Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 12722–12727, November 1997
Plant Biology

Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach

(*Agrobacterium tumefaciens*/insertion mutagenesis/T-DNA)

ELENA BABIYCHUK*, MAYUREE FUANGTHONG*†, MARC VAN MONTAGU*‡, DIRK INZÉ§, AND SERGEI KUSHNIR*

*Laboratorium voor Genetica and §Laboratoire Associé de l'Institut National de la Recherche Agronomique (France), Departement Genetica, Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), Universiteit Gent, B-9000 Gent, Belgium

Contributed by Marc Van Montagu, July 7, 1997

ABSTRACT **Large quantities of DNA sequence information about plant genes are rapidly accumulating in public databases, but to progress from DNA sequence to biological function a mutant allele for each of the genes ideally should be available. Here we describe a gene trap construct that allowed us to disrupt transcribed genes with a high efficiency in *Arabidopsis thaliana*. In the T-DNA vector used, the expression of a bacterial reporter gene coding for neomycin phosphotransferase II (*nptII*) depends on the *in vivo* generation of a translation fusion upon the T-DNA integration into the *Arabidopsis* genome. Analysis of 20 selected transgenic lines showed that 12 lines are T-DNA insertion mutants. The disrupted genes analyzed encoded ribosomal proteins (three lines), aspartate tRNA synthase, DNA ligase, basic-domain leucine zipper DNA binding protein, ATP-binding cassette transporter, and five proteins of unknown function. Four tagged genes were new for *Arabidopsis*. The results presented here suggest that gene trapping, using *nptII* as a reporter gene, can be as high as 80% and opens novel perspectives for systematic gene tagging in *A. thaliana*.**

As in other organisms, two strategies from "phenotype to gene" and from "gene to phenotype" (reverse genetics) had been used for gene isolation and characterization in *Arabidopsis thaliana*, a model dicotyledonous plant. Several combined approaches, such as positional cloning (1), candidate gene approach (2), and insertional mutagenesis with either transposons (3) or T-DNA vectors (4) have been used successfully to isolate *Arabidopsis* genes identified by the phenotype of their mutant alleles. Yet, our knowledge of existing plant genes is accumulating more rapidly than our understanding of their actual function. This lag is largely because of the fact that the majority of plant genes characterized to date have been identified by using conventional library screenings, differential display techniques (5), and both the *Arabidopsis* cDNA and genome sequencing projects (1, 6, 7). Considering that the *Arabidopsis* genome may code for anything from 16,000 to 33,000 genes (5), and that 14,915 expressed sequence tags (ESTs) now have been reported in dbEST (1), the development of reverse genetics approaches is becoming increasingly important. Whereas gene replacement through homologous recombination is now a standard procedure in yeast and mice (8, 9), in plants, including *Arabidopsis*, the rates of homologous recombination in cell types accessible for experimentation are low (10–12). Until now, the most common approach to simulate loss-of-function alleles of plant genes has been through suppression of gene expression by antisense or sense RNA (cosuppression) (13, 14). Recent developments in the reverse genetics of plants are the sequence-based identification of T-DNA or transposon insertion mutations using the PCR-based screenings of large pools of plants (15–17). However, as shown for yeast, *Drosophila* and mammalian cells, another approach with high potential for reverse genetics is that of gene trapping (18–20). Experimental designs for the insertional mutagenesis in *Arabidopsis* by gene trapping constructs based on transposons (21) or T-DNA vectors (22, 23) have been reported, but no estimates of the actual frequencies of gene tagging among all of the events of reporter gene activation have been made, and very few genes have been cloned from such transgenic plants (24, 25). Here, we describe a T-DNA vector with a high efficiency of gene disruption in *Arabidopsis* as exemplified by the isolation of 12 tagged genes.

MATERIALS AND METHODS

Construction of the Gene-Trap Vector. All plasmid manipulations were carried out according to standard procedures (26). The fragment of the gene coding for an apurinic/apyrimidinic endonuclease (*arp*) (27) containing the first and second exons, was PCR-amplified using phosphorylated primers (5′-CTTGTTCCCACTCCAACTCCAAGC-3′; 5′-TCTGTTGGTCTTGAATCTTGA-3′). The generated fragment was treated with Klenow fragment of DNA polymerase I and ligated into the *Bam*HI site of pNPT to give pNARP4. The pNPT is a pBluescript KS (Stratagene) containing a neomycin phosphotransferase II gene (*nptII*) from a binary vector pGA472 (28). The *Sac*I–*Kpn*I fragment from pNARP4 was subcloned after treatment with T4 DNA polymerase into the *Sac*I-digested and T4 DNA polymerase I-treated pGB21, a plasmid containing an *Eco*RI genomic fragment of the *arp* gene. From the resulting plasmid pGB21npt, a *Hin*dIII fragment was joined to the 2.4-kb *arp* downstream sequences in p3GD13 at an *Eco*RI site. The resulting *nptII–arp* fusion was recloned as *Sma*I–*Sal*I fragment in-between the T-DNA borders of the pGSV6 (Plant Genetic Systems, Gent, Belgium) at a *Bam*HI site. The binary vector pNPTARP was introduced into *Agrobacterium tumefaciens* strain C58C1Rif^R(pGV2260) (29).

Plant Transformation. *Arabidopsis* root explants were transformed with *A. tumefaciens* as described (30).

5′/3′Rapid Amplification of cDNA Ends (RACE)–PCR Analysis. 5′RACE–PCR analysis was carried out essentially as described (31), with the minor modifications recommended in a 5′Amplifinder RACE kit for the design of the anchor

Abbreviations: EST, expressed sequence tag; ARP, apurinic/apyrimidinic endonuclease; NPTII, neomycin phosphotransferase II; RACE, rapid amplification of cDNA ends.

Data deposition: The sequences reported in this paper have been deposited in the EMBL database (accession nos. Z86093, Z86094, and Z86095).

†Present address: Laboratory of Biotechnology, Chulabhorn Research Institute, Bangkok 10210, Thailand.

‡To whom reprint requests should be addressed. e-mail mamon@gengenp.rug.ac.be.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

oligonucleotide (CLONTECH). Briefly, total RNA was extracted from leaves of transgenic plants (32). The first-strand cDNA was synthesized by using a SuperScript preamplification system (GIBCO/BRL) and an *nptII* gene-specific primer (5′-CCCTTCCCGCTTCAGTGACAAC-3′). After alkaline hydrolysis of RNA, the phosphorylated and 3′ end-blocked anchor oligonucleotide (5′P-TTCACTATCGATTCTG-GAACCTTCAGAGG-$NH_3$3′) was ligated to the first-strand cDNA with 1 unit of T4 RNA ligase (BioLabs, Beverly, MA) in 50 mM Tris·HCl, pH 8/10 mM $MgCl_2$/10 μg/ml of BSA/25% (wt/vol) polyethylene glycol 8000/1 mM hexamine cobalt chloride/20 μM ATP in a total volume of 10 μl at room temperature for 18 hr. The ligation reaction mixture was diluted with 30 μl of water, and 2-μl aliquots were used for PCR amplification. Standard PCR was done in 50-μl reaction volume on a temperature cycler PTC-200 (MJ Research, Watertown, MA). Thirty cycles at 94°C for 20 s, 62°C for 30 s, 72°C for 1 min plus 3 s for each successive cycle were routinely run. When the next reaction was run with a nested primer, the first PCR was diluted 40-fold in water, and 1-μl aliquots were used for the next run. *Taq* DNA polymerase and reaction buffer supplied by the manufacturer (Boehringer Mannheim) were used. The following primers were used for 5′RACE–PCR: anchor primer 5′-CTGGTTCGGCCCACCTCT-GAAGGTTCCAG-3′, *nptII* gene-specific primers 5′-GCTGCCTCGTCCTGCAGTTCATT-3′, and nested to it 5′-CCCCTGCGCTGACAGCCGGAACACG-3′. Amplified PCR fragments were subcloned into pBluescript KS (Stratagene) ddTTP-tailed at the *Eco*RV site (33). Inserts from selected plasmids were sequenced using the Sequenase II sequencing kit (United States Biochemicals).

The 3′RACE–PCR was done using a 5′/3′ RACE kit (Boehringer). The first-strand cDNA was synthesized on RNA extracted from wild-type *Arabidopsis* C24 plants using oligo(dT)-anchor primer 5′-GACCACGCGTATCGATGTCGAC-$(T)_{16}$V-3′ and buffers and polymerases provided with the kit. cDNAs were amplified by using PCR anchor primer 5′-GACCACGCGTATCGATGTCGAC-3′ and gene-specific primers whose sequences were designed on the basis of the 5′RACE–PCR product sequences. The following primers gave amplification of cDNAs: 5′-TCTCTCTTTCGCTGTCCGAT-TCC-3′ for line SK1-13, 5′-GTGAAACCTCCATTACCT-TCGTC-3′ for line SK1-N2, 5′-GGACGGTGGTATACTTG-GATCCC-3′ for line SK3-1, and 5′-AATCTGCTATTACG-GATC-TTTAATCGG-3′ for line SK2-3. Amplified cDNAs were sequenced, and the compiled sequence information has been submitted to the EMBL database under accession numbers Z86093, Z86094, and Z86095, respectively. The cDNA for line SK2-3 is not published, because it is not full length.

Complementation Analysis of the *dal1* Mutant. The *dal1* genomic sequences were obtained after screening of the *Arabidopsis* genomic library in a λ phage GEM11 (Promega). The genomic library was a kind gift of John T. Mulligan and Ronald W. Davis (Stanford University, CA). For DNA gel blot hybridization analysis, genomic *Arabidopsis* DNA was extracted from leaves as published (34) and additionally purified by centrifugation on a CsCl gradient in the presence of ethidium bromide at a concentration of 0.1 μg per ml. For the complementation analysis of the *dal1* mutant, the full-length cDNA was placed under the control of the cauliflower mosaic virus 35S promoter in a binary vector pGSCDH35S (S.K., unpublished data), which resulted in the binary vector pGSC501-3. The chimeric gene, which conferred hygromycin resistance in transformed plant cells, was present in the T-DNA of the pGSCDH35S in addition to the expression cassette. The root explants of the albino plants were cocultivated with strains of *A. tumefaciens* C58C1Rif^R(pGV2260) harboring either the pGSC501-3, or, as a control, the pGSCDH35S. The regeneration of the green hygromycin-resistant calli was observed only after root cocultivation with the *A. tumefaciens* C58C1Rif^R(pGV2260,pGSC501-3) strain.

RESULTS

To assess the efficiency of gene trapping in *Arabidopsis* either by homologous recombination or by random integration, we made a T-DNA vector, pNPTARP, which contains the bacterial *nptII* gene as a reporter gene (Fig. 1). The coding region of *nptII*, which lacks the first eight amino acids, is flanked by upstream and downstream genomic sequences of the *Arabidopsis arp* gene (27). The *nptII*-coding region was fused in-frame with the *arp* gene coding sequences in the second exon, but the *arp* frame was destroyed in the first exon just downstream from the translation initiation codon. Thus, the majority of T-DNA integration events into the *Arabidopsis* genome will not result in the synthesis of a functional NPTII enzyme, and transformed cells will not have the kanamycin-resistant phenotype. Yet, *nptII* expression may occur in some rare instances, such as T-DNA integration by homologous recombination with the *arp* gene, or by integration into the expressed regions of the *Arabidopsis* genome by classical mechanisms (35, 36), while accompanied by large (at least 570 bp) T-DNA deletions from the left border.

Arabidopsis cells were transformed with the pNPTARP binary vector by using standard protocols (30). Twenty-nine transgenic kanamycin-resistant calli were selected, of which 20 regenerated fertile plants. The kanamycin-resistance phenotype was well expressed at the seedling stage in the progeny of 17 lines, whereas all seedlings in the progeny from the remaining three lines appeared bleached and kanamycin-sensitive. The kanamycin resistance of these lines, however, only became apparent when leaf explants were placed on shoot-inducing medium supplemented with kanamycin.

To identify possible homologous recombination events, we performed DNA gel blot hybridization analysis of the *Sac*I–*Bam*HI-digested genomic DNA extracted from transgenic plants. A radioactively labeled fragment derived from the first intron and second exon of *arp* was used as a probe. In the case of T-DNA integration through homologous recombination, we expected to detect a *Sac*I–*Bam*HI fragment of 420 bp. No such

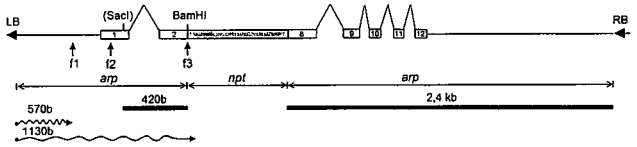

FIG. 1. A schematic drawing of the T-DNA from the pNPTARP binary vector. The left and right borders of the T-DNA are indicated as triangles. The arrowed lines below the T-DNA map indicate sequences derived from the *arp* gene or the *nptII* gene, which lacks the first eight amino acids. Boxed regions are exons. The *arp* gene exons are numbered and the *nptII* gene region is shaded. Appropriate frames are indicated as f1, f2, and f3. At the f3 site is an in-frame fusion of the *arp* second exon-coding frame with the coding region of *nptII*. At this position a *Bam*HI site was generated. The beginning of the *arp*-coding frame with the initiating translation ATG codon is indicated as f2, and a 24-bp downstream frame shift (indicated in brackets) was introduced by deleting a *Sac*I site. The frame shift results in an *arp–nptII* fusion, the frame of which begins at the f1 site, but lacks an ATG start codon. If the T-DNA integrates into the *Arabidopsis* genome by homologous recombination with the *arp* gene over 420-bp upstream and 2.4-kb downstream sequences from *nptII* (areas indicated by black bars below T-DNA), the *nptII* protein can be synthesized from the *arp* first ATG codon. In this case, the *Sac*I site will be derived from a genomic *arp* gene and detection of the 420-bp *Sac*I–*Bam*HI fragment on a DNA gel blots can be used to identify homologous recombination events. Alternatively, before stable integration into the *Arabidopsis* genome, the T-DNA may undergo deletions, from 570 bp to reach the f1 site and up to 1,130 bp to reach the *nptII*-coding region, as indicated by waved lines below the map. In this case the expression of *nptII* may occur as a result of T-DNA integration into *Arabidopsis*-expressed genomic regions, given that an *in vivo* fusion protein will be synthesized in transformed cells as result of either direct exon-exon integrations or after splicing of the generated chimeric intron.

Table 1. T-DNA insertion lines for which disrupted genes have been identified

Line	5′ RACE*	i/e†	EST‡	3′ RACE§	DNA blot¶	Homologous genes‖	Probability‖	Putative phenotype**
SK1-13	530	i	—	1,440	+	*Arabidopsis* DNA-binding protein POSF21 [Q40088]	9.5^{-56}	?
SK1-18	1035	e (≈1,130)	Z25619	NT	+	Human ORF [D38555]	2.6^{-67}	?
SK1-N2	179	e (≈570)	—	1,010	+	*Antirrhinum majus DAG* gene [X95753]	7.0^{-43}	Conditional albino
SK1-B2	220	i	—	NT	+	*Arabidopsis* DNA ligase [X97924]	2.5^{-32}	Lethal
SK2-1	466	i	N37338	NT	+	60S acidic ribosomal protein P2 [O41099]	2.0^{-27}	?
SK2-2	227	i	N97201	NT	+	?	—	?
SK2-3	202	e (≈724)	—	338	+	?	—	Low transmission through pollen
SK2-11	132	i	—	NT	NT	60S ribosomal protein S21E yeast [P05764]	2.9^{-5}	?
SK2-N2	152	i	R30337	NT	+	60S ribosomal protein L31 *Nicotiana glutosa* [P46290]	5.0^{-8}	Narrow leaves
SK3-1	246	i	—	1,020	+	Two domains, one peptidyl-prolyl *cis-trans* isomerase *E. coli* [P39159] and the *Synechocystis sp.* hypothetical protein [g1001676]	7.2^{-12} 2.3^{-30}	?
SK3-3	940	i	N96601	1,150	+	Mice ABC transporter-7 [U43892]	1.8^{-98}	Chlorosis
SK3-7	190	e (≈1,090)	F13974	900	NT	Aspartyl tRNA synthase	2.0^{-10}	?

*The lengths of the cDNA sequences, found in a 5′RACE–PCR fragments and not derived from a T-DNA vector are indicated in base pairs.
†Cases, in which a new sequence began exactly at a 3′ splice site of the first *arp* intron were classified as T-DNA integration in introns (i). If a novel sequence began at any other site of the first or second *arp* exons, 5′ noncoding regions of *arp*, or at the *nptII* gene, the T-DNA integration was considered to occur in an exon (e) of the target gene. In such cases it is possible to estimate the extent of the T-DNA deletion (indicated in brackets in base pairs) required for such fusion to happen.
‡If sequences of the 5′/3′RACE–PCR products showed a match to known *Arabidopsis* EST, database accession number of this EST marker is indicated.
§For gene identification purposes, or to obtain probes for DNA hybridizations, 3′RACE–PCR was done on the first-strand cDNA. Herein, the total length of sequenced cDNAs are indicated. NT, not tested lines.
¶Results of DNA gel blot hybridization were used to confirm T-DNA integration in a gene (+). NT, not tested lines.
‖A BLAST search (NCBI, BLAST network server) was used to find genes homologous to the T-DNA tagged genes. Genes with highest sequence homology (smallest sum probability is indicated in next column) are listed along with their database accession numbers. For two lines, SK2-2 and SK2-3, no significant homology to known DNA or protein sequences was found, maybe because not sufficient sequence information is available.
**Mutant phenotypes associated with gene disruptions were scored in progeny of transgenic lines after third backcross to the wild-type *A. thaliana* C24. Lines for which no obvious phenotype was found under standard growth conditions are indicated by a question mark.

fragment was found and, therefore, in none of the lines examined, was the kanamycin-resistant phenotype due to gene replacement in the *arp* gene by homologous recombination. The efficiency of homologous recombination in somatic plant cells, which occurs at a frequency of 10^{-3} to 10^{-5}, is thought to be proportional to the length of the homologous DNA sequences (11, 12). In our pNPTARP vector the smallest rate-limiting area was fairly short (420 bp), and the total number of transformants necessary to observe such event ($\approx 10^5$) probably was not obtained. Therefore, although gene trap constructs offer the possibility of positive selection for gene replacement events, it remains doubtful whether such an approach really will become practical for routine gene tagging through homologous recombination in plants.

To determine the mechanism behind *nptII* gene expression, we analyzed plants by PCR amplification of the 5′ ends of cDNAs (5′RACE–PCR) using *nptII* gene-specific primers. For each of the lines analyzed, 5′RACE–PCR products were found in which DNA sequences derived from the T-DNA vector were preceded by novel upstream sequences. In all cases an ATG codon was found that could function as a translation initiation codon and was in-frame with the *nptII*-coding region. All the fusions identified fell into two groups. In 10 cases the new sequence began at the 3′ splice site of the *arp* gene intron. Although the genomic organization of the T-DNAs in these lines was not analyzed, we presume that this category of insertions represents T-DNA integration events into genomic sequences, which provide adequate 5′ splice sites, and a functional messenger is produced after the splicing of the chimeric intron. This conclusion is strengthened by the fact that in three of the 10 lines analyzed, two types of 5′RACE–PCR products were found, resulting from splicing at either a normal or a new cryptic 3′ splice site. During the course of the *arp* gene expression study 48 5′RACE–PCR fragments were sequenced, and splicing at the cryptic 3′ splice site never was observed. The splicing events at the cryptic 3′ splice site found in this study are probably the consequence of changes in intron size and sequence in identified transgenic plants. Because the 3′ splice sites are 21 bp apart, alternative splicing does not affect the *nptII* frame fusion in the resultant mRNAs, except for a deletion of seven codons from the second *arp* exon. The other 10 *in vivo* fusions are exon–exon T-DNA integration events.

To identify transformants in which the T-DNA possibly had been integrated into *Arabidopsis* genes, the novel sequences from the 5′RACE–PCR products were compared with the sequences of known genes and dbEST. Six sequences perfectly matched known *Arabidopsis* EST markers, which encode aspartyl tRNA synthase (SK3-7), ATP-binding cassette transporter (SK3-3), human ORF (SK1-18), two ribosomal proteins (SK2-N2, SK2-1), and an unknown protein (SK2-2). The sequence of the 5′RACE products from the lines SK1-B2 and

SK2-11 matched a cDNA of the *Arabidopsis* DNA ligase and a small ribosomal protein S21E from yeast, respectively (Table 1). Lines for which no match was found were analyzed further by 3′RACE–PCR. The amplification of cDNAs was obtained using primers designed on the basis of the 5′RACE product sequences from four lines. The sequences of the amplified cDNAs then were determined, and a database search allowed us to identify novel *Arabidopsis* genes coding for a putative DNA-binding protein from the basic-domain leucine zipper family (line SK1-13), a homologue of the snapdragon *DAG* gene (37) (line SK1-N2), a two-domain protein with homology at its N-terminal part to *Escherichia coli* prolyl-peptidyl *cis-trans* isomerases, and to a hypothetical protein from cyanobacteria in its C-terminal domain (line SK3-1). A cDNA sequence encoded by the tagged gene in line SK2-3 did not show homology to any known gene sequences or ESTs. In the lines in which the identity of the tagged gene was determined, the coding region of the tagged gene is always in-frame with *nptII*.

To obtain additional proof that the respective genes have indeed been disrupted in the generated *Arabidopsis* lines, we carried out DNA gel blot hybridization analysis on DNA extracted from plants after three backcrosses to wild type. For all 10 lines analyzed, new DNA fragments were detected that hybridized with the respective cDNA probes, in addition to fragments with the same electrophoretic mobility as that of the wild type. These data, exemplified in Fig. 2, support the idea that kanamycin-resistance phenotype is due to the *in vivo* expression of the chimeric gene fusions identified above.

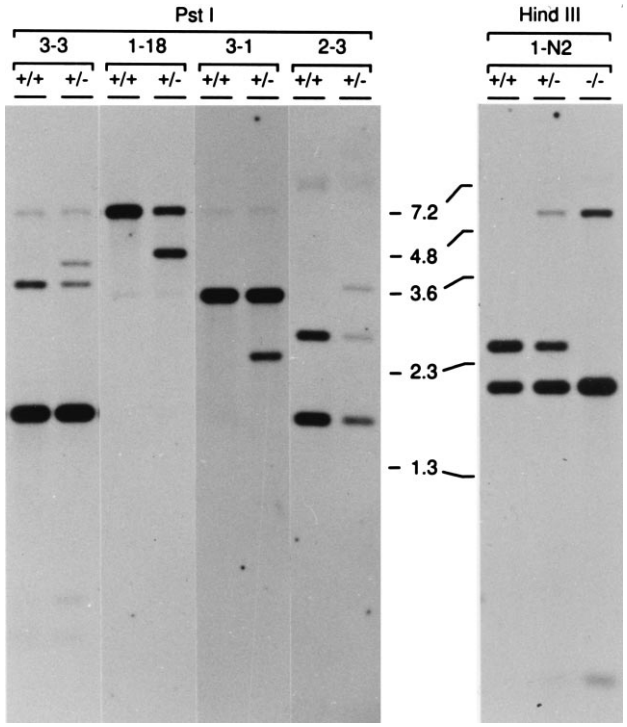

FIG. 2. DNA gel blot analysis of the transgenic *Arabidopsis* lines. DNA was extracted from heterozygous plants (+/−) after the third backcross of the indicated lines and from a wild-type *Arabidopsis* C24 (+/+) used for crosses. For the line SK1-N2, DNA extracted from six albino plants (−/−) segregating in a self-fertilized progeny was analyzed additionally. DNA was digested with *Pst*I or *Hin*dIII; fragments were separated by agarose gel, blotted onto a nylon membrane, and hybridized with DNA probes made from cDNA fragments encoded by the respective T-DNA-tagged genes. In DNA from transgenic plants, new hybridizing fragments can be detected in addition to fragments with a wild-type mobility. The albino plants from the progeny of SK1-N2 appear to be homozygous for the T-DNA insertion and completely lack a 2.4-kb fragment.

Based on the available data, we may conclude that in this gene-trap experiment T-DNA insertions have been identified in 12 *Arabidopsis* genes. The results suggest that overall efficiency of gene tagging is approximately 60%, with an average of 40% of gene tagging events occurring among exon–exon integrations (targeted genes in four of 10 recovered lines) and up to 80% occurring in chimeric intron integrations (eight in 10). The higher efficiency of the gene tagging for the chimeric intron integrations probably results from an additional requirement for a correctly positioned 5′ splice site, besides the promoter and an ORF. In the remaining eight lines for which no targeted genes were identified, the T-DNA probably was integrated into expressed *Arabidopsis* genome regions (intergenic spacers, promoters), which have T-DNA adjoining sequences with enhancer properties.

Mutant phenotypes, which probably result from the disruption of genes identified in this study, were assessed in the progeny of individual plants after three back-crosses to wild-type *Arabidopsis* C24 (Table 1). To prove that the observed phenotypes may be caused by the T-DNA insertion, a complementation analysis was undertaken for line SK1-N2, in which the gene disruption causes a conditional albino phenotype (Fig. 3*A* and *B*). The restriction map and sequence analysis of the genomic sequences showed that of the two *Hin*dIII-hybridizing fragments of 2.4 kb and 2.1 kb on the genomic DNA gel blots (Fig. 2), the 2.4-kb fragment encoded the 5′ part of the *dal1* gene. The T-DNA integrated in the first exon of the *dal1*. This explains that in plants homozygous for the T-DNA insertion, the 2.4-kb fragment is absent. Albino plants can be complemented by the chimeric gene in which the expression of the full-length cDNA is driven by the viral 35S promoter. The protein encoded by the tagged gene is homologous to the *Antirrhinum majus* DAG protein (40% identity and 58.9% similarity at the amino acid level). Transposon integration into the *DAG* gene also results in an albino phenotype (37). Therefore, the identified *DAG*-like *Arabidopsis* gene was designated *dal1*, although it is not known whether these two genes are true orthologs. The albino phenotype is conditional because when plants are grown in liquid media, their leaves turn pale green. The same greening may occur in young leaves of plants growing on a solidified media, but only when leaves are vitrified (Fig. 3*B*). It is possible that under certain growth conditions related proteins can partially complement the loss of function in *dal1*. After a BLAST DNA database search, several *dal1*-homologous EST markers were identified. These ESTs fall into three groups and probably represent three *Arabidopsis* genes: the first group with accession numbers Z25581, T44546, T13916, T42919, and H37475; the second group with accession numbers N65097 and T46596; and the third group represented by one EST (N37309). Plant phytohormones, such as cytokinins and auxins, do not stimulate the greening. The only condition that consistently appears to be associated with a greening process is the absence of air in the leaf intercellular space. The mutant plastids lack thylakoid membranes and normal starch granules and contain electron-dense inclusion bodies (Fig. 3*D*). Plastids from greenish leaves develop thylakoid-like membranous structures, which, however, are not stacked in granas (Fig. 3*F*). Mutation in *dal1* affects the normal morphogenesis of leaf cells and no discernible mesophyll parenchyma cells can be found in a mutant (Fig. 3 *H* and *J*) compared to wild type (Fig. 3 *G* and *I*), which is similar to the *DAG1* mutation. The DAG-related proteins appear to be specific for plants and do not show significant homology to known proteins from other organisms (37); therefore, available data are insufficient to make any sensible speculation about their function.

DISCUSSION

The data presented have important implications for the reverse genetics of *Arabidopsis*. The observed frequencies of

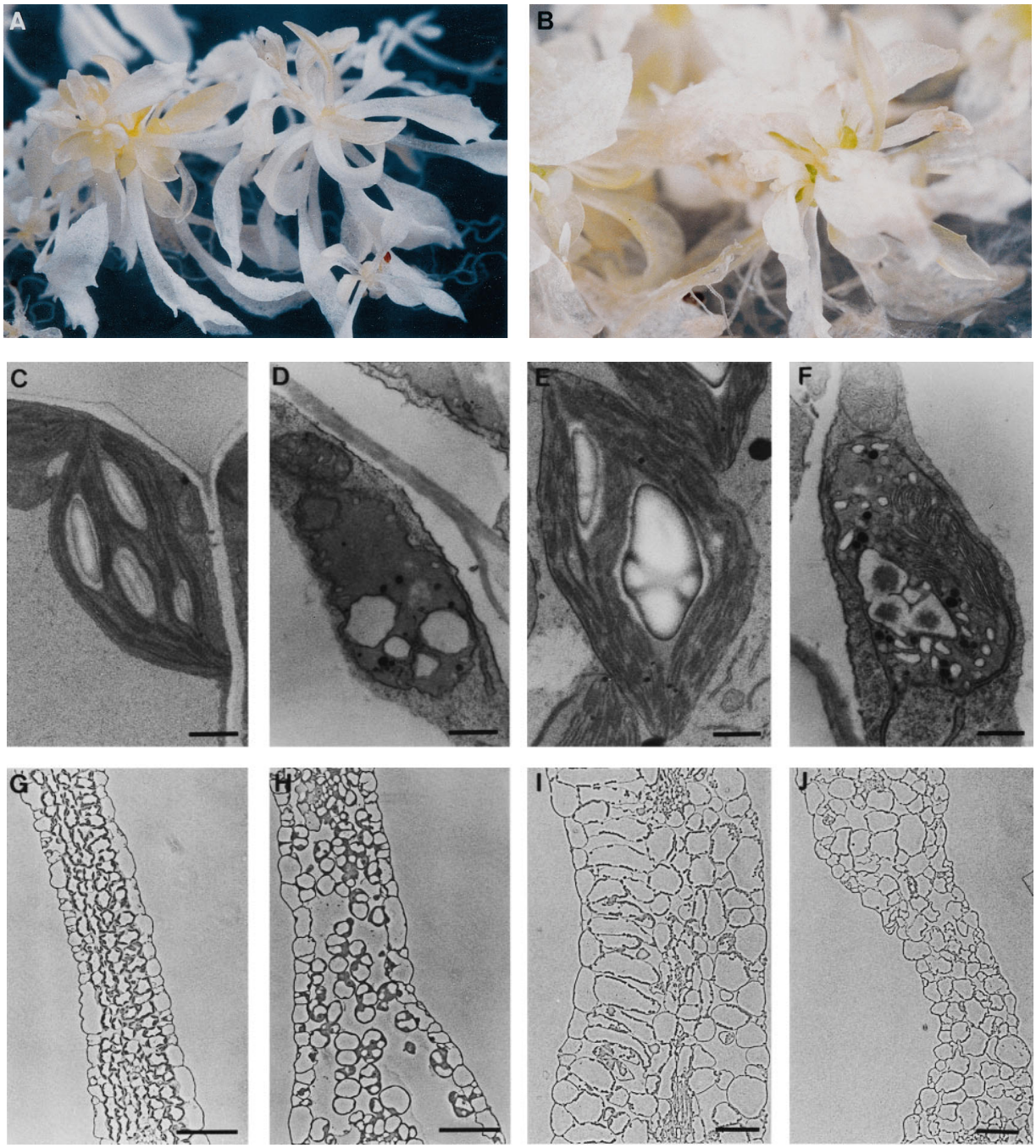

FIG. 3. *dal1* mutant of *Arabidopsis*. (*A* and *B*) Two-month-old *dal1* mutant plants are shown. Mutant plants from the same seed stock were grown *in vitro* in six Petri plates, but only plants from a single plate began to develop greenish young leaves (*B*). The only apparent difference between plants from these different plates is that in *B* plants are vitrified. The ultrastructure of plastids from mutant plants (*D* and *F*) compared to that of the wild type (*C* and *E*). Plants were either grown on agar-solidified Murashige and Skoog medium supplemented with 20 g/liter of sucrose (*C* and *D*) or grown in the same, but liquid medium on a gyratory shaker (*E* and *F*). In plastids of the *dal1* mutant a membranous structure resembling thylakoids can be observed (*F*, but not in *D*). General morphology of the leaf cells for *dal1* (*H* and *J*) compared to that of wild type (*G* and *I*). Young leaves of the plants grown on agar plates (*G* and *H*) or fully expanded leaves of plants grown in a liquid medium (*I* and *J*) were sectioned after imbedding into vinylcyclohexene dioxide (Spurr mixture). In the mutant, no palisade parenchyma cells developed and cells are more rounded. [Bars = 2 μm (*C–F*), 50 μm (*G–H*), and 100 μm (*I–J*).]

gene disruption among transformants recovered is high enough to allow identification of the tagged gene before knowing their associated phenotype, and 5′RACE–PCR is a straightforward method of tagged gene identification. The described vectors have the advantage that both exon (plant gene)—exon (*arp*, *nptII*) and intron (plant gene)—intron (*arp*) fusions might lead to the restoration of NPTII activity. Nevertheless, the total yield of selected transformants was low in our experiments because of the large T-DNA deletions required for the gene trapping events. The deletion of T-DNA from the left border is a well established fact, but in the majority of cases only approximately 10–100 bp of T-DNA are involved (36). Therefore, for gene trap constructs in which no large T-DNA deletions are required, the recovery of transformants should be several-fold and even several orders of magnitude higher (ref. 23; unpublished data). For five of the tagged genes, no *Arabidopsis* EST markers are known, reflecting the fact that even genes with a low expression level can be

tagged with this method. The disadvantage of the use of *nptII* as a reporter is that only genes expressed at the time of selection can be tagged; however, this does not mean that the method is restricted only to genes with constitutive expression. The progeny from three of the lines analyzed display kanamycin sensitivity at the seedling stage, whereas resistance can be shown only for tissue explants placed on kanamycin-containing regeneration media. The presence of plant growth regulators in the media during selection probably contributes to a broadening of the gene spectrum, which is subject to gene trapping. In conclusion, it became obvious that the approach presented here for gene trapping offers the plant scientific community a possibility to initiate a systematic gene tagging program, similar to systematic cDNA sequencing projects, with the advantage that for each gene identified an insertion mutant will be available.

We thank Dr. R. Taylor for the *Arabidopsis* DNA ligase cDNA clones, Riet De Rycke and Gilbert Engler for help with the ultrastructural analysis, Mark Davey, Pierre Rouzé, and Rudy Vanderhaeghen for critical reading of the manuscript, Martine De Cock for help in preparing it, and Karel Spruyt and Christiane Germonprez for artwork. This work was supported by grants from the Interuniversity Poles of Attraction Programme (Belgian State, Prime Minister's Office—Federal Office for Scientific, Technical, and Cultural Affairs; P4/15). D.I. is a Research Director of the Institut National de la Recherche Agronomique (France).

1. Goodman, H. M., Ecker, J. R. & Dean, C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10831–10835.
2. Long, J. A., Moan, E. I., Medford, J. I. & Barton, M. K. (1996) *Nature (London)* **379,** 66–69.
3. Aarts, M. G. M., Dirkse, W. G., Stiekema, W. J. & Pereira, A. (1993) *Nature (London)* **363,** 715–717.
4. Feldmann, K. A. (1991) *Plant J.* **1,** 71–82.
5. Gibson, S. & Somerville, C. (1993) *Trends Biotechnol.* **11,** 306–313.
6. Newman, T., de Bruijn, F. J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. & Somerville, C. (1994) *Plant Physiol.* **106,** 1241–1255.
7. Cooke, R., Raynal, M., Laudié, M., Grellet, F., Delseny, M. *et al.* (1996) *Plant J.* **9,** 101–124.
8. Melton, D. W. (1994) *BioEssays* **16,** 633–638.
9. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hohensel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S. G. (1996) *Science* **274,** 546–567.
10. Puchta, H. & Hohn, B. (1996) *Trends Plant Sci.* **1,** 340–348.
11. Altmann, T., Damm, B., Halfter, U., Willmitzer, L. & Morris, P.-C. (1992) in *Methods in* Arabidopsis *Research*, eds. Koncz, C., Chua, N.-H. & Schell, J. (World Scientific, Singapore), pp. 310–330.
12. Miao, Z.-H. & Lam, E. (1995) *Plant J.* **7,** 359–365.
13. Matzke, M. A. & Matzke, A. J. M. (1995) *Plant Physiol.* **107,** 679–685.
14. Nellen, W. & Lichtenstein, C. (1993) *Trends Biochem. Sci.* **18,** 419–423.
15. McKinney, E. C., Ali, N., Traut, A., Feldmann, K. A., Belostotsky, D. A., McDowell, J. M. & Meagher, R. B. (1995) *Plant J.* **8,** 613–622.
16. Mena, M., Ambrose, B. A., Meeley, R. B., Briggs, S. P., Yanofsky, M. F. & Schmidt, R. J. (1996) *Science* **274,** 1537–1540.
17. Bensen, R. J., Johal, G. S., Crane, V. X., Tossberg, J. T., Schnable, P. S., Meeley, R. B. & Briggs, S. P. (1995) *Plant Cell* **7,** 75–84.
18. Burns, N., Grimwade, B., Ross-Macdonald, P. B., Choi, E.-Y., Finberg, K., Roeder, G. S. & Snyder, M. (1994) *Genes Dev.* **8,** 1087–1105.
19. Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T. & Rubin, G. M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10824–10830.
20. Skarnes, W. C. (1990) *Bio/Technology* **8,** 827–831.
21. Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J. D. G., Dean, C., Ma, H. & Martienssen, R. (1995) *Genes Dev.* **9,** 1797–1810.
22. Kertbundit, S., De Greve, H., Deboeck, F., Van Montagu, M. & Hernalsteens, J.-P. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5212–5216.
23. Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Körber, H., Redei, G. P. & Schell, J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8467–8471.
24. Koncz, C., Mayerhofer, R., Koncz-Kalman, Zs., Nawrath, C., Reiss, B., Redei, G. P. & Schell, J. (1990) *EMBO J.* **9,** 1337–1346.
25. Springer, P. S., McCombie, W. R., Sundaresan, V. & Martienssen, R. A. (1995) *Science* **268,** 877–880.
26. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
27. Babiychuk, E., Kushnir, S., Van Montagu, M. & Inzé, D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3299–3303.
28. An, G., Watson, B. D., Stachel, S., Gordon, M. P. & Nester, E. W. (1985) *EMBO J.* **4,** 277–284.
29. Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M. & Leemans, J. (1985) *Nucleic Acids Res.* **13,** 4777–4788.
30. Valvekens, D., Van Montagu, M. & Van Lijsebettens, M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5536–5540.
31. Troutt, A. B., McHeyzer-Williams, M. G., Pulendran, B. & Nossal, G. J. V. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9823–9825.
32. Shirzadegan, M., Christie, P. & Seemann, J. R. (1991) *Nucleic Acids Res.* **19,** 6055.
33. Holton, T. A. & Graham, M. W. (1990) *Nucleic Acids Res.* **19,** 1156.
34. Michaels, S. D., John, M. C. & Amasino, R. M. (1994) *BioTechniques* **17,** 274–276.
35. Mayerhofer R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K., Redei, G. P., Schell, J., Hohn, B. & Koncz, C. (1991) *EMBO J.* **10,** 697–704.
36. Tinland, B. (1996) *Trends Plant Sci.* **1,** 178–184.
37. Chatterjee, M., Sparvoli, S., Edmunds, C., Garosi, P., Findlay, K. & Martin, C. (1996) *EMBO J.* **15,** 4194–4207.