Plant Molecular Biology **26**: 1989–1993, 1994.

Short communication

Cotton (*Gossypium hirsutum* L.) pollen-specific polygalacturonase mRNA: tissue and temporal specificity of its promoter in transgenic tobacco

Maliyakal E. John* and Michael W. Petersen
*Agracetus, Inc., 8520 University Green Middleton, WI 53562, USA (*author for correspondence)*

Received 26 May 1994; accepted in revised form 7 September 1994

Key words: polygalacturonase, pollen-specific promoter, cotton, transgenics, transformation, β-glucuronidase

Abstract

A gene (G9) expressed during late microsporogenesis in cotton (*Gossypium hirsutum* L.) was isolated. Sequence analysis of the cDNA (1.3 kb) as well as the gene (2.6 kb) revealed an open reading frame of 1233 bases encoding a protein of 43.9 kDa. The coding region of the gene is interrupted by three introns. Northern analysis of the RNA from developing anthers showed that the transcripts appear 12 days before anthesis and that the maximal concentration of RNA occurs in pollen on the day of anthesis. This pattern of gene expression suggests functions in post-anthesis events. Sequence comparisons with other known plant genes indicated that G9 is homologous to polygalacturonases. The G9 promoter conferred tissue and temporal specificity of β-glucuronidase (*GUS*) expression in transgenic tobacco plants. Thus, the G9 promoter can be used to drive gene expression in homologous as well as heterologous plants in a tissue-specific manner.

Cell types that constitute plant tissues exhibit quantitatively and qualitatively different gene expression. The majority of genes that are active in a cell support functions that are common to all cell types. Thus, of the 20000 to 25000 active genes in a tobacco anther, only 10000 are anther-specific [12, 13]. Characterizations of tissue-specific genes, therefore, are important for the study of tissue or organ development. Gene expression in pollen grains, the male gametophytes of the flowering plants, has been studied in a number of plant species including *Zea mays* [1, 9], tomato [4, 8, 17], *Oenothera organensis* [5] and *Tradescantia paludosa* [15]. We are examining the gene expression in cotton (*Gossypium*) pollen to identify genes that are critical to microsporogenesis. The development of cotton pollen is similar to other dicotyledons; however, none of the genes involved in cotton pollen development or function have been reported.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number U09717.

Here we describe the characterization of a gene that is active in the late stages of pollen development. Structural characterization of this gene indicates that it codes for a polygalacturonase-like protein. We have also identified the promoter of this gene which can impart tissue and temporal specificity of expression in heterologous plants.

Five thousand clones from an anther/pollen cDNA library of *G. hirsutum* L. cv. C312 were screened using cDNA probes of petal, ovule, and anther/pollen mRNAs. One cDNA clone, CPA-G9 (insert of 925 bp), was selected for further study. Northern analysis of mRNAs from leaves, roots, ovules, petals, fiber and pollen revealed that G9 expression is confined to pollen in cotton plants and the transcript size is ca. 1.45 kb (Fig. 1A). In order to investigate the expression of the G9 gene during pollen development, RNA was isolated from anthers starting at 15 days preanthesis (DPA) and continuing up to the day of anthesis (day of pollen shedding), and hybridized to the CPA-G9 insert (Fig. 1B). The maximum steady-state level of RNA was found in the pollen grains on the day of anthesis. Low levels of transcripts detected in the RNA of 12 DPA anthers increased during pollen development (Fig. 1B).

Both strands of the CPA-G9 insert (925 bp) were sequenced. Since CPA-G9 is a partial cDNA clone as judged by transcript size, the remaining 5′ end of the G9 mRNA was cloned using primer extension and polymerase chain reaction (PCR). An oligonucleotide primer from the 5′ end of CPA-G9 was used to prime synthesis of cDNAs, and the double stranded cDNAs were cloned into lambda gt11 (pCPAG9-5). The 3′ end of pCPAG9-5 overlapped with the 5′ end of pCPA-G9. Combined sequences of these two clones gave us a near full-length cDNA sequence (1297 bp) which contained a 1233 base open reading frame. Assuming that initiation occurs at the first ATG at position 13, the translation of the mRNA would result in a 407 amino acid protein with an estimated molecular mass of 43.9 kDa and an isoelectric point of 6.5. A search of the sequences in the GenBank (September 1993) revealed that the cotton G9 protein has significant

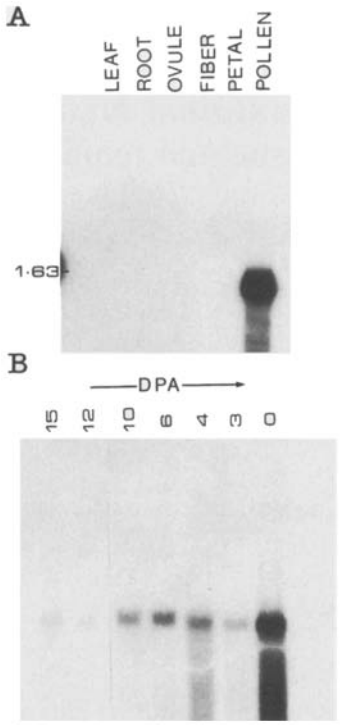

Fig. 1. A. Tissue-specific expression of G9 mRNA. Cotton (*Gossypium hirsutum* L. cv. C312 and *G. barbadense* cv. Sea Island) were greenhouse-grown and anthers and/or pollen were collected starting from 15 days preanthesis (DPA) and stored frozen at −80 °C. Poly(A)$^+$ RNA (0.25 μg) from leaf, fiber, ovule pollen, petal, and root were isolated and size fractionated on formaldehyde/agarose gels and blotted on nitrocellulose. The blot was hybridized to a nick translated insert of pCPA G-9 (1×10^8 cpm/μg; 5×10^5 cpm/ml) and washed under stringent conditions ([11], 0.1× SSC, 53 °C). Autoradiography was done at −70 °C for 12 h. MW standard in kb is marked. B. G9 expression during pollen development. Total RNA (25 μg) from developing anthers was analyzed by northern blot technique. Day of flower opening for each developing bud was determined based on its position on flowering branches. All hybridizations and washing conditions were similar to those described for A. DPA, days preanthesis, are given.

sequence similarities with a number of known polygalacturonases (PGs). Tobacco PG mRNA, Npg1 ([16], accession number X71017 GenBank), and *Oenothera organensis* pollen cDNA, P26 [5], showed 70% and 71% similarity respectively at the protein level (Fig. 2). Maize, PG1 ([1], 63.4%), avocado 5BA 24 ([7], 58%), and tomato PG-2A ([8], 54.6%) also showed significant homology to the cotton G9 protein. In ad-

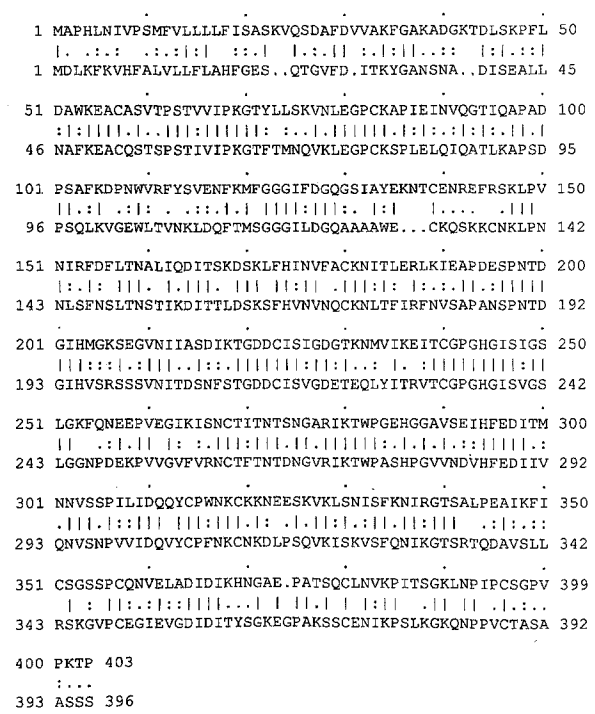

Fig. 2. Sequence similarity between the cotton G9 protein and tobacco (Npg1) polygalacturonases. Nucleotide derived amino acid sequences of cotton G9 (top line) and tobacco Npg1 (bottom line: accession number X71017 GenBank) were aligned to show optimal similarity by the Genetics Computer Group (GCG, University Park, Madison) computer program. Gaps were inserted to maximize the number of matches. Vertical lines indicate identical residues. A 70% similarity is found between these two proteins. Amino acid residues are numbered on the left and right sides. Comparison of nucleotide derived amino acid sequence of *Oenothera organensis* cDNA, P26 [5] to that of the G9 protein also showed a 71% similarity (not shown).

dition, comparison of two fungal PGs, *C. carbonum* [14] and *F. moniliforme* [6], also showed 51% and 44.5% similarity, respectively, at the protein level (not shown). These sequence comparisons suggest that cotton G9 belongs to the PG family of proteins. PGs are cell wall degrading enzymes that hydrolyze polyuronides or pectins and are found in plants as well as plant pathogens [3, 5, 14]. PGs contain a hydrophobic N-terminus signal peptide. The nucleotide derived amino acid sequence of the G9 protein also showed a hydrophobic N-terminus (not shown). Based on the homology of the nucleotide and derived amino acid sequence of G9 with other known PG proteins, and its pattern of gene expression, we conclude that G9 is a pollen-specific PG-like protein in cotton. In a number of plants PGs are encoded by closely related multiple genes. For example, there are multiple PG members in the P2 family in *O. organensis* and maize [1, 5] whereas tomato contains only one PG gene [4]. Southern blot analysis of cotton genomic DNA hybridized to the G9 cDNA insert revealed a simple pattern of hybridization similar to tomato. The cDNA insert hybridized to only a single band indicating the presence of a single PG gene in cotton (not shown).

The promoter of the G9 gene would be useful in directing the expression of heterologous proteins in a tissue-specific and developmentally-regulated manner in cotton. In order to identify the promoter of the PG gene, a genomic library of *Gossypium barbadense* cv. Sea Island (1×10^6 phages) was screened and a hybridizing phage, λ-EMBLSIG-9 was isolated. The insert of λ-EMBLSIG-9 (17 kb) was subcloned into a Bluescript SK+ vector. Subsequent restriction and Southern blot analyses of this clone resulted in the identification and cloning of a 5.6 kb *Bgl* II/*Sal* I fragment containing the G9 gene (SKSIG-9s). The nucleotide sequence of the G9 gene (2.6 kb region) was determined (accession number U09805). Sequence comparison between the *G. hirsutum* cv. C312 cDNA and the *G. barbadense* cv. Sea Island gene showed a 98.9% similarity (not shown). Similarly, at the amino acid level there were only four differences between these two proteins. The Sea Island gene contained three small introns of 92, 124, and 76 bp. There were also two nucleotide substitutions in the 5′-untranslated and one at the 3′-untranslated regions (not shown).

A 1.9 kb *Sal* I/*Nsi* I fragment at the 5′ end of the G9 gene was tested for promoter activity by constructing a transcription fusion gene with β-glucuronidase (GUS) and introducing it into tobacco plants. The unique *Nsi* I site is 695 bases 3′ to the ATG of G9 gene. The 1.9 kb G9 fragment contained the first two introns (92 and 124 bp) of the gene. The GUS gene consisted of

a coding region (1.8 kb), a 38 bp alfalfa mosaic viral (AMV) 5′-untranslated leader, and a 3′ Nos poly(A) addition region (280 bp) of *Agrobacterium* nopaline synthase gene. The mature transcript produced from the G9-GUS fusion gene would be ca. 2600 bases long. However, starting from the G9 context, AACATGG, there are no contiguous open reading frames that are in the frame with the GUS coding region. Therefore, the G9-GUS RNA would be translated from the GUS context, ACCATGG.

The G9-GUS transcription fusion construct was cloned into an *Agrobacterium* transformation vector [2] to generate pTV4G9A-2117. Similarly a second plasmid containing the GUS gene and a cauliflower mosaic viral 35S promoter was made (pTV4-35S2117). These two genes were introduced into *Nicotiana tabacum* cv. H425 by *Agrobacterium*-mediated leaf disk transformation [2]. The transformed tissue was selected and grown on kanamycin [2]. Nine transgenic plants containing the G9 promoter-GUS fusion and one plant containing 35S-GUS fusion were tested for GUS enzyme activity by histochemical staining, in addition to untransformed plants. Seven plants containing the G9-GUS expressed GUS in pollen while stem, leaf, and root showed no GUS activity. Since active GUS enzyme is produced from the G9-GUS fusion gene, we assume that the translation initiation occurred from the GUS initiation codon. Based on our construct, the GUS transcript will have a minimum of 529 bases of untranslated 5′ leader.

Quantitative estimation of GUS expression in the tissues of one transgenic plant containing 35S-GUS, C4473, and two plants containing the G9-GUS, G4573 and G4544, were carried out by fluorometric assays using 4-methylumbelliferyl glucuronide (Fig. 3). Stem, seed, and root tissues of plant C4473 expressed high levels of GUS, while moderate enzyme level was found in the leaves. GUS activity was poor in the pollens (Fig. 3). Transgenic plants containing the G9-GUS gene showed minimal GUS activity in the stem, seed, root, and leaves similar to control plants, whereas pollen grains of both G4573 and G4544 showed strong GUS activity. Enzymatic

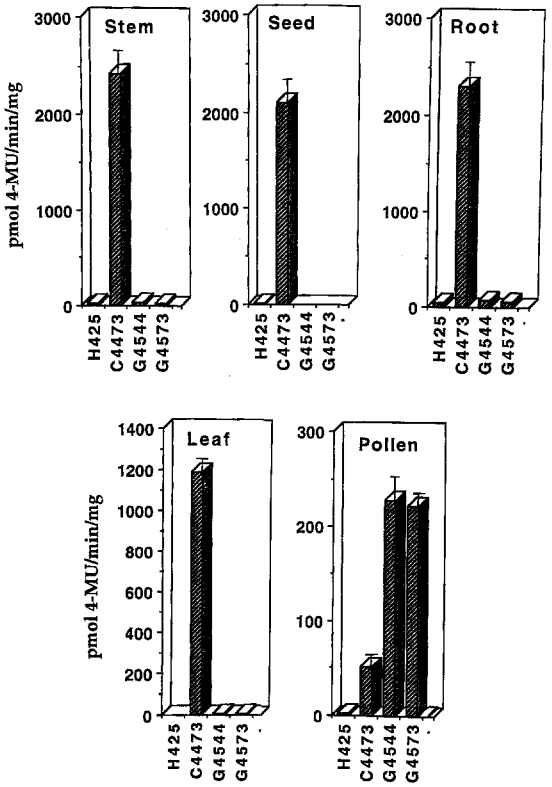

Fig. 3. The G9 promoter confers tissue specific expression of the GUS gene in transgenic tobacco. Transgenic tobacco plants (*Nicotiana tabacum* cv. Havana 425) were generated by *Agrobacterium* infection [2]. The expression of the *GUS* gene either from a 35S promoter or a G9 promoter were then measured in different tissues of transgenic plants by MUG assays [10]. H425 is a control non-transgenic plant, while C4473 is a transgenic tobacco plant containing the 35S-GUS gene. G4573 and G4544 are transgenic plants carrying the G9-GUS gene. The presence of GUS genes was determined by PCR amplification (not shown). These are the average of three or more independent measurements.

activity in pollen grains of G4573 and G4544 was four-fold higher than that of C4473 (Fig. 3). GUS expression was followed in the developing floral buds of G9-GUS and 35S-GUS plants. Weak staining was observed in the circular cell cluster, stomium, and tapetum of stage 1 buds (8 mm) of G4544 and G4573. The staining was barely discernible in these nonreproductive cells. On the other hand, intense staining was observed in the vascular bundle, connective tissue, microspores, circular cell cluster, tapetum, stomium, epidermis, and endothecium of the stage 1 bud of the 35S-

GUS plant C4473. The GUS staining in the non-reproductive cells of G9-GUS plants did not increase in later stages. Shrinking and degeneration of the tapetum and cells adjacent to stomium begins in stage 3 and 4 [13]. GUS staining in the microspores was first observed in stage 3 (15 mm) of the G9 plants. The relative GUS expression level in pollen of G9-GUS plants increased with bud development as judged by increased intensity of staining in later stages of the flower. The most intense staining was observed in the pollen on the day of anthesis (not shown). This is consistent with the pattern seen for the increased G9 mRNA accumulated during cotton pollen development (Fig. 1).

PG is one of the many cell wall degrading enzymes produced by bacteria, fungi, and plants. In plants, fruit specific PGs play a role in ripening [3]. Similarly, pollen specific PGs are thought to take part in the depolymerization of the pectin in its walls during pollen tube elongation, or in that of the pistil during pollination [5]. Cotton G9 may have similar functions in the pollination process. The GUS expression pattern in transgenic tobacco indicates that the G9 promoter is useful in expressing heterologous genes in a tissue-specific manner in dicots.

Acknowledgement

The authors thank Greg Keller, Sandra Cannon, and Michael Miller for technical assistance. Dr Gurdip Brar for assistance with the localization of GUS activity in developing tobacco flowers. We are also thankful to Meg Allen and Peggy Wagner for editorial assistance.

References

1. Allen RL, Lonsdale DM: Molecular characterization of one of the maize polygalacturonase gene family members which are expressed during late pollen development. Plant J 3: 261–271 (1993).
2. Barton KA, Whiteley HR, Yang NS: *Bacillus thuringiensis* δ-endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiol 85: 1103–1109 (1987).
3. Bennett AB, DellaPenna D: Polygalacturonase; its importance and regulation in ripening. In: Thompson WW, Nothnagle E, Huffaker R (eds) Plant Senescence: Its Biochemistry and Physiology, pp. 98–107. American Society of Plant Physiologists, Rockville, MD (1987).
4. Bird CR, Smith CJS, Ray JA, Moureau P, Bevan MW, Bird AS, Hughes S, Morris PC, Grierson D, Schuch W: The tomato polygalacturonase gene and ripening-specific expression in transgenic plants. Plant Mol Biol 11: 651–662 (1988).
5. Brown SM, Crouch ML: Characterization of a gene family abundantly expressed in *Oenothera organensis* pollen that shows sequence similarity to polygalacturonase. Plant Cell 2: 263–274 (1990).
6. Caprari C, Richter A, Bergmann C, Lo Cicero S, Salvi G, Cervone F, DeLorenzo G: Cloning and characterization of a gene encoding the endopolygalacturonase of *Fusarium moniliforme*. Mycol Res 97: 497–505 (1993).
7. Dopico B, Lowe AL, Wilson ID, Merodio C, Grierson D: Cloning and characterization of avocado fruit mRNAs and their expression during ripening and low temperature storage. Plant Mol Biol 21: 437–449 (1993).
8. Grierson D, Tucker GA, Keen J, Ray J, Bird CR, Schuch W: Sequencing and identification of a cDNA clone for tomato polygalacturonase. Nucl Acids Res 14: 8595–8603 (1986).
9. Hanson DD, Hamilton DA, Travis JL, Bashe DM, Mascarenhas JP: Characterization of a pollen-specific cDNA clone from *Zea mays* and its expression. Plant Cell 1: 173–179 (1989).
10. Jefferson RA: Assaying chimeric genes in plants: the *GUS* gene fusion system. Plant Mol Biol Rep 5: 387–405 (1987).
11. John ME, Crow LJ: Gene expression in cotton (*Gossypium hirsutum* L.) fiber: cloning of the mRNAs. Proc Natl Acad Sci USA 89: 5769–5773 (1992).
12. Kamalay JC, Goldberg RB: Organ-specific nuclear RNAs in tobacco. Proc Natl Acad Sci USA 81: 2801–2805 (1984).
13. Koltunow AM, Truettner J, Cox KH, Wallroth M, Goldberg RB: Different temporal and spatial gene expression patterns occur during anther development. Plant Cell 2: 1201–1224 (1990).
14. Scott-Craig JS, Panaccione DG, Cervone F, Walton JD: Endopolygalacturonase is not required for pathogenicity of *Cochliobolus carbonum* on maize. Plant Cell 2: 1191–1200 (1990).
15. Stinson JR, Eisenberg AJ, Willing RP, Enrico PME, Hanson DD, Mascarenhas JP: Genes expressed in the male gametophyte of flowering plants and their isolation. Plant Physiol 83: 442–447 (1987).
16. Tebbutt SJ, Rogers HJ, Lonsdale DM: Characterization of a tobacco gene encoding a pollen-specific polygalacturonase. Plant Mol Biol 25: 283–297 (1994).
17. Ursin VM, Yamaguchi J, McCormick S: Gametophytic and sporophytic expression of anther-specific genes in developing tomato anthers. Plant Cell 1: 727–736 (1989).