Plant Molecular Biology **39:** 749–760, 1999.

Characterization of the subtilase gene family in tomato (*Lycopersicon esculentum* Mill.)

Jurka Meichtry, Nikolaus Amrhein and Andreas Schaller*
*Institute of Plant Sciences, ETH-Zürich, LFW ES1, Universitätstrasse 2, 8092 Zürich, Switzerland (*author for correspondence)*

Received 15 July 1998; accepted in revised form 2 November 1998

Key words: absence of introns, gene family, kexin, plant protease, subtilase, tomato

Abstract

The gene family of subtilisin-like serine proteases (subtilases, SBTs) in tomato (*Lycopersicon esculentum* Mill.) comprises at least 15 members, 12 of which have been characterized in this study. Sequence comparison revealed that tomato subtilases fall into 5 distinct subfamilies. Single genes were shown to exist for *LeSBT1*, *LeSBT2* and *tmp*, while 5 and 6 genes were found in the *LeSBT3/4* and *P69* subfamilies, respectively. With the exception of *tmp*, tomato subtilase genes were found to lack introns. Expression of subtilase genes was confirmed at the mRNA level by northern blot analysis and/or by primer extension experiments. For each of the 5 subtilase subfamilies, a distinctive pattern of expression was observed in tomato organs. At least one of the subtilases was found to be expressed in each organ analysed. Structural features evident from deduced amino acid sequences are discussed with reference to the related mammalian proprotein convertases.

Introduction

Serine peptidases, i.e. proteolytic enzymes that depend on a serine residue for catalytic activity, fall into more than 20 families that can be grouped into six clans [1]. Three of them, the chymotrypsin, subtilisin and carboxypeptidase C clans, share a common reaction mechanism based on the 'catalytic triad' comprising a serine, an aspartic acid, and a histidine residue. The protein folds of these enzymes, however, are quite different and constitute a striking example of convergent evolution. The subtilisin clan is the second largest and to date includes more than 170 proteases (subtilases). The number of known subtilases is steadily growing; more than 100 have been discovered only since 1991 [1, 2]. Interest in this group of proteases has been fuelled by the observation that in animals, subtilases are involved in the maturation of peptide hormones, neuropeptides, growth factors, and receptor proteins, and likewise in the processing of viral envelope proteins and bacterial toxins [3–5]. Mammalian subtilases, just like the yeast kex2 protease, characteristically cleave carboxy-terminal of paired basic residues and they are collectively called proprotein convertases (PCs, [6]). To date, seven human PCs are known many of which process identical substrates *in vitro* at dibasic sequence motives. *In vivo*, however, they exhibit an exquisite specificity since their expression is tightly controlled both spatially and developmentally [4]. In plants, there has initially been circumstantial evidence for the existence of PC-like proteases. In tomato plants, a protein has been identified that specifically interacts with the peptide wound signal systemin. This protein was found to crossreact with an antiserum against a *Drosophila* PC [7]. Furthermore, tobacco plants overexpressing a virally encoded antifungal toxin have been generated and in these plants the toxin was found to be secreted in the correctly processed form. This finding points at the

The nucleotide sequence data reported will appear in the EMBL and GenBank databases under the accession numbers AJ006378 (*LeSBT1*), AJ006379 (*LeSBT2*), AJ006380 (*LeSBT3*), AJOO6377 (*LeSBT4A*), AJ006480 (*LeSBT4B*), AJ006481 (*LeSBT4C*), AJ006482 (*LeSBT4D*), AJ006483 (*LeSBT4E*), AJ005171 (*P69C*), AJ006786 (*P69D*), *AJ005172* (*P69E*) and AJ005173 (*P69F*).

existence of a protease with PC-like specificity in the secretory system of tobacco [8]. The first subtilase to be cloned from a higher plant was cucumisin, an extracellular protease highly abundant in melon fruit [9]. Further subtilase cDNAs have been cloned from *Alnus glutinosa*, *Arabidopsis thaliana* and *Lilium multiflorum* [10, 11]. Three subtilases have been described in tomato two of which (P69A and P69B) seem to be involved in pathogenesis [12, 13]. The third tomato subtilase (tmp) appears to function in microspore development [14]. The only plant subtilase that has been characterized biochemically to some extent is cucumisin. This protease was found to have a broad substrate specificity and not to exhibit the preference for sequence motives of paired basic residues typically encountered in animal PCs [15, 16]. The substrates of plant subtilases and their processing sites have not yet been identified and the function of these enzymes *in planta* is not clear. Therefore, at this time, it is unclear whether or not plant subtilases are proprotein convertases like their animal counterparts.

We wish to learn more about the function of subtilisin-like proteases in higher plants. This study, in which we characterize the gene family encoding these proteases in tomato, is a first step towards this aim.

Material and methods

Probes used in library screening

For the isolation of *LeSBT3/4* clones a partial cDNA was used as a probe that had been isolated serendipitously from a tomato leaf cDNA and given to us by Dr D. Bergey (Montana State University). All other probes were generated by PCR (Cetus; Perkin Elmer, Foster City, CA). All PCR reactions were performed with tomato (*L. esculentum* Mill. cv. Castlemart II) genomic DNA as the template. The oligonucleotide primers comprised *Eco*RI restriction sites to facilitate cloning of the reaction products into the Bluescript pSK(−) vector (Stratagene; La Jolla, CA). Primer design was based on a sequence alignment between animal proprotein convertases and cucumisin from *Cucumis melo* [9]. The conserved regions surrounding the catalytically important His and Asn residues were chosen for the design of degenerate oligonucleotide primers which were supplied by Microsynth (Balgach, Switzerland; 5′ primer CGGAATTCTCAAYGGNCAYGGNACNC; 3′ primer CGGAATTCTTNSGNCCNCCRTTNC, where mixed bases are represented by R=A/G, S=C/G, Y=C/T, and N=G/A/T/C). The hot-start PCR (Taq-beads; Promega, Madison, WI) was performed for 12 cycles at an initial annealing temperature of 58 °C which was then reduced by 0.5 °C in each cycle, followed by 24 cycles at an annealing temperature of 52 °C. Reaction products were cloned into the *Eco*RI site of pSK(−). Products exhibiting sequence similarity with subtilases were used for library screening resulting in the isolation of *LeSBT1* and *LeSBT2*.

The probe specific for the *P69* subfamily was generated by PCR using oligonucleotide primers based on the published P69A cDNA sequence [12]. Hot-start PCR was performed with tomato genomic DNA (150 ng) as template for 30 cycles at an annealing temperature of 55 °C in order to amplify a 543 bp fragment corresponding to nucleotides 89–631 of the P69A cDNA (5′ primer CGGAATTCTTTTGCTCTTTCCCATGGCC; 3′ primer CGGAATTCAGTTGGTAAGACCTGGCTCC). Reaction products were cloned into the *Eco*RI site of pSK(−) and their identity was confirmed by sequence analysis. Similarly, part of the *tmp* gene [14] corresponding to the third exon was amplified and cloned (5′ primer CGGAATTCCCTTGATGTTCTGCGGAAT; 3′ primer CGGAATTCCATGTCCATCAGCATCA).

Screening of genomic DNA and cDNA libraries

The procedure for the screening of lambda phage libraries, like all other molecular biological techniques, was adopted from [17, 18]. The genomic DNA library (Clontech, Palo Alto, CA) contained within the *Bam*HI cloning site of the EMBL3 vector partially *Mbo*I-restricted (12–23 kb) fragments of tomato (*L. esculentum* cv. VFW8) genomic DNA. For each probe, 1.2×10^6 phage were screened on nitrocellulose filters. Hybridization with radiolabelled DNA probes ('Prime-It' system; Stratagene) was performed at 42 °C in 50% formamide, 5× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate), 0.5% SDS, 2× Denhardt's solution (1× Denhardt's solution is 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 50 mM potassium phosphate buffer (pH 7.0), and 200 μg/ml of denatured salmon sperm DNA. Filters were washed in 0.2× SSC, 0.5% SDS at 60 °C and were subsequently exposed to autoradiographic film (Kodak X-Omat AR) using an intensifying screen. Two consecutive rounds of screening were

performed to isolate individual positive lambda phage clones.

The λZAP (Stratagene) cDNA library which was used for the isolation of *Le*SBT4A had been constructed from poly(A)$^+$-RNA isolated from leaves of transgenic tomato plants overexpressing the prosystemin gene [19]. The screening of 2.5×10^5 phage was performed as described for the screening of the genomic DNA library.

Sequence analysis

Lambda phage DNA was isolated using a commercial system (Nucleobond AX, Macherey-Nagel, Oensingen, Switzerland). Sequence analysis was either performed on overlapping fragments of isolated lambda phage DNA after subcloning into Bluescript pSK(−), or else, sequence data were obtained directly from purified lambda DNA by 'primer walking'. For the *P69* subfamily, the latter strategy was employed. For the *LeSBT3/4* subfamily, a combination of both approaches was used while the first strategy was used for the sequencing of the *LeSBT1* and *LeSBT2* genes. All sequences were determined both strands using fluorescent dideoxy chain terminators in the cycle sequencing reaction (Perkin Elmer) and the Applied Biosystems Model 373A DNA sequencer. Sequence data were compiled and analysed with the University of Wisconsin GCG sequence analysis software package [20].

Southern and northern blot analyses

Genomic DNA was isolated from tomato leaves [21], digested with restriction enzymes as indicated, and fragments were separated on 0.8% agarose gels. Processing of the gels and transfer onto nitrocellulose membranes were done according to Sambrook *et al.* [17]. Prehybridization and hybridization conditions were as described above for library screening. Filters were analysed with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). RNA was isolated from tomato leaves ground in liquid nitrogen using a phenol-based extraction procedure. In each experiment, five μg of total RNA were separated on formaldehyde agarose gels, transferred onto nitrocellulose membranes and hybridized to radiolabelled DNA probes using standard laboratory procedures [17]. The specific activities of the probes used in RNA gel blot analyses were estimated to be within a factor of two. Prehybridization, hybridization and evaluation of the gel blots were performed as described above.

Primer extension experiments

Primer extension experiments followed the procedure described by Ausubel *et al.* [18]. Briefly, oligonucleotide primers were synthesized complementary to the nucleotide sequences at the translational start sites of LeSBT4A, 4B, and 4C (4A, GGGAACCAAATGTTTCTTGATCCCA; 4B, CCATAGCACTACACAAAACACTTCC; 4C, TCACTACACACCACTCAAATCTTCA) and were end-labelled using T4 polynucleotide kinase (Promega). Labelled primers were purified by anion exchange chromatography and after ethanol precipitation, they were hybridized to 5 μg of total RNA for 90 min at 65 °C. Primer extension was performed with M-ML V reverse transcriptase (Promega) at 42 °C for 60 min. Extension products were separated on 9% polyacrylamide gels in presence of 7 M urea. After fixation in 10% acetic acid/10% methanol, dried gels were analysed with a PhosphorImager (Molecular Dynamics). For size standards, sequencing reactions were performed using the 'fmol sequencing system' according to the manufacturer's instructions (Promega).

Results

Generation of probes and library screening

Presently, two subtilases (SBTs) are known in tomato, designated P69A and P69B. P69A was first identified as a 69 kDa pathogenesis-related protein in tomato plants [22]. The P69A cDNA was cloned [12] and this subsequently led to the isolation of the related P69B cDNA by Vera and co-workers [13]. According to these authors, P69A and P69B are members of a small gene family in tomato [13]. In addition, the genomic sequence coding for a tomato subtilase can be found in the EMBL nucleotide sequence database (*tmp*, accession number U80583). This sequence is closely related to that of a subtilase in lily which appears to be involved in microspore development [11, 14]. Furthermore, indirect evidence has been obtained that a subtilase may be involved in the regulation of the wound response in tomato [7]. Considering these results and the importance of subtilisin-like proteases in animal systems, we decided to analyse the genes for these proteases in tomato. For the screening of genomic and cDNA libraries molecular probes had to be generated. Based on the published sequences of *tmp* and P69A, oligonucleotide primers were synthesized to amplify by PCR a fragment of tmp corresponding to

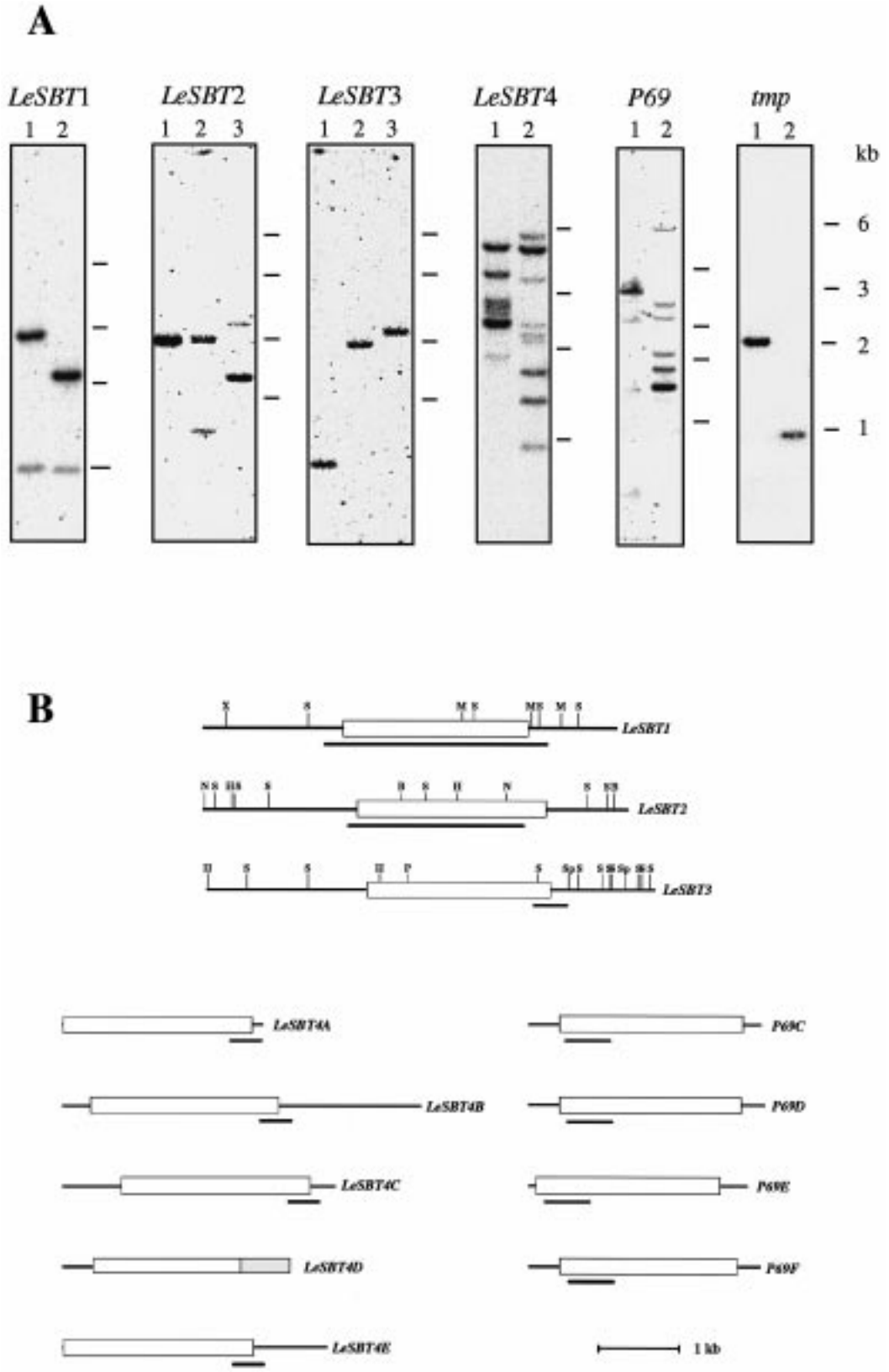
A
LeSBT1
LeSBT2
LeSBT3
LeSBT4
P69
tmp
1 2
1 2 3
1 2 3
1 2
1 2
1 2
kb
6
3
2
1
B
LeSBT1
LeSBT2
LeSBT3
LeSBT4A
LeSBT4B
LeSBT4C
LeSBT4D
LeSBT4E
P69C
P69D
P69E
P69F
1 kb

Figure 1. A. Southern blot analysis of tomato genomic DNA. Fragments from 5 μg of restricted DNA were separated on agarose gels, transferred to nitrocellulose membranes and hybridized to labelled DNA probes. Restriction enzymes were as follows: *LeSBT1*, *Mun*I/*Xba*I (1), *Ssp*I (2); *LeSBT2*, *Spe*I (1), *Bst*XI/*Hin*dIII (2), *Bst*XI/*Nde*I (3); *LeSBT3*, *Ssp*I/*Nde*I (1), *Spe*I/*Pae*I (2), *Spe*I/*Hin*dIII (3); *LeSBT4*, *Ssp*I (1), *Mun*I/*Xba*I (2); *P69*, *Pst*I (1), *Eco*RI (2); *tmp*, *Ssp*I (1), *Mun*I/*Xba*I (2). The positions of size standards (kb) are indicated. B. Schematic representation of the sequenced regions of tomato genomic DNA. The sequenced parts of tomato subtilases are shown. Open boxes indicate the open reading frames of the respective genes. The DNA fragments that were used as labelled probes in Southern analyses (A) are represented by solid lines underneath. The shaded box adjacent to the *LeSBT4D* open reading frame indicates that part of the sequence which resembles part of a gypsy-like retrotransposon (see text). For *LeSBT1–3* restriction sites are indicated for those enzymes that were used in Southern analyses and that were relevant for the size prediction of labelled fragments on DNA gel blots (A; for details, see text).

the third exon, and a 543 bp fragment corresponding to nucleotides 89–631 of the P69A cDNA. Probing of a genomic DNA gel blot with the labeled P69A DNA fragment confirmed the existence of numerous closely related genes in tomato (Figure 1A). The DNA fragment was then used to screen a genomic DNA library of tomato (*L. esculentum* Mill. cv. VFW8) in a lambda phage vector (EMBL3, 1.2×10^6 pfu) yielding 59 positive clones. Nineteen clones were further characterized by sequence analysis and the genomic sequences of 4 new subtilases could be identified which were designated P69C–P69F. The sequence of P69D, however, was found to be extremely similar to that of the P69A cDNA (cf. Figure 3 and Table 1), and these two sequences may in fact be derived from the same gene. The few differences may be due to small variations between the two tomato cultivars (VFW8 and Rutgers) and/or sequencing errors. However, fully identical P69D sequences were determined from three independent lambda phage clones. Thus, sequencing errors in P69D are unlikely to be the cause of the few differences between P69D and P69A. Southern blot analysis of tomato genomic DNA with the *tmp* fragment as a probe revealed that there is only one gene for this protease present in the haploid genome of tomato (Figure 1A). Therefore, we did not attempt to isolate further *tmp*-related clones. For the identification of additional tomato subtilases, degenerate oligonucleotide primers were synthesized on the basis of a sequence alignment between mammalian prohormone convertases and cucumisin, the latter being the only sequence available for plant subtilases at that time [9]. Two highly conserved regions surrounding the catalytically important histidine and asparagine residues (His-204 and Asn-307 of cucumisin) were selected for primer design. Genomic DNA isolated from tomato leaves was used as template in the PCR. Sequence analysis of cloned PCR products revealed the presence of two distinct sequences (*sbt1* and *sbt2*), both of which were related to cucumisin and bacterial subtilisins. Screening of the genomic lambda phage library yielded 14 (*sbt1*) and 13 (*sbt2*) positive clones, respectively. The DNA inserts from individual lambda clones were subcloned into the Bluescript vector and both strands of overlapping fragments were sequenced. The two groups of clones were each derived from single-gene loci designated *LeSBT1* and *LeSBT2*, respectively, which was confirmed by genomic DNA gel blot analysis (Figure 1A). A partial cDNA which appeared to encode another tomato subtilase was isolated serendipitously by D. Bergey (Montana State University) and given to us as a gift. Screening of a tomato leaf cDNA library using the partial cDNA as a probe resulted in the isolation of two full-length cDNAs designated *Le*SBT3 and *Le*SBT4A which had the capacity to encode subtilisin-like proteases. Southern blot analysis using the *Le*SBT4A cDNA as a probe indicated the presence of several closely related genes in the tomato genome (Figure 1A). By screening the tomato genomic DNA library, 33 positive lambda phage clones were isolated. These clones represented at least 5 different genetic loci as revealed by restriction and sequence analysis. The 5 genes were sequenced completely on the sense and antisense strands and were named *LeSBT3* and *LeSBT4B*, *4C*, *4D*, and *4E*. While 4 of these genes had the capacity to encode subtilisin-like proteases, *LeSBT4D* appeared to be non-functional, as the open reading frame was found to be interrupted by a sequence resembling those of gypsy-like retrotransposons [23].

The newly identified subtilase genes do not contain introns

Screening of tomato cDNA and genomic libraries led to the identification of 12 new subtilases designated *Le*SBT1-3, *Le*SBT4A-4E, and P69C-F (Figure 1B). Sequence analysis revealed for each gene, with the exception of *LeSBT4D*, an open reading frame with the capacity to code for a subtilisin-like protease. The relative positions of the open reading frames within the sequenced DNA regions are shown in figure 1B. With the notable exception of *LeSBT4D* (see above), the open reading frames were not interrupted. Therefore,

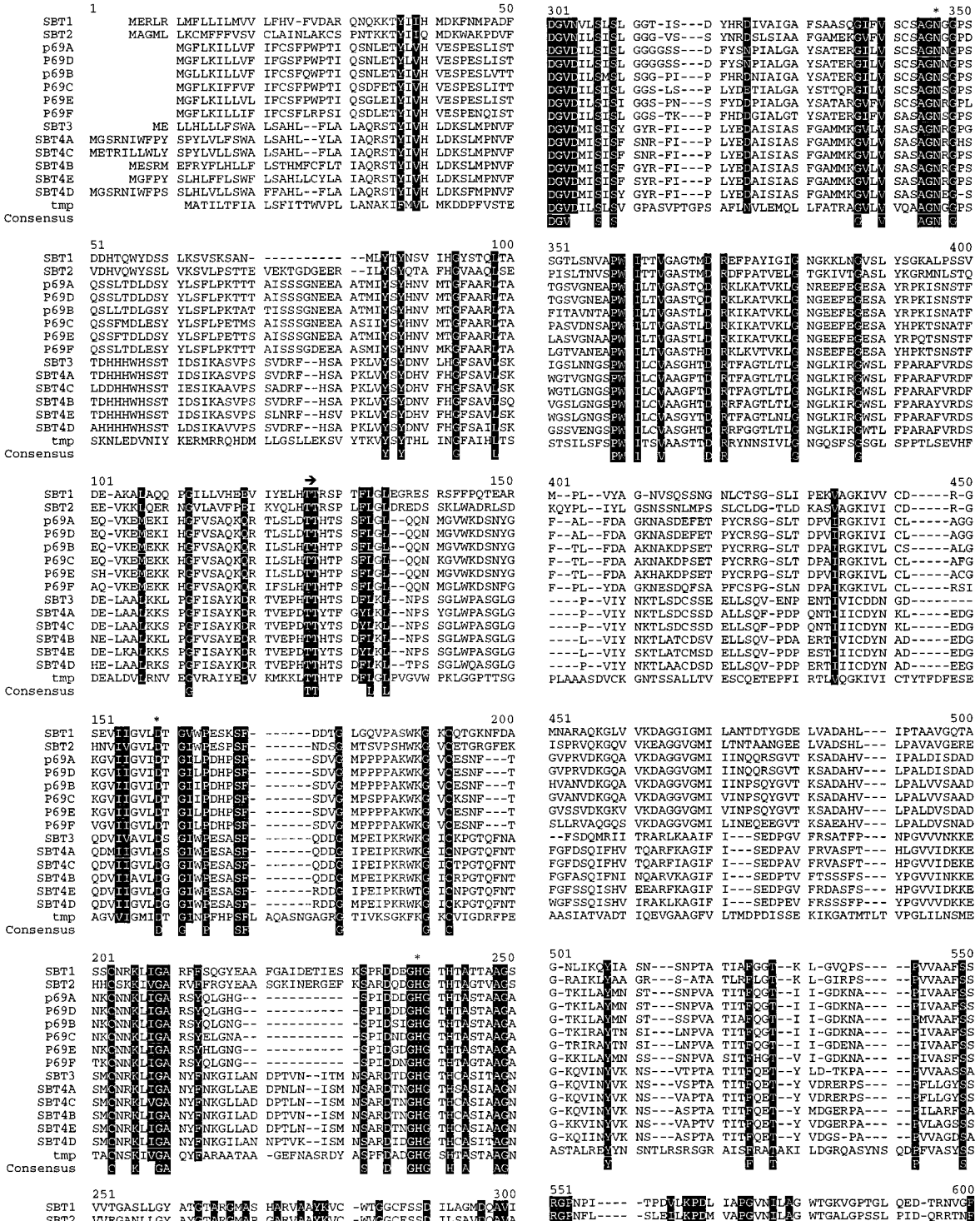

Figure 2.

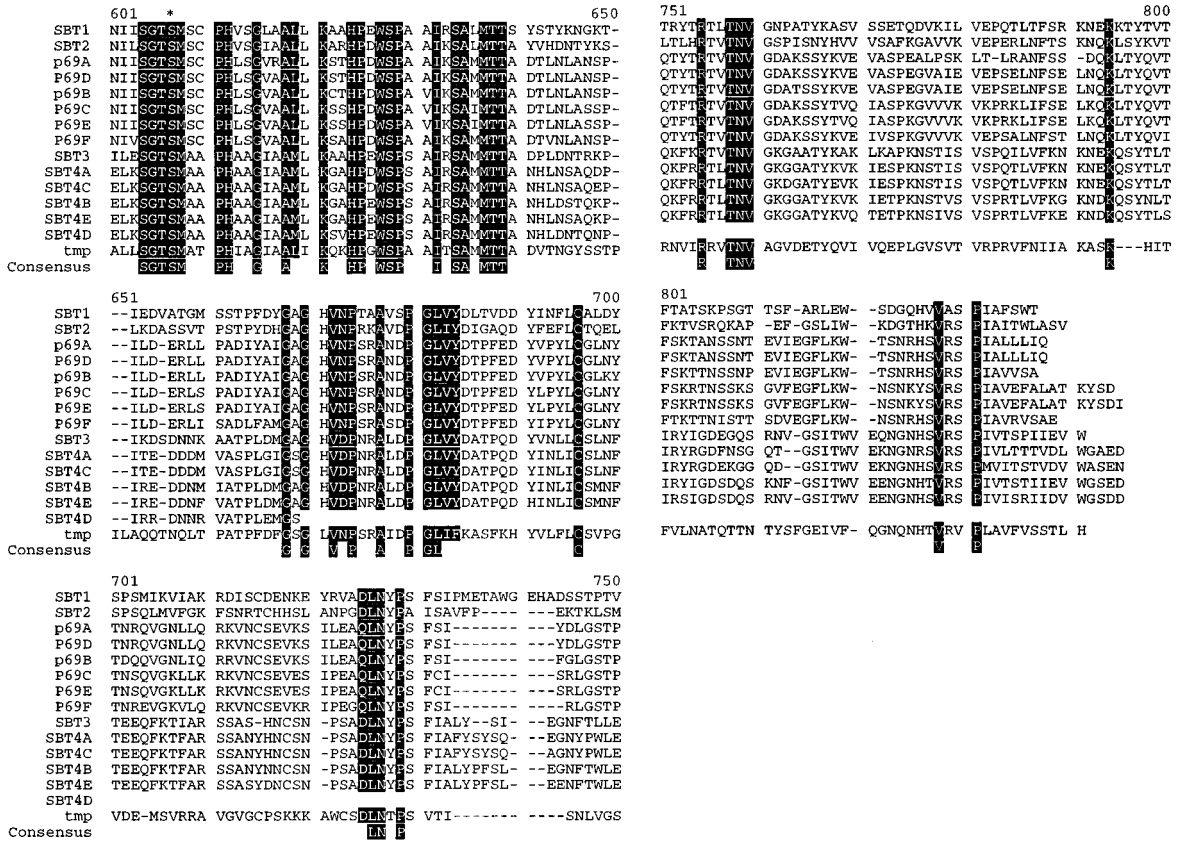

Figure 2. Sequence alignment of tomato subtilases. An alignment of amino acid sequences deduced from tomato subtilase genomic and cDNA sequences is shown. The alignment was created with the program MultAlin [36] on the ExPaSy server (http://expasy.hcuge.ch). The arrow indicates the putative N-termini of the mature enzymes; asterisks indicate active site residues. Conservative replacements and residues identical in at least 90% of the sequences are highlighted. The consensus line shows only those residues that are identical in more than 90% of the sequences.

these genes do not appear to contain any introns. For those genes, where gene-specific probes were available (*LeSBT1–3*), the lack of introns was confirmed by Southern blot analysis. Based on the sequence data obtained, the restriction enzymes for Southern analysis were chosen in a way that allowed to predict the pattern of bands on the DNA gel blot. The restriction enzymes employed, as well as the relative positions of the DNA fragments that were used as probes are indicated in Figure 1B. The corresponding DNA gel blots are shown in Figure 1A. The sizes of the DNA fragments generated from *LeSBT1–3* which hybridized to the labeled probes were consistent with the prediction based on the sequence data (Figure 1B) and thus confirmed the absence of introns from these genes. Therefore, the *tmp* gene appears to be the only subtilase gene in tomato that contains introns.

Primary structure of tomato subtilases

The amino acid sequences deduced from the open reading frames of *LeSBT1–3*, *LeSBT4A–4E* and *P69C–F* are shown in Figure 2 and are compared to those of *P69A*, *P69B*, and *tmp* which had been published previously [12–14]. All tomato subtilases appear to be coded for as preproproteins. At the N-terminus, a stretch of hydrophobic amino acids is found which is typically encountered in signal peptides responsible for targeting the protein to the secretory system. In subtilases, the signal peptide is characteristically followed by a propeptide which may aid in folding of the protein and/or act as an intramolecular inhibitor of enzymatic activity [24, 25]. For P69A, the N-terminus of the mature protease has been determined and thus, the site of propeptide processing could be deduced [12]. Processing occurred amino-terminal of a pair of threonine residues (Thr-115 and

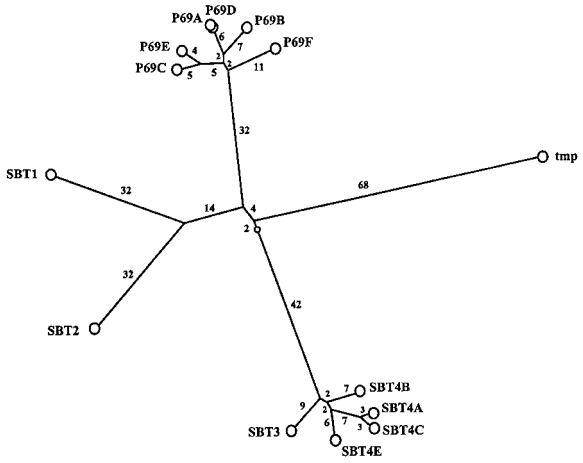

Figure 3. Phylogenetic relationship of tomato subtilases. The unrooted phylogenetic tree was created from the amino acid sequences deduced from tomato subtilase genomic DNA and cDNA sequences using the computational biochemistry research group (CBRG) server at the Swiss Federal Institute of Technology Zurich (http://cbrg.inf.ethz.ch). Numbers indicate the PAM distances (accepted point mutations per 100 residues) between sequences. A 1% mutation matrix was used [37]. The sequence of *Le*SBT4D, which is truncated due to the insertion of a retrotransposon-like sequence, was not included in the analysis.

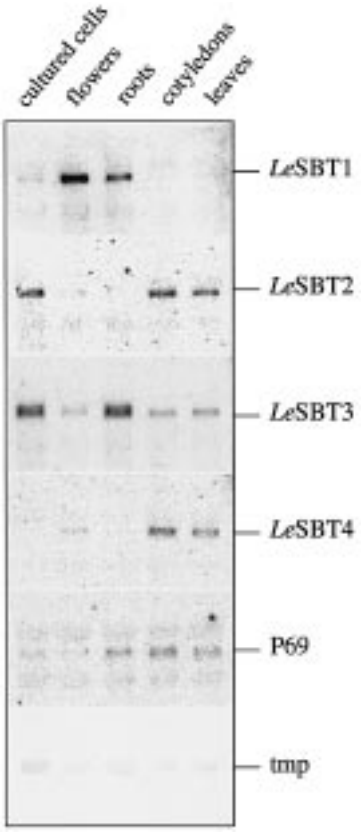

Figure 4. Expression of tomato subtilases in tomato organs. Total RNA isolated from different organs of tomato plants (5 μg) was subjected to electrophoresis through formaldehyde/agarose gels and analyzed on gel blots using radiolabeled DNA probes. The probes used for *LeSBT1–3* and *tmp* were specific for the mRNAs of the corresponding genes, while the probes used for the *LeSBT4* and *P69* subfamilies did not discriminate between individual members in each subfamily.

Thr-116 in P69A) which was also observed in cucumisin [9]. This pair of threonine residues is conserved among all tomato subtilases (Figure 2) and therefore, the site of propeptide processing is likely to be the same for all these enzymes. The Asp, His, and Ser residues, which form the catalytic triad of subtilases, as well as the catalytically important Asn, which is responsible for the stabilization of the transition state oxyanion, are conserved among tomato subtilases (Figure 2). Plant subtilases, in contrast to those of animal and prokaryotic origin, possess a large insertion between the stabilizing Asn and the catalytical Ser residues [1]. This was also observed for all tomato subtilases, where these two residues are separated by 220–248 amino acids as opposed to 60–80 amino acids found in mammalian and bacterial subtilases.

A phylogenetic tree based on the sequence alignment of tomato subtilases (Figure 3) reveals that these enzymes fall into five distinct subfamilies in tomato plants. There are single genes for *Le*SBT1, *Le*SBT2 and *tmp* and numerous genes within the *Le*SBT3/4 and P69 subfamilies. Within the *Le*SBT3/4 group, *Le*SBT4A–4E seem to be more closely related to each other than to *Le*SBT3. These results are consistent with those obtained in Southern blot analyses (Figure 1A). A pairwise comparison of amino acid sequences (Table 1) shows that within a subfamily, subtilase sequences are between 79 and 98% identical. Sequence identity between subfamilies was found to range from 34 to 54%. The tmp protease is the one which is most distantly related to all others which is consistent with the observation that the *tmp* gene is the only intron-containing subtilase gene in tomato thus far analysed.

Expression of tomato subtilases

The expression of subtilases in different organs of tomato plants was analysed by northern blot analysis (Figure 4). Gene-specific probes were used for *Le*SBT1-3 and tmp. The expression of tmp was hardly detectable in the organs analysed. This finding was not surprising, since the closely related protease LIM9

Table 1. Comparison of tomato subtilase sequences. The amino acid sequences deduced from tomato subtilase genomic DNA and cDNA sequences were subjected to pairwise comparison using the program GAP of the GCG package. The degree of sequence identity is given as a percentage. Shaded areas of the table emphasize the high degree of sequence identity observed within the *Le*SBT3/4 and P69 subfamilies.

	SBT2	SBT3	SBT4A	SBT4B	SBT4C	SBT4E	P69A	P69B	P69C	P69D	P69E	P69F	tmp
SBT1	54	41	39	39	39	39	45	46	44	46	44	46	34
SBT2		41	39	39	39	39	43	43	44	44	44	44	36
SBT3			80	84	80	82	39	40	40	40	40	42	36
SBT4A				81	94	81	39	38	39	40	40	39	34
SBT4B					81	87	39	39	40	40	40	40	35
SBT4C						85	38	39	38	39	39	39	35
SBT4E							38	39	38	39	40	40	34
P69A								86	82	98	83	81	37
P69B									85	88	84	80	37
P69C										83	92	79	36
P69D											84	82	38
P69E												80	37
P69F													36

was shown to be expressed only in certain stages of pollen development in tapetal cells of lily flowers [11]. *Le*SBT1–3 each showed a very distinct pattern of expression. While the *Le*SBT3 transcript was found to be present in all organs analysed, being most abundant in roots and cultured cells, a more restricted pattern of expression was observed for *Le*SBT1 and *Le*SBT2. The *LeSBT1* transcript was not detectable in green tissues, while that for *Le*SBT2 was absent from flowers and roots. The expression levels of these three subtilases seem to complement each other, in that the messages for *Le*SBT1–3 add up to comparable levels within the organs analysed. The RNA gel blot for the *Le*SBT4 subfamily (Figure 4) reveals an expression pattern complementary to that of *Le*SBT3: *Le*SBT4 transcript levels were found to be high in those organs where *Le*SB3 expression was low. The probe used for *Le*SBT4 corresponded to a 380 bp fragment of the 3′ end of the *Le*SBT4A cDNA. However, the probe was not specific for *Le*SBT4A, but cross-hybridized with at least one other subtilase of the *Le*SBT4 subfamily as revealed by Southern blot analysis (data not shown). Gene-specific probes have not been generated for subtilases of the LeSBT4 subgroup. Therefore, it is not clear from these data which members of the *Le*SBT4 subfamily are actually expressed and to what extent individual members contribute to the expression pattern observed during northern blot analysis (Figure 4). To confirm the expression of individual members of the *Le*SBT4 subfamily, primer extension experiments were performed using oligonucleotide primers that each hybridized to only one of the *Le*SBT4 subtilases. For *Le*SBT4C a single primer extension product was generated in preparations of total RNA from tomato leaves, cotyledons and flowers but not of RNA isolated from roots (data not shown). This result is consistent with the *Le*SBT4 expression pattern observed during RNA gel blot analysis (Figure 4). For *Le*SBT4A and 4B, a primer extension product was generated from leaf RNA, confirming expression of these proteases in tomato leaf tissue. The transcriptional start sites were determined to be 143, 118, and 133 bp upstream of the ATG start codons of *Le*SBT4A, 4B, and 4C, respectively (data not shown). For the P69 subfamily, the probe used in northern blot analysis was the same that had previously been used for library screening. Hence, this probe hybridized to all members of the P69 subfamily analyzed in this study. Low levels of P69

mRNAs were detectable in all organs analysed, being least abundant in flowers. To what extent individual members of the P69 subfamily contributed to this expression pattern has not been further investigated. The observed pattern of P69 expression (Figure 4) is consistent with results obtained by Vera and co-workers who found low levels of P69A and P69B expression in roots, stems and leaves of healthy tomato plants. These authors found both mRNAs to be induced to very high levels in the leaves and stems of viroid infected tomato plants [12, 13].

Discussion

The gene family of subtilases comprises a minimum of 15 members in tomato plants, twelve of which have been isolated in this study. Based on sequence similarity, the 15 subtilases can be grouped into 5 subfamilies. Single genes exist for *Le*SBT1, *Le*SBT2 and tmp, while 5 and 6 members were found in the *Le*SBT3/4 and P69 subfamilies, respectively. Additional genes are likely to exist in both multi-membered subfamilies since the analysis of genomic clones was not saturating: the genes corresponding to the P69B and *Le*SBT4A cDNAs have not been isolated. Furthermore, the existence of additional distantly related subfamilies cannot be excluded. Therefore, subtilisin-like proteases comprise a large family in tomato plants. The complexity of this family of plant proteases exceeds that of mammalian subtilases with presently seven known proprotein convertases (PCs; [26]).

Northern blot analysis and/or primer extension experiments revealed that most if not all of the tomato subtilase genes are expressed. Expression was found to be organ-specific with at least one of the subtilases being expressed in each organ analysed. In addition to the organ-specific expression, tomato subtilase gene expression has been shown to be regulated developmentally as well as by environmental cues. Expression of P69A and P69B was shown to be induced in viroid-infected tomato plants [12, 13], while tmp was found to be expressed only in anthers at a specific stage of pollen development [14]. The same has been observed for the LIM9 protease in lily flowers [11]. Furthermore, in *Cucumis melo* cucumisin was reported to be highly abundant in the fruit [9], and a subtilase in *Alnus glutinosa* was found to be expressed at early stages of nodule development [10]. In animals, PCs are expressed in a highly tissue-specific manner as well [4]. The restriction of PC expression to certain cell types and developmental stages contributes to the *in vivo* substrate specificity of these enzymes. It has been shown *in vitro* and by co-expression studies that many PCs are able to cleave common substrate proteins at oligobasic sequence motifs [4, 5]. *In vivo*, however, the range of substrates is much more restricted to those protein precursors that are colocalized with the respective protease in the same cell type at the same time in development [4, 27]. The regulated expression of tomato subtilases indicates that a similar situation may exist in tomato plants.

The available information on the function of plant subtilases is scant. In only a single case has a substrate been identified *in vivo*. LRP, a leucine-rich-repeat-containing tomato protein, has been shown to be processed by P69 in diseased tomato plants [28]. The functional significance of this processing event is unknown, however. Indirect evidence has been obtained that in tomato a subtilase may be involved in the processing of the wound signaling peptide systemin which implicates the respective protease in the regulation of the wound response [7]. Clearly, the identification of the *in vivo* substrates, which will help to understand the function of plant subtilases, is one of the major challenges for the future. We plan to approach this problem by overexpression and antisense suppression of these proteases in transgenic tomato plants.

Tomato subtilases share a multi-domain structure with mammalian PCs and other subtilases [1]. The primary structure features an N-terminal signal peptide responsible for targeting of the nascent polypeptide chain to the lumen of the endoplasmic reticulum. The signal peptide is followed by a prodomain which has been shown in subtilisin and other subtilases to be essential for proper folding of the enzyme [29, 30]. In many subtilases, the prodomain is cleaved in an intra-molecular reaction. Nevertheless, the mature protease is not released, but the propeptide remains non-covalently bound to the enzyme as a competitive inhibitor of enzymatic activity. In a second intramolecular reaction, the prodomain is cleaved again (degraded) and the mature enzyme is released [31]. Hence, the amino acid sequence at the junction between prodomain and mature enzyme resembles that of other substrates of the protease. In mammalian PCs, a cluster of basic amino acids is found at the domain junction which reflects the substrate specificity of these subtilases for oligobasic cleavage sites. As far as plant subtilases are concerned, the N-termini of the mature enzymes have been determined for cu-

cumisin [9], for P69A [12] and for LIM9 [11]. In these proteins, the N-termini of the mature enzymes are highly conserved, starting with 2 invariable threonine residues. Sequence conservation in this region extends to all tomato subtilases (Figure 2) and it seems likely, therefore, that in tomato subtilase zymogens as well, processing of the prodomain occurs at this site. It is not known whether in plant subtilases processing of the prodomain is an intra-molecular event as it is in many other subtilases. If this were the case, the sequence at the processing site would be expected to reflect the substrate specificity of the respective enzyme. Only in tmp, an oligobasic sequence (KMKK) is found in proximity (positions -2 to -5) of the putative processing site (cf. Figure 2). Hence, tmp may resemble mammalian PCs in its substrate specificity. In all other plant subtilases, the amino acids which define the C-terminus of the prodomain (His or Asp in the ultimate, Leu or Pro in the penultimate position; cf. Figure 2) are rather well conserved. In the light of the above argument one might expect these proteases to be similar to each other with respect to substrate specificity. Alternatively, processing of the prodomain may occur *in trans* by a different protease as it has been shown for human PC2 [32]. We will address the question of substrate specificity experimentally after overexpression of the recombinant enzymes in a heterologous system.

Surprisingly, tomato subtilase genes, with the exception of the *tmp* gene, were found to lack introns. In plants, intronless genes appear to be more frequent than in animals. In 1.9 Mb of contiguous *A. thaliana* DNA sequence, 15% of structural genes were found to lack introns [33]. The large number of intronless subtilase genes with long open reading frames, however, seems to be exceptional. Based on the presumption that the primaeval genetic material contained introns, the lack of introns in prokaryotic genomes may be explained by the sequential loss of introns as the genomes became streamlined for rapid DNA replication. According to this hypothesis, the introns in eukaryotic genomes are remnants from the ancient genetic material. Alternatively, prokaryotic genes may resemble the ancestral genetic material and introns may have arisen during evolution of the genomes of higher eukaryotes [34, and references therein]. Regardless of which of the two hypotheses is correct, evolutionary pressure must be assumed as the driving force for intronloss and for maintaining the intron-free state, respectively. The nature of this evolutionary pressure in complex eukaryotes is unlikely to be the need for rapid DNA replication [34]. Alternatively, the need for gene expression under conditions in which the splicing machinery is impaired may provide the selective pressure for intron loss. It will be interesting so see if such circumstances exist – for example in diseased tomato plants or under certain stress conditions – when subtilase gene expression is required. The retention of introns in the *tmp* gene, on the other hand, may point to some functional significance of these introns. It has been shown in *Arabidopsis* that introns may be operative at the post-transcriptional level in enhancing gene expression [35].

Acknowledgements

We thank Dr Daniel Bergey for providing a partial cDNA of *LeSBT4A*. This work was supported in part by ETHZ grant 22852/41-2703.5 to A.S. and N.A.

References

1. Siezen RJ, Leunissen JAM: Subtilases: The superfamily of subtilisin-like serine proteases. Prot Sci 6: 501–523 (1997).
2. Siezen RJ, de Vos WM, Leunissen JAM, Dijkstra BW: Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. Prot Enging 4: 719–737 (1991).
3. Steiner DF, Smeekens SP, Ohagi S, Chan SJ: The new enzymology of precursor processing endoproteases. J Biol Chem 267: 23435–23438 (1992).
4. Seidah NG, Chrétien M, Day R: The family of subtilisin/kexin like pro-protein and pro-hormone convertases: Divergent or shared functions. Biochimie 76: 197–209 (1994).
5. Nakayama K: Furin: a mammalian subtilisin/kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. Biochem J 327: 625–635 (1997).
6. Barr PJ: Mammalian Subtilisins: The long-sought dibasic processing endoproteases. Cell 66: 1–3 (1991).
7. Schaller A, Ryan CA: Identification of a 50-kDa systemin-binding protein in tomato plasma membranes having Kex2p-like properties. Proc Natl Acad Sci USA 91: 11802–11806 (1994).
8. Kinal H, Park C-M, Berry JO, Koltin Y, Bruenn JA: Processing and secretion of a virally encoded antifungal toxin in transgenic tobacco plants: evidence for a kex2p pathway in plants. Plant Cell 7: 677–688 (1995).
9. Yamagata H, Masuzawa T, Nagaoka Y, Ohnishi T, Iwasaki T: cucumisin, a serine protease from melon fruits, shares structural homology with subtilisin and is generated from a large precursor. J Biol Chem 269: 32725–32731 (1994).
10. Ribeiro A, Akkermans ADL, van Kammen A, Bisseling T, Pawlowski K: A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. Plant Cell 7: 785–794 (1995).
11. Taylor AA, Horsch A, Rzepczyk A, Hasenkampf CA, Riggs CD: Maturation and secretion of a serine proteinase is as-

sociated with events of late microsporogenesis. Plant J 12: 1261–1271 (1997).
12. Tornero P, Conejero V, Vera P: Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: similarity of functional domains to subtilisin-like endoproteases. Proc Natl Acad Sci USA 93: 6332–6337 (1996).
13. Thornero P, Conejero V, Vera P: Identification of a new pathogen-induced member of the subtilisin-like protease family from plants. J Biol Chem 272: 14412–14419 (1997).
14. Riggs CD, Horsch A: Molecular cloning of an anther specific gene from tomato. Plant Physiol 108: 117 (1995).
15. Uchikoba T, Yonezawa H, Kaneda M: Cleavage specificity of cucumisin, a plant serine protease. J Biochem 117: 1126–1130 (1995).
16. Yonezawa H, Uchikoba T, Kaneda M: Identification of the reactive histidine of cucumisin, a plant serine protease: modification with chloromethyl ketone derivative of peptide substrate. J Biochem 118: 917–920 (1995).
17. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
18. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: Current Protocols in Molecular Biology. Greene Publishing Associates/Wiley-Interscience, New York (1987).
19. Heitz T, Bergey DR, Ryan CA: A gene encoding a chloroplast-targeted lipoxygenase in tomato leaves is transiently induced by wounding, systemin, and methyl jasmonate. Plant Physiol 114: 1085–1093 (1997).
20. Genetics Computer Group: Program Manual for the Wisconsin Package, Vers. 8.0. GCG, Madison, WI (1994).
21. Murray MG, Thompson WF: Rapid isolation of high molecular weight plant DNA. Nucl Acids Res 8: 4321–4325 (1980).
22. Vera P, Conejero V: Pathogenesis-related proteins of tomato. P-69 as an alkaline endoproteinase. Plant Physiol 87: 58–63 (1988).
23. Royo J, Nass N, Matton DP, Okamoto S, Clarke AE, Newbigin E: A retrotransposon-like sequence linked to the S-locus of *Nicotiana alata* is expressed in styles in response to touch. Mol Gen Genet 250: 180–188 (1996).
24. Ujwal S, Masayori I: Propeptide-mediated folding in subtilisin: the intramolecular chaperone concept. In: Richard B, Christian B (eds) Subtilisin Enzymes: Practical Protein Engineering, Vol. 379, pp. 147–154. Plenum Press, New York (1996).
25. Siezen RJ, Leunissen JAM, Shinde U: Homology analysis of the propeptides of subtilisin-like serine proteases (subtilases). In: Shinde U (ed) Intramolecular Chaperones and Folding, pp. 231–253. R. G. Landes Company, Austin, TX (1995).
26. Creemers JWM, Jackson RS, Hutton JC: Molecular and cellular regulation of prohormone processing. Sem Cell Dev Biol 9: 3–10 (1998).
27. Gensberg K, Shamem J, Matthews GM: Subtilisin-related serine proteases in the mammalian constitutive secretory pathway. Sem Cell Dev Biol 9: 11–17 (1998).
28. Tornero P, Mayda E, Gómez MD, Canas L, Conejero V, Vera P: Characterization of LRP, a leucine-rich repeat (LRR) protein from tomato plants that is processed during pathogenesis. Plant J 10: 315–330 (1996).
29. Shinde U, Li Y, Inouye M: Propeptide mediated protein folding: intramolecular chaperones. In: Shinde U, Inouye M (eds) Intramolecular Chaperones and Protein Folding, pp. 1–9. R.G. Landes Company, Austin, TX (1995).
30. Eder J, Rheinnecker M, Fersht A: The pro-sequence assists subtilisin BPN to fold from an intermediate to the native state: a common mechanism for proteases, but not for all proteins with pro-sequences. In: Shinde U, Inouye M (eds) Intramolecular Chaperones and Protein Folding, pp. 35–60. R.G. Landes Company, Austin, TX (1995).
31. Thomas G, Molloy SS, Anderson ED, Thomas L: Multi-step activation of furin: a model for the eukaryotic proprotein convertases. In: Shinde U, Inouye M (eds) Intramolecular Chaperones and Protein Folding, pp. 157–179. R.G. Landes Company, Austin, TX (1995).
32. Shennan KIJ, Taylor NA, Jermany JL, Matthews G, Docherty K: Differences in pH optima and calcium requirements for maturation of the prohormone convertases PC2 and PC3 indicates different intracellular locations for these events. J Biol Chem 270: 1402–1497 (1995).
33. Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, van Staveren M, Stiekema W *et al.*: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391: 485–488 (1998).
34. Gilbert W, Marchionni M, McKnight G: On the antiquity of introns. Cell 46: 151–154 (1986).
35. Rose AB, Last RL: Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene *PAT1*. Plant J 11: 455–464 (1997).
36. Corpet F: Multiple sequence alignment with hierarchical clustering. Nucl Acids Res 16: 10881–10890 (1988).
37. Benner SA, Badcoe I, Cohen MA, Gerloff DL: *Bona fide* prediction of aspects of protein conformation. Assigning interior and surface residues from patterns of variation and conservation in homologous protein sequences. J Mol Biol 235: 926–958 (1994).