ORIGINAL PAPER

ps-2, the gene responsible for functional sterility in tomato, due to non-dehiscent anthers, is the result of a mutation in a novel polygalacturonase gene

**Benoit Gorguet · Danny Schipper ·
André van Lammeren · Richard G. F. Visser ·
Adriaan W. van Heusden**

Abstract The recessive mutation *ps-2*, which appeared spontaneously in tomato, confers functional male sterility due to non-dehiscent anthers. In this study, we isolated and characterized the *PS-2* gene. A single nucleotide mutation in a novel tomato polygalacturonase gene is responsible for the *ps-2* phenotype. The mutation in *ps-2* is responsible for an alternative splicing during maturation of the pre-mRNA, which leads to an aberrant mRNA. Differentiation between *ps-2* and wild type (*PS-2*) anthers only appears in the final developmental stage in which the stomium remains closed in the mutant. To our knowledge, this is the first functional sterility gene isolated in the Solanaceae family. The specific expression of the *Arabidopsis* homolog of *PS-2* in the anther dehiscence zone suggests a conserved mode of action over the plant kingdom, which means that the repression of *PS-2* homologs may be a potential way to introduce functional sterility in other species.

B. Gorguet · D. Schipper · R. G. F. Visser ·
A. W. van Heusden (✉)
Graduate School of Experimental Plant Sciences,
Wageningen UR Plant Breeding, PO Box 386,
6700 AJ Wageningen, The Netherlands
e-mail: sjaak.vanheusden@wur.nl

B. Gorguet
De Ruiter Seeds, Leeuwenhoekweg 52,
PO Box 1050, 2660 BB Bergschenhoek, The Netherlands
e-mail: benoit.gorguet@deruiterseeds.com

A. van Lammeren
Graduate School of Experimental Plant Sciences,
Wageningen UR Laboratory of Plant Cell Biology,
Arboretumlaan 4, 6703 BD Wageningen, The Netherlands

Introduction

In higher plants mature pollen are released from the anther by dehiscence, which consists of a succession of cell destructions occurring successively in the tapetum, the septum and ultimately in the stomium. In tomato, after degeneration of the tapetum, exothecium cells are reinforced, creating a breaking force in the stomium that eventually leads to pollen release (Keijzer 1987).

Recent progress was made in understanding the molecular control of anther dehiscence, which involves mainly the discovery of the implication of jasmonic acid (JA) and ethylene. Several mutants affecting diverse steps in the synthesis of JA in the anthers were identified in *Arabidopsis*. These mutants displayed a delay of anther dehiscence (reviewed by Scott et al. 2004). The role of ethylene signaling in this phenomenon was highlighted by Rieu et al. (2003) who observed a delay in the dehiscence of anthers of ethylene insensitive tobacco plants.

Polygalacturonases (PGs) belong to one of the largest hydrolase families (Torki et al. 2000; Markovic and Janecek 2001). The PG activities are associated with a wide range of plant developmental programs (reviewed by Hadfield and Bennett 1998), among them, anther dehiscence. The activity of PGs has been observed in the dehiscence zone of anthers of maize, tobacco, oilseed rape and *Arabidopsis* (Dubald et al. 1993; Sander et al. 2001). However their role in the dehiscence process has never been studied in detail.

We have recently fine-mapped the *positional sterility-2* gene, conferring non-dehiscent anthers in tomato (Gorguet et al. 2006). Here, we report the isolation of the *ps-2* gene

by positional cloning. A single mutation in the coding sequence of a novel polygalacturonase gene is responsible for this phenotype. We found that this mutation affects one of the intron splicing recognition sites of the gene, giving rise to an aberrant mRNA, lacking one of the exons. Microscopic observation of the *ps-2* phenotype revealed similarities with what was already found in *Arabidopsis* and tobacco mutants (reviewed by Scott et al. 2004; Rieu et al. 2003). This new *PG* gene, is also expressed in maturing fruits, which opens the way to speculation on the possible role of *Ps-2* in fruit maturation.

Materials and methods

Plant material

The F$_2$ recombinant sub-population developed from the cross between *ps-2*ABL (*S. lycopersicum*, homozygous for the *ps-2* mutation) and *S. pimpinellifolium*, was used for genetic mapping. This population, segregating for *ps-2*, is composed of 146 F$_2$ recombinant plants in the *ps-2* locus region (Gorguet et al. 2006).

Another 176 advanced breeding lines (ABLs), 8 of them being *ps-2/ps-2*, were used to test the association between the identified SNP and the *ps-2* locus.

Anthers of *ps-2*ABL and cv. Moneymaker were used for microscopic observations.

Microscopy

Plant material was fixed for 24 h at 4°C in 0.1 M phosphate buffer, pH 7.0, containing 4% paraformaldehyde. Samples for scanning electron microscopy (SEM) were processed as described in Dornelas et al. (2000), and digital images were obtained using an Orion Framegrabber (Matrox Electronic Systems, Unterhaching, Germany). Samples for light microscopy were embedded in Technovit 7100 (Hereaus Kulzer, Wehrheim, Germany), sectioned, stained with toluidine blue, and mounted in Euparal (Chroma-Gesellschaft, Kongen, Germany).

BAC library screening and contig construction

We used the tomato *Hin*dIII BAC library constructed from genomic DNA of the cultivar accession Heinz 1706. The Heinz library is a 15-genome equivalent with an average insert size of 114.5 kb (Budiman et al. 2000). Screening of the BAC library was performed by PCR amplification, first on plate pools and then on individuals. Plasmid DNA of the positive BAC clones was then isolated and used for further analysis.

BAC ends sequences of the positive BAC clones were obtained from the SGN database (Mueller et al. 2005).

Conversion of the BAC ends sequences into CAPS or dCAPS markers was performed as described by Gorguet et al. (2006). Details of the PCR markers derived from BAC ends sequences are presented in Table 1.

Fingerprinting patterns of individual BAC clones were generated essentially as described by Brugmans et al. (2006), using the *Hin*dIII/*Taq*I enzyme combination.

BAC DNA sequencing and candidate gene analysis

The size of BAC clone 143M15 was estimated by pulse field gel electrophoresis. The BAC clone 143M15 was sequenced via the shotgun-sequencing method by Greenomics (The Netherlands). PCR markers derived from the candidate genes identified on the BAC sequence were developed as described by Gorguet et al. (2006). To identify putative genes, the final BAC DNA sequence was scanned against the tomato Unigene database from SGN (Mueller et al. 2005) and the *Arabidopsis* gene models database from TAIR (http://www.Arabidopsis.org; Huala et al. 2001), using the TBLASTX interface of SGN (Mueller et al. 2005), with a significance threshold of $1E^{-10}$. The PCR markers based on putative ORF sequences were developed and screened on the recombinant population as described by Gorguet et al. (2006). Details of these PCR markers are given in Table 1.

The complete genomic DNA sequence of ORF4, as well as up-stream and down-stream sequences containing, respectively, promoter and terminator was amplified and sequenced from Moneymaker and *ps-2*ABL using several successive overlapping primer pairs giving products of around 900 bp. The resulting DNA sequences of Moneymaker and *ps-2*ABL were assembled with DNAStar. The FGENESH software was used to identify the putative exons and introns of the candidate gene (Salamov and Solovyev 2000).

Phylogenetic analysis

Tomato polygalacturonase protein sequences with known tissue expression, as well as the best hits of a protein BLAST search with the candidate protein, were selected to conduct a phylogenetic analysis. Only the Pfam glycosyl hydrolase 28 domains, of the selected protein sequences, were used for the analysis. The Pfam GH28 domain of each protein sequence was identified with the protein BLAST interface of NCBI. Selected amino acid sequences of Pfam GH28 domains were aligned using ClustalW multiple sequence alignment software (Higgins et al. 1994).

PAUP software package version 4 (Swofford 2002) was used to construct a 50% majority-rule consensus phylogenic tree using maximum parsimony (1,000 bootstrap replicates and 250 addition sequences replicates). Cedar PG protein sequence was used and defined as outgroup.

Table 1 PCR markers used in the genetic linkage maps

Marker name	Use	Primer sequence	Size (bp)	Restriction enzyme
Markers developed from BAC ends sequences				
67F23-T	CAPS	Fw: CTACTCTTCCGCCATAACTG	599	*Hin*cII
		Rv: GATCCAAACGAACAAAAGTCA		
67F23-S	CAPS	Fw: TCATTCCGTTGCTGAATGAGA	413	*Dra*I
		Rv: ATAACTTATATCACTCCCAATCA		
69C22-T	dCAPS	Fw: TCTTTCGATATTTTTCAGAACTAA	200	*Dde*I
		Rv: TGAGATGTTTGCAATAACATTCT		
143M15-S	CAPS	Fw: CATCGAAGTAACAGAGATATTA	369	*Mwo*I
		Rv: CCATAGGGATTATGATGTGTA		
114C15-T	CAPS	Fw: GCACTGAAGAATGGATAGACTC	457	*Mnl*I
		Rv: GGAATTGACCAAAAAGGATAGC		
118A17-T	CAPS	Fw: GGCATGGTGAAGTCCACATT	739	*Hae*III
		Rv: GTGTCACAGGTTTGGTTCAT		
15N23-T	CAPS	Fw: GGCAGATATCTGCAATACGT	576	*Taq*I
		Rv: ATCATGAACAGCAAAACAACCA		
Markers developed from BAC 143M15 sequence				
ORF1	CAPS	Fw: CTGTATCTATGACGAGGAGA	625	*Hae*III
		Rv: GATCCTGAAGCTGAAGCTT		
ORF2	CAPS	Fw: AATATTTTCAACTTTCAAATCTCTT	206	*Mnl*I
		Rv: ACGAAGGCATGATTGTCGTTA		
ORF3	CAPS	Fw: GTTGAACTTATACCACTAGGA	937	*Nde*I
		Rv: GTGCGGTCTCATCAACTCAA		
ORF4	dCAPS	Fw: GAACACTTAGGTTAAAATATAGC	217	*Alu*I
		Rv: CCTACTATCCTTCTTGTAATCT		
ORF5	CAPS	Fw: CTTAAAGGCACACTTAGATTCA	962	*HpyCH*4IV
		Rv: CTGAGAATTCTCTTGACTGCA		
ORF4(1)	SCAR	Fw: GCTTTATTCATAGTAAATTCTGT	805/ 885	
		Rv: TCAGACAAATCATCGTATATTGA		
ORF4(2)	SCAR	Fw: TCCATTTGTAGTTTCATAAAGC	465/ 515	
		Rv: CCAAGCGGATAATTAATGTCA		
Marker developed for *ps-2*ABL allele identification within *S. lycopersicum*				
ps-2 marker	CAPS	Fw: CAAATTGGATGAGAGTTTTGAA	695	*HpyCH*4IV
		Rv: CATTTTACAAGTGTAACAACTTG		

The accession numbers of the amino acids sequences in the phylogenetic tree (Fig. 5) are as follows; Kiwi fruit (AAC14453), grape berry fruit (AAK81876), soybean pods (AAL30418), peach fruit (CAA54448), apple fruit (AAA74452), pear fruit (BAC22688), *Arabidopsis* dehiscence zone ADPG1 (CAA05525), oilseed rape dehiscence zones RDPG1 (CAA65072), oilseed rape pod (CAA90272), turnip silique valve desiccation (CAD21651), bell pepper fruit (BAE47457), tomato fruit TFPG (CAA32235), tomato pistil (AAC70951), tomato abscission zones TAPG1, 2, 4, 5 (AAC28903, AAB09575, AAB09576, AAC28906), tomato wound leaf (AAD17250) and tomato seed (AAF61444). The accession number of the tomato *β*-actin is: BT013524.

Total RNA isolation, cDNA synthesis and quantitative PCR analysis

The total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Between 50 and 100 mg of each tissue (anthers, fruit, flower abscission zone and leaf abscission zone) was used per RNA isolation reaction. Only 1 μg of total RNA, after DNase I treatment (Boehringer Manheim), was used per sample for the synthesis of cDNA. First strand cDNA template was synthesized using random hexamers as primers and Multiscribe reverse transcriptase (Applied Biosystems). The nearly complete coding sequence of ORF4 was amplified to study the intron splicing using the forward primer: TAGCTCCAAAGCTATCC

ACAT, located on the first exon, starting 47 nucleotides down-stream the start codon and the reverse primer: TGGAGAATGTGAAATTGTTAGG, located on the last exon, stopping 100 nucleotides up-stream the stop codon (Fig. S1). Nearly complete CDS of ORF4 was amplified with standard PCR reaction (55°C annealing temperature and 35 cycles).

Real-time experiments were conducted in an iCycler MyiQ detection system (Bio-Rad), using the SYBR green PCR master mix kit (Applied Biosystems). The primer sequences were: forward primer 5′-TTTTGCCATTGCCA TTGATA-3′, reverse primer 5′-TGTGGTGTCCCAGAAC AAGA-3′ (ORF4); and forward primer 5′-ACCACTTTCC GATCTCCTCTC-3′, reverse primer 5′-ACCAGCAAAT CCAGCCTTCAC-3′ (β-actin). Relative quantification of the ORF4 transcript level was calculated with the internal β-actin control by applying the $2^{-\Delta CT}$ formula. The purity of the PCR products was verified with their melting curves. The experiment was performed twice with true biological replicates. The PCR controls were performed in the absence of added reverse transcriptase to ensure that the RNA samples were free of DNA contamination. In addition, the ORF4 primers were designed on either sides of an intron so that the PCR product obtained on genomic DNA was significantly longer than on cDNA and that the difference could easily be observed on gel.

The sequence of the tomato *positional sterility-2* (*Ps-2*) genomic DNA was deposited in GenBank (accession number EU111748).

Results

Microscopic observation of the *ps-2* phenotype

Transverse sections of anthers of wild type (Moneymaker) and mutant (*ps-2*ABL) were prepared and stained with toluidine blue to identify the stage at which anther development/dehiscence in the mutant is blocked. Anthers at anthesis were prepared for SEM.

Early anther development and the fusion of the locules were similar in mutant and wild type: breakage of the septum, by which the locules containing the microspores fused, did occur in the mutant at a similar stage as in the wild type at pre-anthesis. Crystals appeared in the septum cells of mutant and wild type, and pollen development appeared normal (Fig. 1a). Breakage of the septum did occur in the mutant at similar stage than the wild type. However, at anthesis, the mutant stomium did not degenerate and the pollen remained in the anthers. After anthesis, the exothecium in the wild type was still discernable as a rigid structure, whereas the exothecium of the mutant was flattened and bent. Thus, it is likely that in the mutant, at

anthesis, the exothecium cells lacked the rigidity needed to create a breaking force on the stomium.

Physical mapping and candidate gene identification

We have previously mapped the *ps-2* locus to an interval of 1.65 cM defined by the COS-derived CAPS markers T0958 and T0635 on the short arm of chromosome 4 (Gorguet et al. 2006). A physical map for the *ps-2* locus region was built using the Heinz BAC library. The library was screened by PCR amplification with the closest markers relative to the *ps-2* locus and by computational means using the sequences of those markers. Positive BACs were then anchored to the genetic map by converting BAC ends into PCR markers (Fig. 2) and by screening these markers in the recombinant population. The BAC fingerprints were also compared to evaluate the overlapping of BACs from the same contig and verify whether BACs were part of the same contig. The closest BAC end relative to the *ps-2* locus was then used for a second round of screening. Eventually, the entire contig spanned 1.70 cM (32 recombinants) from the COS-derived CAPS marker T1070 to BAC end 15N23-T (Fig. 2). BAC 143M15 did obviously span the *ps-2* locus and was therefore sequenced.

In order to identify the genes present in BAC clone 143M15, the BAC DNA sequence was scanned against the Tomato SGN Unigene database, using a BLASTN interface, and against the *Arabidopsis* gene models database, using a TBLASTX interface. Two tomato coding sequences and five *Arabidopsis* genes matched the BAC clone sequence (Table 2). The five candidate genes were named ORF1 to ORF5. In addition, six sequences of a retrotransposon family were also identified but were not taken into account in the further study. The positions of the five corresponding *Arabidopsis* genes were not contiguous in the *Arabidopsis* genome, which showed the absence of synteny between the two species in this specific region. Moreover, none of them was homologous to one of the functional male sterility genes identified in *Arabidopsis* (listed in Gorguet et al. 2006).

In order to localize the candidate genes in the high-resolution map, we converted the putative gene sequences (or the sequences nearby) into PCR markers and mapped them in the recombinant population. Every candidate gene mapped at a different position in the high-resolution linkage map and therefore we could easily identify the likeliest candidate for *ps-2* based on their positions in the genetic map. ORF4, a putative polygalacturonase gene, mapped the closest to the *ps-2* locus (Fig. 2). Subsequently, the putative introns and exons were identified using the FGENSH software of Softberry. The candidate gene ORF4 is composed of nine exons and eight introns, covering a genomic distance of 6,716 nucleotides from putative start to stop codon,
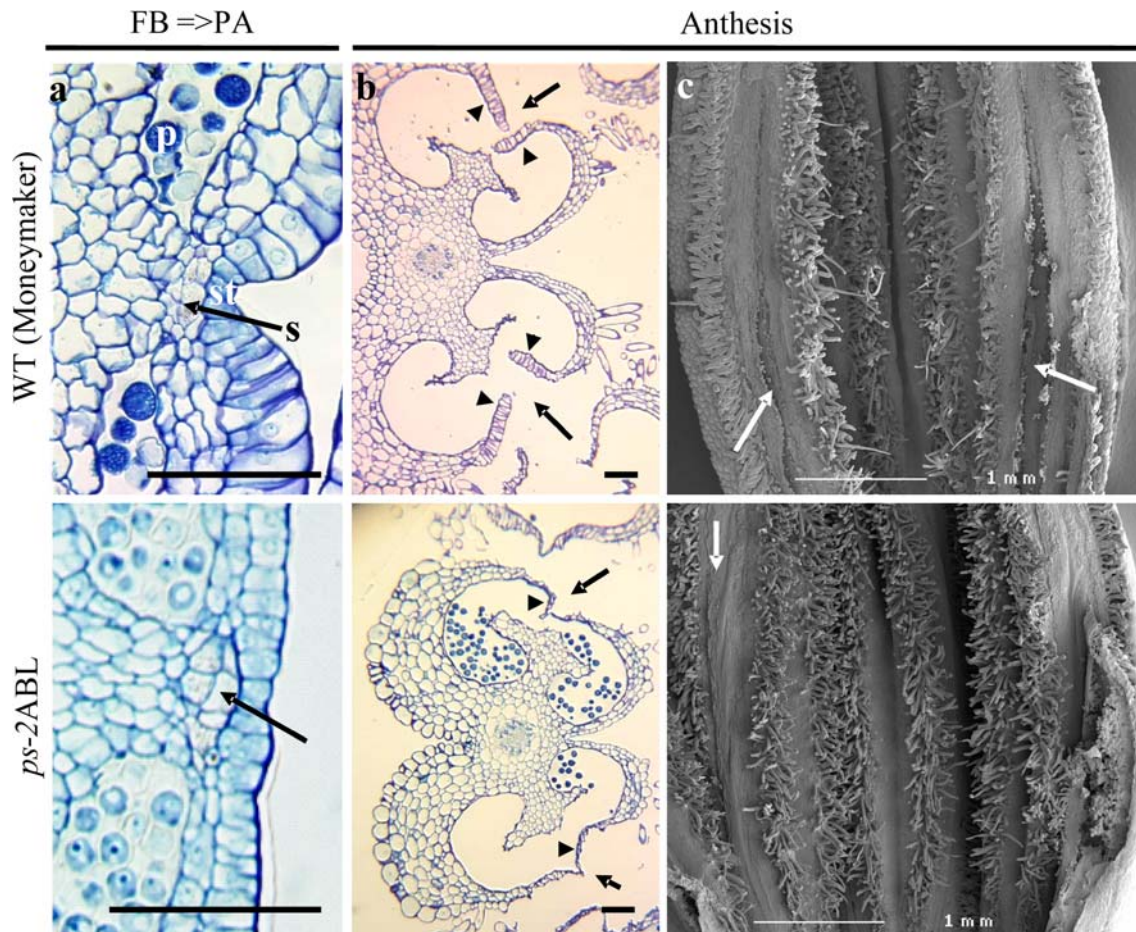
Fig. 1 Microscopic observation of anthers of *ps-2*ABL and wild type (*WT*; Moneymaker). **a** Cross-section of anthers at late flower bud (*FB*), early pre-anthesis (*PA*) stained in toluidine blue. *Arrows* refer to the crystals that can be seen in the septum cells of WT and *ps-2*ABL. *s* septum, *st* stomium, *p* pollen. **b** Cross-section of anthers at anthesis stained in toluidine blue. *Arrows* indicate the opening of the anther in WT and the stomium that remains closed in *ps-2*ABL because stomium cells are not torn apart. Note the rigid locule walls in wild type and the flattened ones in the mutant (*arrowheads*). **c** Longitudinal view of anther cones at anthesis observed by SEM. *Arrows* indicate the longitudinal opening of the anthers in WT with pollen presented along the opening edges, and the closed anthers in *ps-2*ABL. Some pollen is visible in the mutant anther at the right-hand side, due to cutting during preparation

for a coding sequence of 1,179 nucleotides. The SNP used to develop the PCR marker ORF4 was located in the second putative intron. We developed and mapped two extra PCR markers, one based on a deletion of 76 bp in the first intron in the *S. pimpinellifolium* allele [ORF4(1)] and one based on an insertion of 38 bp in the sixth intron in the *S. pimpinellifolium* allele [ORF4(2)]. The three PCR markers, ORF4(1), ORF4 and ORF4(2) mapped at an interval of one recombinant between each other, indicating that at least two recombinations had occurred within the candidate gene ORF4 in the recombinant population (Fig. 2). The ORF4(2) co-segregated with the *ps-2* locus on the high-resolution map.

Mutation in ORF4 and molecular marker development

To consolidate the hypothesis that ORF4 corresponds to the *ps-2* gene, we searched for sequence alterations in the *ps-2*ABL allele. The entire 9 kb, from the putative promoter to the putative terminator of ORF4 was sequenced from cv. Moneymaker and *ps-2*ABL. One single mutation was identified in the last nucleotide of the fifth putative exon of ORF4, in which the nucleotide guanine was replaced by cytosine. To test the association between the identified SNP in ORF4 and the *ps-2* trait, we developed a molecular marker based on that SNP, in such a way that the *ps-2*ABL allele and the wild type allele in *S. lycopersicum* could easily be differentiated on gel. This marker was tested on a set of 176 ABLs among which eight were *ps-2*/*ps-2*. These eight male sterile plants showed the same marker pattern, distinct from the other ABLs, which confirms that the SNP is present in the *ps-2* lines tested. This marker can now easily be used for molecular-assisted introduction of the *ps-2* trait into modern tomato lines.
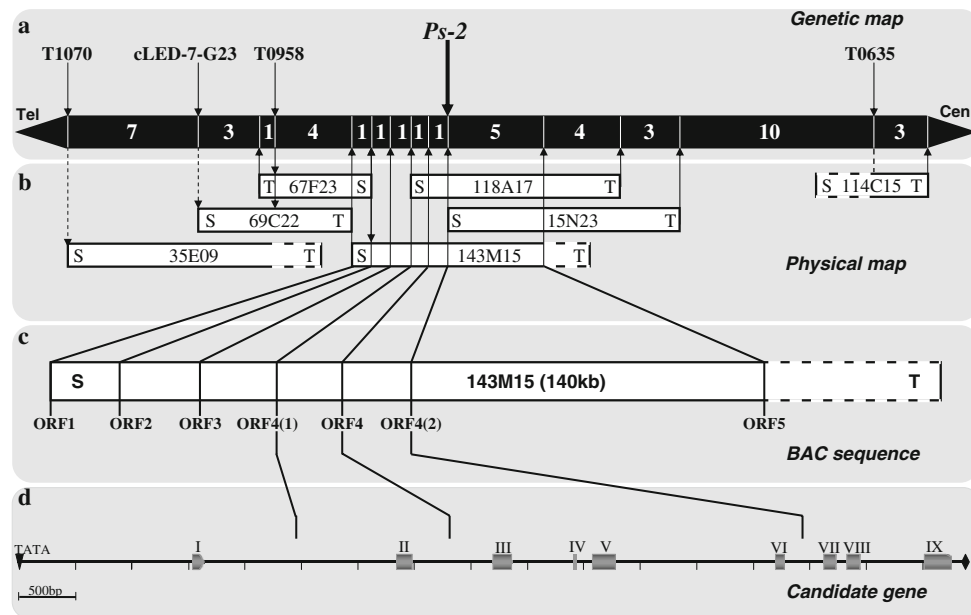
Fig. 2 Cloning of the *ps-2* gene, from genetic map to gene structure. **a** Genetic map developed in a recombinant F$_2$ population (*ps-2*ABL × *S. pimpinellifolium*; Gorguet et al. 2006). In *white* are the numbers of recombinant plants between each marker. **b** Physical map: *arrows* in *dashed* represent the computational anchoring of BACs. *Arrows* in full line represent the anchoring of BACs by molecular markers (*S*: SP6; *T*: T7). **c** Genetic positions of the ORFs on BAC clone 143M15. **d** Structure of candidate gene ORF4. *Cylinders* with *roman numbers* represent exons. Positions of the putative TATA box and PolyA signal sequence are indicated by an inverted *triangle* at the beginning of the sequence and a *rhombus* at the end of the sequence, respectively

Table 2 Identification of candidate genes present on BAC 143M15 based on *Arabidopsis* gene models

Candidate gene	Tomato Unigenes acc. no.	*Arabidopsis*						
		Accession no.	Gene function	Chr.	AGI coordinates (bases)	TBLASTX		
						E value	Score	
ORF1	U323899	AF014399	Magnesium-chelatase	1	2696415–2700961	0	214	
ORF2		NM112785	Transcriptional factor B3 family	3	6548875–6551847	3E-21	105	
ORF3	U317249	AF326883	Remorin family protein	2	17477944–17480014	9E-37	110	
ORF4		NM111676	Polygalacturonase	3	21294315–21296918	1E-79	108	
ORF5		AB017502	Glycosyl hydrolase family 3	5	7107378–7111311	0	276	

Alternative intron splicing

Because the sequence mutation in the *ps-2*ABL allele of ORF4 is located in one of the intron recognition splice sites, we hypothesized that this mutation could affect the pre-mRNA splicing of the gene. In order to verify this hypothesis, we designed primers to amplify the nearly full-length cDNA clone of ORF4 (1,032nt out of 1,179nt): the forward primer was designed on the first exon of ORF4 and the reverse primer on the last exon (Fig. S1). RT-PCR was performed on cDNA made from RNA of anthers at post-anthesis from *ps-2*ABL and cv. Moneymaker. The obtained amplified product for cv. Moneymaker was of the expected size (1,032 bp), which was confirmed by sequencing. The amplified product of *ps-2*ABL was significantly smaller

than the Moneymaker product, which suggested an alteration in intron splicing (Fig. 3). The sequencing of the fragment amplified on cDNA of *ps-2*ABL showed that the fifth exon, on which the SNP is present, was absent in the cDNA sequence. This exon was removed together with the two flanking introns during the pre-mRNA maturation process (Fig. 4). The wild type 5′ sequence of this exon–intron splice junction is CAG/GTATCG, which is identical to one of the splice junction sequence identified in *Solanum tuberosum* (Brown 1986). The mutation found in the *ps-2*ABL allele induces the following sequence: CA**C**/GTACG, which is not present in the list of intron splicing recognition sites. The absence of the fifth exon in the mature mRNA represents a deletion of 208 nucleotides, which induces a frame-shift in the remaining coding
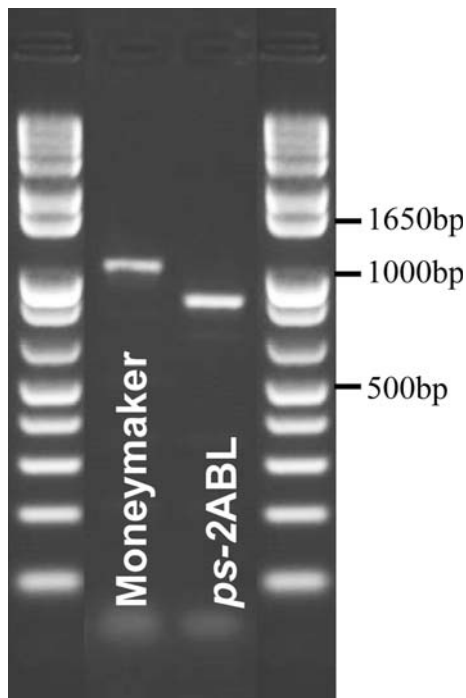
Fig. 3 Size of the nearly complete coding sequence of ORF4 in Moneymaker and *ps-2*ABL. The positions of the primers on the gene sequence are indicated on the Fig. S1

sequence down-stream. This frame-shift causes a premature termination of translation after 14 aa due to a newly framed stop codon. The complete putative mutated protein is therefore 154 aa long in comparison to 392 aa for the wild type protein, and is likely to be non-functional.

ORF4 sequence analysis

A BLAST search with the putative candidate protein sequence of ORF4, in the protein database of NCBI, resulted in a list of PG proteins from several plants. The identified proteins have functions in fruit ripening and

siliques/pods dehiscence. Among them, the ADPG1 protein had been found to be expressed in the dehiscence zone of siliques of *Arabidopsis* as well as the dehiscence zone of anthers (Sander et al. 2001). Amino acids of the PG protein domain (Pfam GH28), of the best BLAST hits were aligned together with the sequences of the candidate protein and the already known tomato PG proteins with identified functions, and one gymnosperm PG (cedar). A phylogenetic analysis was performed on the final alignment in order to place the candidate protein in one of the referenced PG clades. The ORF4 was identified as a PG of clade B (Fig. 5). Clade B is composed of cloned genes that encode PG expressed in fruit and dehiscence zone, as previously characterized by Hadfield and Bennett (1998). This was also observed in the phylogenetic tree. TFPG, the only tomato PG known to be expressed in fruits, was also part of the same clade. Alignment of Pfam GH28 domains of ORF4 and TFPG is presented in Fig. 6. The ORF4 and TFPG have a similarity of 59% in the entire protein sequence.

The putative derived protein of ORF4 contains the four conserved domains characteristic of PG proteins, as presented by Rao et al. (1996; Fig. 6). The first conserved domain is located on the fifth exon, which is absent in the mutant protein and the three others domains are located more C-terminally and therefore not in frame in the mutant sequence. Thus, the mutant protein does not contain any of the four conserved domains that play a major role in the function of the protein.

Expression of ORF4 in anther, fruit and other tissues

We tested the presence of ORF4 transcript in several tissues including abscission zones of leaf and flowers, mature fruits and anthers at anthesis. No ORF4 transcript was detected in the abscission zones (Fig. 7). The presence of the ORF4 transcript was confirmed in anthers as well as in mature fruit. In order to study the transcription level of ORF4 over

Fig. 4 Intron splicing between exon IV and VI of ORF4, in Moneymaker (WT) and in *ps-2*ABL. The mutated nucleotide ($G \rightarrow C$) is *circled* by a *dashed rectangle*
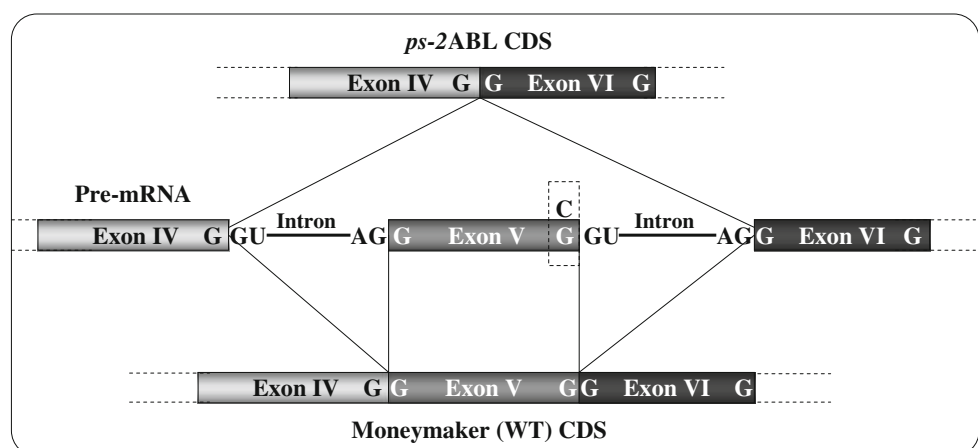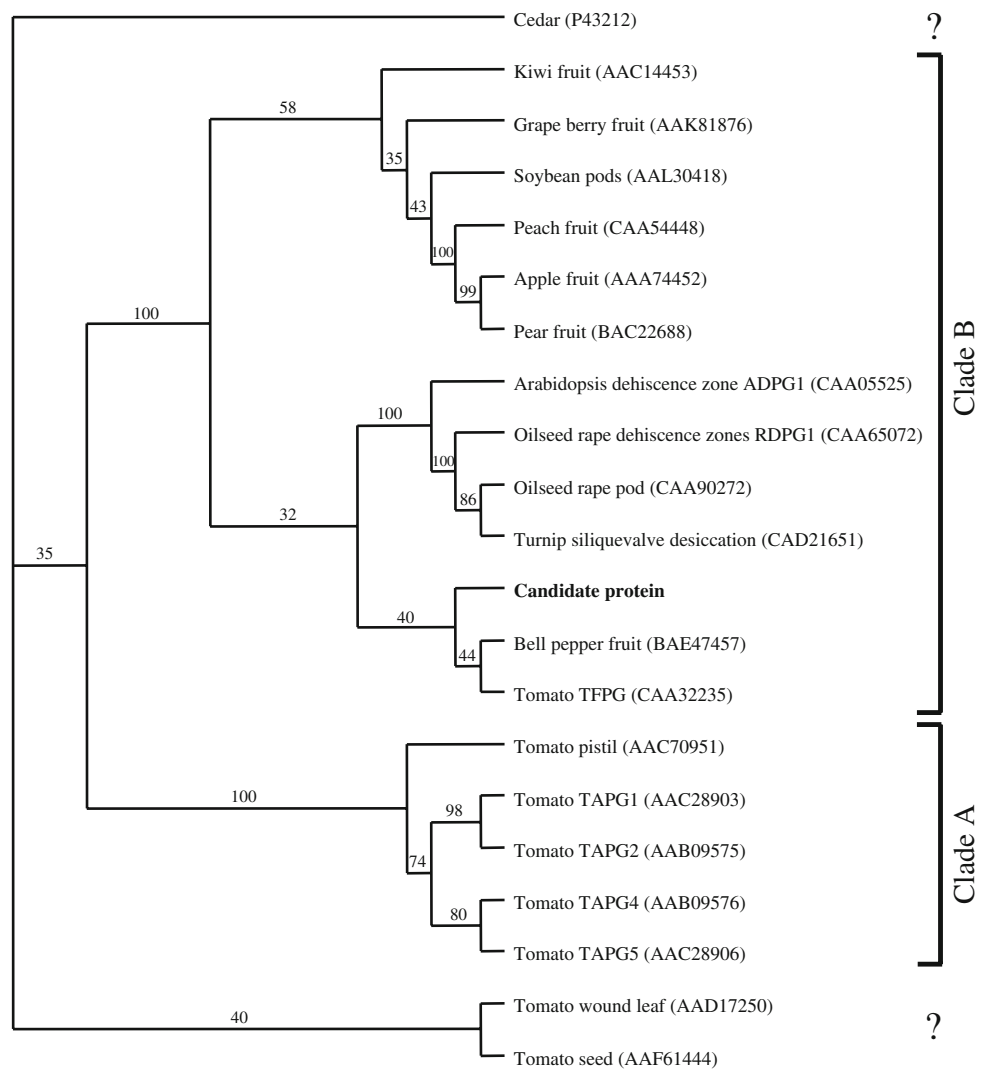
Fig. 5 Phylogenetic tree for 20 angiosperm and one gymnosperm PG aa sequences. The phylogram is the result of a 50% majority-rule consensus of 3,296 trees (using tree weights). The PG sequences segregate into two major clades identified as the well characterized clades A and B. The candidate protein (PS-2) is part of clade B

different stages of anther and fruit development, we performed a quantitative expression analysis of ORF4 at four developmental stages of anthers: flower bud, pre-anthesis, anthesis, post-anthesis, and eight developmental stages of the fruit, from 5 DAP (days after pollination) to 57 DAP (mature fruit); 47 DAP corresponded to breaker stage. In anthers, the transcription level of ORF4 was tested on Moneymaker and *ps-2*ABL, and in the fruits only on Moneymaker. The results are presented in Fig. 8. In anthers of cv. Moneymaker, ORF4 transcripts were already detected at flower bud stage. At pre-anthesis the level of ORF4 transcript was similar to the level in flower buds, and the ORF4 transcript accumulation increased at anthesis and reached a maximum at post-anthesis. In anthers of *ps-2*ABL, ORF4 transcripts were also detected, except at flower bud stage. The transcript level in *ps-2*ABL anthers was lower than in the wild type anthers at anthesis and post-anthesis (Fig. 8). In fruit, no ORF4 transcript was detected before 37 DAP. From 37 DAP, ORF4 transcripts were detected and

increased significantly over time to reach a maximum at mature stage (57 DAP).

Discussion

The *ps-2* ABL produces anthers that do not undergo dehiscence. In this study, we showed that anther dehiscence of *ps-2*ABL is blocked in the ultimate phase.

This phenotypic mutation is recessive and under the control of one single locus. In a previous study we fine-mapped the *ps-2* gene on the short arm of chromosome 4 (Gorguet et al. 2006). Here, we report the isolation and functional characterization of the *ps-2* gene. This is the first functional sterility gene isolated in the Solanaceae family. The high resolution of the linkage map developed on the recombinant population allowed us to narrow down the position of the *ps-2* gene to one single candidate gene on the BAC sequence. The presence of at least two recombinations

Fig. 6 Alignment of Pfam GH28 domain of TFPG, ORF4 and *ps-2* (mutant of ORF4). *Color code* is only effective for the comparison between TFPG and ORF4: In *reverse color* are the identical amino acids. In *gray* are the conserved substitutions. In *bold underlined* are the four conserved domains of polygalacturonase as defined by Rao et al. (1996), identified by *roman numbers*
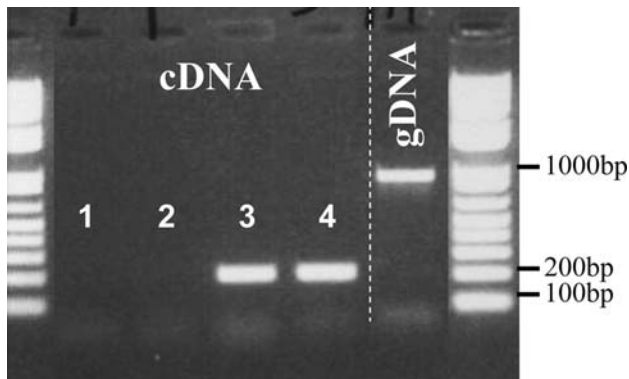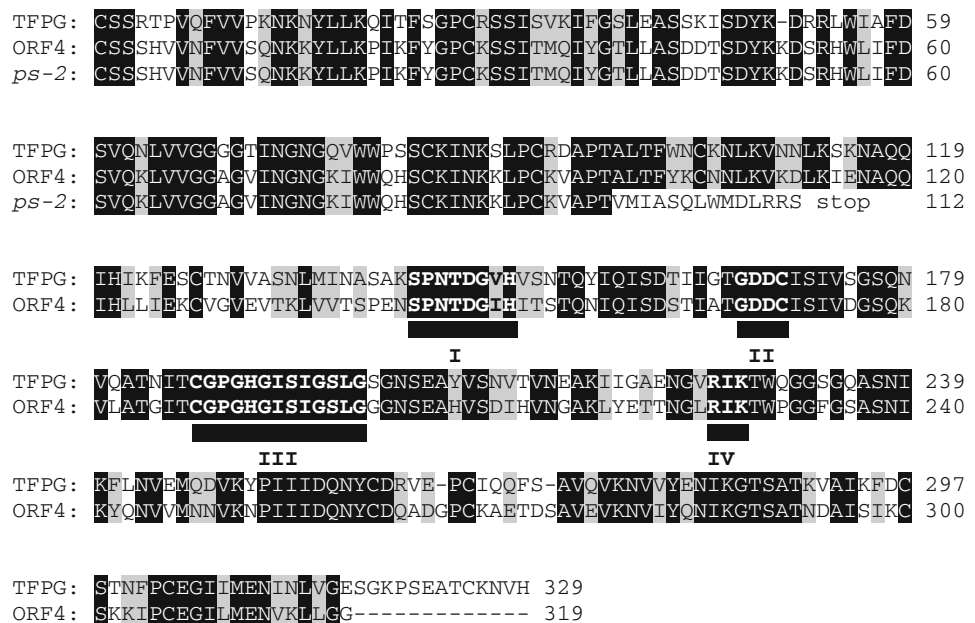
```
TFPG:  CSSRTPVQFVVPKNKNYLLKQITFSGPCRSSISVKIFGSLEASSKISDYK-DRRLWIAFD 59
ORF4:  CSSSHVVNFVVSQNKKYLLKPIKFYGPCKSSITMQIYGTLLASDDTSDYKKDSRHWLIFD 60
ps-2:  CSSSHVVNFVVSQNKKYLLKPIKFYGPCKSSITMQIYGTLLASDDTSDYKKDSRHWLIFD 60

TFPG:  SVQNLVVGGGGTINGNGQVWWPSSCKINKSLPCRDAPTALTFWNCKNLKVNNLKSKNAQQ 119
ORF4:  SVQKLVVGGAGVINGNGKIWWQHSCKINKKLPCKVAPTALTFYKCNNLKVKDLKIENAQQ 120
ps-2:  SVQKLVVGGAGVINGNGKIWWQHSCKINKKLPCKVAPTVMIASQLWMDLRRS stop     112

TFPG:  IHIKFESCTNVVASNLMINASAKSPNTDGVHVSNTQYIQISDTIIGTGDDCISIVSGSQN 179
ORF4:  IHLLIEKCVGVEVTKLVVTSPENSPNTDGIHITSTQNIQISDSTIATGDDCISIVDGSQK 180
                             I                        II
TFPG:  VQATNITCGPGHGISIGSLGSGNSEAYVSNVTVNEAKIIGAENGVRIKTWQGGSGQASNI 239
ORF4:  VLATGITCGPGHGISIGSLGGGNSEAHVSDIHVNGAKLYETTNGLRIKTWPGGFGSASNI 240
                  III                          IV
TFPG:  KFLNVEMQDVKYPIIIDQNYCDRVE-PCIQQFS-AVQVKNVVYENIKGTSATKVAIKFDC 297
ORF4:  KYQNVVMNNVKNPIIIDQNYCDQADGPCKAETDSAVEVKNVIYQNIKGTSATNDAISIKC 300

TFPG:  STNFPCEGIIMENINLVGESGKPSEATCKNVH 329
ORF4:  SKKIPCEGILMENVKLLGG------------- 319
```


Fig. 7 Tissue-specific expression of ORF4. *cDNA* of Moneymaker from diverse tissues was subjected to PCR amplification using primers dedicated to quantitative PCR analysis. *1* Leaf abscission zone, *2* flower abscission zone, *3* anthers at anthesis, *4* fruit at mature stage, *gDNA* genomic DNA control

within ORF4 allowed us to restrict the location of the *ps-2* mutation to the end of the gene sequence. Still, we analyzed the complete gene and putative promoter sequence.

We found that the *ps-2* phenotype is the result of a single nucleotide mutation in a polygalacturonase gene unknown to date, composed of nine exons. This single nucleotide mutation is located on the last nucleotide of the 3′ end of the fifth exon, affecting the intron splicing recognition site, which is changed from CA**G**/GTATCG to CA**C**/GTATCG (exon 3′/intron 5′). Though the cytosine base is present in 11% of the intron splice sites at this specific position in plants, the combination "CAC" at the exon 3′end has never been detected in any splice site in plants (Brown 1986). The fifth exon is spliced out together with the two flanking introns. Analysis of *Arabidopsis* mutants with mutations around splice sites has revealed several examples of exon

skipping in plant splicing (reviewed by Brown and Simpson 1998). Most of these mutations are located in the intron part of the recognition splice sites. To our knowledge, in plants, the only mutant showing exon skipping due to a mutation in the exon part of a recognition splice site, to date, was the *spy-1* mutant in *Arabidopsis* (Jacobsen et al. 1996) in which the CA**G**/GTTTGA (exon 3′/intron 5′) recognition splice site at the end of the eighth exon was mutated into CA**A**/GTTTGA. The exon skipping observed in the mutated allele of ORF4 induces a frame-shift in the rest of the sequence, which has as consequence to create an early stop codon 14 aa further. The complete mutant protein is therefore 154 aa long in comparison to 392 aa long for the wild type, and does not contain any of the four domains characteristic of PGs.

We did not clone the genomic sequence of ORF4 with its respective promoter, in order to perform a complementation experiment with the *ps-2*ABL. A complementation test would be the ultimate proof to determine whether the mutation identified in ORF4 is causing the non-dehiscent phenotype of *ps-2*ABL. However, based on the accuracy of the high-resolution map, the specific expression pattern of ORF4 in anthers and the non-functionality of ORF4 in *ps-2* ABL we believe to have proved that the mutation found in ORF4 is responsible for the functional sterility observed in *ps-2*ABL.

ps-2 is a PG of clade B

The isolated gene responsible for the *ps-2* phenotype is a *PG* unknown to date. The phylogenetic analysis of PS-2 amino acids sequence revealed a close similarity with PGs of clade B as defined by Hadfield and Bennett (1998).
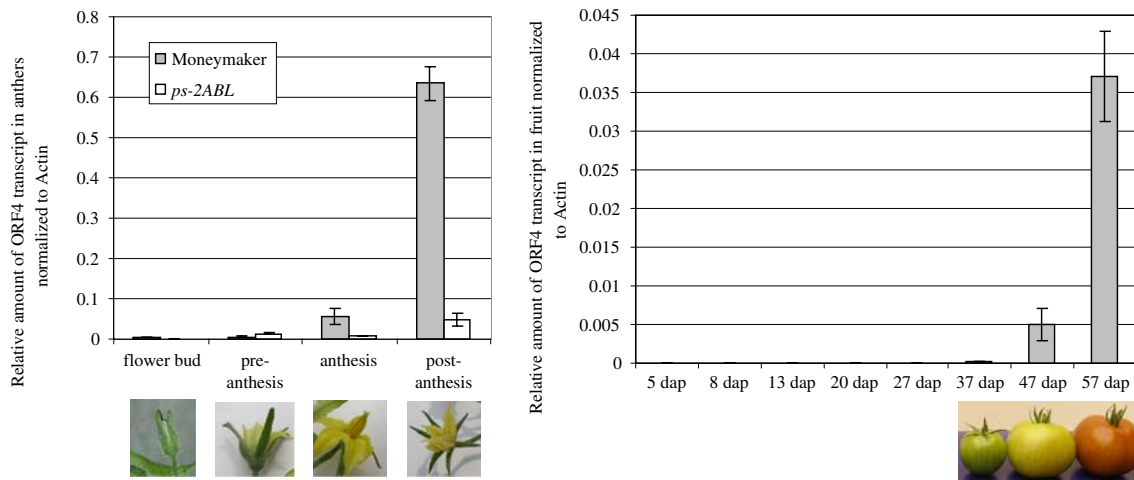
Fig. 8 Real-time PCR expression data of ORF4 at different stages of anthers development on Moneymaker and *ps-2*ABL (*left*) and during fruit development and ripening of Moneymaker (*right*). *DAP* days after pollination

Clade B is here composed of fruit PGs, among them the tomato fruit PG, as well as silique or pod dehiscence PGs. *ADPG*, the *Arabidopsis* homolog of *PS-2* is also expressed in the anther dehiscence zone (Sander et al. 2001). This suggests a conserved role of *PS-2* homologs over the plant kingdom. We suggest that the repression of *PS-2* homologs in other plant species may procure functional sterility in species where such a trait does not exist.

The other tomato PGs cluster in different clades. In accordance with the assertion that the divergence of PG family members occurred prior to the separation of the angiosperm species (Hadfield and Bennett 1998), PS-2 is here more closely related to proteins of the same clade, from other species, than to tomato PGs from other clades. Most of the other tomato PGs is related to abscission (TAPG). In our study, expression of *PS-2* was not detected in flower and leaf abscission zones. However, in addition to anther tissues, we detected mRNA transcript of *PS-2* in fruits.

PS-2 transcript accumulation increases along with the development of the anthers

We measured the relative level of *PS-2* transcript at different stages in anthers. The *PS-2* transcript is already detected in the anthers of Moneymaker at the flower bud stage. The transcript level increases over stages to reach a maximum at post-anthesis, when anthers are dehisced. This increase of *PS-2* transcript accumulation is parallel to the septum and stomium degeneration observed in the anthers. The *ps-2* transcript level in *ps-2*ABL anthers was detected from pre-anthesis on, but the transcript level remained very low in comparison to Moneymaker, at anthesis and post-anthesis. Very likely the mutant mRNA is recognized as non-sense and degraded by non-sense-mediated mRNA

decay (NMD). The NMD functions as a quality control mechanism to eliminate abnormal transcripts (Lejeune and Maquat 2005). Although unlikely, our findings do not exclude the possibility that the mutant phenotype is caused by the truncated protein derived from the mutant mRNA.

PS-2 may also play a role in tomato fruit ripening

TFPG, the only tomato fruit polygalacturonase identified to date, has been characterized as one of the main actors in the process of fruit softening. Anti-sense repression of *TFPG* has lead to the production of tomato fruit with longer shelf life but the fruits did undergo ripening indicating that other actors also play a relevant role in the process of fruit softening (Smith et al. 1988). The PS-2 may well be one of these actors by contributing to the fruit cell wall degradation. The *PS-2* transcript was detected in the late stages of fruit development in Moneymaker (Fig. 8). The maximum transcripts were found at mature stage (57 DAP). Similarly, *TFPG* has been detected only at ripening stages, starting at mature green or at breaker stage (Thompson et al. 1999; Eriksson et al. 2004). The level of *TFPG* was also found to increase over time from breaker stage to mature fruit, in cv. AC and Liberto (Thompson et al. 1999).

The expression of *Ps-2* in fruits does not prove that this gene has an effect in fruit maturation. The comparison of fruit maturation and shelf life, between *ps-2*ABL after hand pollination and Moneymaker or any other normal tomato line is unreliable due to the difference of genetic background. Knock-out of *PS-2* by anti-sense RNA or RNAi in tomato, or double mutants in which both *PS-2* and *TFPG* are impaired, could provide new insights to answer that question. The time between hand pollination to mature fruit stage and fruit shelf life could be measured and compared to the untransformed control in order to evaluate the effect

of PS-2 in the fruit. Knock-outs of homologs of *PS-2* in other plants species are also of valuable interest to verify whether the control of anther dehiscence in its ultimate phase is conserved among species.

References

Brown JWS (1986) A catalogue of splice junction and putative branch point sequences from plant introns. Nucleic Acids Res 14:9549–9559

Brown JWS, Simpson CG (1998) Splice site selection in plant pre-mRNA splicing. Annu Rev Plant Physiol Plant Mol Biol 49:77–95

Brugmans B, Hutten RGB, Rookmaker ANO, Visser RGF, Van Eck HJ (2006) Exploitation of a marker dense linkage map of potato for positional cloning of a wart disease resistance gene. Theor Appl Genet 112:269–277

Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. Genome Res 10:129–136

Dornelas M, van Lammeren A, Kreis M (2000) *Arabidopsis thaliana* SHAGGY-related protein kinases (AtSK11 and 12) function in perianth and gynoecium development. Plant J 21:419–429

Dubald M, Barakate A, Mandaron P, Mache R (1993) The ubiquitous presence of exopolygalacturonase in maize suggests a fundamental cellular function for this enzyme. Plant J 4:781–791

Eriksson EM, Bovy A, Manning K, Harrison L, Andrews J, De Silva J, Tucker GA, Seymour GB (2004) Effect of the colorless non-ripening mutation on cell wall biochemistry and gene expression during tomato fruit development and ripening. Plant Physiol 136:4184–4197

Gorguet B, Schipper D, van Heusden AW, Lindhout P (2006) High resolution fine mapping of *ps-2*, a mutated gene conferring functional male sterility in tomato due to non-dehiscent anthers. Theor Appl Genet 113:1437–1448

Hadfield KA, Bennett AB (1998) Polygalacturonases: many genes in search of a function. Plant Physiol 117:337–343

Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, Mueller LA, Bhattacharyya D, Bhaya D, Sobral BW, Beavis W, Meinke DW, Town CD, Somerville C, Rhee SY (2001) The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res 29:102–105

Jacobsen SE, Binkowski KA, Olszewski NE (1996) SPINDLY, a tetratricopeptide repeat protein involved in gibberellin signal transduction in *Arabidopsis*. Proc Natl Acad Sci USA 93:9292–9296

Keijzer CJ (1987) The processes of anther dehiscence and pollen dispersal. I. The opening mechanism of longitudinally dehiscing anthers. New Phytol 105:487–498

Lejeune F, Maquat LE (2005) Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. Curr Opin Cell Biol 17:309–315

Markovic O, Janecek S (2001) Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. Protein Eng 14:615–631

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL genomics network. A comparative resource for Solanaceae biology and beyond. Plant Physiol 138:1310–1317

Rao MN, Kembhavi AA, Pant A (1996) Implication of tryptophan and histidine in the active site of endo-polygalacturonase from *Aspergillus ustus*: elucidation of the reaction mechanism. Biochim Biophys Acta 1296:167–173

Rieu I, Wolters-Arts M, Derksen J, Mariani C, Weterings K (2003) Ethylene regulates the timing of anther dehiscence in tobacco. Planta 217:131–137

Salamov AA, Solovyev VV (2000) ab initio gene finding in *Drosophila* genomic DNA. Genome Res 10:516–522

Sander L, Child R, Ulvskov P, Albrechtsen M, Borkhardt B (2001) Analysis of a dehiscence zone endo-polygalacturonase in oilseed rape (*Brassica napus*) and *Arabidopsis thaliana*): evidence for roles in cell separation in dehiscence and abscission zones, and in stylar tissues during pollen tube growth. Plant Mol Biol 46:469–479

Scott RJ, Spielman M, Dickinson HG (2004) Stamen structure and function. Plant Cell 16:S40–S60

Smith CJS, Watson CF, Ray J, Bird CR, Morris PC, Schuch W, Grierson D (1988) Anti-sense RNA inhibition of polygalacturonase gene expression in transgenic tomatoes. Nature 334:724–726

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland

Thompson AJ, Tor M, Barry CS, Vrebalov J, Orfila C, Jarvis MC, Giovannoni JJ, Grierson D, Seymour GB (1999) Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant. Plant Physiol 120:383–390

Torki M, Mandaron P, Mache R, Falconet D (2000) Characterization of a ubiquitous expressed gene family encoding polygalacturonase in *Arabidopsis thaliana*. Gene 242:427–436