migrates as an 8.2-kb band.

**(B)** In *ap1::(XIX-6-E)* progeny, the *AP1* allele is represented by the 2.3-kb band, and the mutant allele by the 4.5-kb band.

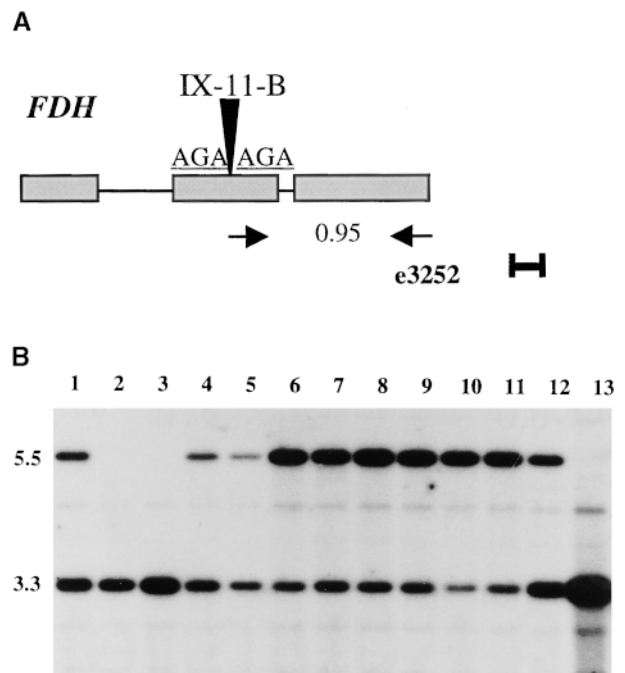
allele. On average, ~10% germinal revertants were obtained that had lost a *fdh::I* allele. In the phenotypically wild-type progeny (lanes 1 to 5) of a revertant, a majority in the wild-type 3.3-kb fragment (65% in lanes 1, 4, and 5) is evident. This band shift in revertant progeny, along with the other DNA gel blot data, confirms the correlation between the *fdh* mutant phenotype and the presence of the insert in the *FAE1* homolog. Recently, the complete sequence of the *FDH* gene (GenBank accession number AJ010713) was deposited in the database and offers corroboratory evidence that our insertional mutant indeed occurs in the *FDH* gene (Yephremov et al., 1999).

**DISCUSSION**

Insertional mutagenesis has proven to be a powerful tool for generating knockout mutations in Arabidopsis. Such muta-

tions often give rise to phenotypes that provide clues to the function of the gene in the organism. With the advent of large-scale genome sequencing, the number of putative genes has increased greatly, and there is now a prevalent interest to verify their functions. In a systematic gene function search strategy, insertional mutants that display a phenotype give a straightforward way to describe the potential function of a gene.

In endogenous transposon systems, such as those that exist in *Drosophila*, *petunia*, and *maize*, multiple transposons are present per individual, and so genome saturation with insertions can be attained with a relatively small number of individuals. In contrast, heterologous T-DNA or transposon insertions into Arabidopsis usually must begin with a few transformed copies per individual. The *En-1* transposon system nevertheless has proven capable of accumulating in copy number in Arabidopsis by virtue of its high



**Figure 7.** Reverse Genetics Analysis of an *FAE1* Homolog.

**(A)** Diagram showing the *FDH* gene (drawn from the sequence of GenBank accession number AJ010713) with position of the primer e3252 used for screening, the *I* insert in the second exon, and the TSD (underlined). The large arrowhead indicates the *I* insertion position; horizontal primers show PCR fragment boundaries, with size in kilobases. Bar = 200 nucleotides.

**(B)** Cosegregation analysis of an *I* element insertion with the *fdh* phenotype. HindIII-digested DNA of wild-type progeny of a revertant (lanes 1 to 5), *fdh* mutant progeny (lanes 6 to 11), a putative germinal revertant (lane 12), and wild-type *Ler* (lane 13), hybridized with EST clone ATTS3252.

independent transposition frequency (Aarts et al., 1995a). Here, we have exploited this characteristic to generate a population of plants containing multiple (20 to 25) independent *I* transposons per line, from which a workable strategy for genome saturation in a relatively small number of plants was developed. Our approach to genome saturation, moreover, has yielded a verifiable means for reverse genetics strategies. Subpools of our population will be available through the Nottingham Arabidopsis Stock Centre (Nottingham, UK) for production of such insertion libraries.

To assess the randomness of the *I* element distribution in the Arabidopsis population that we generated, we compiled an ITS database of the original 17  $S_1$  lines. A total of 99 independent insertions were identified in transposase-free  $S_1$  segregants from which our population descended, originating from seven transposon-bearing lines, each of which was estimated to contain seven to 10 inserts. Thus, in three generations of crossing and selfing, the original 49 to 70 inserts had accumulated to 99 independent inserts, which translates into an overall transposition frequency of 40 to 100% (or between 13 and 33% per generation), in good agreement with previous observations (Aarts et al., 1995a). The independence of our observed transpositional events is evident by the  $S_6$  generation, in which ITSs were distributed over all chromosomes, suggesting that further transposition would likely access most of the genome. It is nevertheless clear that transposing *I* elements can show strong biases for inserting in specific chromosomal regions, as evidenced by the accumulation of 10% of observed ITSs in a 10-cM interval to one side of the T-En5 locus. It should be pointed out that such biases could prove advantageous in certain studies of chromosomal regions.

An apparent preference for insertion into A-T-rich regions also was observed. Specifically, 73% of the nucleotides that comprise the compiled TSDs were either A or T, and not a single one of the 117 ITSs created a TSD consisting solely of G and C nucleotides. Interestingly, analysis of a 1.9-Mb sequence of the Arabidopsis genome showed that the A+T content of introns and intergenic regions is 66 to 67%, whereas that of exons is 56% (Bevan et al., 1998). Additionally, ~25% of the Arabidopsis genomic DNA represented by a number of sequenced BAC clones consists of exclusively A+T-containing trinucleotide sequences, but only ~5% of the cloned sequences similarly manifest G+C-containing trinucleotide sequences (A. Pereira, unpublished data). Thus, the transposon TSDs may be only moderately biased for A-T-rich sites. Although such an extensive sequence analysis of random *En-I* target sites has not been described previously in any organism—despite much speculative pondering in the literature—the preference for insertion into A-T-rich sequences has been noted in the more limited number of *En-I* (*Spm-dSpm*) inserts characterized in maize (Klösigen et al., 1986; Schwarz-Sommer et al., 1987; Menssen et al., 1990) and tobacco (Pereira and Saedler, 1989). Whether this is a specific characteristic of the *En-I* system or a more general phenomenon remains to be seen.

Our ITS database was compared with public databases, and many ITSs were found to correspond to deposited gene sequences (see Table 1). Many of these genes already have been found to be essential to Arabidopsis, and so insertion alleles, such as those apparently generated in our present work, deserve further study. An especially intriguing dimension to mutant analysis that is engendered by the insertional mutants created here is the possibility for revertant analysis potentially made possible by use of transposase-containing lines. It is in any event clear that the generation of an ITS database, such as that compiled here, represents a powerful novel method to create inserts in a variety of genes so that their essentiality and function may be assessed.

To demonstrate the applicability of our mutagenic regime to specific genes of interest, we searched for the *ap1::I* and *stm::I* mutations in our population. The three-dimensional pooling strategy that we used to find these mutations within our population of 2592 multiple-transposon lines was described and successfully used previously for petunia (Koes et al., 1995), maize (Das and Martienssen, 1995), and Caenorhabditis (Zwaal et al., 1993). We successfully identified two independent *ap1::I* alleles, *ap1::I(XIX-6-E)* with an insertion in an intron and an average excision rate of 12%, and *ap1::I(V-11-G)* with an insertion in the seventh exon and an average excision rate of 35%. Such frequencies of excision have been seen before for different insertion sites with the *En-I* system conferred by the T-En5 transposase locus but not heretofore documented for two insertion sites of the same gene. Either the position of the insert in the gene or interactions of the T-En5 transposase with other genetic factors may have determined these excision frequencies.

We also recovered a mutant that bears an insertion in an *FAE1* homolog and displays a *fdh* phenotype, thereby demonstrating a role for long-chain lipid synthesis in epidermis development. This finding directly exemplifies the use of our system for identifying inserts in members of multigene families. Interestingly, an *FAE1* homolog recently has been ascribed a function in cuticular wax biosynthesis by sense suppression-mediated reverse genetics (Millar et al., 1999).

Part of our rationale in choosing to generate a population of lines containing multiple *I* elements per plant was to decrease the number of lines necessary for PCR analysis relative to mutagenic regimes that depend on T-DNA. Implementation of this choice clearly did not interfere in the identification of mutations carrying an insertion in a gene of interest. While our population was being generated over several generations, PCR screens were performed with ~150 genes by collaborators (data not shown). Insertions were found for ~100 genes and shown to be heritable in 50% of the initial positives. These insertions currently are being used in functional analyses in several laboratories worldwide.

It also should be noted that our multiple-transposon population, manifesting a number of phenotypes among the 2592 lines, also can be subjected to traditional forward genetic screens. Definite proof for a correlation between a par-

ticular phenotype and an insertion in a given gene can be obtained by cosegregation analysis in transposase-free plants, the identification of revertant sectors, and independent insertion alleles with similar phenotypes. These procedures help distinguish insertionally tagged mutants from the background of "footprint" mutants that inevitably arise in transposon-mediated mutagenesis.

In conclusion, our population of multiple-transposon Arabidopsis lines can be used to identify insertions in any gene or DNA sequence and is a valuable addition to the available collection of T-DNA insertions (Koncz et al., 1992; Krysan et al., 1996; Winkler et al., 1998) and other transposon-based populations of Arabidopsis (Baumann et al., 1998; Wisman et al., 1998). Gene insertion searches now can be initiated in our population, as described herein, for gene sequences of interest that are now available. The availability of the whole genome sequence will further require the systematic generation of multiple mutations, including knockout mutations. The creation of an ITS database as described here represents a cost-effective asset to whole-genome functional analysis, so that every isolated insertion can be systematically incorporated into collaborative schemes as our knowledge of the genes and the organism grows.

## METHODS

### Generation of Lines Containing Multiple *I* Elements

The two-element (in *cis*) *Enhancer-Inhibitor* (*En-I*) transposon system was introduced into *Arabidopsis thaliana* ecotype Landsberg *erecta* (*Ler*) as described by Aarts et al. (1995a). Specifically, a T-DNA construct was used that contained an unmarked 2.2-kb *I* element within the open reading frame of a *neomycin phosphotransferase II* (*NPTII*) marker gene; in this way, excision could be monitored phenotypically. The same construct also harbored an *En* element (i.e., a stable source of transposase) from which the 5' and 3' ends had been replaced by the cauliflower mosaic virus (CaMV) 35S promoter and terminator sequences, respectively. A primary transformant displaying active transposition subsequently was used as the source of transposon in generating mutants. The primary transformant contained two T-DNA loci (*T-En*) harboring the transposase: *T-En5* on chromosome II (bacterial artificial chromosome [BAC] T1B8; see Figure 2) mediating *I* transposition at a rate of 10 to 30% and *T-En2* on chromosome I (near *CER1*) conferring *I* transposition at a rate of 2 to 5%. The primary transformant was selfed over several generations, during which multiple transpositions of the original *I* elements occurred (Aarts et al., 1995a). Seven of the resulting lines, each containing seven to 10 independent *I* insertions, were selected and intercrossed to increase the element copy number additively (Figure 1). Twenty-three resultant single crosses were double crossed in 40 nonparental combinations, and their progeny subsequently were selfed. From a total of 600 selfed  $S_1$  progeny sown, 216 *T-En5* homozygotes were selected based on DNA gel blots with a *T-En5* flanking probe. From these 216 lines, 12 plants were sown per line to produce a primary population of 2592  $S_2$  plants that were harvested individually and used to generate the multiple-transposon

population. The individually harvested 2592 lines were propagated by a single-seed descent procedure for four generations of selfing ( $S_3$  to  $S_6$ ). The resulting populations from the 2592 individual plants were planted in 27 trays/blocks, and young inflorescence material was harvested in a three-dimensional pooling strategy (Figure 3). From each plant, two to four young flower buds were harvested at intervals of a few days from the bottom, middle, and top of the inflorescence for the three-dimensional pools to represent the gametes for the following generation, which resulted from selfed seed of the inflorescence.

### Isolation of Genomic DNA

Single-plant DNA isolations (mini-preps) were performed as described previously (Pereira and Aarts, 1998). The larger pools of inflorescence material were ground with a mortar and pestle in liquid nitrogen and 2.5 mL of DNA extraction buffer (0.3 M NaCl, 50 mM Tris-HCl, pH 7.5, 20 mM EDTA, 2% Sarkosyl [v/v], 0.5% SDS [w/v], 5% phenol [v/v], and 5%  $H_2O$  [v/v]) in a scale-up of the described procedure. After RNase treatment and a second phenol/chloroform extraction, the DNA was pelleted and dissolved in 200 to 400  $\mu$ L TE (10 mM Tris-HCl and 1 mM EDTA, pH 8.0) to a final concentration of 50 to 100 ng/ $\mu$ L DNA.

### Generation of *Inhibitor*-Tagged Sites

$S_1$  plants containing multiple *I* inserts were selected by DNA gel blot analysis for the absence of the *T-En5* transposase locus. The progeny of a set of 17 such  $S_1$  plants (siblings labeled C, D, and H in Table 1), representing the original lines used for generating the multiple-transposon population, were used for the primary analysis of insertion sites. To characterize the insertion sites of advanced-generation progeny, we selected a set of genotypes free of the *T-En5* transposase locus as described below. DNA from these plants was digested with the enzyme *HinfI* and used for inverse polymerase chain reactions (PCRs), as described by Pereira and Aarts (1998). The PCR products were digested with *BglIII* (restriction site present in the primers), cloned into the *BamHI* site of pBluescript SK+, and sequenced via an automated sequencing system (model 373 DNA Sequencer; Applied Biosystems, Foster City, CA). The DNA sequences were cropped to represent transposon flanking DNA, entered into an *Inhibitor*-tagged site (ITS) database, and analyzed with the BLAST algorithm at the National Center for Biological Information. These sequences were entered into GenBank (accession numbers AF184768 to AF184882).

### Development of Stable *I*-Containing Lines

To stabilize the lines containing multiple *I* elements by segregation of the transposase locus carrying a hygromycin resistance marker, we outcrossed the multiple-transposon lines to the wild-type *Ler* ecotype, and the resultant  $F_2$  populations were sown in the greenhouse. Three to 4 days after first leaf emergence,  $\sim 1$   $\mu$ L of a hygromycin solution (25 mg/mL in 0.001% Triton X-100) was added to one true leaf. At 5 to 7 days after application, sensitive seedlings were identified by a halo of necrotized tissue and tagged as hygromycin sensitive. From each  $F_2$  family, the inflorescence tissue was harvested and used for DNA isolation. The DNA was used in PCRs to confirm the transposase genotype. The primers used were the *T-DNA* (*En*)-specific

primer Ten: CACCAGTTGCTATCATTGCGGCC and the flanking (BAC T1B8) genomic DNA primers Ten3: GGACCACAAAACATGGAA-CATG and Ten5: CACAGCACATGTACACAACCTGGCG. The PCRs (20  $\mu$ L) contained 20 ng of genomic DNA, with 100 ng of each of the three primers, and 100  $\mu$ M of each nucleotide with 0.5 units of Taq DNA polymerase (Supertaq; HT Biotechnology Ltd., Cambridge, UK). The reactions were performed in a PTC-200 Peltier Thermal Cycler (MJ Research, Waltham, MA) by using one incubation at 94°C for 4 min followed by 35 cycles of 45 sec at 94°C, 45 sec at 60°C, and 2 min at 68°C, followed by a 7-min extension step at 72°C. A PCR product of 0.5 kb indicates the presence of the wild-type (T-En5 free) allele, and the presence of a 164-base fragment indicates the presence of the T-En5 locus. A sample of *I*-containing lines was used to generate the ITS database. From the S<sub>6</sub> generation, seven random plant genotypes (labeled S<sub>6</sub> in Table 1) were selected, as well as two additional plants (labeled R- in Table 1) from the S<sub>5</sub> generation produced as part of a mutant segregation analysis.

### Detection and Analysis of *I* Insertions

To detect *I* insertions in specific target genes, we used two *I*-specific primers: Itir2 (CTTTGACGTTTTCTGTAGTG) and Itir3 (CTTGCC-TTTTTCTGTAGTG; Life Technologies, Gaithersburg, MD), complementary to the 5' and 3' terminal inverted repeats of *I*. Both primers were used in one PCR together with one gene-specific primer. For the target genes *STM* and *AP1*, two sets of gene-specific primers were tested, complementary to either the 5' or 3' end of the coding region. Primers 1074 (5'-ATGGAGAGTGGTTCCAACAGC) and 1075 (3'-CTCCGACGGCTTCCAATGCCG) were used to test for *I* insertions in *STM*, and primers 1078 (5'-GGTTCATACCAAAGTCTGAGC) and 1079 (3'-TTGGATGAAAAGAGCCTAGCC) were used to test for insertions in *AP1*.

To screen for insertions in genes homologous to *FAE1*, we requested a number of expressed sequence tag (EST) clones with homologies to *FAE1* from the Arabidopsis Biological Resource Center (Ohio State University, Columbus) and completely sequenced them. Five *FAE1* homologous clones (ATTS1282, ATTS3218, ATTS3252, G7F6T7, and G11D6T7) were chosen for further analysis; primers complementary to their 3' ends were synthesized and used for reverse genetics PCR screens. The primer complementary to the EST sequence ATTS 3252, termed e3252: CCTGGTTGGCTTCTCACCTCCTC, was used in further analysis described here.

PCRs were performed in a total volume of 50  $\mu$ L (0.5  $\mu$ g of each primer, 250  $\mu$ M of each nucleotide, and 1  $\times$  Supertaq buffer), containing 50 to 100 ng of DNA and 1 unit of Taq DNA polymerase (Supertaq; HT Biotechnology Ltd.). The reactions were performed in a PTC-200 Peltier Thermal Cycler (MJ Research): one incubation at 94°C for 4 min, followed by 31 cycles of 45 sec at 94°C, 45 sec at 60°C, 3 min at 72°C, followed by a 7-min extension step at 72°C.

Amplification products were size separated on 1% TBE (45 mM Tris-borate and 1 mM EDTA) agarose gels containing ethidium bromide, alkali treated, then vacuum transferred onto Hybond-N<sup>+</sup> membranes and hybridized to a gene-specific probe. Primer combinations 1074 and 1075 for *STM* and 1078 and 1079 for *AP1* (Figure 5) were used to amplify genomic DNA products for use as probes. For the *FAE1* homolog gene screens, the cDNA fragments from the EST clones were used as probes. After identification of a positive three-dimensional insertion address, ~30 seeds of the positive plant were grown in vitro on 0.5  $\times$  Murashige and Skoog medium (Duchefa, Haarlem, The Netherlands) and harvested as a pool in an Eppendorf

tube when the primary leaves of the seedlings started to appear. After reconfirmation of an insertion after DNA extraction, PCR, and DNA gel blotting and hybridization, 12 to 24 seeds were sown again as individuals in the greenhouse.

Amplification products from pooled DNA as well as from DNA of individual plants, obtained with each gene-specific primer and both *I* primers, were cloned into the PGEM-T vector (Promega) and sequenced. Individuals were screened for visible phenotypes, and floral material was harvested for DNA extractions. The individual plants were analyzed further for insertion sites by using PCRs as described above and genomic DNA gel blot analysis.

### DNA Gel Blot Analysis

Genomic DNA (~0.5  $\mu$ g) of progeny to be analyzed with respect to *AP1*, *STM*, and *FDH* was digested with the restriction enzymes SacI, HindIII, and HindIII, respectively; these enzymes do not cut in the *I* element, and they produce a gene-specific fragment of reasonable size. The digested genomic DNA was resolved on a 0.8% agarose gel in 1  $\times$  TAE (40 mM Tris-acetate and 1 mM EDTA), blotted onto Hybond-N<sup>+</sup> membranes, and hybridized to a gene-specific probe. For *STM*, a genomic fragment PCR-amplified with primers 1074 and 1075 was used as probe. For *AP1*, a genomic fragment was amplified with primers 1078a (CACTGCTCTTAAGCACATCCGC) and 1079c (CCACTGCTCCTGTGAGCCCT) and used as a probe (Figure 5B). For the elongase homolog ATTS3252, the cDNA insert was used as a probe. After hybridization, filters were analyzed to establish the somatic reversion frequency in the *ap1::I* and *fdh::I* mutant alleles, by using a bioimager (Fujix Bas 2000; Fuji Film, Hasca, IL) and a scanning software program (TINA 2.0; Isotopenmessgeräte GmbH, Straubenhardt, Germany). The intensities of the *AP1* wild-type band and the mutant *ap1::I* band both were measured in Bkg/mm<sup>2</sup>. Somatic reversions subsequently were expressed as *AP1/I* (*AP1* + *ap1::I*)  $\times$  100%. Similar estimates were made for the *fdh* alleles.

### ACKNOWLEDGMENTS

We thank Raffaella Greco, Hailing Jin, Marten Denekamp, Dr. John Ward, Sonja Gazzarini, Dr. Anja Schneider, and Andreas Möller for their help in harvesting and DNA extractions of several populations. We also thank Dr. Mark Aarts for selecting the initial transposon lines, Marjo van Staveren for technical assistance, and Dr. Marty Yanofsky for providing the genomic sequence of *AP1*. The Arabidopsis Biological Resource Center is gratefully acknowledged for the EST clones, and the European Union Biotech Project AIM (Grant No. BIO4-CT95-0183) is thanked for financial support.

Received July 14, 1999; accepted August 24, 1999.

### REFERENCES

- Aarts, M.G.M., Dirkse, W.G., Stiekema, W.J., and Pereira, A. (1993). Transposon tagging of a male sterility gene in *Arabidopsis*. *Nature* **363**, 715–717.
- Aarts, M.G.M., Corzaan, P., Stiekema, W.J., and Pereira, A.

- (1995a). A two-element *Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **247**, 555–564.
- Aarts, M.G.M., Keijzer, C.J., Stiekema, W.J., and Pereira, A. (1995b). Molecular characterization of the *CER1* gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **7**, 2115–2127.
- 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.
- Bancroft, I., and Dean, C. (1993). Transposition pattern of the maize element *Ds* in *Arabidopsis thaliana*. *Genetics* **134**, 1221–1229.
- Bancroft, I., Jones, J.D.G., and Dean, C. (1993). Heterologous transposon tagging of the *DRL1* locus in *Arabidopsis*. *Plant Cell* **5**, 631–638.
- 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.
- Bevan, M., et al. (1998). Analysis of a 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.
- Camilleri, C., Lafleur, J., Macadré, C., Varoquaux, F., Parmentier, Y., Picard, G., Caboche, M., and Bouchez, D. (1998). A YAC contig map of *Arabidopsis thaliana* chromosome 3. *Plant J.* **14**, 633–642.
- Cardon, G.H., Frey, M., Saedler, H., and Gierl, A. (1993). Mobility of the maize transposable element *En/Spm* in *Arabidopsis thaliana*. *Plant J.* **3**, 773–784.
- C. elegans Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282**, 2012–2018.
- Choi, S., Creelman, R.A., Mullet, J.E., and Wing, R.A. (1995). Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*. *Plant Mol. Biol. Rep.* **13**, 124–128.
- Das, L., and Martienssen, R. (1995). Site-selected transposon mutagenesis at the *hcf106* locus in maize. *Plant Cell* **7**, 287–294.
- Endrizzi, K., Moussian, B., Haecker, A., Levin, J.Z., and Laux, T. (1996). The *SHOOT MERISTEMLESS* gene is required for maintenance of undifferentiated cells in *Arabidopsis* shoot and floral meristems and acts at a different regulatory level than the meristem genes *WUSCHEL* and *ZWILLE*. *Plant J.* **10**, 967–979.
- Feldmann, K.A. (1991). T-DNA insertion mutagenesis in *Arabidopsis*: Mutational spectrum. *Plant J.* **1**, 71–82.
- Goodman, H.M., Ecker, J.R., and Dean, C. (1995). The genome of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **92**, 10831–10835.
- Höfte, H., et al. (1993). An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* **4**, 1051–1061.
- James, D.W., Lim, E.J., Keller, J., Plooy, I., Ralston, E., and Dooner, H.K. (1995). Directed tagging of the *Arabidopsis* *FATTY ACID ELONGATION1 (FAE1)* gene with the maize transposon *Activator*. *Plant Cell* **7**, 309–319.
- Jones, D.A., Thomas, C.M., Hammond-Kosack, K.E., Balint-Kurti, P.J., and Jones, J.D.G. (1994). Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science* **266**, 789–793.
- Kaiser, K., and Goodwin, S.F. (1990). Site-selected transposon mutagenesis 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. 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.
- Klößgen, R.B., Gierl, A., Schwarz-Sommer, Z.s., and Saedler, H. (1986). Molecular analysis of the *waxy* locus of *Zea mays*. *Mol. Gen. Genet.* **203**, 237–244.
- 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.
- Krysan, P.J., Young, J.F., 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.
- Liu, Y.G., Mitsukawa, N., Vazquez-Tello, A., and Whittier, R.F. (1995). Generation of a high quality P1 library of *Arabidopsis thaliana* suitable for chromosome walking. *Plant J.* **7**, 351–358.
- Lolle, S.J., Cheung, A.Y., and Sussex, I.M. (1992). *Fiddlehead*: An *Arabidopsis* mutant constitutively expressing an organ fusion program that involves interactions between epidermal cells. *Dev. Biol.* **152**, 383–392.
- Long, J.A., Moan, E.I., Medford, J.I., and Barton, M.K. (1996). A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* **379**, 66–69.
- Mandel, M.A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**, 273–277.
- McClintock, B. (1954). Mutations in maize and chromosomal aberrations in *Neurospora*. *Carnegie Inst. Washington Yearbook* **53**, 254–260.
- McKinney, E.C., Ali, N., Traut, A., Feldmann, 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.
- Mena, M., Ambrose, B.A., Meeley, R.B., Briggs, S.P., Yanofsky, M.F., and Schmidt, R.J. (1996). Diversification of C-function activity in maize flower development. *Science* **274**, 1537–1540.
- Menssen, A., Höhmann, S., Martin, W., Schnable, P.S., Peterson, P.A., Saedler, H., and Gierl, A. (1990). The *En/Spm* transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a *dSpm* element in the *A2* gene. *EMBO J.* **9**, 3051–3057.

- Mewes, H.W., Albermann, K., Bähr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S.G., Pfeiffer, F., and Zollner, A. (1997). Overview of the yeast genome. *Nature* **387** (suppl.), 7–9.
- Millar, A.A., Clemens, S., Zachgo, S., Michael Giblin, A., Taylor, D.C., and Kunst, L. (1999). *CUT1*, an Arabidopsis gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell* **11**, 825–838.
- Mozo, T., Fischer, S., Meier-Ewert, S., Lehrach, H., and Altmann, T.S.O. (1998). Use of the IGF BAC library for physical mapping of the *Arabidopsis thaliana* genome. *Plant J.* **16**, 377–384.
- Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Somerville, C. (1994). Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones. *Plant Physiol.* **106**, 1241–1255.
- Okuley, J., Lightner, J., Feldmann, K., Yadav, N., Lark, E., and Browse, J. (1994). *Arabidopsis FAD2* gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. *Plant Cell* **6**, 147–158.
- Pereira, A., and Aarts, M.G.M. (1998). Transposon tagging with the *En-I* system. In *Arabidopsis Protocols*, J. Martinez-Zapater and J. Salinas, eds (Totowa, NJ: Humana Press), pp. 329–338.
- Pereira, A., and Saedler, H. (1989). Transpositional behavior of the maize *En/Spm* element in transgenic tobacco. *EMBO J.* **8**, 1315–1321.
- Peterson, P. (1970). The *En* mutable system in maize. *Theor. Appl. Genet.* **40**, 367–377.
- Rounsley, S.D., Glodek, A., Sutton, G., Adams, M.D., Somerville, C.R., Venter, J.C., and Kerlavage, A.R. (1996). The construction of Arabidopsis expressed sequence tag assemblies. *Plant Physiol.* **112**, 1177–1183.
- Rushforth, A.M., Saari, B., and Anderson, P. (1993). Site-selected insertion of the transposon *Tc1* into a *Caenorhabditis elegans* myosin light chain gene. *Mol. Cell. Biol.* **13**, 902–910.
- Schmidt, R., West, J., Cnops, G., Love, K., Balestrazzi, A., and Dean, C. (1996). Detailed description of four YAC contigs representing 17 Mb of chromosome 4 of *Arabidopsis thaliana* ecotype Columbia. *Plant J.* **9**, 755–765.
- Schmidt, R., Love, K., West, J., Lenehan, Z., and Dean, C. (1997). Description of 31 YAC contigs spanning the majority of *Arabidopsis thaliana* chromosome 5. *Plant J.* **11**, 563–573.
- Schwarz-Sommer, Z.s., Shepherd, N., Tacke, E., Gierl, A., Rohde, W., Leclercq, L., Mattes, M., Berndtgen, R., Peterson, P.A., and Saedler, H. (1987). Influence of transposable elements on the structure and function of the *A1* gene of *Zea mays*. *EMBO J.* **6**, 287–294.
- 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. (1998). The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis. *Plant Mol. Biol.* **37**, 989–999.
- Yephremov, A., Wisman, E., Huijser, P., Huijser, C., Wellesen, K., and Saedler, H. (1999). Characterization of the *FIDDLEHEAD* gene of Arabidopsis reveals a link between adhesion response and cell differentiation in the epidermis. *Plant Cell* **11**, in press.
- Zachgo, E.A., Wang, M.L., Dewdney, J., Bouchez, D., Camilleri, C., Belmonte, S., Huang, L., Dolau, M., and Goodman, H.M. (1996). A physical map of chromosome 2 of *Arabidopsis thaliana*. *Genome Res.* **6**, 19–25.
- Zwaal, R.R., Broeks, A., van Meurs, J., Groene, 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.