ELSEVIER

Plant Science 138 (1998) 217–228

PLANT SCIENCE

T-DNA trapping of a cryptic promoter identifies an ortholog of highly conserved *SNZ* growth arrest response genes in *Arabidopsis* ☆

László Ökrész [a], Csaba Máthé [a], Éva Horváth [a], Jeff Schell [b], Csaba Koncz [a,b], László Szabados [a,*]

[a] *Institute of Plant Biology, Biological Research Center, Hungarian Academy of Sciences, Temesvári krt. 62, H-6726 Szeged, Hungary*

[b] *Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné Weg 10, D-50829 Köln, Germany*

Received 7 April 1998; received in revised form 14 August 1998; accepted 17 August 1998

Abstract

A T-DNA tagged *Arabidopsis* locus, *A37*, identified by a promoter-trap *aph*(*3′*)*II* reporter gene fusion expressed in calli and roots, encodes an ortholog of evolutionarily conserved SNZ growth arrest response proteins. Gene *A37* is located on chromosome 3–35, lacks introns, and shares considerable sequence identity with *HEVER1* from rubber tree, *SLEXORFA-1* from *Stellaria longipes*, *SNZ1* from yeast, and *SNZ*-homologs from bacteria and archaebacteria. Southern DNA hybridization and physical mapping data show that *A37* is a single copy gene, but sequence similarity to expressed sequence tags (ESTs) suggests that at least two other *SNZ*-homologs are present in *Arabidopsis*. The *A37* gene is abundantly expressed in cultured callus tissues and at lower levels in leaves, stems and roots. In the promoter-trap locus *a37*, the T-DNA-linked *aph*(*3′*)*II* reporter gene is transcribed oppositely to the *A37* gene by a cryptic promoter located 0.52 kb upstream of the *A37* coding region. Promoter deletion studies with *uidA*-reporter gene constructs show that the cryptic promoter consists of regulatory sequences located in both promoter and transcribed regions of the *A37* gene that activate transcription only in roots. The *a37* promoter trap is thus controlled by transcriptional regulatory sequences that function as an active promoter only in linkage with a promoterless reporter gene introduced artificially into the *Arabidopsis* genome by a T-DNA tag. © 1998 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: T-DNA insertional mutagenesis; Reporter gene fusion; Cryptic promoter; Growth arrest response proteins; *Arabidopsis*

☆ GenBank accession numbers of DNA sequences described in this paper are AF029980 and AF029981.

* Corresponding author. Tel.: + 36-62-432080 ext. 239; Fax: + 36-62-433434; e-mail: szabados@rosi.szbk.u-szeged.hu.

0168-9452/98/$ - see front matter © 1998 Elsevier Science Ireland Ltd. All rights reserved.

PII: S0168-9452(98)00163-0

1. Introduction

T-DNA mediated insertional mutagenesis is widely used for genetic identification and functional characterization of plant genes [1,2]. To isolate T-DNA insertions in transcribed chromosomal loci, several gene fusion and enhancer trap technologies were developed. Thus, integration of promoter or enhancer trap T-DNA tags into actively transcribed plant genes is monitored by transcriptional activation of promoterless reporter genes coding for β-glucuronidase (*uidA*), kanamycin phophotransferase (*aph*(*3*′)*II*), and luciferase (*luc* or *luxF*) in vitro or in vivo [2–6]. In *Arabidopsis* and *Nicotiana*, about one third of T-DNA integration events was observed to produce active transcriptional reporter gene fusions suggesting that the T-DNA is frequently integrated into chromosomal loci that are potentially transcribed [4]. In contrast, translational gene fusions generated by reporter genes lacking the ATG start codon occurred at significantly lower (15–20%) frequencies [3,4]. Therefore, it was noted that activation of T-DNA-linked promoterless reporter genes carrying a translational start codon does not necessarily demonstrate that a T-DNA tag is indeed located in a transcribed plant gene [4]. In fact, enhancer trapping data showed that certain T-DNA tags in the genome link the promoterless reporter genes to transcriptional regulatory sequences that are apparently not associated with RNA polymerase II transcribed genes, although they display cryptic promoter or enhancer activity [7,8]. Since such T-DNA inserts do not cause gene mutations, the efficiency of gene isolation by promoter or enhancer tagging was questioned. To clarify the function and location of potential cryptic promoter elements in the genome, we have characterized a T-DNA-tagged locus identified by promoter trapping in a transgenic *Arabidopsis* line, *Ath621-37* [4,9]. The *a37* line was generated by a transcriptional gene fusion vector, pPCV621, carrying a promoterless *aph*(*3*)*II* marker gene with an ATG start codon linked to the right T-DNA border. Plant DNA sequences activating the *aph*(*3*)*II* reporter gene fusion were rescued from *a37* and shown to promote transcription of an *uidA* reporter gene in callus, root tips and root hairs in tobacco and *Arabidopsis* [4]. Here we demonstrate that promoter trap *a37* is transcribed by a cryptic promoter which was activated by a T-DNA integration event. The activity of this cryptic promoter was found to be controlled by sequence elements located in the promoter and coding region of the *A37* gene that codes for an ortholog of SNZ proteins showing high degree of evolutionarily conservation between plants and unicellular organisms [10]. This data illustrates that T-DNA insertions can in fact detect, or even activate, artificial promoters, sequence elements of which may nevertheless be assigned to transcribed genomic loci in *Arabidopsis*.

2. Experimental

2.1. Plant growth and transformation

Arabidopsis thaliana (ecotype Col-1) was grown in glass jars containing seed-germination medium (SG) under controlled culture conditions (22°C, 16 h light and 8 h dark cycle) as previously described [5]. Tissue culture and *Agrobacterium*-mediated root transformation of *Arabidopsis*, as well as tobacco leaf disk transformation were according to Koncz et al. [4,5]. Root-derived callus and cell suspension cultures were established and subcultured weekly in MSAR-medium containing 1 mg/l 2,4-D (2,4-dichloro-phenoxyacetic acid) as described [5]. Seedlings (2 weeks old) were subjected to ABA and NaCl treatments by transferring them into 250 ml Erlenmeyer flasks containing 30 ml liquid SG-medium containing either 1 μM ABA (abscisic acid) or 200 mM NaCl.

2.2. DNA cloning and sequencing

Arabidopsis λ-ZapII cDNA (obtained from I. Somssich, MPI, Cologne) and λ-EMBL4 genomic libraries [11] were screened using a radiolabeled *Hin*dIII–*Bcl*I plant DNA fragment from plasmid pPCV621-37 [4,9] as described [12]. Two full-length cDNAs were sequenced in pBluescript SK^{+}. From several overlapping genomic clones one clone was subjected to detailed physical map-

ping followed by subcloning of DNA fragments into pBluescript SK$^+$, to determine the nucleotide sequence of *A37* gene. DNA fragments were sequenced using an ABI Prism™ Dye Terminator Cycle Sequencing kit and an automatic ABI377 sequencer. Nucleotide and amino acid sequence analyses were carried out using the GCG program package (Genetics Computer Group Inc., Wisconsin), whereas multiple sequence alignments were obtained with the ClustalW software.

2.3. Southern hybridization analysis and chromosomal mapping

Genomic DNA was prepared from 4-week-old Columbia (Col-1) and Landsberg (Ler-0) plants for Southern DNA hybridization analysis [12,13] using either the *A37* cDNA or a *Hin*dIII genomic DNA fragment of 1.9 kb from the *A37* gene as probe. Following hybridization in a buffer containing 50% formamide, the filters were washed twice with $2\times$SSPE, 0.1% SDS, and once with $0.1\times$ SSPE, 0.1% SDS at 65°C as described [12]. *Arabidopsis* yeast artificial chromosome (YAC) libraries CIC and yUP [14,15] were screened using the same *A37* gene probe according to Matallana et al. [16]. In order to confirm the map position of *A37* locus, DNA was purified from YAC clones detected by the *A37* probe and subjected to Southern blot analysis as previously described [17].

2.4. RNA hybridizations

RNA was extracted from 3 to 5 g of stems, leaves and roots of plants grown in sterile culture, as well as from calli as described [18]. To prepare Northern blots, 20 μg of total RNA was subjected to electrophoresis in denaturing agarose gels and blotted onto Hybond N$^+$-filters. Hybridization of RNA blots was carried out with radiolabeled probes overnight at 42°C, then the filters were washed twice with $3\times$ SSC, 0.1% SDS, and twice with $1\times$SSC, 0.1% SDS at 65°C as described [12].

2.5. Promoter fusions and deletions

Fragments from *A37* and cryptic promoters were cloned into unique restriction endonuclease cleavage sites of promoter testing vector pPCV812 [5,19]. Thus, the *A37* promoter was cloned as a *Hin*dIII–*Bam*HI genomic DNA fragment, whereas a deletion series from the cryptic promoter was created by cloning *Ssp*I–*Bcl*I, *Xba*I–*Bcl*I, *Bam*HI–*Bcl*I, and *Hin*dIII–*Bcl*I fragments of the cryptic promoter from plasmid pPCV621-37 into pPCV812 (Figs. 1 and 6) [4,9]. Vector constructs carrying the promoter fragments were introduced into *Arabidopsis* and tobacco by *Agrobacterium*-mediated transformation [5]. Expression of the *uidA* reporter gene was followed by quantitative measurement of GUS-enzyme activity in four to six transgenic plants, as well as by histochemical staining [20].

3. Results

3.1. Characterization of the A37 gene using a promoter trap T-DNA insertion

In a T-DNA-tagged *Arabidopsis* line, *Ath621-37*, carrying a promoter trap *aph(3′)II* reporter gene fusion expressed in calli and roots, a single T-DNA insertion was previously identified [4,9]. The promoter trap *aph(3′)II* gene fusion was rescued in plasmid pPCV621-37 carrying a plant genomic DNA fragment of 0.9 kb upstream of the reporter gene [4]. This rescued plant DNA fragment was sequenced and used as hybridization probe to isolate genomic and cDNA clones from wild type λEMBL4 and λZapII *Arabidopsis* genomic and cDNA libraries, respectively. From several overlapping genomic clones one was subjected to physical mapping with different restriction endonucleases (Fig. 1). Further mapping of this genomic clone by Southern hybridization identified an *Eco*RI fragment of 3.35 kb which hybridized with the *Arabidopsis* DNA fragment of 0.9 kb flanking the *aph(3′)II* reporter gene in the promoter trap clone pPCV621-37. This *Eco*RI fragment, as well as its *Sac*I–*Eco*RI subfragments of 1.3 and 2.2 kb, detected a single transcript of

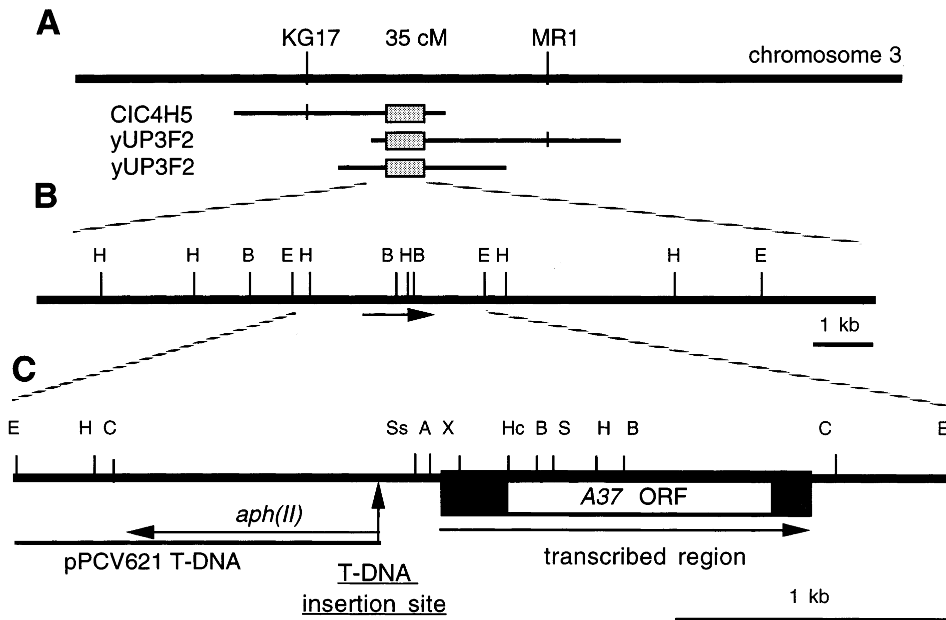

Fig. 1. Physical mapping and chromosomal localization of the *A37* gene in *Arabidopsis thaliana*. (A) Determination of map position of *A37* gene on chromosome 3 by YAC hybridization. Regions of YAC clones CIC4H5, yUP3F2 and yUP3F2 hybridizing with the *A37* cDNA probe are labeled by gray box. The position of RFLP markers KG17 and MR1 within YAC clones CIC4H5 and yUP3F2 is indicated. (B) Physical map of the *A37* locus. (C) Structure of the A37 gene. The transcribed region and open reading frame of *A37* are indicated. The T-DNA insertion site and orientation of the *aph(3′)II* reporter gene in the promoter trap *Arabidopsis* line *Ath621-37* are shown by arrows. H, *Hin*dIII; E, *Eco*RI; X, *Xba*I; B, *Bam*HI; C, *Cla*I; S, *Sac*I; Ss, *Ssp*I; Hc, *Hin*dII; A, *Acc*I; Bg, *Bgl*II.

about 1.3 kb on Northern blots carrying RNA purified from calli (Fig. 2b). Further resolution of this region by hybridization (data not shown) indicated that the whole *Eco*RI fragment of 3.35 kb carried only one transcribed region which was located upstream of the T-DNA insertion and transcribed opposite to the *aph(3′)II* reporter gene in the promoter tap *Arabidopsis* line *Ath621-37*.

This conclusion was confirmed by sequencing the *Eco*RI genomic DNA fragment of 3.35 kb (Fig. 3, GenBank AF029981). Sequence comparison between the wild type (*A37*) and T-DNA tagged (*a37*) loci revealed that the T-DNA insertion occurred 0.52 kb upstream of an open reading frame (ORF) by dividing the *Eco*RI fragment of 3.35 kb to two segments of 1.3 and 2.2 kb (Fig. 1). The genomic DNA sequence located upstream of this ORF (and thereby upstream of the T-DNA target site defined by the plant DNA junction with the promoterless *aph(3′)II* reporter gene in *Ath621-37*, [4,9]) did not reveal any ORF. Moreover, sequence homology searches in the GenBank/EMBL database did not identify any homology to this genomic DNA segment which showed very low coding probability (data not shown). Therefore, it appeared that the T-DNA tag landed in a non-coding sequence located upstream of a gene, termed *A37*, in the *Ath621-37* promoter trap line. Because the *A37* ORF of 942 bp showed opposite polarity to the *aph(3′)II* gene fusion identified in *Ath621-37*, it became evident that the expression of this reporter gene fusion was driven by a cryptic promoter. Since the distance between the ATG codons of *aph(3′)II* reporter gene and *A37* ORF was only 520 bp, the sequence data indicated that the T-DNA tag landed in the upstream promoter region of *A37* gene which was able to activate transcription with a polarity opposite to that of *A37* ORF. In fact, sequence analysis predicted that the distance be-

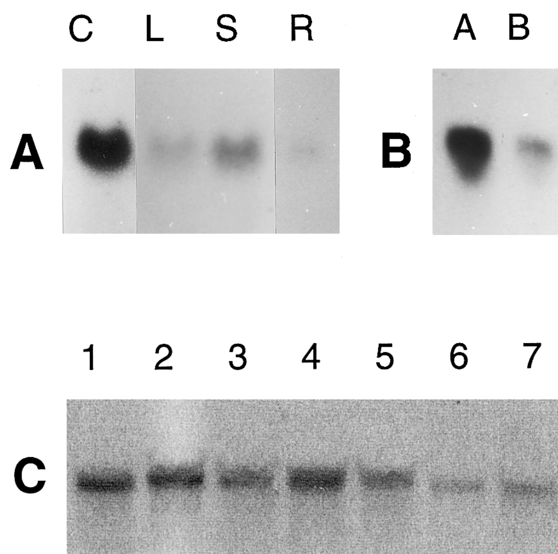

Fig. 2. Analysis of *A37* gene expression. (A) Northern blot hybridization of RNA samples purified from different organs of wild type *Arabidopsis* plants. Total RNA samples (20 μg) were resolved by electrophoresis in denaturing agarose gels, blotted onto Hybond nylon membranes, and hybridized with radiolabelled *A37* cDNA. C, callus; R, root; S, stem; L, leaf. (B) Northern blot hybridization of callus RNA samples with probes derived from the *A37* genomic *Eco*RI fragment of 3.35 kb. (A) *Eco*RI–*Sst*I fragment of 1.3 kb; (B) *Eco*RI–*Sst*I fragment of 2.2 kb. (C) Hybridization of *A37* cDNA probe with RNA samples prepared from *Arabidopsis* (Col-1) cell suspension culture (samples 1–5) and seedlings (samples 6–7) following treatments with 1 μM ABA and 200 mM NaCl. Cell suspension culture (1) no treatment; (2) ABA treatment for 20 min; (3) NaCl treatment for 20 min; (4) NaCl treatment for 6 h; (5) ABA treatment for 6 h. Seedlings (6) no treatment; (7) NaCl treatment for 6 h.

tween transcription initiation sites of *aph(3′)II* and *A37* genes was only 302 bp (Fig. 3). To support these data, an *Arabidopsis* cDNA library was screened with probes carrying plant DNA sequences located upstream and downstream of the T-DNA integration site in *Ath621-37*. The cDNA screening yielded clones showing identity solely with the *A37* gene. A full-length *A37* cDNA of 1352 bp (GenBank AF029980) carried an untranslated 5′-leader of 220 bp, an open reading frame of 942 bp, and a 3′-untranslated sequence of 175 bp followed by a poly(A) track. The *A37* cDNA and genomic sequences showed a perfect identity (except for the polyA track) indicating that the *A37* gene contained no intron. The 5′-end of *A37* cDNA corresponded in fact to the position of a putative transcription initiation motif predicted by sequence analysis of genomic DNA (Fig. 3).

3.2. The A37 gene encodes an ortholog of evolutionarily conserved SNZ proteins

The *A37* ORF was found to encode a putative polypeptide of 314 amino acids with a predicted molecular mass of 33.8 kDa and an isoelectric point of 5.47. The amino acid sequence showed a surprisingly high level of identity with several known protein sequences from plants, yeast, bacteria and archaebacteria. Homologs of *A37* included proteins encoded by the ethylene-induced *HEVER-1* gene in *Hevea brasiliensis* (HEV-ERP1A, [21]), *SLEXORFA-1* in *Stellaria longipes* (STLONG_ORF, GenBank X71601), *SNZ* genes in yeast (YFF9_YEAST, YMR09_YEAST, [10]), and SNZ-like genes from *Bacillus subtilis* (YAAD_BACSU, GenBank D26185), *Haemophilus influenzae* (YG47_HAEIN, GenBank U32837) and *Methanococcus vanielii* (MVAN_ORF, [22]) (Fig. 4). ClustalW analysis of phylogenic relationships showed, as expected, that *HEVER-1* from rubber tree is the closest known homolog of *A37* (Fig. 4). In addition, two *Arabidopsis* expressed sequence tags (ESTs, GenBank C72151 and Z17740) sharing significant sequence similarity with *A37* were detected in the database indicating that at least two other *SNZ*-like genes are present in the *Arabidopsis* genome.

SNZ homologs were found to represent one of the most evolutionarily conserved protein families [10] that share some domain similarity with an indole-3-glycerol-phosphate synthase from *Haloferax volcanii* (GenBank P18307) and small subunits of bacterial Δ-amino acid dehydrogenases. Nonetheless, database searches offered no reliable information about the function of A37 and other SNZ proteins. Two potential casein kinase phosphorylation sites (a SVAD motif at amino acid position 128–131, and a TPAD motif between positions 238 and 241) in conserved SNZ-domains and a *N*-myristylation site in a similarly conserved hydrophobic domain

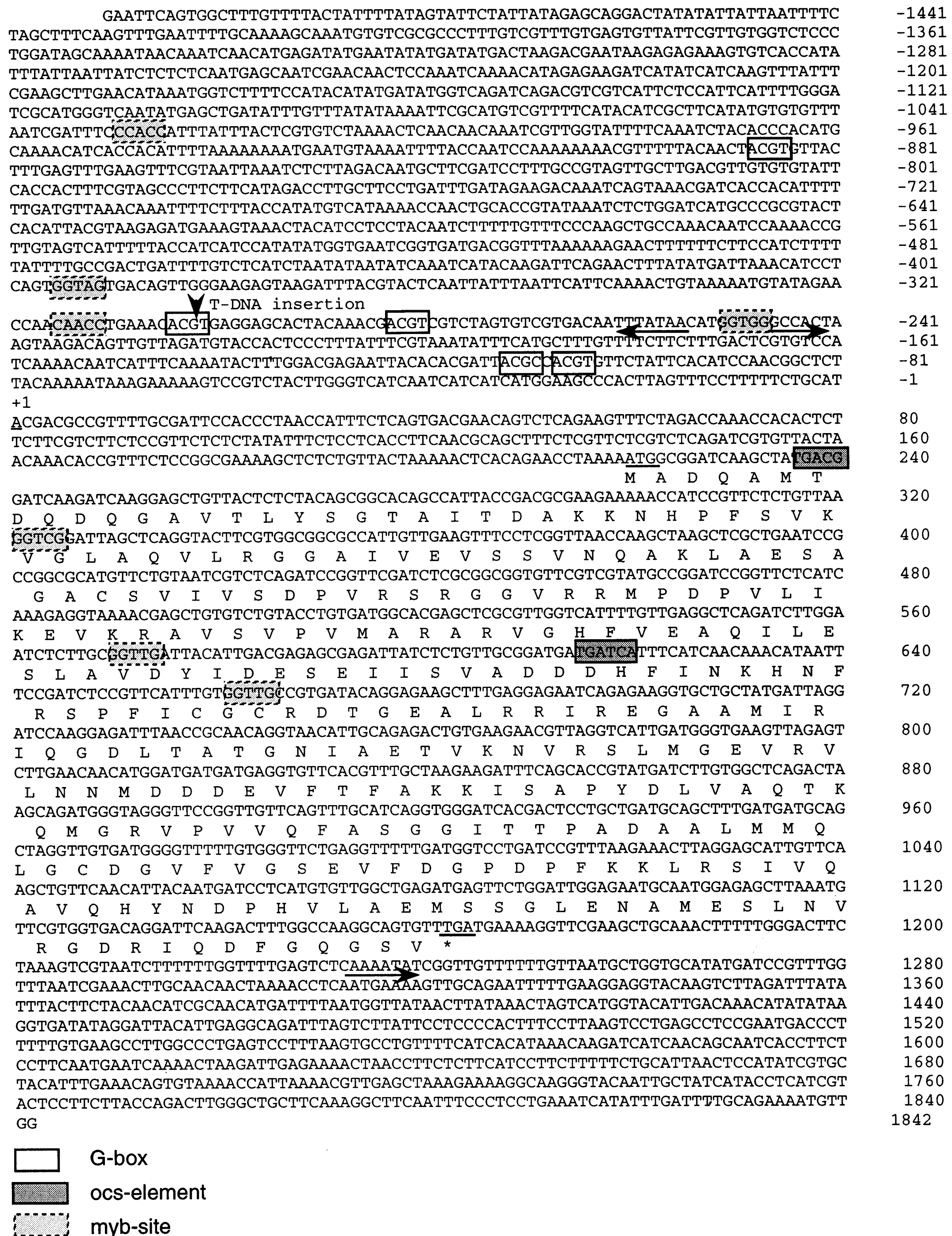

Fig. 3. Nucleotide sequence of the *A37* gene and amino acid sequence deduced from the *A37* cDNA. Target site of the T-DNA insertion is marked by a vertical arrow. Putative TATA and CAAT boxes, as well as polyadenylation signals, are labeled by arrow-headed lines below the sequence. Potential regulatory sequences are framed, translation initiation and termination codons are underlined.

(GVFVGS motif between amino acid positions 252–257) were characteristic for A37 and all other SNZ-like proteins. Recognizing an acylation consensus sequence and hydrophobic nature of the C-terminus, computer analyses predicted that A37 may be membrane-associated protein. Other apparent organelle or membrane localization signals were however not identified in the A37 amino acid sequence using the PSORT method (http://psort.nibb.ac.jp).

3.3. Chromosomal mapping and characterization of the A37 gene

Southern hydridization with the *A37* cDNA probe detected under stringent conditions only a single genomic DNA fragment in both *Arabidopsis* ecotypes Columbia (Col-1) and Landsberg (Ler-0) indicating that *A37* is a single copy gene (Fig. 5A). Since no RFLP was detected between Col-1 and Ler-0, the chromosomal position of *A37* locus was determined by hybridization of the *A37* cDNA probe to filters carrying the *Arabidopsis* yeast artificial chromosome (YAC) libraries CIC and yUP [14–17]. The *A37* probe detected three YAC clones: CIC4H5, yUP3F2, and yUP3E10 (Fig. 5B). CIC4H5 and yUP3F2 were already known to map to chromosome 3-35 cM, a position flanked by RFLP markers KG17 and MR1 (http://genome-www.stanford.edu/Arabidopsis/, Fig. 1). The *A37* probe hybridized to identical DNA fragments in all three YAC clones and genomic DNAs from both Col-1 and Ler-0 *Arabidopsis* ecotypes (Fig. 5). Because the YAC clones overlapped, an apparent gap between Contigs defined by RFLP markers KG17 and MRI could be closed by A37 on the available physical map.

RNA hybridization analyses with the *A37* cDNA probe detected characteristic tissue specific differences in the expression pattern of *A37* gene. The highest level of steady-state *A37* mRNA was observed in callus cultures indicating that expression of the *A37* gene is enhanced in dividing plant cells. The levels of *A37* mRNA were significantly lower in leaves, stems, and roots (Fig. 2). To characterize in more detail the expression pattern of *A37* gene, the promoter region (extending from position −1197 to +465) was fused as a *Hin*dIII–*Bam*HI fragment to an *uidA* reporter gene in pPCV812 [5] and introduced into *Arabidopsis*. Supporting the results of transcript analysis, high GUS activities were detected in meristematic tissues of vascular system in leaf, stem, root and inflorescence, as well as in calli, especially in regenerating tissues composed of fast proliferating and differentiating cells (data not shown). Inspection of the *A37* promoter sequence has identified potential bZIP-binding G-box motives (at positions −111, −286, −308, −892, and −1151), and other putative binding sites for HD-bZIP (at positions −252 and −362), myb (at positions −253, −317, −397, and −1131), and silencer T-box-binding SBF1 (at positions −416, and −719) transcription factors (http://transfac.gbf-braunschweig.de, [23,24]). A TATA-box consensus sequence occurred only once in reverse orientation at position −269 followed by a CAAT-box element at position −250. This TATA-element was thus flanked by two perfect G-box motives (at −111 and −286) suggesting that the promoter region between −286 and −111 may initiate transcription bidirectionally when placed in a proper context of initiation sequences. Such proper initiation sequence was however found to occur only at a position corresponding to the 5′-end of *A37* transcript in the genomic DNA sequence.

Because bZIP-binding sites were found to be essential for transcriptional regulation of numerous plant promoters in response to stress stimuli, we have tested the inducibility of *A37* gene expression in proliferating cell cultures and seedlings treated with ABA and NaCl (Fig. 2C). Expression of *A37* showed however no response to brief or prolonged ABA or NaCl treatments in cell cultures and seedlings indicating that, if transcription of *A37* was indeed controlled by stress in analogy to other *SNZ*-genes, factors other than ABA and salt are involved in the regulation of *A37* gene.

A

```
ATA37        MADQAMTDQDQGAVTLYSGTAITDAKKNHPFSVKVGLAQVLRGGAIVEVSSVNQAKLAES
HEVERP1A     -------MAGTGVVAVYGNGAITETKKS-PFSVKVGLAQMLRGGVIMDVVNPEQARIAEE
STLONG_ORF   ------------------------------------------------------------
YAAD_BACSU   --------------------MAQTGTER----VKRGMAEMQKGGVIMDVINAEQAKIAEE
YFF9_YEAST   ----------------------MSEFK-----VKTGLAQMLKGGVIMDVVTPEQAIIAER
YMR09_YEAS   --------------------MTGEDFK-----IKSGLAQMLKGGVIMDVVTPEQAKIAEK
YG47_HAEIN   --------------------MAENRYE-----LNKNLAQMLKGGVIMDVQNPEQARIAEA
MVAN_ORF     ------------------------------------------------------------
                                                    ..    .. .  .     .  ..

ATA37        AGACSVIVSD--P--VRSRGGVRRMPDPVLIKEVKRAVSVPVMARARVGHFVEAQILESL
HEVERP1A     AGACAVMALERVPADIRAQGGVARMSDPQLIKEIKQSVTIPVMAKARIGHFVEAQILEAI
STLONG_ORF   ----------------------------------------MAKARIGHFVEAQILESI
YAAD_BACSU   AGAVAVMALERVPADIRAAGGVARMADPTIVEEVMNAVSIPVMAKARIGHIVEARVLEAM
YFF9_YEAST   AGACAVMALERIPADMRKSGQVCRMSDPRMIKEIMEAVSIPVMAKVRIGHFVEAQILEEL
YMR09_YEAS   SGACAVMALESIPADMRKSGKVCRMSDPKMIKDIMNSVSIPVMAKVRIGHFVEAQIIEAL
YG47_HAEIN   AGAAAVMALERIPADIRAVGGVSRMSDPKMIKEIQGAVSIPVMAKVRIGHFVEAQILEAI
MVAN_ORF     ------------------------------------------MAKARIGHFVEAQVLESI
              ..  .      .   .  .  .  .. ..          .  ..**..*.**.***...*.

                       P
ATA37        AVDYIDESEIISVADDDHFINKHNFRSPFICGC---RDTGE-ALRRIREGAAMIRIQGDLT
HEVERP1A     GIDYVDESEVLTPADEENHINKHNFRIPFVCGC---RNLGE-ALRRIREGAAMIRTKGEAG
STLONG_ORF   GVDYVDESEVLTPADEDHHINKHNFQIPFVCGLSIPRGAPPAYRRGYAGHDTGPRVRKP-
YAAD_BACSU   GVDYIDESEVLTPADEEFHLNKNEYTVPFVCGC---RDLGE-ATRRIAEGASMLRTKGEPG
YFF9_YEAST   QVDYIDESEVLTPADWTHHIEKHNFKVPFVCGA---KDLGE-ALRRINEGAAMIRTKGEAG
YMR09_YEAS   EVDYIDESEVLTPADWTHHIEKDKFKVPFVCGA---KDLGE-ALRRINEGAAMIRTKGEAG
YG47_HAEIN   EIDYIDESEVLSPADNRFHVDKKEFQVPFVCGA---KDLGE-ALRRIAEGASMIRTKGEPG
MVAN_ORF     GVDMIDESEVLTPADEINHINKKAFTAPFVCGA---RNLGE-ALRRIDEGAAMIRTKGEAG
              .*..****....**    .. *  .  **.**     .  ..  * **. .*    .  .

ATA37        ATGNIAETVKNVRSLMGEVRVLNNMD----DDEVFTFAKKISA-PYDLVAQTKQMGRVPV
HEVERP1A     -TGNVIEAVRHVRSVMGDIRLLRNMD----DDEVFTFAKKIAA-PYDLVMQTKQLGRLPV
STLONG_ORF   STGNVVEAVRHIRSVMGEIRLLRNMD----DDEVFAYAKKISA-AYDLVMQTKQLGRLPV
YAAD_BACSU   -TGNIVEAVRHMRKVNAQVRKVVAMS----EDELMTEAKNLGA-PYELLLQIKKDGKLPV
YFF9_YEAST   -TGDVSEAVKHITKIKAEIQQYKENL--KTESDFAAKATELRV-PVDLLKTTLSEGKLPV
YMR09_YEAS   -TGDVSEAVKHIRRITEEIKACQQ-L--KSEDDIAKVAEEMRV-PVSLLKDVLEKGKLPV
YG47_HAEIN   -TGDIVQAVRHMRMMSQEIRRIQN----LREDELYVAAKDLQV-PVELVQYVHKHGKLPV
MVAN_ORF     -TGNVVEAVKHMRAVNEGIARVVGYHEMGLEAELVQMARNELKVPMEIILEVAKLKRLPV
              **  . ..*.... .    .          .   *     .   .   ..    ...**

                        P                M
ATA37        VQFASGGITTPADAALMMQLGCDGVFVGSEVFDGPDPFKKLRSIVQAVQHYNDPHVLAEM
HEVERP1A     VQFAAGGVATPADAALMMQLGCDGVFVGSGVFKSGDPARRARAIVQAVTHYSDPDMLAEV
STLONG_ORF   VNFAAGGVATPADAALMMQLGCDGVFVGSGVFKSGDPAKRARAIVQAVTHYSDPDLLGRG
YAAD_BACSU   VNFAAGGVATPADAALMMQLGADGVFVGSGIFKSDNPAKFAKAIVEATTHFTDYKLIAEL
YFF9_YEAST   VNFAAGGVATPADAALLMQLGCEGVFVGSGIFKSSDPEKLACAIVEATTHYDNPAKLLQV
YMR09_YEAS   VNFAAGGVATPADAALLMQLGCDGVFVGSGIFKSSNPVRLATAVVEATTHFDNPSKLLEV
YG47_HAEIN   VNFAAGGIATPADAALMMQLGAEGVFVGSGIFKSGDPIKRASAIVKAVTNYRNPQILAQI
MVAN_ORF     VNFAAGGIATPADAALMMQMGCDGVFVGSGIFKSGNPEIRAKAIVEATYNFDKPELIGEV
             *.**.**..*******.**.*..******..*..  *      ..* *    ..     .

ATA37        SSGLENAMESLNVRGDRIQDFG--QGSV------
HEVERP1A     SCGLGEAMVGINLNDKKVERFA--NRSE------
STLONG_ORF   EFGLGEAMVGIIVRMRRLRGTP--IVLNEVI---
YAAD_BACSU   SKELGTAMKGIEISNLLPEQRMQERGW-------
YFF9_YEAST   SSDLGDLMGGISIQSINEAGGK--NGARLSEIW-
YMR09_YEAS   SSDLGELMGGVSIESISHAS----NGVRLSEIGW
YG47_HAEIN   SEDLGEAMVGINENEIQILMAE--RGK-------
MVAN_ORF     SKNLGEAMVGINIDQIPEEMLLAKRGI-------
             .   *.   *  ..
```

B

Fig. 4.

3.4. Deletion mapping of the cryptic promoter in *Ath621-37*

Because both physical mapping and sequencing data indicated that the *aph(3′)II* reporter gene was transcribed in the promoter trap line *Ath621-37* by a cryptic promoter located within the *A37* gene, a series of deletions was constructed to exactly map DNA sequences required for this cryptic promoter activity. *A37* sequences extending from position + 680 to the T-DNA insert junction at position − 302 (Fig. 3) were shown to function as promoter driving *aph(3′)II* and *gus* reporter gene expression in a callus and root specific fashion in *Arabidopsis* [4,9]. In fact, a putative TATA-sequence localized previously in the vicinity of T-DNA integration site proved to be identical with the TATA-consensus found at position − 269 in the *A37* promoter. Thus, sequences of the cryptic promoter appeared to overlap with that of the *A37* promoter in reverse orientation.

A *Hin*dIII–*Bcl*I fragment of 0.9 kb (Fig. 6, H/T) carrying the cryptic promoter between positions + 680 and − 302 was cloned upstream of an *uidA* reporter gene in pPCV812 and transformed into tobacco. As expected, the cryptic promoter drove GUS expression only in roots. The level of GUS enzyme activity conferred by this cryptic promoter was about one order of magnitude lower in comparison to that measured in control plants expressing an *uidA* reporter gene under the control of Cauliflower Mosaic Virus 35S promoter (data not shown). A 5′-deletion shortening the cryptic promoter to position + 465 marked by a *Bam*HI site within the *A37* coding sequence (Fig. 6, B/T) resulted in a reduction of GUS expression by 50%. This deletion removed from the cryptic promoter an *as-1* motif (Fig. 3, TGATCA at position + 620, [25]) and two putative myb-binding sites (at positions + 668 and + 576). Further deletions extending to *Xba*I (at position + 50) and *Ssp*I (at position − 196, Fig. 3) sites in *A37* inactivated the cryptic promoter and reduced GUS expression to a background level (Fig. 6, X/T and Ss/T). Between the *Bam*HI

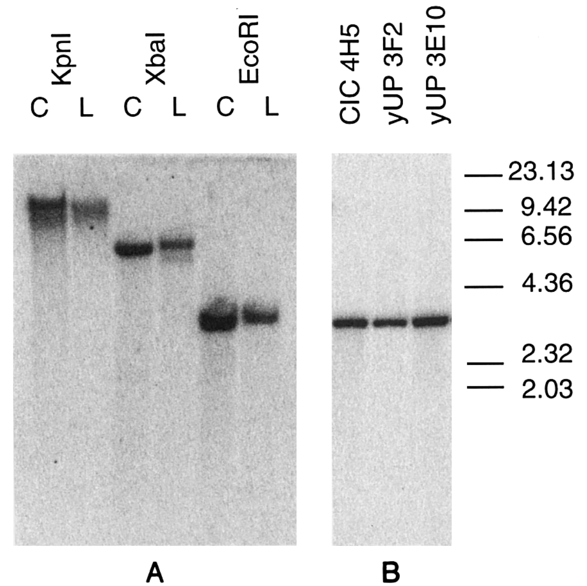

Fig. 5. Confirmation of *A37* chromosomal mapping data by Southern hybridization. (A) Hybridization of *A37* cDNA probe with *Arabidopsis* genomic DNAs isolated from Columbia (Col-1, C) and Landsberg *erecta* (Ler-0, L) ecotypes after digestion with *Kpn*I, *Xba*I, and *Eco*RI. (B) Hybridization of A37 cDNA probe to *Eco*RI-digested DNAs from YAC clones CIC4H5, yUP3F2, and yUP3E10.

Fig. 4. (A) Multiple alignment of predicted A37 amino acid sequence (ATA37) with related plant, yeast and bacterial SNZ-like proteins. A37 shares an amino acid sequence identity of: 62.9% with HEVER1PA (HEVER-1 from *H. brasiliensis*), 60.3% with STLONG-ORF (SLEXORFA-1 from *S. longipes*), 48.9% with YAAD_BACSU (*B. subtilis* SNZ homolog), 48.9% with YFF9_YEAST (Snz3p from yeast), 45.5% with YMR09_YEAST (Snz1p from yeast), 47.7% with YG47_HAEIN (*H. influenzae* SNZ-homolog), and 48.5% with MVAN_ORF (SNZ protein encoded by a gene in the rRNA region of *M. vanielii*). Identical (*) and similar (.) amino acids are indicated, conserved sequence domains are framed. Potential casein kinase II phosphorylation sites (P) and a *N*-myristylation sequence (M) are marked by gray boxes. (B) ClustalW analysis of phylogenetic relationship between SNZ orthologs: *Mycobacterium leprae* MLCL581.12c (EMBL Z96801), *Mycobacterium tuberculosis* MTCY1A10 (EMBL Z95387), *Haemophilus influenzae* YG47-HAEIN (Swissprot P45293), *Saccharomyces cerevisiae* YFF9-YEAST (Swissprot P43545, Snz3p), YN73_YEAST (Swissprot P53824, Snz2p), and YMR09 YEAST (Swissprot Q03148, Snz1p), *Methanococcus jannaschii* MJ0677 (TIGR U67515), *Methanococcus vannielii* MJ0677 (Swissprot Q50841), *Bacillus subtilis* YADD-BACSU (Swissprot P37527), *Schizosaccharomyces pombe* SPAC29B12 (EMBL Z99164), *Stellaria longipes* STLONG ORF (Swissprot Q41348), *Hevea brasiliensis* HEVERP1A (GenBank M88254) and *Arabidopsis thaliana* ATA37 (GenBank AF029980).

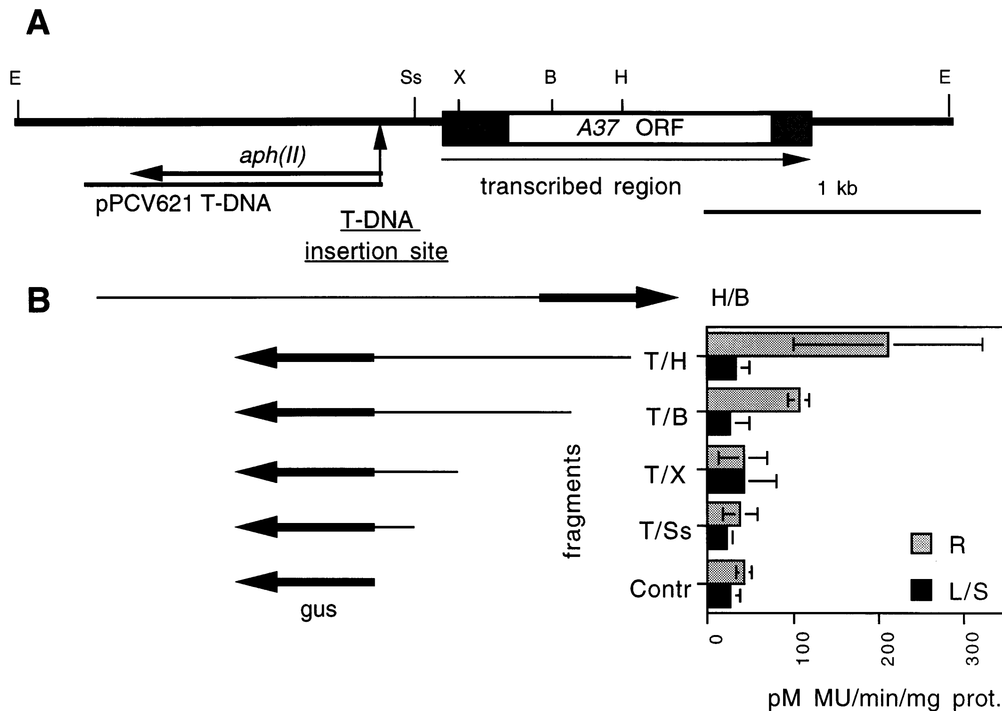

Fig. 6. Deletion mapping of the cryptic promoter region within the *A37* gene. (A) Schematic physical map of the *A37* promoter region in the promoter trap line *Ath621-37*. Abbreviations are identical with those in Fig. 1. (B) Promoter fragments cloned upstream of an *uidA* reporter gene in pPCV812 and their activity in transgenic tobacco plants. A37 promoter: H/B, *Hin*dIII–*Bam*HI fragment. The cryptic promoter constructs carried a common 3′-end corresponding to the T-DNA insertional target site (T), whereas their 5′-ends were marked by *Hin*dIII (T/H), *Bam*HI (T/B), *Xba*I (T/X), and *Ssp*I (T/Ss) restriction endonuclease cleavage sites located in the *A37* gene as described in the text. Bars show average GUS activities measured in root (R) and combined leaf and stem (L/S) tissues of four to six randomly chosen transgenic tobacco plants. GUS enzyme activity units are given in pM MU/min per mg protein units. MU, 4-methylumbelliferone.

and *Xba*I sites in the *A37* transcribed region a single palindromic repeat for an *ocs*-binding site (TGACG, at position +230, Fig. 3) was found which was previously shown to direct root specific expression of reporter genes when placed upstream of a minimal TATA-box promoter [25,26]. This region, including the *ocs* sequence motif just dowstream of the *A37* translational initiation site, appeared thus to be essential for root specific activity of the cryptic promoter identified by the promoter trap T-DNA tag in *Ath621-37*.

4. Discussion

The data described above direct attention to a common problem of in vivo transcriptional gene fusion studies using T-DNA insertional mutagenesis in plants. Previously, we isolated a T-DNA tag derived from the pPCV621 gene fusion vector which has integrated into a transcribed chromosomal locus in the promoter trap *Arabidopsis* line *Ath621-37* [4]. The T-DNA of pPCV621 vector was designed to detect transcriptional gene fusions by monitoring the activation of a promoterless *aph*(*3*′)*II* gene in vivo. In *Ath621-37*, a single T-DNA insertion was observed to result in callus and root specific activation of the *aph*(*3*′)*II* marker gene [4,9]. However, sequence analysis of wild type and T-DNA tagged genomic loci did not confirm the prediction that the *aph*(*3*′)*II* gene was fused to a plant ORF in *Ath621-37*. Instead, the data demonstrated that the T-DNA-linked *aph*(*3*′)*II* reporter gene was transcribed by a cryp-

tic promoter which was mapped to promoter and coding domains of *Arabidopsis* gene *A37* located in the vicinity of T-DNA insertion. Sequences constituting this cryptic promoter contributed to transcription of the *A37* gene with reverse polarity. This data supports the previous observations [7] indicating that the activation of promoterless reporter genes in enhancer traps may not always identify T-DNA tags in transcribed plant genes. Genomic sequences functioning as cryptic promoters were also reported in animals [27]. In contrast to previous studies, our data demonstrate that DNA sequences functioning as a cryptic promoter can be assigned to an actively transcribed chromosomal locus located in the vicinity of a promoter trap T-DNA insert. In case of *Ath621-37*, sequence analysis identified several known *cis*-regulatory elements in the cryptic promoter. Intriguingly, the cryptic promoter appeared to share a common TATA-box with the *A37* promoter that promoted transcription with opposite polarity. Although a G-box sequence [24] occurred in proper distance upstream of the TATA-box, deletion analysis revealed that *cis*-regulatory sequences required for root specific activity of the cryptic promoter, including a consensus *ocs* element [25,26], were in fact located within the coding region of *A37* gene.

Further characterization of the *A37* gene and derived cDNAs showed that *A37* codes for an ortholog of SNZ proteins in *Arabidopsis*, the sequence of which is conserved between plants, yeast, bacteria and archaebacteria. Conservation of sequence and structure of SNZ homologs suggests that the A37 protein performs an essential function. Synthesis of SNZ proteins in budding yeast is induced in cells entering the stationary phase, whereas an SNZ homolog in *Bacillus subtilis* plays a role during initiation of sporulation. It is thus probable that the function of conserved SNZ proteins is related to an ancient growth regulatory pathway controlling cellular responses to nutrient limitation [10]. *HEVER-1*, a gene coding for an SNZ protein in rubber tree, is known to be activated by ethylene and salicylic acid during senescence and in response to pathogenic infection [21]. Similarly to SNZ proteins of yeast, HEVER1-homologs are encoded by a gene family in rubber tree. Although *A37* appears to be a single copy gene, the occurrence of homologous EST sequences in the database suggests that SNZ-like proteins may also encoded by a small divergent gene family in *Arabidopsis*. Although the function of A37 protein remains elusive, it is possible that SNZ homologs in *Arabidopsis*, rubber tree and other plants perform a similar function as their counterparts in prokaryotes and yeast. Further elucidation of the function of A37 protein is therefore expected to answer many intriguing questions about the role played by SNZ-proteins in unicellular and multicellular organisms.

Acknowledgements

This work was supported by OTKA (Grant: T13182), the TéT foundation (M-5/94), the EU Phare-Accord programme (Grant: 9112-0148), and the DFG (Grant: 436 UNG 113/117, Sche 220/5-1) in frame of scientific exchange agreement between the Max-Planck Institut für Züchtungsforschung (Cologne, Germany) and the Biological Research Center (Szeged, Hungary). L. Ökrész was supported by the Ph.D. program of József Attila University, Szeged, Hungary.

References

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

[2] C. Koncz, K. Németh, G.P. Rédei, J. Schell, T-DNA insertional mutagenesis in *Arabidopsis*, Plant. Mol. Biol. 20 (1992) 963–976.

[3] S. Kertbundit, H. De Greeve, F. Deboeck, M. Van Montagu, J.-P. Hernalsteens, In vivo random β-glucuronidase gene fusions in *Arabidopsis thaliana*, Proc. Natl. Acad. Sci. USA 88 (1991) 5212–5216.

[4] C. Koncz, N. Martini, R. Mayerhofer, et al., High-frequency T-DNA-mediated gene tagging in plants, Proc. Natl. Acad. Sci. USA 86 (1989) 8467–8471.

[5] C. Koncz, N. Martini, L. Szabados, et al., Specialized vectors for gene tagging and expression studies, in: S. Gelvin, R. Schilperoort (Eds.), Plant Molecular Biology Manual, vol. B2, Kluwer Academic, Dordrecht, The Netherlands, 1994, pp. 1–22.

[6] T.H. Teeri, L. Herrera-Estrella, A. Depicker, M. Van Montagu, E.T. Palva, Identification of plant promoters in situ by T-DNA-mediated transcriptional fusions to the npt-II gene, EMBO J. 5 (1986) 1755–1760.

[7] P.R. Fobert, H. Labbe, J. Cosmopoulos, et al., T-DNA tagging of a seed-coat-specific cryptic promoter in tobacco, Plant J. 6 (1994) 567–577.

[8] K. Lindsey, W. Wei, M.C. Clarke, et al., Tagging genomic sequences that direct transgenic expression by activation of a promoter trap in plants, Transgenic Res. 2 (1993) 33–47.

[9] R. Mayerhofer, Z. Koncz-Kálmán, C. Nawrath, et al., T-DNA integration: a mode of illegitimate recombination in plants, EMBO J. 10 (1991) 607–704.

[10] E.L. Braun, E.K. Fuge, P.A. Padilla, M. Werner-Washburne, A stationary-phase gene in *Saccharomyces cerevisiae* is a member of a novel, highly conserved gene family, J. Bacteriol. 178 (1996) 6865–6872.

[11] C. Nawrath, J. Schell, C. Koncz, Homologous domains of the largest subunit of eukaryotic RNA polymerase II are conserved in plants, Mol. Gen. Genet. 223 (1990) 65–75.

[12] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[13] S.L. Dellaporta, J. Wood, J.B. Hicks, A plant DNA mini-preparation: version II, Plant. Mol. Biol. Rep. 1 (1983) 19–21.

[14] F. Creusot, E. Fouilloux, M. Dron, et al., The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*, Plant J. 8 (1995) 763–770.

[15] J.R. Ecker, PFGE and YAC analysis of the *Arabidopsis* genome, Methods 1 (1990) 186–194.

[16] E. Matallana, C.J. Bell, P.J. Dunn, M. Lu, J.R. Ecker, Genetic and physical linkage of the *Arabidopsis* genome: Methods for anchoring yeast artificial chromosomes, in: C. Koncz, N.-H. Chua, J. Schell (Eds.), Methods in *Arabidopsis* Research, World Scientific, Singapore, 1992, pp. 144–169.

[17] S. Gibson, C. Somerville, Chromosome walking in *Arabidopsis thaliana* using yeast artificial chromosomes, in: C. Koncz, N.-H. Chua, J. Schell (Eds.), Methods in *Arabidopsis* Research, World Scientific, Singapore, 1992, pp. 119–143.

[18] K. Pawlowski, R. Kunze, S. de Vries, T. Bisseling, Isolation of total poly(A) and polysomal RNA from plant tissues, in: S. Gelvin, R. Schilperoort (Eds.), Plant Molecular Biology Manual, vol. D5, Kluwer Academic, Dordrecht, The Netherlands, 1994, pp. 1–13.

[19] R. Walden, C. Koncz, J. Schell, The use of gene vectors in plant molecular biology, Methods Mol. Cell. Biol. 1 (1990) 175–194.

[20] R.A. Jefferson, T.A. Kavanagh, M.W. Bevan, GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants, EMBO J. 6 (1987) 3901–3907.

[21] S. Sivasubramaniam, V.M. Vanniasingham, C.-T. Tan, N.-H. Chua, Characterization of *HEVER*, a novel stress-induced gene from *Hevea brasiliensis*, Plant Mol. Biol. 29 (1995) 173–178.

[22] G. Wich, M. Jarsch, A. Böck, Apparent operon for a 5S ribosomal RNA gene and for tRNA genes in the archaebacterium *Methanococcus vannielii*, Mol. Gen. Genet. 196 (1984) 146–151.

[23] E. Lam, P.N. Benfey, P.M. Gilmartin, R.-X. Fang, N.-H. Chua, Site-specific mutations alter in vivo factor binding and change promoter expression pattern in transgenic plants, Proc. Natl. Acad. Sci. USA 86 (1989) 7890–7894.

[24] A.E. Menkens, U. Schlindler, A.R. Cashmore, The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins, Trends Biochem. Sci. 20 (1995) 506–510.

[25] P.C. Fox, V. Vasil, I.K. Vasil, W.B. Gurlrey, Multiple *ocs*-like elements required for efficient transcription of the mannopine synthase gene of T-DNA in maize protoplasts, Plant Mol. Biol. 20 (1992) 219–233.

[26] H. Fromm, F. Katagiri, N.-H. Chua, An octopine synthase enhancer element directs tissue-specific expression and binds ASF-1, a factor from tobacco nuclear extracts, Plant Cell. 1 (1989) 977–984.

[27] R. Al-Shawi, J. Burke, H. Wallace, et al., The herpes simplex virus type 1 thymidine kinase is expressed in the testes of transgenic mice under the control of a cryptic promoter, Mol. Cell. Biol. 11 (1991) 4207–4216.