Plant Molecular Biology **14**: 17–28, 1989.

Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen *Erwinia*

Rod A. Wing, Judy Yamaguchi, Susan K. Larabell, Virginia M. Ursin and Sheila McCormick*
Plant Gene Expression Center, USDA/ARS-UC-Berkeley, 800 Buchanan St., Albany, CA 94710, USA
(author for correspondence)*

Received 22 June 1989; accepted 7 August 1989

Key words: allergen, *Erwinia*, genetic linkage, pectate lyase, pollen, tomato

Abbreviations: LAT, late anther tomato; bp, base pairs; MA, mature anther; PL, pectate lyase; kb, kilobase (pairs)

Abstract

A set of cDNAs that are expressed in tomato anthers were isolated [24]. We further characterized two of these cDNAs (LAT56 and LAT59) and their corresponding genomic clones. LAT56 and LAT59 show low levels of steady-state mRNA in immature anthers and maximal levels in mature anthers and pollen. The LAT56 and LAT59 genes are single-copy in the tomato genome, and are linked on chromosome 3, approximately 5 cM apart. Although these cDNAs did not cross-hybridize, their deduced protein sequences (P56 and P59) have 54% amino acid identity. The LAT56 and LAT59 genes each have two introns, but they are located in non-homologous positions. P56 and P59 show significant protein sequence similarity to pectate lyases of plant pathogenic bacteria. The similarity of P56 and P59 to the bacterial pectate lyases is equivalent to the homology described for different pectate lyase sequences of the genus *Erwinia*. We suggest that the pollen expression of LAT56 and LAT59 might relate to a requirement for pectin degradation during pollen tube growth.

Introduction

Determining the genetic and molecular mechanisms of male gametophyte development can yield information on the coordination of sporophytic and gametophytic gene expression, cell-cell recognition mechanisms and tissue-specific gene expression. As a first step towards this goal, we used differential screening of cDNA libraries prepared from tomato anthers to generate a set of cDNAs specifically expressed in anthers [24]. Further characterization of five of these clones (LAT51, LAT52, LAT56, LAT58 and LAT59) showed that their corresponding mRNAs were localized in the anthers and were maximally expressed in pollen ([35 36], this paper).

The RNAs corresponding to these five pollen-

The nucleotide sequences reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X15500 (LAT56) and X15499 (LAT59).

expressed cDNAs are present in relatively high abundance, and might code for proteins needed during pollen development or tube growth. We hoped that a search of computer databases might help identify functions for the pollen-expressed genes. We recently reported the molecular characterization of one of these anther-expressed cDNAs, LAT52 [35]. The deduced protein sequence of LAT52 (P52) is cysteine-rich and shows 32% amino acid identity to a deduced protein sequence of a pollen specific gene from corn (Zmc13) [9]. No function is known for the gene products of LAT52 or Zmc13.

In this article we show that the sequences of two other anther-specific cDNAs (LAT56 and LAT59) have sequence similarity to pectin-degrading enzymes (pectate lyases) of plant pathogenic bacteria. Because pollen 'invades' the female tissue in order to fertilize the egg cell, analogies between host-pathogen interactions and pollen-pistil interactions have been made [3]. This is the first report, to our knowledge, of sequence similarities between pollen proteins and plant pathogen proteins. The degree of similarity between these pollen proteins and the bacterial pectate lyases is similar to the level of homology seen among the several bacterial pectate lyases.

Plant pathogenic bacteria are successful pathogens at least in part because they secrete large amounts of pectic enzymes that macerate the host plant, as reviewed by Kotoujansky [15]. It is probable that pectin degradation is important for pollen function. Pollen grains hydrate, a tube emerges, and the pollen tube grows through the intercellular spaces of the style in order to achieve fertilization. Rapid wall synthesis is required for growth of the pollen tube. The products from degradation of pectin in stylar secretions could be used for pollen tube wall synthesis [16]. Heslop-Harrison [10] has also suggested that pectin-degrading enzymes might be needed for the localized weakening of the pollen wall required for pollen tube emergence.

Materials and methods

Northern analysis

RNA was isolated according to Rochester *et al.* [29], and Northern blots prepared according to Maniatis *et al.* [21]. Hybridizations were done using random-primer-labeled probes [7] in 5x SSPE, 2x Denhardt's solution, 0.1% SDS, 100 μg/ml denatured salmon sperm DNA, and 100 μg/ml poly(A) at 68 °C. Filters were washed in 0.3x SSPE, 0.1% SDS at 68 °C.

Southern analysis

DNA was isolated according to Bernatzky and Tanksley [2]. DNA was digested with restriction endonucleases, separated by electrophoresis through 0.8% agarose gels and blotted (to Nytran membranes) as described [21]. Hybridization and washes were done as for northern blots. The chromosomal location of LAT56 and LAT59 was determined with Southern blots of chromosome addition lines of *Lycopersicon esculentum* and *Solanum lycopersicoides* (gift of J. DeVerna and R. Chetelat, Campbell Institute).

Genomic clone isolation

An amplified EMBL3 genomic library of tomato (gift of C. Gasser, Monsanto) was screened according to Benton and Davis [1].

DNA sequencing

Both strands of the cDNAs and corresponding genomic clones were sequenced using the United States Biochemicals Co. Sequenase DNA sequencing kit, [^{35}S]dATP, and specific DNA sequencing primers.

RNase protection for intron mapping

RNase protection was performed essentially as described by Melton *et al.* [25]. *In vitro* [^{32}P]-

labeled antisense transcripts of a 600 bp *Ase* I-*Bgl* II fragment of gLAT56 in pGEM7zf(+) were annealed for 16 hours at 42 °C to 1 μg of poly(A)$^+$ RNA from mature anthers. Unannealed probe was digested with RNase A (30 μg/ml) and RNase T1 (840 u/ml) for 60 minutes at 37 °C. The reaction product was phenol-extracted, ethanol-precipitated and then separated on a 4% acrylamide sequencing gel.

Primer extension cDNA cloning

A 20 bp synthetic oligonucleotide corresponding to the 5′ end of pLAT59 was synthesized (5′-CACTCTTTGTACGTATACCA-3′). One picomole of this primer was annealed to 2 μg of poly(A)$^+$ RNA from mature tomato anthers and extended with AMV reverse transcriptase as described [8]. The cDNA was phenol-extracted and purified on a G-50 (Sephadex) spun column [21]. Second-strand synthesis, *Eco* RI linker addition, and ligation into *Eco* RI-digested λgt10 arms were performed as described [4]. The ligation mix was packaged *in vitro* using a kit (Promega). Recombinant bacteriophage were screened with radioactively labeled pLAT59 as described above. The DNA inserts from positive clones were subcloned into the *Eco* RI site of pGEM7zf(+) and sequenced.

PCR cDNA cloning

First-strand synthesis was performed as above using an 18-mer oligonucleotide (5′-GAGCCCTAACTTCGAAAT-3′) homologous to the antisense strand 65 bp 5′ to the *Eco* RI site of gLAT59 at base 483. The first-strand reaction was phenol-chloroform-extracted, ethanol-precipitated and washed with 70% ethanol. Second-strand synthesis and amplification of the cDNA used the polymerase chain reaction [30], priming with the first-strand oligonucleotide and a 20-mer (5′-CCTATTGTCTGCCCAATTAGG-3′) homologous to a region 50 bp 3′ to the transcription start site.

5′ end mapping of LAT56 and LAT59

Primer extension and RNase protection were performed as previously described [35]. For primer extension mapping LAT56 a 20-mer oligonucleotide (5′-GTTCTATAAGAATACTCCAT-3′) whose 3′ end is at the putative initiator methionine was annealed to 2 μg of poly(A)$^+$ RNA from mature anthers. For LAT59, an 18-mer oligonucleotide (5′-ATTTCGAAGTTAGGGCTC-3′) homologous to the anti-sense strand of the LAT59 mRNA 183 nucleotides upstream of the putative initiator methionine was annealed to 2 μg of poly(A)$^+$ RNA from mature anthers. For RNAse protection of LAT56 the 1.4 kb *Eco* RI DNA fragment was transcribed into anti-sense RNA in pGEM7zf(+) and protected with 2 μg of poly(A)$^+$ RNA from mature anthers. For LAT59, a 1.4 kb *Sst* I-*Hpa* II DNA fragment containing the transcription start site was transcribed into anti-sense RNA and protected with poly(A)$^+$ RNA as above.

Computer analysis

DNA and protein sequences were analyzed using programs from PCGene, Intelligenetics, Genetics Computer Group, and DNA strider 1.0 [23]. Genetic linkage analysis was done with the Linkage-1 program [31]. We used the NBRF Protein Information Resource, version 16 (3/88), and the GenBank Genetic Sequence Data Bank, version 57 (9/88) databases.

Results

Developmental expression of LAT56 and LAT59

Differential screening of the cDNA library with seedling RNA suggested that LAT56 and LAT59 were specifically expressed in anthers [24]. Northern analysis confirmed the anther specificity, with the exception that LAT56 and LAT59 mRNA could be detected in roots, at levels approximately 100-fold less than that in mature

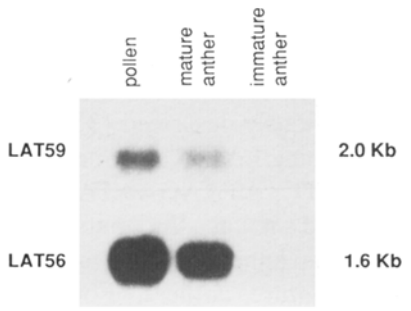


Fig. 1. Developmental expression of LAT56 and LAT59 mRNA. Northern analysis of 2 μg poly(A)$^+$ RNA isolated from pollen, mature anthers, and immature anthers. The samples were hybridized with ^{32}p-labeled probes. The size of the hybridizing transcripts is indicated in kb.

anthers (data not shown). Fig. 1 shows the results of a northern blot containing RNA prepared from immature anthers (containing tetrad stage microspores), mature anthers (containing mature pollen), and isolated pollen, hybridized with LAT56 and LAT59. LAT56 hybridized to a 1.6 kb transcript, while LAT59 hybridized to a 2 kb transcript. Both genes showed no detectable steady-state mRNA in immature anthers, showed weak hybridization in slightly older anthers (not shown), and showed maximal levels of hybridization in mature anthers and pollen, indicating that both clones are expressed late in flower development.

Genetic analysis

LAT56 and LAT59 each hybridized to distinct single DNA fragments on Southern blots of *Eco* RI-digested tomato genomic DNA, which indicate that they are each encoded by single nuclear genes. Restriction fragment length polymorphisms (RFLPs) between *L. esculentum* and *L. pennellii* were identified for LAT56 and LAT59 in order to map the position of these genes in the tomato genome. pLAT56 hybridizes to a 2.9 kb band in VF36 and a 8.4 kb band in LA716 when the DNAs are digested with *Bst* NI (data not shown). pLAT59 hybridizes to bands of 3.6 kb and 2.2 kb in VF36 and to bands of 3.15 kb and 2.5 kb in LA716 when the DNAs are digested with *Hae* III (data not shown). Segregation analysis in a F2 population derived from a cross of *L. esculentum* and *L. pennellii* demonstrated that the two genes are linked, *ca.* 5 cM apart (Table 1). Alien addition lines between *L. esculentum* and *S. lycopersicoides* [6] were used to demonstrate that LAT56 and LAT59 are on chromosome 3 (data not shown).

Gene structure of LAT56 and LAT59

Partial restriction maps and the organization of the cDNAs and genomic clones of LAT56 and LAT59 are shown in Figs. 2a and 3a. The nucleotide and deduced protein coding sequences of LAT56 and LAT59 are shown in Figs. 2b and 3b.

End mapping

The transcription start sites of LAT56 and LAT59 were determined by primer extension and RNase protection (data not shown). The results of LAT56 showed that there are two equivalent major transcription start sites between 189 and 184 nucleotides upstream of the putative translation initiation codon. The sequence TATATAA occurs 32 nucleotides upstream of the largest extension product. This sequence appears to be a double 'TATA' [26] which may account for the presence of two transcription start sites. The results for LAT59 showed a major transcription start site 236 nucleotides upstream of the putative initiator methionine. A 'TATA'-like sequence [26] is located 28 nucleotides upstream of the transcription start site.

The 3′ ends of LAT56 and LAT59 were determined by DNA sequencing of the corresponding cDNAs. Both cDNAs contain poly(A) tails that indicate the positions of the poly(A) addition sites (Figs. 2b, 3b).

Intron mapping

LAT56 and LAT59 were analyzed for the presence of introns either by RNase protection or

Table 1. Linkage analysis between LAT56 and LAT59.

	LAT59 RFLP		
LAT56 RFLP	VF36	VF36/LA716	LA716
VF36	12	4	0
VF36/LA716	0	14	1
LA716	0	0	17
N	12	18	18

$X^2(1:2:1) = 71.02$

Map distance = 5.3 ± 2.4 cM

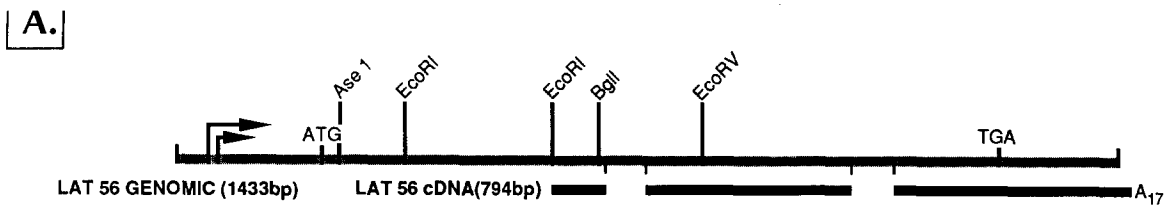
Cosegregation of LAT56 and LAT59 restriction fragment length polymorphisms in F2 plants derived from crosses between *L. esculentum* (VF36) and *L. pennellii* (LA716). Map distance (cM) was calculated with Linkage-1 (Suiter et. al., 1983).

by comparison of cDNA sequences with genomic DNA sequences. Comparison of the LAT56 cDNA with the corresponding genomic DNA sequence revealed the presence of two small introns of 86 and 71 bp, beginning at positions 579 and 949 respectively (Fig. 2b). The genomic DNA 5′ to pLAT56 was analyzed by RNase protection, and no additional introns were found.

No introns were found in the genomic DNA sequence of gLAT59 in the region that overlapped with the LAT59 cDNA. Upon further sequencing of the gLAT59 5′ to the cDNA, a putative 102 bp intron was discovered by sequence comparison with gLAT56. This intron was later confirmed by sequencing a specific cDNA overlapping this region (primer extension cDNA) (Fig. 3A). RNase protection was done on a 630 bp *Msp* I-*Eco* RI fragment from gLAT59, and was found to not be fully protected indicating the presence of intron(s). The second intron was identified by synthesizing and sequencing a cDNA that spanned the 5′ intron of the LAT59 gene (PCR cloning product, Fig. 3a). Thus LAT59 has two small introns of 73 and 102 bp, beginning at positions 578 and 951 respectively, 5′ to the cDNA (Fig. 3b).

LAT56 encodes a putative 356 amino acid protein (P56) of 40,561 Da with a predicted pI of 8.28. The predicted protein contains 3 potential N-linked glycosylation sites (Asp-X-Ser/Thr) at amino acid positions 134, 164 and 227. Fig. 2c shows the hydropathy plot of P56, indicating that P56 is relatively hydrophilic. The N-terminal 27 amino acids have the properties of a signal sequence [37]. There are 2 predicted signal sequence cleavage sites in this region that conform to the −3, −1 rule [37]; one potential cleavage site is between amino acids 26 (Thr) and 27 (Ala) (S value of 6.68), while the second is between two alanine residues at amino acids 22 and 23 (S value of 6.17).

LAT59 encodes a putative 449 amino acid protein (P59) of 50,893 Da with a predicted pI of 8.36. There are three potential N-linked glycosylation sites, at amino acids 56, 80 and 81. Fig. 3c shows the hydropathy plot of P59, and shows that P59 is also relatively hydrophilic. The N-terminal region is hydrophobic, and has a predicted signal



B.

```
-222  TATATAAATAAGGTGATTGTATGAATGAAAATAAGCACAAAGAGGAACAAATAATGTTGC
-162  TTGTTTAGGAATAATTAATTATCCTATTTTTCTCACAACACTTTTTTAAGGGTAGTAGTG
-102  GTGTAAAGGATTGACCTTACATCAACTACCATCAACACCCCTAGCAATTTAACAATAACC
 -42  CCAACTTACCTACCAACCTACCTACCTTATTAATTTATAATAATGGAGTATTCTTATAGA
                                                  MetGluTyrSerTyrArg
  18  ACAAAAATTAATGTATTATTCATTGTACTCATTCTCTTTGTATTTGCTGCACTCTTTACA
      ThrLysIleAsnValLeuPheIleValLeuIleLeuPheValPheAlaAlaLeuPheThr
  78  GCAATAAACGCTCCGAGGAGGAAACTAACTAAGAAATACAGAGGCCCGTGTATGGCCGTG
      AlaIleAsnAlaProArgArgLysLeuThrLysLysTyrArgGlyProCysMetAlaVal
 138  AATTCAATTAAAAAGTGCTGGAGATGTGACCCTTTCTGGGCTGAAAGATGCGGCCAAATG
      AsnSerIleLysLysCysTrpArgCysAspProPheTrpAlaGluArgCysGlyGlnMet
 198  CTTTGTGCGTTGGGCTTTGGTATCAATGCAATGGGAGGAAAGTATGGTCCATACTACATT
      LeuCysAlaLeuGlyPheGlyIleAsnAlaMetGlyGlyLysTyrGlyProTyrTyrIle
 258  GTCACAGATAATTCTGATGATGACGTTGTTGATCCTAAGCCTGGAACTCTCCGATTCGGG
      ValThrAspAsnSerAspAspAspValValAspProLysProGlyThrLeuArgPheGly
 318  GTGATCCAAAAAGGACCATTGTGGATCAGATTTGCCAGAAGTATGAGAATCCGATTGACC
      ValIleGlnLysGlyProLeuTrpIleArgPheAlaArgSerMetArgIleArgLeuThr
 378  AGGGAACTTATTGTTAGTAGCAATAAAACAATTGATGGACGTGAAAAATACGTTCACATA
      ArgGluLeuIleValSerSerAsnLysThrIleAspGlyArgGluLysTyrValHisIle
 438  GCAAATGGAGCAGGAATTAAGATACAAAGCGCATCTAATGTTATTATCTCTAATCTTTCG
      AlaAsnGlyAlaGlyIleLysIleGlnSerAlaSerAsnValIleIleSerAsnLeuSer
      . *
 498  AATTCACAATATTGTACCTACCGCAGTGGCCTGCTAAGGGAGTCTGACGACCACCTTGGC
      AsnSerGlnTyrCysThrTyrArgSerGlyLeuLeuArgGluSerAspAspHisLeuGly
 558  CTAAGGGGCGCAGATGAAGgtaatataaatgactatatttattgattggaaaaaagagaa
      LeuArgGlyAlaAspGluG

 618  gacaatttaaataatttaattatgcatacagGTGATGCCATCAGTATCTTCAATTCTCAT
                                      lyAspAlaIleSerIlePheAsnSerHis
 678  GATATATGGATTGACCACATCTCTATGTCTCGTGCCACTGATGGTCTGATAGATGCTGTT
      AspIleTrpIleAspHisIleSerMetSerArgAlaThrAspGlyLeuIleAspAlaVal
 738  GCTGGTTCCACTAATATTACGATATCCAACTGTCACTTCACTGATCATGAAAAGGTAATG
      AlaGlySerThrAsnIleThrIleSerAsnCysHisPheThrAspHisGluLysValMet
 798  TTGTTTGGTGCTAACGATCATGCGGAAGAGGACAGGGGAATGAAAATAACATTGGCGTAT
      LeuPheGlyAlaAsnAspHisAlaGluGluAspArgGlyMetLysIleThrLeuAlaTyr
 858  AACCACTTTGGAAAGAGATTGGATCAAAGGATGCCTAGATGCAGGTTTGGATTTTTCCAT
      AsnHisPheGlyLysArgLeuAspGlnArgMetProArgCysArgPheGlyPhePheHis
 918  CTGGTGAACAATGATTACACTCATTGGGAAAGgtacatacgtattaaataatattggtaa
      LeuValAsnAsnAspTyrThrHisTrpGluAr
 978  gtattgaatgatgattgatgtaagcgcgccatggtttaatgtgtgtgtatggttgatgca
1038  gATACGCTATTGGAGGAAGCAGTGGAGCGACAATCATCAGCCAAGGCAATAGGTTTATTG
       gTyrAlaIleGlyGlySerSerGlyAlaThrIleIleSerGlnGlyAsnArgPheIleA
1098  CGGAGGATAAATTGTTGGTGAAAGAGGTGACATATAGAGAAAAAAGTACATCAAGTGTTG
      laGluAspLysLeuLeuValLysGluValThrTyrArgGluLysSerThrSerSerValG
1158  AAGAATGGATGAAATGGACATGGATAACAGATGGTGATGATTTCGAAATGGTGCTACATT
      luGluTrpMetLysTrpThrTrpIleThrAspGlyAspAspPheGluMetValLeuHisP
1218  TCACCATCTGGTGACCAGAATTTACTCTCTAAAATTGACCATCTCAATTTGATTCAACCA
      heThrIleTrpEnd
1278  GAGCCCTCTTCTAAAGTTGGATTACTAACTAAGTTTTCTGGTGCTTTGTCTTGCAAGATT
1338  CGACGCCCATGTTAACTTCTTTCTTTTTGGTAATTGGGGTCACAAACAGGACTTTAATGT
1398  AACAATATTCCTCATATATAAAGCAATTATTCTTTT 1433
```

C.

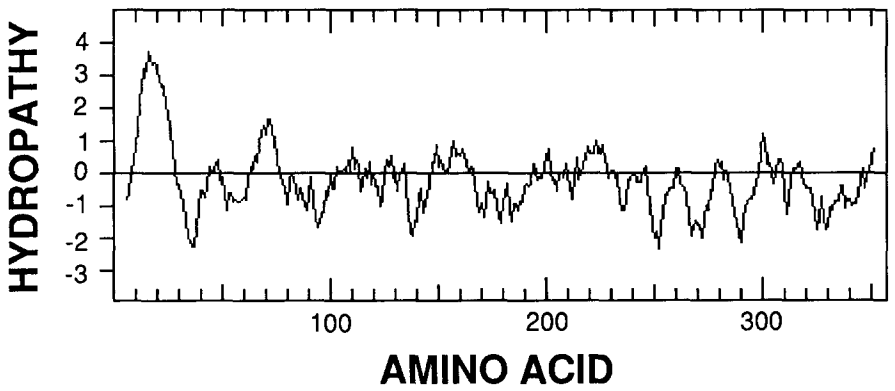
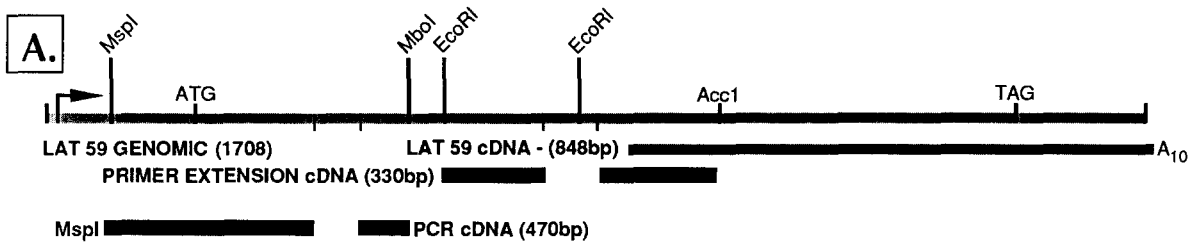


Fig. 2. Structure of LAT56. A. Graphical representation of LAT56 genomic and cDNA sequences indicating the transcription start sites (arrows), the initiator methionine (ATG), the TGA stop codon, the intron positions and the relevant restriction endonuclease sites. B. Sequence of LAT56 genomic DNA and its predicted translation product. Lower-case DNA indicates the sequences of the introns. *indicates the 5′ end of the original cDNA. Arrows indicate the position of transcription start sites. C. Hydropathy plot according to Kyte and Doolittle [17] (window size: 9 residues) of the predicted LAT56 protein sequence.



B.

```
-263  TATTTAAACAACATTTCACAAAGTGCTAAAAAAAAAAAAAAAAAAAAAAAAAACCTTTCAT
-203  ACTTGTTGTCTCATTGAGACGAGCCCTAACTTCGAAATAAAACTTCTTGTTGTCACGACC
-143  GGTCAATAAAGGGCCAGAACCCCTAGAGATTAAAAGGTTCGACCCTTGACCTTGGTAACA
 -83  ACAACAGGAAGGTTTAAAGAAAACGAAAAAAAATAAAATATCAATATAAAAAATTGTTTT
 -23  AAAAAAAAGGAAAAAAAATAGAGATGGGAGGTCCTAAGATAAAGTATAGCTTCCTGTTTC
                              MetGlyGlyProLysIleLysTyrSerPheLeuPheL
  37  TATGTATCACTTTTGCTACAATAATACCAAGTCTTATGGCTCATATTGGTCACTACGACG
      euCysIleThrPheAlaThrIleIleProSerLeuMetAlaHisIleGlyHisTyrAspG
  97  AGGTGTGGAGGAGAAGAGCCGAAGAAGCAAAGGAGTATGCCCGTAACATTTATGAACCAC
      luValTrpArgArgArgAlaGluGluAlaLysGluTyrAlaArgAsnIleTyrGluProH
 157  ATCCTGAAAATGTCACGCTTGCATTTAACCAAAAACTTCGGGAgtaagttataacttttat
      isProGluAsnValThrLeuAlaPheAsnGlnLysLeuArgAs
 217  ttattcttcatttttcatgtttgtctttatgtattcacgttatggtttgtattgcagCAC
                                                              pTh
 277  GATGAAGGAATTAAAGAAAGTAAAAGGTACACATAATAATAGCACAAGGAGGGGTCTAGG
      rMetLysGluLeuLysLysValLysGlyThrHisAsnAsnSerThrArgArgGlyLeuGl
 337  GACAAAAAAGTACACTGGACCATGCATGGTCACAAATCCAATTGATAAGTGTTGGACATG
      yThrLysLysTyrThrGlyProCysMetValThrAsnProIleAspLysCysTrpArgCy
 397  TGATCCTAATTGGGCAGACAATAGGAAAAAATTAGCAGATTGTGCAATGGGATTTGGATC
      sAspProAsnTrpAlaAspAsnArgLysLysLeuAlaAspCysAlaMetGlyPheGlySe
 457  TAAGGCCATTGGTGGCAAAGACGGTGAATTCTATGTTGTCACAGATAATTCTGATGATTA
      rLysAlaIleGlyGlyLysAspGlyGluPheTyrValValThrAspAsnSerAspAspTy
 517  TAATGATCCAAAACCCGGAACTCTTCGTCATGCTGTTATCCAAAAGGAGCCATTGTGGAT
      rAsnAspProLysProGlyThrLeuArgHisAlaValIleGlnLysGluProLeuTrpIl
 577  CATATTTAAAAGGGGTATGAACATTAGgtaagtacgtacgtttctctatatatatatatg
      eIlePheLysArgGlyMetAsnIleAr
 637  caatgcaatgcaatgcaaatagaattcaatgctcatatatatatatatatatatatgtt
 697  tcatgaatagGTTGCACCAGGAGATGATCATGCAGAGTGACAAGACGATAGATGCTCGTG
                gLeuHisGlnGluMetIleMetGlnSerAspLysThrIleAspAlaArgG

 757  GTGTTAATGTTCACATTACTAAAGGTGCTGGTATAACCCTCCAGTACATCAAGAATGTGA
      lyValAsnValHisIleThrLysGlyAlaGlyIleThrLeuGlnTyrIleLysAsnValI
 817  TCATCCACGGACTTCACATTCATGACATCGTTGAGGGTAATGGTGGAATGGTTCGGGATG
      leIleHisGlyLeuHisIleHisAspIleValGluGlyAsnGlyGlyMetValArgAspA
 877  CTGTTGACCATATTGGTATACGTACAAAGAGTGATGGAGATGGTATTTCAATCTTCGGTG
      laValAspHisIleGlyIleArgThrLysSerAspGlyAspGlyIleSerIlePheGlyA
 937  CATCATATATATGGATTGATCATGTGTCTATGCAACGTTGTTATGATGGTTTGATAGATG
      laSerTyrIleTrpIleAspHisValSerMetGlnArgCysTyrAspGlyLeuIleAspA
 997  CTGTAGAAGGATCAACAGGTATCACTATATCAAATGGACATTTCACTGATCACAACGAGG
      laValGluGlySerThrGlyIleThrIleSerAsnGlyHisPheThrAspHisAsnGluV
1057  TGATGTTGTTTGGTGCAAGTGATAGTTCTTCGATTGATCAAGTTATGCAAATAACATTAG
      alMetLeuPheGlyAlaSerAspSerSerSerIleAspGlnValMetGlnIleThrLeuA

1117  CTTTCAATCATTTTGGAAAGAGATTGATACAAAGAATGCCAAGATGTCGATGGGGATATA
      laPheAsnHisPheGlyLysArgLeuIleGlnArgMetProArgCysArgTrpGlyTyrI
1177  TACATGTTGTTAACAATGATTATACTCATTGGAATATGTATGCCATTGGTGGTAGCATGC
      leHisValValAsnAsnAspTyrThrHisTrpAsnMetTyrAlaIleGlyGlySerMetH
1237  ATCCTACTATCATTCACCAGGGTAATCGTTTTATAGCACCACCAGACATCTTCAAGAAAC
      isProThrIleIleHisGlnGlyAsnArgPheIleAlaProProAspIlePheLysLysG
1297  AGGTAACAAAAAGGGAGTACAATCCAGAATCAGTTTGGATGCAATGGACATGGAGATCAG
      lnValThrLysArgGluTyrAsnProGluSerValTrpMetGlnTrpThrTrpArgSerG
1357  AAGGAAATCTATTCATGAATGGTGCATACTTCACTGAATCAGGAGATCCAGAATGGTCAA
      luGlyAsnLeuPheMetAsnGlyAlaTyrPheThrGluSerGlyAspProGluTrpSerS
1417  GCAAACACAAAGATCTTTATGATGGAATATCAGCAGCTCCAGCAGAAGATGTCACTTGGA
      erLysHisLysAspLeuTyrAspGlyIleSerAlaAlaProAlaGluAspValThrTrpM
1477  TGACTAGATTTGCAGGTGTACTTGGCTGCAAACCAGCAAAACCTTGTTAGATCACTAATT
      etThrArgPheAlaGlyValLeuGlyCysLysProGlyLysProCysEnd
1537  CGAGCTCTTCTTTTTTTGTATCAGATCACCAACTGAAAATTCATTTTTTCCTCTCACTAA
1597  ATTATAACCCTTCGTTTTTCTTTTTCTCCTCAATGAAGCCTTAACTTTGTACATTATTTT
1657  TCTATGTACCTTCAAACAATTAGTTTCAAATAACAACAGATAGCTCTTTAT  1708
```

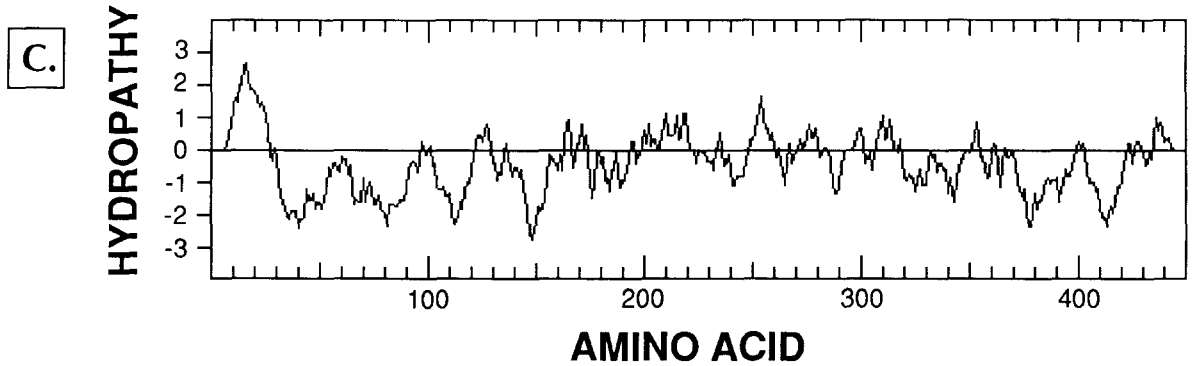


Fig. 3. Structure of LAT59. A. Graphical representation of LAT59 genomic and cDNA sequences, indicating the transcription start site, the initiator methionine (ATG), the TAG stop codon, the intron positions and the relevant restriction endonuclease sites. B. Sequence of LAT59 genomic DNA and its predicted translation product. Lower-case DNA indicates the sequences of the introns. *indicates the 5′ end of the original cDNA clone. The arrow indicates the position of the start of transcription. C. Hydropathy plot (window size as in Fig. 2) of the predicted LAT59 protein sequence.

sequence cleavage site between amino acid 22 (Pro) and 23 (Ser). However the S value for this cleavage site is 4.17, which is lower than the scores for any known signal sequence [37].

LAT56 and LAT59 were isolated independently and did not cross hybridize at standard or low-stringency (T_m = − 20 °C) hybridization and wash conditions (data not shown). Surprisingly, a comparison of the nucleic acid and deduced protein sequences of LAT56 and LAT59 indicated that these two genes are homologous. The protein sequences are aligned in Fig. 4a. The protein sequence homology between LAT56 and LAT59 extends over 328 amino acids; in this region 195 (54.7%) of the amino acids in P56 can be aligned identically with those in P59. The DNA sequence homology in the region of protein sequence homology is 62.6%. The introns in LAT56 and LAT59 (Fig. 4b) are in non-homologous positions, and the potential glycosylation sites are not conserved between the two proteins.

P56 and P59 show sequence similarity to Erwinia *pectate lyases*

Comparison of P56 and P59 with all of the protein sequences in the NBRF-protein information resource and the GenBank databases showed that P56 and P59 share significant sequence similarity with pectate lyase PLe and PLb from the bacterial plant pathogen *Erwinia chrysanthemi*

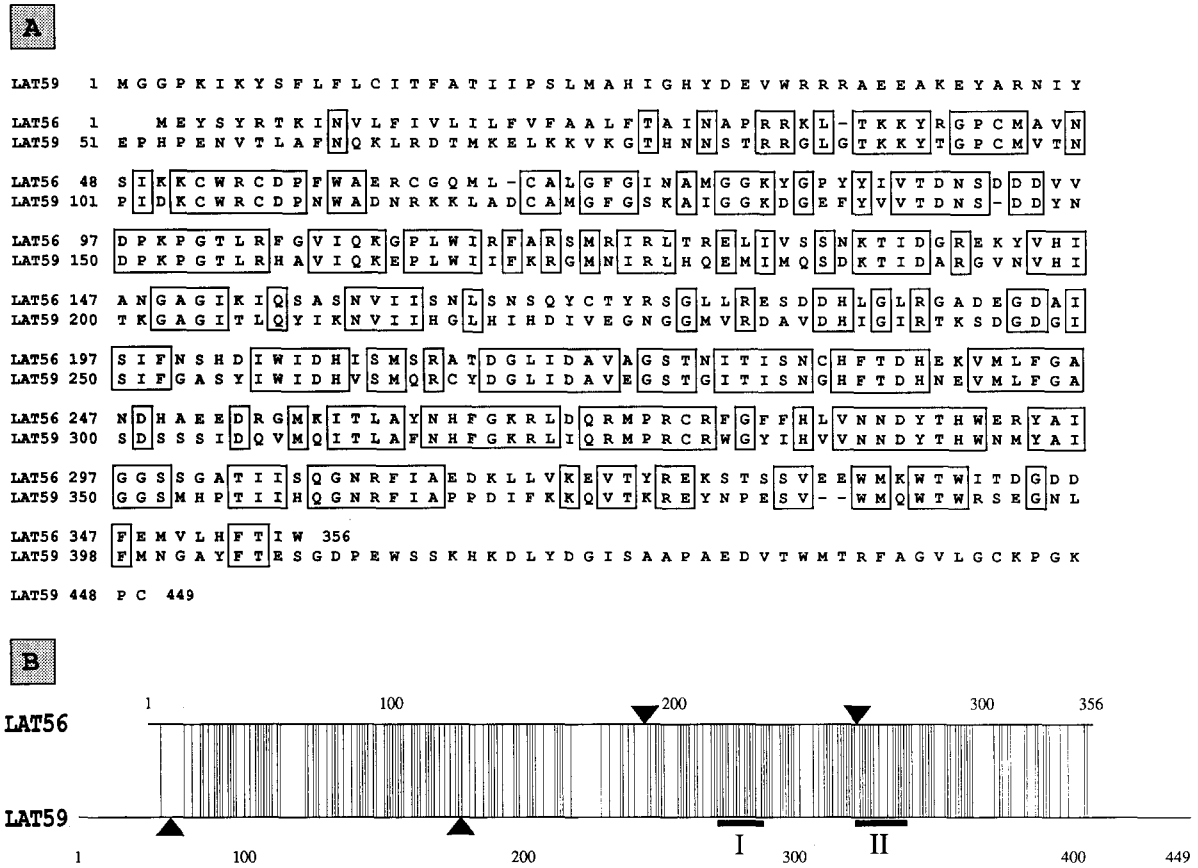


Fig. 4. Homology between LAT56 and LAT59. A. The complete deduced amino acid sequences (in one-letter code) of P56 and P59, and the alignment between the sequences. Identical amino acids are boxed. B. Graphical representation of A. The horizontal lines represent the protein sequences. The vertical lines represent amino acid identities between LAT56 and LAT59. The arrowheads show the position of the introns. The horizontal bars indicate the regions homologous to regions I and II of the *Erwinia* pectate lyases.

[13]. The protein sequence from LAT59 was similar to both PLe and PLb encoded by the *pelE* and *pelB* genes respectively. The similarity (identical amino acid and/or conservative amino acid substitution [27]) was 36.5% in a 74 amino acid overlap for PLe and 27.1% identity over a 96 amino acid overlap with PLb. The LAT56 protein sequence showed a 27.1% similarity in a 96 amino acid overlap with PLb. Further analysis of the P56 and P59 sequences showed that the similarities encompassed the so-called region I and region II homologies that were described for 7 pectate lyases of *E. chrysanthemi* and *E. carotovora* [13, 32, 18, 19, 12]. Notably, regions I and II are the only sequences that are conserved between all these *Erwinia* pectate lyases.

The extent of the amino acid similarities of P56 and P59 with region I and region II of the pectate lyase sequences is shown in Fig. 5. In region I there is a 61% amino acid sequence identity between either P56 or P59 and the PL sequences from *Erwinia*. Region II shows a 54.5% amino acid sequence identity. Both P56 and P59 have 33 amino acids in the interregion between region I and region II. The spacing between region I and region II in the *Erwinia* proteins varies between 29 and 34 amino acids. The P56 and P59 interregion also contains significant similarity to the various *Erwinia* pectate lyases (*ca.* 51%), equivalent to the level of homology seen between, for example, PLa and PLe of *E. chrysanthemi*.

By introducing either a 17 amino acid gap for PLa and PLe of *E. chrysanthemi* or a 13 amino acid gap for the rest of the *Erwinia* pectate lyases, another region of amino acid similarity (30 amino acids) arises between P56 and P59 and all of the pectate lyases; the similarity is comparable to the similarity between Ple and Plb from *E. chrysanthemi*.

If conservative amino acid changes are considered [27] the similarities between the pectate lyase genes of *Erwinia* and P56 and P59 increase. P59 is then 75% (54/72 amino acids) similar in the region I-interregion-region II interval, while LAT56 is 71% (51/72) similar. P56 and P59 show no other regions of significant similarity to the *Erwinia* pectate lyases. Pectin methylesterase from *E. chrysanthemi* [28] also contains a sequence similar to region I of the PL genes, and to LAT56 and LAT59 (Fig. 5).

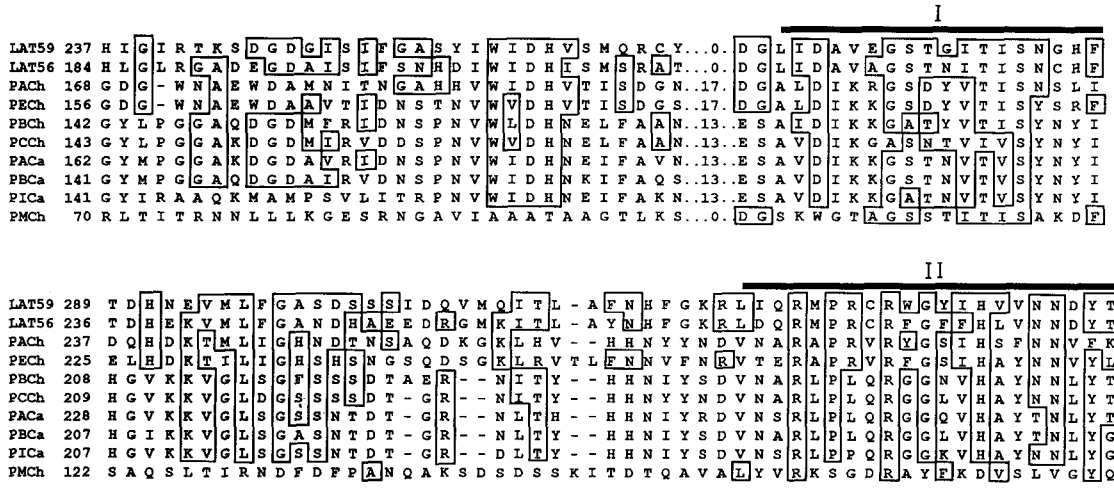


Fig. 5. Amino acid identities between LAT56, LAT59 and the pectate lyases and pectin methylesterase from *Erwinia* spp. Boxes are drawn around amino acids that are identical in either LAT56 or LAT59, and at least one of the PL sequences. PACh, PBCh and PCCh and PECh are pectate lyases A, B, C and E from *E. chrysanthemi*, PACa, PBCa, and PICa are pectate lyases A, B and I from *E. carotovora*, and PMCh is pectin methylesterase from *E. chrysanthemi*. The horizontal bars indicate the consensus regions I and II found in the *Erwinia* PLs. Either a 17 or 13 amino acid gap was introduced in the PL proteins in order to align the amino acid region immediately before region I.

Sequence similarity to cedar pollen allergen

The major pollen allergen of japanese cedar is a basic protein of about 50 kDa. The N-terminal 20 amino acids of this protein are D N P I D S? W R G D S N W A Q N R M K [33]. P59 shows 65% amino acid identity to this sequence, beginning at amino acid 99, while P56 shows 35% amino acid identity, beginning at amino acid 46.

Discussion

LAT56 and LAT59 show similar tissue specificity and timing of expression in developing anthers and pollen (Fig. 1). *In situ* hybridizations [36] showed that transcripts homologous to LAT56 and LAT59 are present in microspores, pollen, pollen tubes and in the anther wall. This expression pattern is essentially the same as previously seen for LAT52. Because LAT56, LAT59 and LAT52 appear to be expressed coordinately it is possible that a regulatory protein would bind a similar sequence shared between all three genes. However, comparison of the 5′ flanking DNA of all three genes has not yet revealed any obvious sequence similarities (unpublished).

That LAT56 and LAT59 are genetically linked (Table 1) and share sequence homology might suggest that these two genes arose via an ancestral gene duplication. The first intron in LAT59 interrupts a region of homology between LAT56 and LAT59 (Fig. 4b), but the other three introns occur in non-conserved regions. Intron positions in genes with functional homologies are usually conserved within species [20]; so the differing intron positions in LAT56 and LAT59 might suggest that the introns were inserted after a gene duplication event.

LAT56 and LAT59 code for homologous proteins. However, there are significant sequence and protein differences, including the amino and carboxy terminal extensions in LAT59 and the non-conserved glycosylation sites, that may reflect different functions, or different cellular or extracellular targets for the proteins. Both P56 and P59 have hydrophobic amino termini suggesting that these proteins may be targeted for secretion (Figs. 2c and 3c, respectively). The homology seen between P59, P56 and the N-terminal region of the cedar pollen allergen [33] lends additional support to this hypothesis. Pollen allergens are generally considered to be associated with the pollen grain wall [14]. Antibodies to the P56 and P59 proteins will be required to definitively test whether these proteins are secreted to the pollen wall.

P56 and P59 are 70% similar to bacterial pectate lyases in the region I-interregion-region II interval (Fig. 5). P56 and P59 maintain the spacing between region I and II that is seen in the *Erwinia* pectate lyases, and have an additional region of homology (5′ of region I) to the *Erwinia* pectate lyases. Because of the sequence conservation between P56 and P59 and the *Erwinia* PL proteins, it is tempting to consider that region I and region II have some functional significance, and that LAT56 and LAT59 code for pectate lyases. However, even in *Erwinia* it is not clear which domains of the pectate lyase proteins are required for function. A pectate lyase isolated from the non-plant pathogenic bacterium *Yersinia pseudotuberculosis* has no sequence similarity to the *Erwinia pel* genes [22]. However, an additional pectate lyase gene from *E. carotovora* that has homology to the *Yersinia* pectate lyase was recently characterized [34]. From these results it is clear, even within *Erwinia*, that there are multiple protein sequences that can exhibit pectate lyase activity. Pectate lyases have also been characterized from other plant pathogenic bacteria [4] and from fungal plant pathogens [5], although nucleotide or protein sequences have not yet been reported for these proteins. It will be interesting to see if these proteins share sequence homology with the *Erwinia* pectate lyases, or with P56 and P59. Gene constructions that replace the region I-region II of an *Erwinia* pectate lyase gene with that portion of LAT56 or LAT59 might help determine whether this region is important for enzyme function.

Acknowledgements

We thank Noel Keen for helpful discussions and Joe DeVerna and Roger Chetelat for the gift of chromosome addition line Southern blots. We thank Barbara Baker, Mike Christianson and Reinhard Hehl for critically reading this manuscript. We also thank Brian Osborne for PCR expertise, Chuck Gasser for the tomato genomic library, and Stephen Goff and Karen Suslow for synthesis of oligonucleotides. This work was supported by CRIS grant 5335-22230-002-00D.

References

1. Benton WD, Davis RW: Screening lambda gt recombinant clones by hybridization to single plaques. Science 196: 180–182 (1977).
2. Bernatzky R, Tanksley SD: Genetics of actin related sequences in tomato. Theor Appl Genet 72: 314–321 (1986).
3. Clarke AE, Gleeson PA: Molecular aspects of recognition and response in the pollen-stigma interaction. Rec Adv Phytochem 15: 161–211 (1981).
4. Daniels MJ, Collinge DB, Dow JM, Osbourn AE, Roberts IN: Molecular biology of the interaction of *Xanthomonas campestris* with plants. Plant Physiol Biochem 25: 353–359 (1987).
5. Dean RA, Timberlake WE: Production of cell wall -degrading enzymes by *Aspergillus nidulans*. Plant Cell 1: 275–284 (1989).
6. DeVerna JW, Chetelat RT, Rick CM: Cytogenetic, electrophoretic, and morphological analysis of sesquidiploid *Lycopersion esculentum* – *Solanum lycopersicoides* hybrids x *L. pennellii*. Biol Zentrbl 106: 417–428 (1987).
7. Feinberg AP, Vogelstein, S: A technique for radiolabelling DNA fragments to high specific activity. Anal Biochem 137: 266–267 (1984).
8. Gidoni D, Bond-Nutter D, Brosio P, Jones J, Bedbrook J, Dunsmuir P: Coordinated expression between two photosynthetic petunia genes in transgenic plants. Mol Gen Genet 211: 507–514 (1988).
9. Hanson DD, Hamilton DA, Travis JL, Bashe DM, Mascarenhas JP: Charaterization of a pollen-specific cDNA clone from *Zea mays* and its expression. Plant Cell 1: 173–179 (1989).
10. Heslop-Harrison J: Aspects of the structure, cytochemistry and germination of the pollen of rye (*Secale cereale* L.). Ann Bot 44 (suppl 1): 1–47 (1979).
11. Huynh TV, Young RA, Davis RW: Constructing and screening cDNA libraries in lambda gt10 and gt11. In: Glover D (ed) DNA Cloning Techniques – A Practical Approach, Vol. 1, pp. 49–78. IRL Press, Oxford (1985).
12. Ito K, Kobayashi R, Nikaido N, Izaki K: DNA structure of pectate lyase I gene cloned from *Erwinia carotovora*. Agric Biol Chem 52: 479–487 (1988).
13. Keen NT, Tamaki S: Structure of two pectate lyase genes from *Erwinia chrysanthemi* EC16 and their high-level expression in *Echerichia coli*. J Bact 168: 595–606 (1986).
14. Knox RB, Heslop-Harrison J, Reed C: Localization of antigens associated with the pollen grain wall by immunofluorescence. Nature 225: 1066–1068 (1970).
15. Kotoujansky A: Molecular genetics of pathogenesis by soft-rot Erwinias. Annual Rev Phytopath 25: 405–430 (1987).
16. Kroh M, Miki-Horosige H, Rosen W, Loewus F: Incorporation of label into pollen tube walls from myoinositol-labeled *Lilium longiflorum* pistils. Plant Physiol 45: 92–94 (1970).
17. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132 (1982).
18. Lei S-P, Lin H-C, Wang S-S, Callaway J, Wilcox G: Characterization of the *Erwinia carotovora pelB* gene and its product pectate lyase. J Bact 169: 4379–4383 (1987).
19. Lei S-P, Lin H-C, Wand S-S, Wilcox G: Characterization of the *Erwinia carotovora pelA* gene and its product pectate lyase A. Gene 62: 159–164 (1988).
20. Ludwig SR, Oppenheimer DG, Silflow CD, Snustad DP: Characterization of the α1-tubulin gene family of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 84: 5833–5837 (1987).
21. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
22. Manulis S, Kobayashi DY, Keen NT: Molecular cloning and sequencing of a pectate lyase gene from *Yersinia pseudotuberculosis*. J Bact 170: 1825–1830 (1988).
23. Marck C: 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucl Acids Res 16: 1829–1836 (1988).
24. McCormick S, Smith A, Gasser C, Sachs K, Hinchee M, Horsch R, Fraley R: Identification of genes specifically expressed in reproductive organs of tomato. In: Nevins DJ, Jones RA (eds) Tomato Biotechnology, pp. 255–265. Alan R. Liss, New York (1987).
25. Melton DA; Krieg PA, Rebagliatti MR, Maniatis T; Zinn K, Green MR: Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl Acids Res 12: 7035–7056 (1984).
26. Messing J, Geraghty D, Heidecker G, Hu N-T, Kridl J, Rubenstein I: Plant gene structure. In: Kosuge T, Meredith CP, Hollaender A (eds) Genetic Engineering of Plants, pp. 211–227. Plenum, New York (1983).
27. Pearson W, Lipman DJ: Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448 (1988).

28. Plastow GS: Molecular cloning and nucleotide sequence of the pectin methyl esterase gene of *Erwinia chrysanthemi* B374. Mol Microbiol 2: 247–254 (1988).
29. Rochester DE, Winter JA, Shah DM: The structure and expression of maize genes encoding the major heat shock protein, hsp70. EMBO J 5: 451–458 (1986).
30. Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, Horn G, Mullis K, Erlich H: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487–491 (1988).
31. Suiter K, Wendel J, Case J: Linkage-1: a PASCAL computer program for the detection and analysis of genetic linkage. J Heredity 74: 203–204 (1983).
32. Tamaki SJ, Gold S, Robeson M, Manulis S, Keen NT: Structure and oranization of the *pel* genes from *Erwinia chrysanthemi* EC16. J Bact 170: 3468–3478 (1988).
33. Taniai M, Ando S, Usui M, Kurimoto M, Sakagushi M, Inouye S, Matuhasi T: N-Terminal amino acid sequence of a major allergen of Japanese cedar pollen (*Cry j* I) FEBS Lett 239: 329–332 (1988).
34. Trollinger D, Berry S, Belser W, Keen NT: Cloning and characterization of a pectate lyase gene from *Erwinia carotovora* EC153. Mol Plant Microbe Interact 2: 17–25 (1989).
35. Twell D, Wing RA, Yamagushi J, McCormick S: Isolation and expression of an anther-specific gene from tomato. Mol Gen Genet, 217: 240–245 (1989).
36. Ursin VM, Yamagushi J, McCormick S: Gametophytic and sporophytic expression of anther-specific genes in developing tomato anthers. Plant Cell, 1: 727–736 (1989).
37. von Heijne G: A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14: 4683–4690 (1986).