 Cytometry 33:460–468 (1998)

Nuclear Expressed Sequence Tag (NEST) Analysis: A Novel Means to Study Transcription Through Amplification of Nuclear RNA

Jiri Macas,[1,2] Georgina M. Lambert,[1] David Dolezel,[1,3] and David W. Galbraith[1]*

[1]Department of Plant Sciences, University of Arizona, Tucson, Arizona
[2]Institute of Plant Molecular Biology, Academy of Sciences of the Czech Republic, Ceske Budejovice, Czech Republic
[3]University of Southern Bohemia and Institute of Parasitology AS CR, Ceske Budejovice, Czech Republic

Received 17 March 1998; Revision Received 2 July 1998; Accepted 29 July 1998

We describe a novel concept and corresponding methods for the analysis of transcription in higher plant cells. The concept is that an examination of the presence of different polyadenylated transcripts within isolated nuclei reflects the state of gene expression at a given moment more precisely than do conventional techniques using total cellular mRNA. The methods involve isolation of polyadenylated nuclear transcripts from flow-sorted nuclei, reverse transcription, amplification using the polymerase chain reaction, and analysis of the products through gel electrophoresis and sequencing. By using specific primers, we have demonstrated detection of selected gene products in nuclei from transgenic plants. We also employed a technique for analysis of individual transcripts based on the length polymorphisms of restriction fragments derived from their 3′ ends. Because the technique does not require a priori knowledge of the analyzed sequences, it is suitable for displaying the complete spectra of RNA transcripts present in nuclei at the moment of their isolation. These fragments can be easily isolated and sequenced and the sequence information used for assignment of putative function of corresponding genes. These techniques have been used to identify leaf-, root-, and cell cycle-specific transcripts. In principle, they should be applicable to the tissues of any eukaryotic species that contain transcriptionally active nuclei. Cytometry 33:460–468, 1998. © 1998 Wiley-Liss, Inc.

Key terms: flow cytometry; cell sorting; nuclear mRNA; gene expression; nuclei; transcription; polymerase chain reaction

It is accepted that an understanding of gene expression is central to elucidating the relationship between the phenotype of an organism and the genes that it contains. Gene expression describes a complex, multipart process in which the information contained within the genome is manipulated by the organism to ultimately give rise to a measurable phenotype. It involves several major steps, including transcription and transcript processing, mRNA export from the nucleus, movement of mRNA to appropriate sites of translation, translation, protein processing and transport, and activation/inhibition of enzymatic functions (5). A common and important means to control gene expression involves regulation of the magnitude of transcription. There is historical emphasis on the use of mRNA concentrations in cells or tissues as a measurement of gene transcription; in part this has been due to the successful development of accurate and sensitive methods to estimate mRNA types and concentration. Because factors other than transcription rate are also involved in regulating the concentrations of cellular mRNA (5,7,31), there has been interest in the measurement of gene expression at a level that more closely tracks transcription per se. In particular, methods for analysis of run-off transcription have been widely used (14,19,30). However, nuclear run-off transcription as a method suffers from several drawbacks. The principal of these are requirements for large numbers of transcriptionally active nuclei, for high amounts of radioactivity, and for the prior availability of cloned genes for analysis.

One of the research questions being addressed in our laboratory concerns defining the spectrum of genes that are transcriptionally active in different cell types. We are particularly interested in the analysis of this process at a

Sequence deposition: dbEST accession numbers 1168193–1168208 and GenBank accession numbers AA523553–AA523568.

Contract grant sponsor: U.S.D.A.–N.R.I.C.G.P. (Plant Genome); Contract grant sponsor: D.O.E.–U.R.I.; Contract grant sponsor: NSF; Contract grant number: STI 96–20092.

*Correspondence to: David W. Galbraith, University of Arizona, 303 Forbes, Tucson, AZ 85721.
E-mail: galbraith@arizona.edu

point that is as close as possible to the act of transcription. In the present report, we outline a method, based on flow cytometric analysis and sorting of nuclei, followed by amplification of polyadenylated nuclear RNAs, that provides a quick and facile means for the analysis of these transcripts. Because the amplification is performed in a sequence-independent manner similar to amplified fragment length polymorphism (AFLP) (2) and differential display (16,17), it can detect a wide variety of transcripts with very high sensitivity. Our technique requires relatively few nuclei and no radioactivity and provides for the ultimate analysis of individual transcripts through direct cloning and/or sequencing. Because the cloned transcripts provide expressed sequence tags, we have termed this method nuclear expressed sequence tag (NEST) analysis. In principle, NEST analysis is applicable to any eukaryotic organism or cell type that contains transcriptionally active nuclei. The ways in which flow cytometry and sorting can be combined with NEST analysis is discussed in terms of the analysis of gene expression during development and in response to biotic and abiotic stimuli.

MATERIALS AND METHODS

Enzymes and Nucleic Acid Manipulations

Unless stated otherwise, all enzymes were purchased from Gibco BRL (Gaithersburg, MD) and used according to the manufacturer's recommendations. Basic nucleic acid manipulations were done according to the method of Sambrook et al. (27). The solutions used for RNA isolation were treated with diethylpyrocarbonate (DEPC) as described by Sambrook et al. (27).

Flow Cytometric Analysis and Sorting of Nuclei

Nuclei were isolated from leaves or roots of tobacco (*Nicotiana tabacum* L. Xanthi) plants cultivated in vitro in Murashige-Skoog media (22). Three-day-old seedlings of pea (*Pisum sativum* L. Little Marvel) grown in liquid Hoagland medium (11) were employed for isolation of nuclei from root apical meristems. The nuclei were released by chopping the tissues in isolation buffer (3) at 0°C, filtered through 60-µm nylon mesh, and stained with propidium iodide (PI) at final concentration of 50 µg/ml. Flow cytometric analysis and sorting involved a Coulter Elite flow cytometer, using DEPC-treated isolation buffer lacking Triton X-100 as sheath fluid, according to the general methods described by Galbraith and Lambert (10).

Preparation of cDNA

All experimental manipulations were done at room temperature (ca. 23°C) unless otherwise noted. Sorted nuclei were collected by centrifugation at 600*g* for 7 min at 0°C. Most of the supernatant was removed, leaving approximately 50 µl containing the sedimented nuclei. Isolation and purification of polyadenylated RNA was done using the Dynabeads mRNA Direct Kit (Dynal, Oslo, Norway). First, the nuclei were lysed by addition of 300 µl of lysis/binding buffer (100 mM Tris-HCl, pH 8.0, 500 mM LiCl, 10 mM EDTA, 1% lithium dodecyl sulfate [LiDS], and 5 mM dithiothreitol). Ten units of RNase inhibitor were added to the samples, and the lysate was vortexed and incubated for 5 min. The samples were centrifuged for 3 min in an Eppendorf microcentrifuge, and 300 µl of the supernatant were transferred into a new tube containing 0.25 mg of Dynabeads Oligo$(dT)_{25}$. After incubation for 5 min, the beads were retrieved using the magnetic particle concentrator and washed three times with 200 µl washing buffer (100 mM Tris-HCl, pH 8.0, 150 mM LiCl, 1 mM EDTA) supplemented with 0.1% LiDS and twice with 200 µl washing buffer. Polyadenylated mRNA retained on the beads was eluted into 20 µl of DEPC-treated water, mixed with 10× DNase buffer, and incubated with 1 unit of DNAase I (amplification grade) for 15 min. After addition of 2 µl of 25 mM EDTA and heat inactivation of the DNAase, the samples were mixed with 130 µl of lysis/binding buffer and 0.25 mg Dynabeads Oligo$(dT)_{25}$ and incubated for 5 min. Samples were washed twice in 200 µl of washing buffer containing 0.1% LiDS, once in 200 µl of washing buffer, and once in 100 µl of polymerase chain reaction (PCR) buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl) supplemented with 2.5 mM $MgCl_2$. Reverse transcription was done using the SuperScript Preamplification System for First-Strand cDNA Synthesis (Gibco BRL) for 50 min at 42°C. Before adding reverse transcriptase, each sample was divided into two equal parts, one of which served as a control from which the reverse transcriptase was omitted. After completion of first cDNA strand synthesis, aliquots were taken for PCR detection of transcripts by using sequence-specific primers. If intended for NEST analysis, the samples were resuspended in 50 µl of the second-strand cDNA synthesis mix (20 mM Tris-HCl, pH 7.4, 50 mM KCl, 5 mM $MgCl_2$, 5 mM $[NH_4]_2SO_4$, 0.15 mM dNTPs, 0.5 unit RNAase H, 10 units DNA polymerase I) and were incubated for 90 min at 16°C and then for 10 min at 65°C. The samples were washed twice in 100 µl of TET (10 mM Tris-HCl, 1 mM EDTA, pH 8.0 [TE buffer], supplemented with 0.1% Tween 20), once in 100 µl TE, once in 30 µl of *Escherichia coli* ligase buffer, and incubated overnight with 2.5 units of *E. coli* ligase at 4°C.

PCR Detection of Specific Transcripts

For detection of the NLS-GFP-GUS (13) transcript, 1-µl aliquots of reverse transcription mix, taken after first cDNA strand synthesis (from a total volume of 20 µl), were used for PCR amplification. The reaction mix (30 µl volume) contained 1× PCR buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 1.2 units of Taq polymerase, and 0.2 µM each of primers GUS-EF2 (5′ AAG CAA CGC GTA AAC TCG AC 3′) and GUS-ER2 (5′ AGG CTG TAG CCG ACG ATG 3′). The primers amplify a 564-bp DNA fragment from the β-glucuronidase (GUS) coding sequence. PCR involved 35 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min. Cycling was preceded by an initial denaturation step (94°C, 2 min) and followed by a final extension step (72°C, 15 min). The amplification products were then analyzed by agarose (1.5%) gel electrophoresis.

NEST Analysis

Double-stranded cDNA attached to the beads was digested by the addition of 10 units of *Nla*III (New England Biolabs, Beverly, MA) for 30 min. The enzyme was heat inactivated, and the beads were washed twice in 100 µl TET, once in 100 µl TE, and once in 50 µl of ligase buffer. The 3′ cDNA ends remaining on the beads were then ligated to a linker prepared by annealing of oligonucleotides FNla-U (5′ TTT TGC AGC TTA TTC AAT TCG GTC TGG ATG CAT G 3′) and FNla-LP (5′ CAT CCA GAC CGA ATT GAA TAA GCT GCA 3′, 5′ phosphorylated). The ligation was performed in 25 µl of the ligation mix (1× reaction buffer, 10 µM linker, 5 U of T4 DNA ligase) for 3 h at 16°C. After washing twice in 100 µl TET and once in 100 µl TE, the samples were resuspended in 15 µl of TE and stored at 4°C until used for PCR amplification.

The PCR amplification consisted of two steps. First, 3 µl aliquots of the samples were preamplified by using the linker-derived primer LD-1 (5′ GCA GCT TAT TCA ATT CGG TC 3′) and one of the GT24 primers (GT24-A, GT24-C, or GT24-G) in three separate reactions. All GT24 primers have the same sequence $G(T)_{24}$, except for the last (3′) selective base. The reaction was performed in a final volume of 50 µl composed of 1× PCR buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM primers, and 2.5 units of Taq polymerase and involved the hot-start technique using AmpliWax beads (Perkin-Elmer, Oakbrook, IL) instead of a mineral oil overlay. After an initial denaturation at 94°C for 2 min, 17 PCR cycles were done. Cycles consisted of incubation at 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min. Cycling was followed by incubation at 72°C for 15 min. Second, the samples were then diluted 10-fold in TE buffer, and 3-µl aliquots were used for a second round of amplification using primers with two selective bases. The reaction mix (30 µl) was the same as that used for preamplification, except that it contained primers Nla-XX (5′ GTC GGT CTG GAT GCA TGX X 3′, where XX represent selective bases extending into the amplified fragments) and GT24-XX ($G[T]_{24}$ with two selective bases). The selective base at the first position of the GT24-XX primer was the same as that employed in the GT24-X primer used in preamplification. Both the Nla-XX and GT24-XX primers contain a G residue at the 5′ ends to avoid appearance of double bands from individual PCR products caused by incomplete addition of an extra nucleotide due to the terminal transferase activity of Taq polymerase (33). The PCR cycling profile for second-round amplification was identical to that used for preamplification, except for the annealing temperature over the first 20 cycles, which was 65°C for first two cycles but was subsequently decreased by 1°C every two cycles (''touchdown'' PCR) (8). Amplification then proceeded for another 20 cycles with an annealing temperature of 55°C. The samples were mixed with an equal volume of formamide loading buffer (27), subjected to electrophoresis on 6% denaturing polyacrylamide gels, and visualized by silver staining using Silver Sequence DNA Staining Reagents (Promega, Madison, WI). The stained gels were photographed using Kodak Electrophoresis Duplicating Film (Eastman Kodak, Rochester, NY).

Sequence Analysis of NESTs

Silver-stained gels were soaked in distilled water for 30 min. Selected bands were then excised with a sterile scalpel and stored in 10 µl of sterile water in −20°C until use. Part of the excised gel was transferred into a total volume of 50 µl PCR mix (1× PCR buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM respective primers, and 2 units of Taq polymerase). PCR amplification involved 15 or 20 cycles (94°C for 1 min, 55°C for 1 min, and 72°C for 1 min), accompanied by prior denaturation and final extension steps as described above. Amplification of single products was confirmed by electrophoresis on 1.5% agarose and 6% polyacrylamide gels. The products were purified with the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and were sequenced automatically (cycle sequencing system from Applied Biosystems, Fost City, CA) by using the appropriate Nla-XX oligonucleotide as the sequencing primer. The NEST sequences were searched for homologies with nucleic acid and protein databases with the FASTA (24) (European Bioinformatics Institute, http://www.ebi.ac.uk) and BLAST (1) (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov) algorithms.

RESULTS

For setting up and trouble-shooting NEST analysis, nuclei were isolated from leaves and roots of tobacco plants transformed with a chimeric construct comprising the coding sequence of the green-fluorescent protein (GFP), translationally fused to a nuclear localization se-

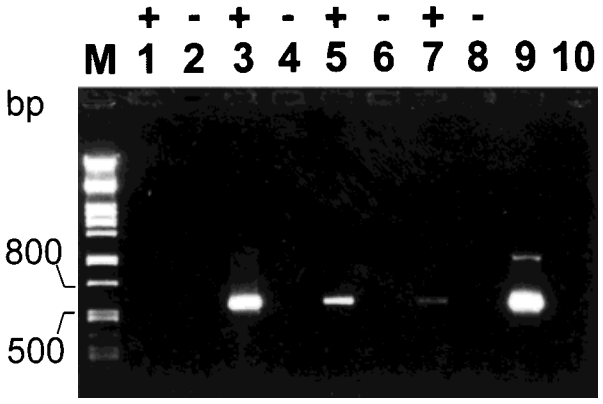

FIG. 1. Detection of the NLS-GFP-GUS transcript in flow-sorted tobacco nuclei by polymerase chain reaction (PCR) using sequence-specific primers. **Lanes 1 and 2** contain samples prepared from 10^5 nuclei from a nontransgenic (wild-type) plant; **lanes 3–8** contain samples prepared from different numbers of nuclei from transgenic plants expressing the NLS-GFP-GUS chimeric gene: 10^5 nuclei (lanes 3, 4), 10^4 nuclei (lanes 5, 6), and 10^3 nuclei (lanes 7, 8); **lane 9** contains 0.2 ng of a plasmid (pBGF-0), containing the cloned GUS sequence (positive PCR control); **lane 10** contains no DNA (negative PCR control). The samples marked with plus signs were initially subjected to reverse transcription; reverse transcription was omitted from samples marked with minus signs. The amplification products were resolved on a 1.5% agarose gel; **lane M** contains λDNA digested with *Pst*I.

quence (NLS) and to the *E. coli* GUS gene, transcriptionally regulated by the CaMV 35S promoter (13). These plants accumulate the chimeric NLS-GFP-GUS protein within the nuclei of most cells (12). Whereas leaf and root tissues were expected to differ in a variety of transcripts, the same tissues obtained from nontransformed tobacco plants provided experimental material theoretically lacking only the transgene-coded mRNA.

For isolation of nuclear transcripts, nuclei were purified from plant tissue homogenates by flow cytometry and sorting, based on fluorescence emission following staining with PI. The nuclei were then lysed to release the nascent RNA transcripts, which were immediately captured on oligo-dT-linked magnetic beads and converted into double-stranded cDNA (see Materials and Methods). The first series of experiments was done to optimize these steps through the detection of specific transcripts within the nuclei. The second series of experiments then allowed analysis of the spectrum of transcripts contained by the nuclei (NEST analysis).

Detection of Specific Nuclear Transcripts

To assess the quality of the RNA isolated from flow-sorted nuclei, experiments were done to determine whether detection could be achieved of the NLS-GFP-GUS transcript. The cDNA prepared from RNA of nuclei from transgenic plants expressing the NLS-GFP-GUS gene fusion or from nontransgenic controls was used as a template for PCR amplification by employing primers that were specific for a 564-bp portion of the GUS sequence. The products were then analyzed by agarose gel electrophoresis (Fig. 1). A DNA fragment of the appropriate length was observed, being only present in lanes containing transgenic samples that had been initially subjected to reverse transcription. The fragment was absent from lanes containing nontransgenic samples (Fig. 1, lanes 1, 2). To determine the limit of sensitivity, various numbers of nuclei were sorted and the resultant cDNAs amplified (Fig. 1, lanes 3–8). It appears that as few as 1,000 sorted nuclei provide sufficient RNA to be detected by this approach; the amplification requires only 35 PCR cycles.

The NEST Principle

The principle underlying NEST analysis involves the following steps (Fig. 2). (a) Nuclei are flow sorted, and polyadenylated nuclear transcripts are captured on magnetic beads via hybridization to poly-dT oligonucleotides that subsequently serve as primers for reverse transcription. (b) Conversion of mRNA to double-stranded cDNA results in corresponding sequences covalently attached to the beads. (c) After digesting the cDNA with a frequently cutting restriction enzyme, the resultant fragments of 3′ ends remaining attached to the beads are ligated to synthetic linkers. (d) These fragments are PCR amplified by using primers derived from the linker and poly-dT sequences. (e) Because the distance of the restriction site from the transcript end differs between individual transcripts, each cDNA results in a characteristic fragment of defined length that can be identified by using analytical polyacrylamide gel electrophoresis.

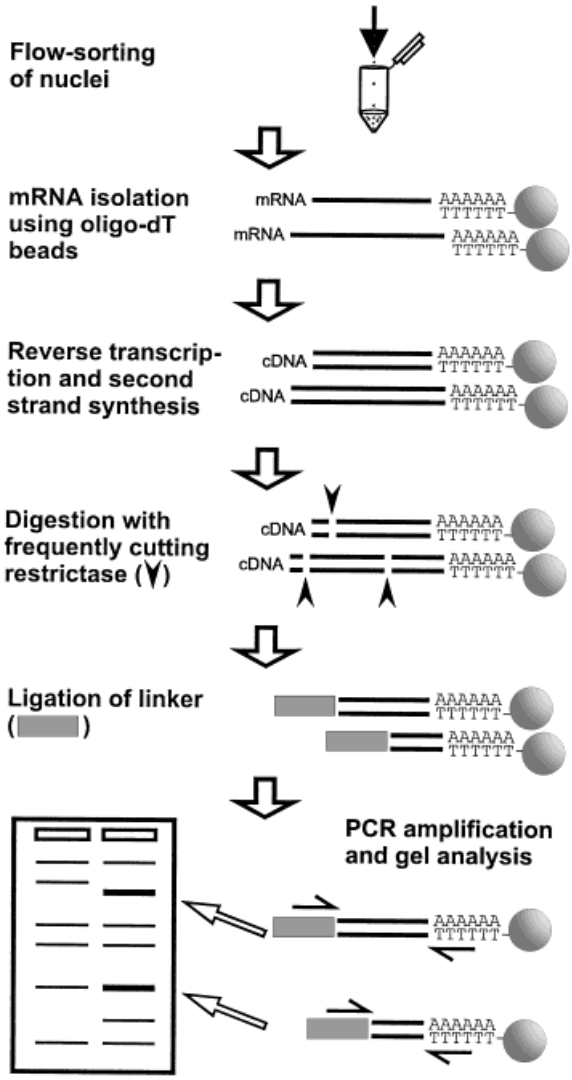

FIG. 2. The NEST principle.

NEST Analysis of the Spectrum of Transcripts Within Nuclei

The bead-linked cDNA was subjected to restriction digestion by using the frequently cutting enzyme *Nla*III to leave short diagnostic stretches of cDNA covalently attached to the magnetic beads. Linkers were then ligated to the free ends of the cDNA. PCR amplification of resultant recombinant molecules involved two steps. First, the total cDNA was preamplified in three separate reactions, each using the linker-based primer and one of three poly-dT primers having a *one-base* (A, G, or C) extension at the 3′ ends. The one-base extension was necessary to ''anchor'' the primer and to prevent smears generated by amplification (smearing occurred when homopolymeric poly-dT was used as the primer; data not shown). Second, the

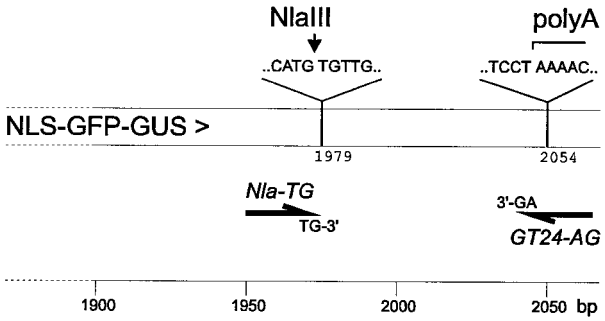

FIG. 3. Schematic representation of the 3′ end of the NLS-GFP-GUS gene. The positions and adjacent sequences of the polyadenylation site and the upstream *Nla*III site are indicated (base pairs are numbered with respect to the first base of NLS-GFP-GUS start codon, i.e., 1). Two specific NEST primers (shown as arrows) are required for amplification of the corresponding transcript (primers Nla-TG and GT24-AG; their selective 3′ ends are indicated). The resulting fragment (117 bp) is observed on the sequencing gel (Fig. 4A).

products of the first reaction were diluted, followed by reamplification with different primer sets, all of which contained *two-base* 3′-extensions. This approach, also used in the AFLP technique (33), leads to selective amplification of a subset of the original population of DNA fragments and thereby simplifies analysis. For analytical purposes, the DNA fragments were separated on large polyacrylamide gels and their positions were identified with silver staining.

Most of the amplified fragments were 50–400 bp in length. The band patterns were highly reproducible when comparing separate amplifications of the same samples and different amplifications using separate samples from the same type of tissue and were specific for each primer combination (Fig. 4). The amplification patterns were also stable when different numbers (50,000–500,000) of nuclei were used (data not shown). On average, 50–100 scorable bands were produced by individual primer combinations when the samples were subjected to reverse transcription. The control samples processed without treatment with reverse transcriptase typically produced no bands, except for PCR amplification artifacts (primer dimers) that did not exceed 50–70 bp in length.

To test whether the method could be employed to detect differential transcription of a single gene, we compared samples of leaf nuclei isolated from control plants and from transformed plants expressing NLS-GFP-GUS (12). In this transgene, the position of polyadenylation is defined by the CaMV terminator sequence (28). This sequence predicted the size of the fragment that would be amplified from the 3′ end of transgene mRNA; because the first *Nla*III site is 75 bp upstream from the site of polyadenylation, the fragment length, including the primers, should be 117 bp (Fig. 3). A product of this predicted size was observed for the transgenic sample and was absent from the control (nontransgenic) sample (Fig. 4A). Amplification of this band only occurred for the combination of 3′-selective bases predicted by the genomic sequences adjacent to the *Nla*III and polyA sites. The double band presumably represents the two antiparallel strands of the DNA molecule migrating in sequence-dependent manner, a phenomenon typically observed for gel electrophoresis of nucleic acids under denaturing conditions.

Comparison of lanes containing samples from leaf and root nuclei demonstrated many tissue-specific bands. However, the great majority of bands were present in both tissues, suggesting that they represent constitutively expressed, house-keeping genes (Fig. 4B). We tested 36 primer combinations, which produced 37 leaf-specific and 40 root-specific bands (representing about 1–1.5% of all bands scored). Because tobacco roots contain large proportions of nuclei in both G_1 and G_2 phases of the cell cycle (9), we were able to analyze these classes of nuclei separately following flow sorting. A number of transcripts specific for the nuclei in either G_1 or G_2 phase could be identified, as could a majority that was present in both classes (Fig. 4B).

In the next series of experiments, we analyzed G_1 and G_2 nuclei flow sorted from apical root meristems of pea (*Pisum sativum* L.). As expected, the NESTs generated from these samples showed considerably fewer differentially expressed transcripts as compared with the experiments using leaf and root tissues (Fig. 4D). Only two G_1-specific and no G_2-specific transcripts were found after screening 60 primer combinations (of a possible total of 192 combinations).

Sequence Analysis of NESTs

To characterize the sequence composition of selected bands efficiently, we first optimized the procedure for their isolation from the gel and subsequent sequencing. The bands were readily excised from silver-stained polyacrylamide gels and could be reamplified by using 15 or 20 PCR amplification cycles. Part of the reaction mixes were then reanalyzed on polyacrylamide gels to confirm amplification of single fragments (Fig. 4C). The remainder was purified and used for dideoxy sequencing by using linker-derived primers. Preliminary experiments indicated that direct sequencing of fragments produced data identical to those by sequencing of the same fragments after cloning but in a much shorter period of time (data not shown). As a control, analysis of the band presumably derived from transcription of the NLS-GFP-GUS transgene (Fig. 4A, lane 7) provided a sequence identical to that known for the 3′ end of this transgene (data not shown).

By using these experimental procedures, several fragments corresponding to the tissue or cell cycle-specific transcripts were sequenced. Sequences 66–219 bp in length were obtained. All sequences contained the appropriate 3′ polyA-tail, as would be predicted for amplification when using the Nla-XX and GT24-XX primers. Details of the sequence analyses are summarized in Table 1. For 16 NESTs that were tested, three displayed significant homology to sequences in the databases. Leaf-specific transcript 0521–1 showed a very high homology to spinach gene *PsbW*, from which is transcribed a nuclear-encoded protein that forms part of the photosystem II reaction center

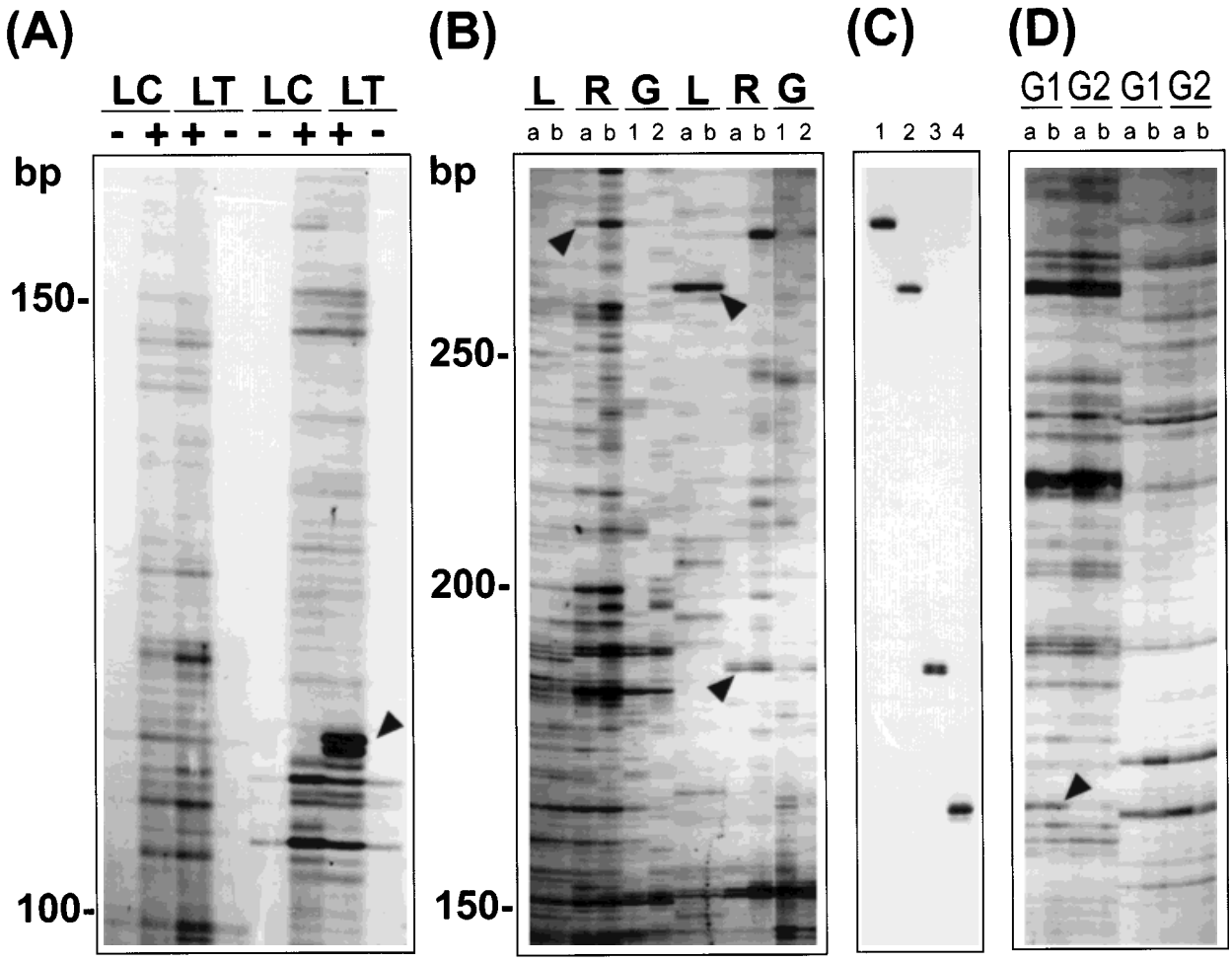

FIG. 4. Nuclear expressed sequence tag (NEST) analysis. **A:** Detection of the NEST corresponding to the transcript from the NLS-GFP-GUS transgene in leaf nuclei. LC, control plant; LT, transformed plant; plus and minus signs indicate whether or not, respectively, samples were subjected to reverse transcription. The primer combinations used for amplification were GT24-GA and Nla-TG (lanes 1–4) and GT24-AG and Nla-TG (lanes 5–8). The transgene NEST is indicated by the arrowhead. **B:** Comparison of leaf and root NESTs. Samples of leaf (L) and root (R) nuclei were prepared in two independent experiments, designated as a and b. The root samples contain nuclei in both G_1 and G_2 of the cell division cycle. G_1 and G_2 nuclei were sorted according to DNA content and analyzed separately (lanes designated G1 and G2, respectively). Primers used for the amplifications were GT24-GC and Nla-AA (lines 1–6) and GT24-AC and Nla-AA (lines 7–12). Some tissue-specific bands, individually marked by the arrowheads, were isolated from the gel. They were reamplified and are shown in C. **C:** Analysis of individual NESTs excised from the gels shown in B and D after reamplification. **D:** Analysis of NESTs from pea root meristem nuclei sorted according to DNA content. Samples are shown from two independent experiments, designated a and b, comparing G_1 (G1) and G_2 (G2) nuclear transcripts. The amplification primers were GT24-GC and Nla-TC (lanes 1–4) and GT24-CA and Nla-AA (lanes 5–8). The fragment differentially expressed in G_1 nuclei (arrowhead) is shown after isolation and reamplification in C, lane 4.

(18). Two other leaf-specific NESTs, 0622–5 and 0622–7, showed homology to glycine-rich proteins (GRP) from tobacco (32) and tomato (29), respectively. These NESTs are also homologous to each other (72% identity), suggesting that they belong to the same gene family. Because putative upstream regulating sequences homologous to those of the ribulose bisphosphate carboxylase small subunit gene have been found in tobacco GRP genes (32), it is likely that these are expressed in a leaf-specific manner, consistent with their NEST assignment.

The expression patterns of two newly isolated transcripts (root-specific 05093 and leaf-specific 05211; Table 1) were also examined using reverse transcriptase–PCR with specific primer pairs based on their NEST sequences. RNA isolated from flow-sorted nuclei or total cellular polyA$^+$ mRNA was used as a template. In both cases, the reverse transcriptase–PCR confirmed the results of NEST analysis in showing preferential expression of 05093 in roots and 05211 in leaves, respectively (Fig. 5).

DISCUSSION

The ability to measure transcriptional activities is critical toward developing an understanding of the control mechanisms that govern gene expression in eukaryotic cells. Much information has come from measurement of the types and amounts of mRNAs that accumulate within the cell. Such measurements predominantly report the cytoplasmic contribution because this is the location of the bulk of the cellular mRNA (4). Because cytoplasmic

Table 1
*Summary of Sequence Information From Individual Nuclear Expressed Sequence Tags (NESTs)**

Species	NEST ID	Leaf	Root			Sequence length (bp)	Homology			
			W	G_1	G_2		Length[a]	%[b]	Accession number[c]	Description[d]
Tobacco	0509-2	+	−	−	−	107				
	0521-1	+	−	−	−	219	150	83	X85038	*S. oleracea* 6.1-kDa polypeptide of photosystem II
	0521-2	+	−	−	−	219				
	0618-2	+	−	−	−	218				
	0622-5	+	−	−	−	135	94	72	M37152	*N. tabacum* glycine-rich protein
	0622-7	+	−	−	−	90	71	71	X55689	*L. esculentum* glycine-rich protein
	0509-1	−	+	n.t.	n.t.	147				
	0509-3	−	+	n.t.	n.t.	170				
	0513-2	−	+	n.t.	n.t.	66				
	0516-1	−	+	+	+	125				
	0516-2	(+)	+	n.t.	n.t.	100				
	0516-3	−	+	+	+	121				
	0521-3	−	+	n.t.	n.t.	134				
	0618-6	−	+	+	+	128				
	0622-6	−	+	+	+	105				
Pea	0513-3	n.t.	n.t.	+	−	117				

*The presence or absence of corresponding transcripts in specific individual tissues is indicated by + or −, respectively (n.t., not tested). Nuclei from tobacco roots were isolated either in toto (W), or G_1 and G_2 nuclei were separated. For tobacco, differentiated root tissues were used for preparation of nuclei; for pea, only root apical meristems were used. The sequences showing significant homology to NESTs were identified by FASTA searches of the EMBL nucleic acid database. Only the best scoring sequences are listed: [a]the length of similarity region, [b]percentage of matching nucleotides in this region, [c]sequence accession number in the EMBL database, and [d]sequence description are indicated. *S. oleracea*, *Spinacia oleracea* (spinach); *N. tabacum*, *Nicotiana tabacum* (tobacco); *L. esculentum*, *Lycopersicon esculentum* (tomato). NEST sequences are deposited in the dbEST (accession numbers 1168193–1168208) and GenBank (accession numbers AA523553–AA523568) databases.

degradation mechanisms are an important factor regulating mRNA half-lives (31), we speculated that an examination of the types and amounts of nuclear pre-mRNA transcripts, prior to export into the cytoplasm, might provide a more accurate read-out of the transcriptional activity of the genome than would analysis of the total cellular mRNA species. We recognized that flow cytometric analysis and sorting could provide a convenient and rapid means for the isolation of large numbers of nuclei from which such transcripts might be prepared. PCR-based analysis of nuclear transcripts can theoretically be accomplished in two ways, by either detection of individual mRNAs by using sequence-specific primer pairs or displaying the entire population of transcripts by using sequence-independent amplification. This work demonstrates that both approaches can be done with nuclei obtained by flow-sorting from different plant tissues.

The procedure for direct isolation of transcribed RNA from sorted nuclei, which involves its capture on oligo-dT-coupled magnetic beads followed by reverse transcription, proved capable of generating high-quality cDNA, even from extremely limited amounts of starting material. For direct detection of selected transcripts, synthesis of the first cDNA strand was sufficient to provide a template for PCR with gene-specific primers. Even for the lowest number of sorted nuclei that was tested (1,000), the NLS-GFP-GUS transcript was readily detected within the transgenic samples. Because this low number of nuclei can be obtained in only a few minutes by flow sorting, preparation of multiple samples is trivial. Moreover, because reverse transcription is done by using a single

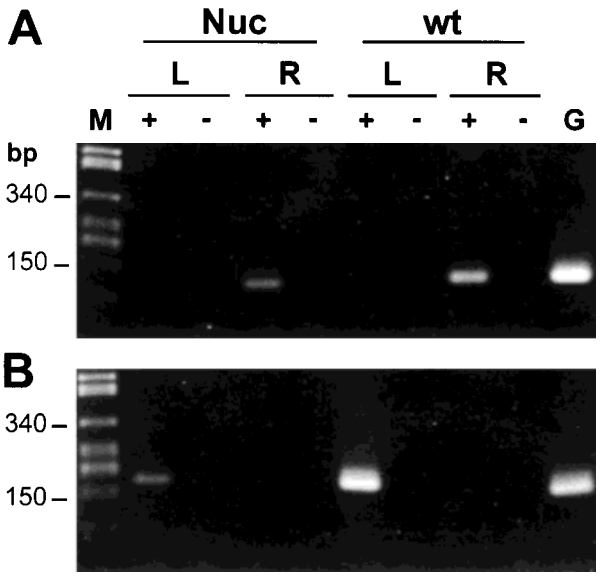

FIG. 5. Reverse transriptase–polymerase chain reaction (RT-PCR) detection of nuclear expressed sequence tags (NESTs) using specific primers. Primer pairs derived from sequences of root-specific NEST 05093 (**A**, expected PCR product size = 128 bp) and leaf-specific NEST 05211 (**B**, expected PCR product size = 187 bp) were employed in RT-PCR with mRNA isolated from flow-sorted nuclei (Nuc) or with mRNA extracted from whole tissues (wt). Pairwise samples prepared from tobacco leaves (L lanes) and roots (R lanes) were compared. For reactions with nuclear mRNA, 10,000 nuclei were sorted and processed as described in Materials and Methods for detection of the NLS-GFP-GUS transcript but with 40 PCR cycles. For reactions with total cellular mRNA (wt samples), 250 ng of mRNA were used, and 25 cycles of PCR were performed. The samples were either subjected to reverse transcription (+) or not (−). The G lanes are PCR reaction products with tobacco genomic DNA as a positive control; the M lanes contain DNA size markers (λ digested with *Pst*I).

poly-dT primer, every sample prepared in that way can be conveniently employed in separate reactions with different gene-specific primers.

An accurate simultaneous display of the different transcripts present within nuclei requires consistent amplification of the cDNAs produced from these transcripts. In the present study, we employed digestion with a frequently cutting restriction endonuclease to generate short NESTs from isolated polyadenylated nuclear RNAs based on the distance between the polyA tract and the nearest 5′ *Nla*III recognition site. The length variability of these fragments is sufficient for identification of most transcripts, and this is facilitated by amplification of subsets by using primers with different selective bases. Moreover, different restriction enzymes can be used in additional experiments to ensure detection of cDNAs having *Nla*III sites too distant or too close to the site of polyadenylation. Theoretically, the average length of the NESTs should approximate the random rate of occurrence of a 4-bp restriction site (256 bp). The majority of the amplified fragments were 50–400 bp in length, which is in agreement with this prediction. It should be noted that specific amplification of the NLS-GFP-GUS transcript employed a 5′ primer located approximately 800 bp from the 3′ end. Therefore, we conclude that the cDNA is of good quality and that RNA degradation does not seriously influence the pattern of amplification.

Compared with other published PCR-based methods for mRNA display, the present technique offers several advantages. First, it reflects the state of gene transcription at any given moment more accurately than conventional techniques because only nuclear mRNAs are analyzed. Second, the technique employs long primers and stringent PCR conditions. This improves the reliability of amplification and renders it more quantitative than other PCR-based methods, such as arbitrarily primed RNA fingerprinting and differential display techniques (16,17,20). Similar AFLP-based protocols have recently been described for use with total cDNA (2,15,21), but these protocols have used 0.1–1.0 μg of starting mRNA/cDNA and radioactive detection of reaction products. In contrast, the present assay requires only minute amounts of mRNA extracted from flow-sorted nuclei, and the resulting bands are visualized by silver staining. This avoids the need for radioactivity and facilitates precise isolation of bands from the analytical polyacrylamide gel. We noticed the frequent occurrence of double bands, representing the individual strands of single DNA molecules. This phenomenon, which has also been described for silver-stained AFLP gels (6), does not seriously affect gel readability. The procedure used for staining the sequencing gels offers excellent sensitivity and, in contrast to the problems encountered by other investigators (6,34), it was easy to reamplify DNA from excised gel pieces in a single round of PCR.

NEST sequence information, because it is obtained from the 3′ ends of RNA transcripts, is generally expected to be less efficient than its 5′ counterpart in comparative searches for homologies in DNA databases of known sequences because the 3′ information includes noncoding sequences that are less highly conserved than their corresponding open-reading frames. Nonetheless, our results have provided examples for which sequence homologies are found, and these homologies are consistent with the tissue sources of the RNA samples. It should also be noted that the public databases are underrepresented in plant sequences (e.g., 87% of the entries in the dbEST database are of human or mouse origin), which reduces the probability of finding homologies. However, in the case of identification of new transcripts, NESTs provide enough sequence information for efficient isolation of corresponding genes by a variety of genome walking techniques (23,25,26).

The results of the present study indicate that a number of cell type-specific nuclear transcripts can be easily identified and isolated from different plant tissues. This is an advantage over nuclear run-off methods, which are limited to monitoring transcripts that have already been characterized. The NEST procedure can also be employed, with slight modifications, for the study of stability of individual transcripts and their accumulation within cell compartments; work is in progress on a comparison of nuclear and polysomal pools of mRNAs. However, the main strength of NEST analysis is anticipated to center around the generation of populations of expressed sequence tags from nuclei sorted according to characteristics that reflect different patterns of gene expression. For example, flow cytometry can be employed to separate nuclei according to DNA content and, as indicated by our initial results (Fig. 4D), this allows studies of the regulation of gene expression within the different phases of the plant cell division cycle. An important point is that NEST analysis avoids the need to employ synchronized populations of cells and the possible artifacts introduced by chemical synchronization treatments. NEST analysis would also be particularly useful in situations involving the analysis of the interactions of mixtures of different organisms, e.g., in the interactions of pathogens with their hosts. Flow sorting would allow discrimination between pathogen and host nuclei, and NEST analysis using these sorted nuclei would rapidly provide information about genes whose expression is altered during the interaction. In other work, we have shown that nuclear accumulation of GFP can be employed as a criterion for the flow sorting of nuclei (12). Through selection of appropriate developmentally regulated or tissue-specific promoters, which should lead to fluorescent nuclei only within specific cell types of transgenic plants, it should be possible to prepare NESTs representative of transcription within these cells. From this information, an understanding of coordinate gene regulation should emerge. Finally, the principle of flow sorting followed by NEST analysis can be applied to any eukaryotic organism that has transcriptionally active nuclei; many of the suggested uses of this technology also appear generally applicable. The ability to sort nuclei according to DNA content, followed by NEST analysis, could be particularly useful in the study of malignancies and of apoptosis. Nuclei could also be extracted for NEST analysis from cells that had been presorted by using cell surface markers as a sort criterion.

ACKNOWLEDGMENTS

This work was supported by a grant to D.W.G. from the U.S.D.A.–N.R.I.C.G.P. (Plant Genome). Purchase of the flow cytometer was made possible by a grant from the D.O.E.–U.R.I. program. D.D. was supported by NSF grant STI 96–20092.

LITERATURE CITED

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 215:403–410, 1990.
2. Bachem CWB, van der Hoeven RS, de Bruijn SM, Vreugdenhil D, Zabeu M, Visser RGF: Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: Analysis of gene expression during potato tuber development. Plant J 9:745–753, 1996.
3. Bharathan G, Lambert G, Galbraith DW: Nuclear DNA contents of monocotyledons and related taxa. Am J Bot 81:381–386, 1994.
4. Brandhorst BP, McConkey EH: Stability of nuclear RNA in mammalian cells. J Mol Biol 85:451–463, 1974.
5. Calkhoven CF, Ab G: Multiple steps in the regulation of transcription-factor level and activity. Biochem J 317:329–342, 1996.
6. Chalhoub BA, Thibault S, Laucou V, Rameau C, Hofte H, Cousin R: Silver staining and recovery of AFLP amplification products on large denaturing polyacrylamide gels. Biotechniques 22:216–220, 1997.
7. Chianale J, Vollrath V, Wielandt AM, Miranda S, Gonzales R, Fresno AM, Quintana C, Gonzales S, Andrade L, Guzman S: Differences between nuclear run-off and mRNA levels for multidrug resistance gene expression in the cephalocaudal axis of the mouse intestine. Biochim Biophys Acta 1264:369–376, 1995.
8. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS: 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008, 1991.
9. Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E: Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science 220:1049–1051, 1983.
10. Galbraith DW, Lambert G: Advances in the flow cytometric characterization of plant cells and tissues. In: Flow Cytometric Applications in Cell Culture, Al-Rubeai M, Emery AN (eds). Marcel Dekker, New York, 1995, pp. 311–326.
11. Gamborg OL, Wetter LR: Plant Tissue Culture Methods. N.R.C. of Canada, Saskatoon, 1975.
12. Grebenok RJ, Lambert G, Galbraith DW: Characterization of the targeted nuclear accumulation of GFP within the cells of transgenic plants. Plant J 12:685–696, 1997.
13. Grebenok RJ, Pierson EA, Lambert G, Gong F-C, Afonso CL, Haldeman-Cahill R, Carrington JC, Galbraith DW: Green-fluorescent protein fusions for efficient characterization of nuclear localization signals. Plant J 11:573–586, 1997.
14. Greenberg ME, Bender TP: Identification of newly transcribed RNA. In: Current Protocols in Molecular Biology, Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Stahl K (eds). John Wiley & Sons, New York, 1997, pp. 4.10.1–4.10.11.
15. Habu Y, Fukadatanaka S, Hisatomi Y, Iida S: Amplified restriction fragment length polymorphism-based mRNA fingerprinting using a single restriction enzyme that recognizes a 4-bp sequence. Biochem Biophys Res Commun 234:516–521, 1997.
16. Liang P, Averboukh L, Pardee AB: Distribution and cloning of eukaryotic mRNAs by means of differential display: Refinements and optimization. Nucleic Acids Res 21:3269–3275, 1993.
17. Liang P, Pardee AB: Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971, 1992.
18. Lorkovic ZJ, Schroder WP, Pakrasi HB, Irrgang KD, Herrmann RG, Oelmuller R: Molecular characterization of PsbW, a nuclear-encoded component of the photosystem II reaction center complex in spinach. Proc Natl Acad Sci USA 92:8930–8934, 1995.
19. Marzluff WF, Huang RCC: Transcription of RNA in isolated nuclei. In: Transcription and Translation: A Practical Approach, Hames BD, Higgins SJ (eds). IRL Press, Oxford, 1984, pp. 89–130.
20. McClelland M, Welsh J: RNA fingerprinting by arbitrary primed PCR. PCR Methods Appl 4:566–581, 1994.
21. Money T, Reader S, Qu LJ, Dunford RP, Moore G: AFLP-based mRNA fingerprinting. Nucleic Acids Res 24:2616–2617, 1996.
22. Murashige T, Skoog F: A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473–497, 1962.
23. Parker JD, Rabinovitch PS, Burmer GC: Targeted gene walking polymerase chain reaction. Nucleic Acids Res 19:3055–3060, 1991.
24. Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448, 1988.
25. Rosenthal A, Jones DSC: Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. Nucleic Acids Res 18:3095–3096, 1990.
26. Rosenthal A, MacKinnon RN, Jones DSC: PCR walking from microdissection clone M54 identifies three exons from the human gene for the neural cell adhesion molecule L1 (CAM-L1). Nucleic Acids Res 19:5395–5401, 1991.
27. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.
28. Sanfacon H, Brodmann P, Hohn T: A dissection of the cauliflower mosaic virus polyadenylation signal. Genes Dev 5:141–149, 1991.
29. Showalter AM, Zhou J, Rumeau D, Worst SG, Varner JE: Tomato extensin and extensin-like cDNAs: Structure and expression in response to wounding. Plant Mol Biol 16:547–565, 1991.
30. Srivastava RAK, Schonfeld G: Measurements of rate of transcription in isolated nuclei by nuclear ''run-off'' assay. In: Methods in Molecular Biology, Harwood AJ (ed). Humana Press, Totowa, NJ, 1994, pp. 281–288.
31. van Hoof A, Green PJ: Control of mRNA decay in plants. In: mRNA Metabolism and Post-Transcriptional Gene Regulation, Harford JB, Morris DR (eds). Wiley-Liss, New York, 1997, pp. 201–216.
32. van Kan JA, Cornelissen BJ, Bol JF: A virus-inducible tobacco gene encoding a glycine-rich protein shares putative regulatory elements with the ribulose bisphosphate carboxylase small subunit gene. Mol Plant Microbe Interact 1:107–112, 1988.
33. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M: AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414, 1995.
34. Weaver KR, Caetano-Anolles G, Gresshoff PM, Callahan LM: Isolation and cloning of DNA amplification products from silver-stained polyacrylamide gels. Biotechniques 16:226–227, 1994.