Genome-Wide Analyses of Subtilisin-Like Serine Proteases on Solanum tuberosum

Natalia Sigrid Norero, Martin Alfredo Castellote, Laura de la Canal & Sergio Enrique Feingold

American Journal of Potato Research
The Official Journal of the Potato Association of America

ISSN 1099-209X
Volume 93
Number 5

Am. J. Potato Res. (2016) 93:485-496
DOI 10.1007/s12230-016-9525-5

Springer

Am. J. Potato Res. (2016) 93:485–496
DOI 10.1007/s12230-016-9525-5

Genome-Wide Analyses of Subtilisin-Like Serine Proteases on *Solanum tuberosum*

Natalia Sigrid Norero[1] · **Martin Alfredo Castellote**[1] · **Laura de la Canal**[2] · **Sergio Enrique Feingold**[1]

Published online: 10 June 2016

Abstract Plant subtilisin-like serine proteases or subtilases constitute a large expanded gene subfamily in potato. We identified 82 potato subtilases encoded by 74 genes. All subtilases present a characteristic S08 domain, and the majority of them have an association protein domain, an inhibition_I9 domain and a signal peptide that consists of a signal sequence for targeting to the secretory pathway. Phylogenetic studies revealed five subtilase groups named *St*SBT1 to *St*SBT5. A genome expansion of potato subtilase subfamily was reflected in two phylogenetic clusters, *St*SBT1 and *St*SBT4 both characterized by intronless genes in tandem arrays mainly in chromosome I and VIII. Twenty nine of the identified subtilases co-localize with six out of 24 metaQTLs related to late blight resistance previously described in potato. These metaQTLs includes subtilase genes up regulated in detached potato leaves inoculated with *P. infestans*, some of which are homologous to p69 subtilases genes from tomato.

Resumen Las serin proteasas tipo subtilisinas o subtilasas constituyen una gran subfamilia de genes en papa. En este trabajo se identificaron 82 subtilasas codificadas por 74 genes. Todas las subtilasas presentan un dominio S08 característico, y la mayoría de ellas poseen un dominio asociado a proteasa, un dominio inhibidor I9 y un péptido señal que consiste en una señal que conduce a la vía secretora. En los estudios filogenéticos se describen cinco grupos de subtilasas nombrados de *St*SBT1 a *St*SBT5. La expansión en la subfamilia de subtilasas se ve reflejada en dos grupos, *St*SBT1 y *St*SBT4, caracterizados por la presencia de genes sin intrones con disposición en tándem localizados en el cromosoma I y VIII. Veintinueve de las subtilasas identificadas co-localizan con seis de los 24 metaQTLs de resistencia a tizón tardío previamente descriptos en papa. Estos metaQTLs incluyen genes de subtilasas con regulación positiva en ensayos de hojas de papa desprendidas inoculadas con *P. infestans*, algunos de los cuáles son homologos a genes de subtilasas p69 descriptos en tomate.

Keywords Subtilases · Potato · *Phytophthora_infestans* · metaQTL · Intronless

Electronic supplementary material The online version of this article (doi:10.1007/s12230-016-9525-5) contains supplementary material, which is available to authorized users.

✉ Natalia Sigrid Norero
norero.natalia@inta.gob.ar

✉ Sergio Enrique Feingold
feingold.sergio@inta.gob.ar

[1] Laboratorio de Agrobiotecnología, Instituto Nacional de Tecnología Agropecuaria (INTA) EEA - Balcarce, Ruta 226, Km 73,5. C.C. 276, (7620) Balcarce, Argentina

[2] Instituto de Investigaciones Biológicas (IIB), Universidad Nacional de Mar del Plata (UNMdP), Dean Funes 3350. C.C. 722, (7600) Mar del Plata, Argentina

Introduction

Subtilisin-like serine proteases (or subtilases, SBTs) constitute a large subfamily across all kingdoms. Subtilases belong to the clan SB of serine peptidases, family S8, subfamily S8A (S08.001, according to MEROPS database) (Rawlings et al. 2006). The S8A subfamily comprises subtilisin, thermitase, lantibiotic leader peptidase, proteinase K and pyrolysins-related enzymes as homology groups (Siezen and Leunissen 1997). Plant subtilases have been found only at the pyrolysin group (Schaller et al. 2012). They differentiate from other proteases by a specific arrangement of the catalytic triad with the amino

acid residues from the active site: aspartate (Asp, D), histidine (His, H) and serine (Ser, S) in the sequence Asp-Thr/Ser-Gly; His-Gly-Thr-His and Gly-Thr-Ser-Met-Ala-X-Pro arranged in an α/β protein scaffold. In same cases, an asparagine (Asn, N) is present between D and H residues. The S8A subfamily has been described in *Arabidopsis*, rice and more recently in grape comprising 56, 63 and 80 members respectively (Beers et al. 2004, Rautengarten et al. 2005; Tripathi and Sowdhamini 2006; Cao et al. 2014). There are also at least fifteen subtilases genes described in tomato (Jordá et al. 1999; Meichtry et al. 1999; Jordá et al. 2000) organized in clusters in a tandem array. These tomato subtilases are coded by intronless genes with the exception of TMP (Riggs et al. 2001). Intronless genes have been found to be present in eukaryotic genomes belonging to large gene families whose members appear in a tandem array and would have evolved by gene duplication or reverse transcription/integration (Jain et al. 2008).

Subtilases comprehend diverse functions such as control of development, protein degradation and signaling (Rautengarten et al. 2005). These functions includes xylem differentiation (Zhao et al. 2000), air root development (AIR3) (Neuteboom et al. 1999), epidermal surface formation (Tanaka et al. 2001), stomata density and distribution (SDD1) (Berger and Altmann 2000), microsporogenesis (TMP) (Riggs et al. 2001), senescence, programmed cell death (Coffeen and Wolpert 2004; Fernández et al. 2015), plant defense (Jordá et al. 1999; Tian et al. 2004; Tornero et al. 1997; for review see Antão and Malcata 2005) and more recently immune priming (Ramirez et al. 2013).

The first subtilase related to plant defense was a 69 KDa protein isolated from tomato which was induced after *Phytophthora infestans* infection and consequently it was classified as a defense related protease and named PR-P69 (Tornero et al. 1996, 1997). Among p69 subtilases, p*69b* and p*69c* genes are both specifically induced following tomato infection with *Pseudomonas syringae* (Jordá et al. 1999) and *Phytophthora infestans* (Tian et al. 2004) and are inhibited by *P. infestans* effectors EPI1 and EPI10 respectively (Tian et al. 2004, 2005).

P. infestans is the most devastating pathogen of potato and tomato plants worldwide. This is a heterothallic oomycete responsible for late blight disease that can cause up to 100 % yield losses (Reviewer by Nowicki et al. 2012). A comparative genome study of *Phytophthora* (Haas et al. 2009) and host phylogenetic analyses reveals that *P. infestans* has a dynamic genome and an accelerated gene evolution that includes host jumps (Dong et al. 2014; Raffaele et al. 2010).

In the last two decades, many efforts have been related to potato QTL mapping looking for *P. infestans* resistance genes in different *Solanum* populations. Danan et al. (2011) analyzed and grouped this information by clustering 144 late blight resistance QTLs into 24 metaQTLs redefining important genomic regions across species. However, most of the late blight responsible genes for resistance of each potato metaQTL has not been completely identified yet. Previous evidences in our laboratory suggest that there are potato subtilase genes linked to molecular markers that co-localize with QTLs of late blight resistance in potato (data not published).

In 2011, the potato genome of a doubled monoploid *S. tuberosum* Group Phureja (DM1-3 516 R44; DM) genotype was released (Potato Genome Sequence Consortium; PGSC 2011). In this publication, 39,031 protein-coding genes were annotated along with RNA-seq based expression data (PGSC 2011).

The aim of this work is to identify subtilisin-like serine proteases on the potato genome (PGSC 2011), characterize them based on phylogenetic analyses, gene and protein primary structure and RNASeq expression profile identifying gene subtilases candidates for late blight resistance.

Materials and Methods

SBT Identification and Characterization

A profile of hidden Markov Model (HMMER2.3.2) was created based on P69 amino acid sequences from *Solanum Lycopersicum* P69A (JC6119), P69B (CAA71234.1), P69C (CAA76726), p69D(CAA76727), P69E (CAB67119) and P69F (CAB67120) identified by Jordá et al. (1999, 2000), P69C (CAA06412.1) and P69E (CAA06413.1) described by Meichtry et al. (1999). These sequences were aligned with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2), the resulting alignment was trimmed on both ends, removing regions where the alignment was poor due to the variable nature. Using this profile, a HMMER search was performed (Eddy 1998) against PGSC DM peptide database (Version 3.4). HMMER profile was validated on Arabidopsis TAIR database (http://www.arabidopsis.org, TAIR10_pep_20101214). Finally, a HMMER logo was obtained using LogoMat (Schuster-Böckler et al. 2004).

Domain Structure Analysis and Subcellular Localization

SBT proteins domains and active sites were analyzed in Pfam database (Punta et al. 2012, 2014) and InterproScan (Jones et al. 2014); signal peptide was detected with Phobius program (Käll et al. 2007) and TargetP1.1 with default settings (Emanuelsson et al. 2007). PredoTar V1.3 (https://urgi.versailles.inra.fr/Tools/Predotar) and ProtComp V9.0 (http://www.softberry.com/berry.phtml?topic=protcompplandgroup=programsandsubgroup=proloc) were used to predict signal sequences to organelles and other subcellular localizations. Protein structure was plotted with iTOL (Letunic and Bork 2006, 2011).

Gene Structure Analysis

SBT genes were checked for intron and exon structure based on primary transcripts from the PGSC data (http://solanaceae.plantbiology.msu.edu). A schematic representation of introns/exons structure was visualized in the phylogenetic tree with iTOL program (Letunic and Bork 2006, 2011).

Phylogenetic Analyses

Two phylogenetic analyses were carried out. The first one with all the potato subtilase amino acid sequences identified and the second one combining these amino acid sequences with 60 SBTs amino acid sequences from *Arabidopsis thaliana* that includes those analyzed by Beers et al. (2004) and Rautengarten et al. (2005) included four *At*SBTs obtained with our HMMER profile on Arabidopsis TAIR database, and 16 *Solanum lycopersicum* subtilases described in Meichtry et al. (1999).

The sequence aminoacids alignments were performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).

Phylogenetic tree analyses were generated with MEGA version 6 (Tamura et al. 2013) with the method maximum likelihood considering 1000 bootstrap replications. The dendrogram was plotted with iTol (Letunic and Bork 2006, 2011). Specificity-determining positions (SDPs) were analyzed for each *St*SBT group with SDPFox (Mazin et al. 2010).

A phylogenetic dendrogram with nucleotide sequences from representative transcripts of genes with one o more "number fragments per kilobase of exon per million fragments mapped" (FPKM) in DM potato libraries (http://solanaceae.plantbiology.msu.edu) was performed. The nucleotide alignment was performed with Multiple Alignment using Fast Fourier Transform (MAFFT; http://www.ebi.ac.uk/Tools/msa/mafft/). Phylogenetic tree analyses were generated with MEGA version 6 (Tamura et al. 2013) with the method maximum likelihood considering 1000 bootstrap replications. The dendrogram was plotted with iTol (Letunic and Bork 2006, 2011).

Expression Profile http://solanaceae.plantbiology.msu.edu

RNA-Seq expression data from doubled monoploid *Solanum tuberosum* group Phureja DM1-3 516 R44 clone (DM) (http://solanaceae.plantbiology.msu.edu) were integrated into a local software (developed by staff at Laboratorio de Agrobiotecnología, INTA EEA- Balcarce), analyzed and exported to iTol in order to integrate expression results with a phylogenetic dendrogram from nucleotide sequences. These data include libraries from different tissues and plant organs: root, stem, leaf, flower, stolon, primary and secondary tuber and immature and mature fruits as well as detached leaf inoculated with 0.5 to 0.7 ml containing 3×10^4 sporangia/ml of *P. infestans* (isolated US8: Pi02-007), 2 mg/ml of DL-β-aminobutiric acid (BABA), a resistance gene inductor, and 100 μg/ml of benzoic-(1,2,3)-thiadiazol-7-carbotioic acid S-methyl ester (BTH) an analog to salicylic acid (PGSC 2011). Expression data were expressed in "number fragments per kilobase of exon per million fragments mapped" or FPKM (PGSC 2011) considering at 0 hs and a pool of samples at 24, 48 and 72 h post inoculation (hpi).

Genetic map with *St*SBT and Late Blight Resistance metaQTL Localizations

We compared a genetic map with a physical map constructed with potato subtilase gene localizations obtained from the potato sequenced genome PGSC database (http://potato.plantbiology.msu.edu/data/PGSC_DM_V403_genes.gff.zip) and potato molecular markers with sequence information, base pair (bp) scale chromosome position (Sharma et al. 2013) as well as centimorgan (cM) position information obtained from the consensus potato map (www.solgenomics.net, Danan et al. 2011). The genetic map includes molecular markers which flanks late blight resistance metaQTLs in potato (Danan et al. 2011). A specific program was develop to rapidly look for each of the 74 SBT genes in the 24 metaQTLs of resistance to Pi described by Danan et al. (2011). The final figure was visualized with a viewer developed with HTML5, JQuery and Kinetics.js framework developed in the Laboratorio de Agrobiotecnología from INTA EEA-Balcarce.

Results and Discussion

Identification and Characterization of Subtilisin-Like Serine Protease Genes in Potato

Based on subtilase genes identified in tomato, we built a HMMER profile upon tomato P69 proteins (presented in the electronic supplementary material ESM 1, ESM 2: Fig. 1), to look for subtilases on potato DM peptide PGSC database (Version 3.4). The HMMER search was able to identify 82 subtilase-like serine proteases (*St*SBT) coded by 74 genes with a score ranged from 1762.2 to 32.3, and the e-value from 0 to 1.10^{-5} (ESM 3). To verify the specificity of the HMMER profile used, we perform a HMMER search on the peptide database of *Arabidopsis thaliana* (http://www.arabidopsis.org) TAIR10_pep_20101214) to contrast these results with previous published by Beers et al. (2004) and Rautengarten et al. (2005). Hence, we identified 60 *At*SBT in *A. thaliana* that include all subtilases previously described by Beers et al. (2004) and Rautengarten et al. (2005) and four additional subtilase sequences: AT1G32980, AT5G59110, AT1G71950 and AT2G39851 (ESM 4) not reported elsewhere. These proteins

were verified in Interproscan site to be subtilases. This result confirms the specificity and power of the HMMER profile used.

The number of subtilases found in the potato genome was higher than the 56 reported in *Arabidopsis* (Rautengarten et al. 2005) and the 63 in rice (Tripathi and Sowdhamini 2006) that could suggests function diversification or redundancies. However, the potato SBT number was similar to grape with 80 subtilases (Cao et al. 2014) and to tomato that presents 76 subtilases in its genome (result based on a HMMER search performed on tomato peptidase database from Sol Genomic Network, with a score ranged from 2426.4 to 89 and the e-value from 0 to $5{,}6.10^{-23}$, data not shown). The protein length ranged from 295 to 1141 amino acids (PGSC003DMP400029746 and PGSC003DMP400007009 respectively) with a mean of 702 ± SD 117 amino acids long.

Phylogenetic Analyses of Potato Subtilases

Phylogenetic analysis of the 82 potato subtilases was performed and the consensus phylogeny obtained was shown in two figures, in the first one we included gene location and intron/exon structure prediction (Fig. 1) while in the second figure we presented protein domain structure information (Fig. 2). We also performed a phylogenetic analysis with 82 *St*SBT, 60 *At*SBT and 16 tomato subtilases (Meichtry et al. 1999) to relate information of these species (Fig. 3).

Based on phylogenetic relationships of potato subtilases found, we have divided them into five *St*SBT groups named from *St*SBT1 to *St*SBT5 and an ungrouped subtilase PGSC0003DMP0057105 (Figs. 1 and 2).

*St*SBT1 group comprise seventeen subtilases from chromosome I, IV, X and XII characterized mainly by intronless genes (Fig. 1). Subtilases from chromosome X presented primary structural changes with protein associated domian (PA), I9_inhibitor and signal peptide deficiencies (Fig. 2b).

*St*SBT2 group includes sixteen subtilases coded by thirteen genes located on nine chromosomes. This is the most diverse group based on chromosome localization, primary transcript structure (Fig. 1), protein length and domain structure (Fig. 2).

*St*SBT3 group includes eleven subtilases. The majority of these subtilases are coded by intronless genes, located in seven chromosomes, mostly with a similar domain structure.

*St*SBT4 is the largest group which includes 35 subtilases with 77 % of their genes located on chromosome VIII and 21 of them consist of intronless genes.

Protein domain structures are similar with some exceptions characterized by short amino acid sequences (PGSC0003DMP0007009 with 295 and PGSC0003DMP0041860 with 532 amino acids long) or long amino acid sequences as the case of PGSC00003DMP0029746 with 1141 aminoacids. In the case of PGSC0003DMG400003913, presented a double length when it is compared others potato subtilase genes. Analysis of the corresponding chromosome sequence (including 500 bp flanking regions upstream and downstream of this gene) on FGENESH gene annotation software evidenced the presence of two different genes (data not shown). This case should be analyzed in more detail to determine if it corresponds to an assembly or an annotation error.

Finally, *St*SBT5 is the smallest group with three subtilases, two of them coded by intronless genes localized on chromosome VIII.

There is no clear domain or intron structure that defines each group. Although, intron presence is more frequent in *St*SBT2. In analyzing amino acid sequences that can justify grouping, we found 29 possible specificity-determining positions (SDPs) with Z-score ranging from 4.28 to 7.61 and *p*-value from −196.91 to −24.60 (ESM 5: Table 1). These SDPs could represent important amino acid positions for protein structure or related to specific functions (Mazin et al. 2010).

Phylogenetic Tree of Potato, Tomato and Arabidopsis Subtilases

Biological functions of plant subtilases remain largely unknown however some efforts are being carried out to understand their role. The phylogenetic analysis of 158 amino acid sequences including potato, tomato and Arabidopsis subtilases evidenced seven clusters named from I to VII which are illustrated on Fig. 3.

- *Cluster I* includes all potato subtilases from *StSBT*4 group and *St*SBT5 (Figs. 1 and 2) and all P69 tomato subtilases described by Jordá et al. (1999 and 2000). The phylogenetic tree evidenced twelve *St*SBTs closely related to eight P69 subtilases. Some *St*SBTs joined in pairs with P69 tomato subtilases which could represent homologous genes such as: PGSC0003DMP0028116 with P69E (CAB67119); PGSC0003DMP0006964 with P69D (CAA76727); PGSC0003DMP0007010 with P69C (CAA76726); PGSC0003DMP0006965 and PGSC0003DMP0006967 (both coded by PGSC0003DMG400003913) with P69A (JC6119). While P69B (CAA71234.1) related with PGSC0003 DMP0007009 (coded by PGSC0003DMG400003938) and PGSC0003DMP0007007 (coded by PGSC0003 DMG401003937) and with three more subtilases possible paralogs among them: PGSC0003DMP0056894 (coded by PGSC0003DMG400034790), PGSC0003 DMP0066275 (coded by PGSC0003DMG400044171) and PGSC0003DMP0067339 (coded by PGSC000 3DMG40G0045235). Thus, subtilases members of cluster I are related to P69B and P69C, and could be involved in defense functions as will be analyzed later. This cluster also includes the ungrouped subtilase PGSC0003

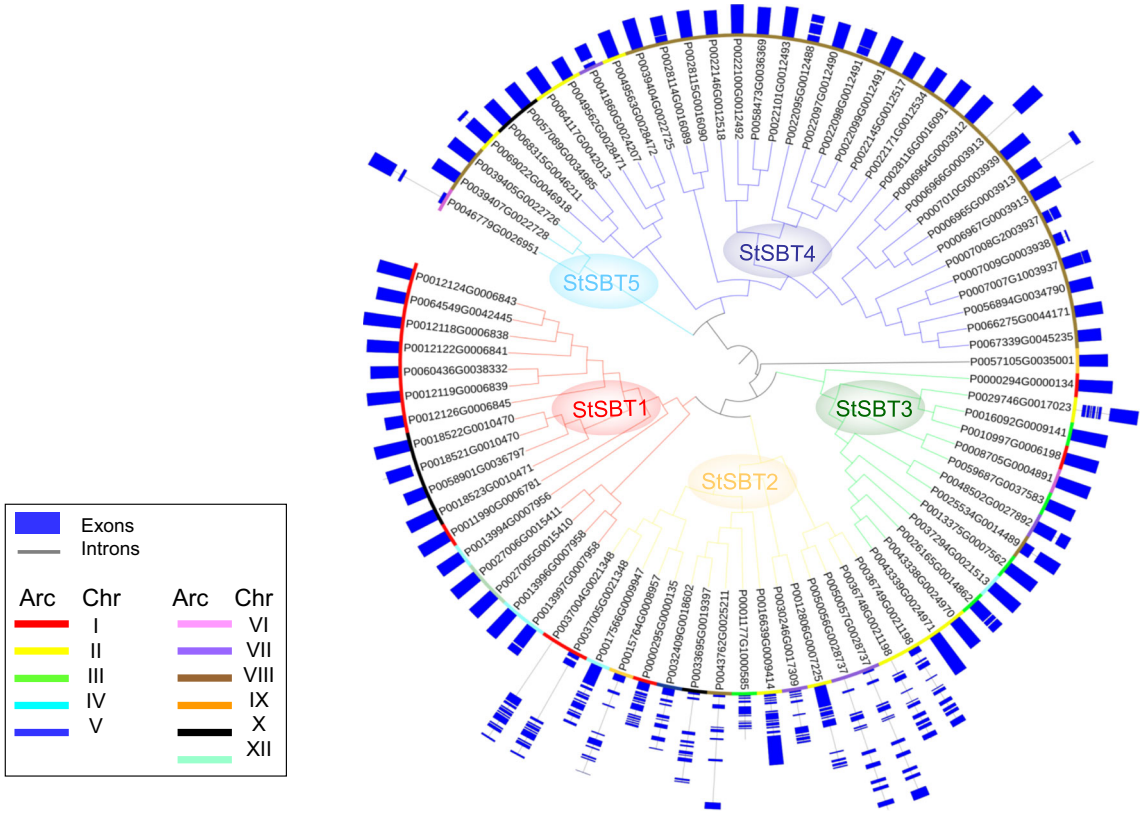

Fig. 1 Phylogenetic tree of 82 subtilisin-like serine proteases from *Solanum tuberosum* group *Phureja*. The phylogenetic tree was built with MEGA version 6 with maximum likelihood analyses, 1000 bootstrap replications. The phylogram was built with iTol program (Letunic and Bork 2006, 2011). This is a neighbor joining tree based on the program ClustalO (default parameters). The circular phylogram includes schematic representation of exon and introns structure of predicted primary transcripts (boxes and lines respectively). Around the tree is indicated the chromosome number with a colored curved line. Cluster assignments are indicated by color-coding in the tree. For brevity, potato protein and gene names have been shortened and joined i.e. P#######G#######, instead of PGSC0003DMP####### and PGSC0003DMG40#######, where # represent the digit number assigned by PGSC (2011)

DMP0057105 illustrated on Figs. 1 and 2. This potato subtilase was closely related to At1g04110 (SDD1) which is involved in stomatic development, meristemoid and guard mother cells precursor (Von Groll et al. 2002).

- *Cluster II* includes *St*SBT3 (Figs. 1 and 2) and *At*SBT1 (Rautengarten et al. 2005). This cluster includes members related to development and abiotic stress. Here we found PGSC0003DMP0026165 which showed high expression in tubers, roots and stems (Fig. 4) and it is a possible homolog to At5g67360 (ARA12) which is involved in nodule development (Ribeiro et al. 1995). Also, this cluster includes PGSC0003DMP0010997, which is expressed in all plant organs and is induced by wounding and BTH (Fig. 4) and appear as a possible homolog to At4g34980 (SLP2) that is strong induced upon abiotic stress (Golldack et al. 2003).
- *Cluster III* groups all *St*SBT1 (Figs. 1 and 2) and Atg67090 (*At*ABT1.9, Rautengarten et al. 2005) with unknown function. This cluster includes *St*SBTs homologs to *Le*SBT1-3 and *Le*SBT4A-E most of them presented in pairs (Meichtry et al. 1999).
- *Cluster IV* includes four *St*SBT2 and all members from *At*SBT2 (which includes ALE1, involved in cuticle formation; Tanaka el al. 2001) and *At*SBT6 (described by Rautengarten et al. 2005) and TMP from tomato (Riggs et al. 2001).
- *Cluster V* includes five *St*SBT2 (Figs. 1 and 2) and members from *At*SBT4 and *St*SBT5 (Rautengarten et al. 2005). PGSC0003DMP0000295 could be the homolog of At2g00230 (XSP1) involved in xylem differentiation (Zhao et al. 2000).
- *Cluster VI* includes three *St*SBT2 (Figs. 1 and 2) and four members of *At*SBT5 (Rautengarten et al. 2005). In this cluster PGSC0003DMP0030246 is closely related to Arabidopsis subtilases linked with plant development: At2g04160 (AIR3) involved on the lateral roots emergency and At5g59810 implicated on the maintenance of the apical meristem (Neuteboom et al. 1999)
- *Cluster VII* groups four *St*SBT2 (Figs. 1 and 2) and *At*SBT3 (Rauntegarten et al. 2005).

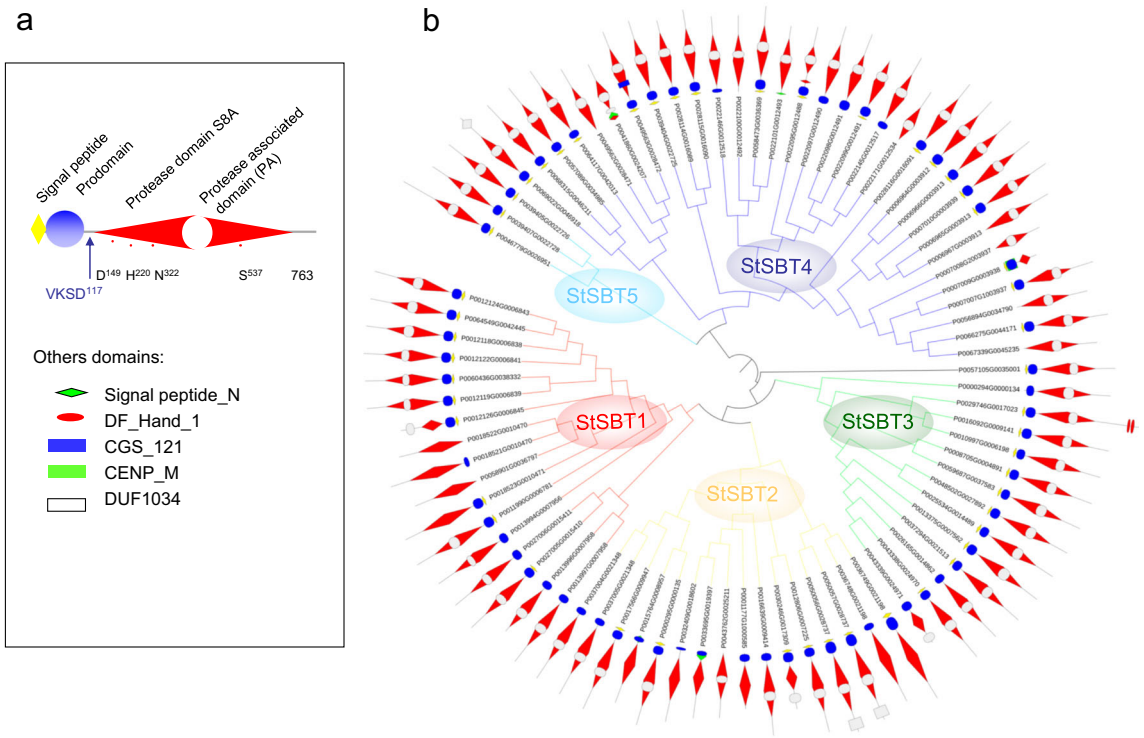

Fig. 2 Phylogenetic tree of 82 subtilisin-like serine proteases from *Solanum tuberosum*. (**a**) Plant subtilisin-like serin protease typical primary structure. The *arrow* indicates the site where the prodomain is cleavaged and the red points indicate the amino acids of the catalytic triad. (**b**) Potato subtilases phylogenetic tree and multi-domain structures. The phylogenetic tree was built with *MEGA* version 6 with maximum likelihood analyses, 1000 bootstrap replications. Cluster assignments are indicated by color-coding in the tree. The phylogram was built with iTol program (Letunic and Bork 2006, 2011). For brevity, potato protein and gene names have been shortened and joined i.e. P########G########, instead of PGSC0003DMP######## and PGSC0003DMG40########, where # represent the digit number assigned by PGSC

This phylogenetic tree showed a gene expansion of the subtilase family on potato genome evidenced in cluster I where most potato subtilases genes are located in chromosome VIII. These *St*BST could have evolved separated from Arabidopsis subtilases. Multigenic families frequently present gene expansions characterized by duplication events that appeared in the genome in tandem arrays (Page and Holmes 1998; Jain et al. 2008). Gene expansion has been described as a characteristic feature on potato and tomato genomes (PGSC 2011). This could be accompanied by functional diversification as an important adaptation role (Page and Holmes 1998). For instance, gene expansion has been described in gene families related with defense such as the nucleotide binding site leucine-rich repeats (NBS-LRR) genes that encoded a large class of disease resistance (R) proteins in plants (Arya et al. 2014; Zhang et al. 2014).

Gene Structure Analyses

Intronless genes are a characteristic feature of prokaryotes and constitute a significant portion of eukaryote genomes (Sakharkar and Kangueane 2004; Jain et al. 2008). To gain further insights into the structural diversity of subtilases, we compare exon-intron organization based on the predicted primary transcript structure of *St*SBT. We found 47 intronless subtilase genes on the potato genome. This represents almost the 63 % of the total *St*SBT genes identified (Fig. 1) which is significantly higher than the percentages observed in Arabidopsis and grape that reach 16 % and 25 % of intronless subtilases genes respectively (Cao et al. 2014; Beers et al. 2004; Rautengarten et al. 2005). Intronless genes were grouped in four clusters *St*SBT1, *St*SBT3, *St*SBT4 and *St*SBT5, while most genes with numerous exons and introns, varying from 7 to 11 exons, constituted another cluster, *St*SBT2 (Fig. 1). It can be noticed that members from clusters *St*SBT1 and *St*SBT4 are mostly coded by intronless genes localized mainly in chromosome I and VIII, respectively. The first cluster includes the potato homologues to tomato *Le*SBT reported by Meichtry et al. (1999) while the second cluster presents the homologues to tomato p69 subtilases (Jordá et al. 1999, 2000) equally arranged in tandem.

Intronless genes have been found in *A. thaliana* and rice belonging to large gene families whose members appear in tandem arrays and would have evolved by gene duplication or reverse transcription/integration (Jain et al. 2008). They have

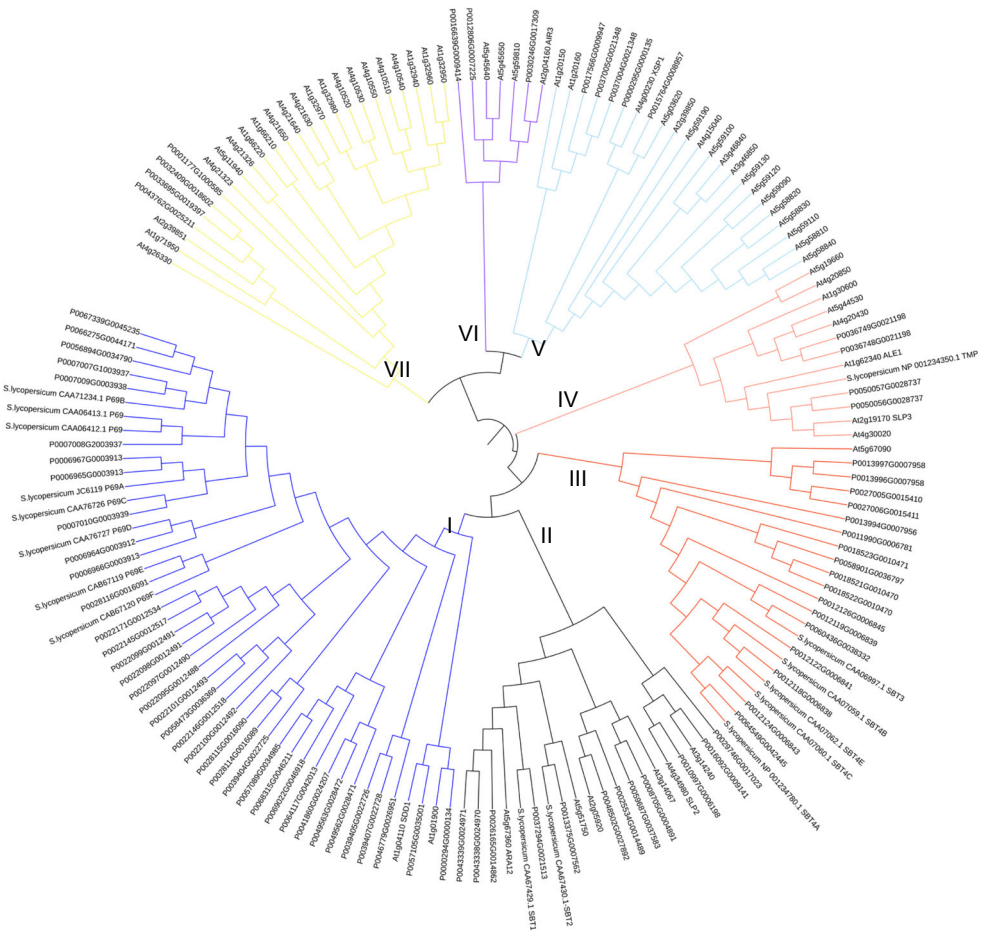

Fig. 3 Phylogenetic tree of 82 subtilisin-like serine proteases from *Solanum tuberosum*, 16 from *Solanum lycopersicum* and 60 from *A. thaliana*. Phylogenetic tree built with MEGA version 6 with maximum likelihood analyses, 1000 bootstrap replications Jones-Taylor-Thornton (*JTT*) model and Nearest-Neighbor-Interchange method. The phylogram was built with iTol program (Letunic and Bork 2006, 2011)

been related to protein synthesis, protein turn over (ribosomal proteins, F-box proteins), signal transduction (protein kinases leucine-rich repeat proteins (LRR), DNA binding (zinc finger proteins, AP2 domain proteins), metabolism (cytocrome P450 proteins) and disease resistance proteins (Jain et al. 2008).

Domain Structure Analyses

Subtilases are characterized by a multi-domain structure which is illustrated in Fig. 2a (Rautengarten et al. 2005; Cao et al. 2014). Most plant subtilases comprise mainly a signal peptide, a pro-peptide, a protease domain and a protease associated domain (Siezen et al. 2007). The signal peptide usually consists of a secretion signal sequence for targeting to the secretory pathway or other subcellular compartments (Vartapetian et al. 2011). The pro-domain N terminus, propeptide or inhibitor_I9 (PF05922) is involved after removal, on the pro-enzyme activation, working as a molecular chaperone in the folding of the mature peptidase. Thus, the inhibitor_I9 prevents the access of the substrate to the active site and activates the peptidase when it is removed either by autocatalytic cleavage or by interaction with a secondary peptidase (Siezen 1996). The protease domain (S08 domain) that define the subfamily S8A (PF00082) includes the catalytic triad and a protease associated domain (PA) (PF02225) which is about 170-210 amino acids long, implicated in protein - protein interactions or substrate specificity (Mahon and Bateman 2000; Luo and Hofmann 2001).

Potato subtilases presented all the typical multi-domain structure including the S08 domain with the characteristic active sites of subtilases varying slightly along subtilases but with the expected order for the S8A subfamily classification. Fifty-eight *St*SBTs presented the inhibitor_I9 domain or prodomain (Fig. 2b) while the PA domain was present in 72 *St*SBTs. The absence of the PA domain in subtilases is not a common feature in the S8A subfamily. However some of the subtilases reported in *At*SBTs (Rautengarten et al. 2005) and grape subtilases (Cao et al. 2014) also lack this domain.

In addition we found some subtilases with less represented domains in S8A subfamily (Tripathi and Sowdhamini 2006) such us PGSC003DMP400029746 with two EF_Hand_1 domains related to calcium binding (PDOC00018) and PGSC003DMP400050057 and PGSC003DMP400050056 both with an Unknown Function Domain (DUF1034) with

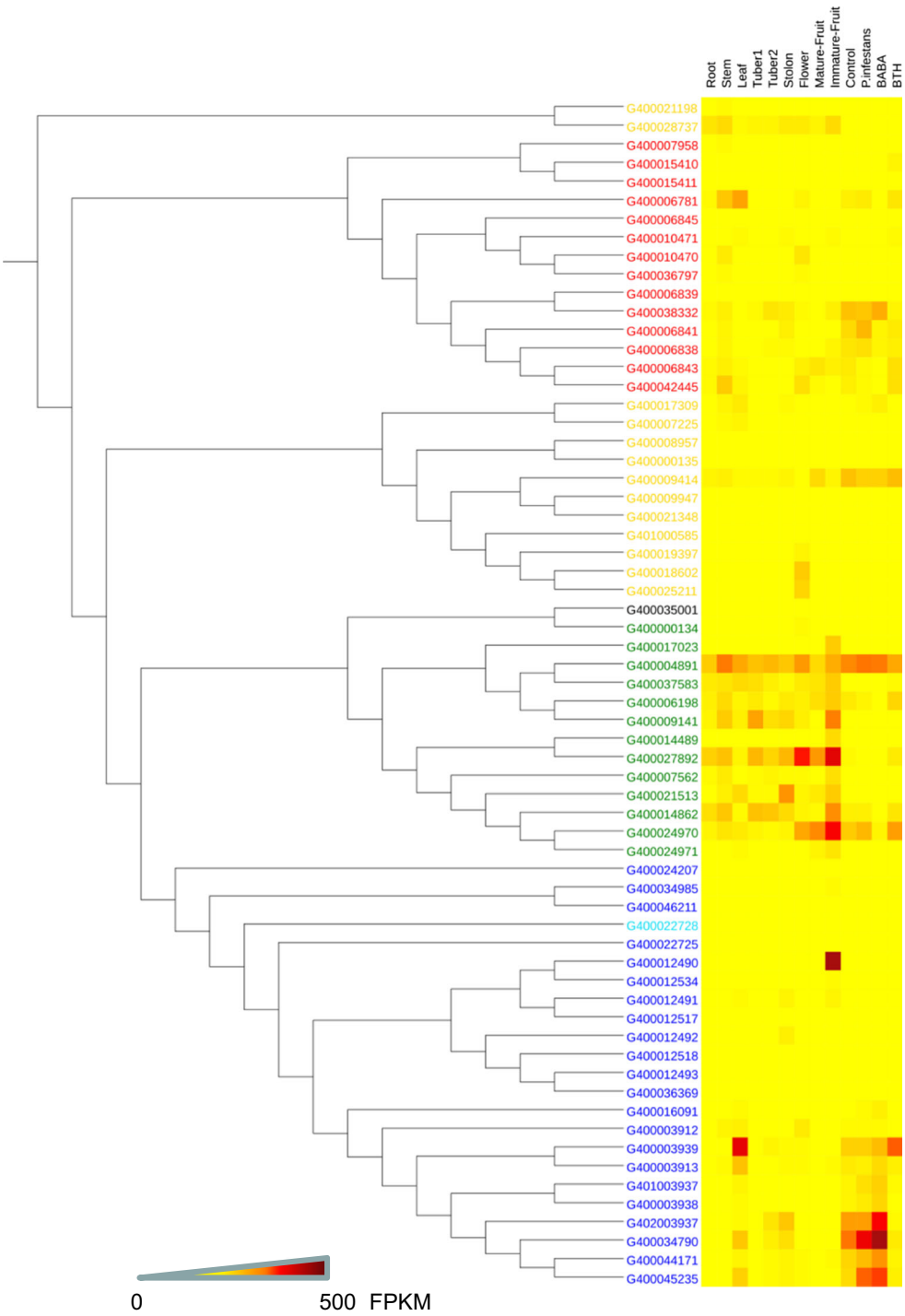

Fig. 4 Plant organs and biotic expression heatmap of 64 potato subtilases genes. The phylogenetic tree from aligned nucleotide sequences was built with *MEGA* version 6 with maximum likelihood analyses. RNA-Seq data from different libraries from doubled monoploid *Solanum tuberosum* group Phureja DM1-3 516 R44 clone (DM) are indicated as "number fragments per kilobase of exon per million fragments mapped" (FPKM) by color scale. The phylogram was built with iTol program (Letunic and Bork 2006, 2011). Gene names are indicated as G#########, instead of PGSC0003DMG######### while color names represent the color from the *St*SBT group from Fig. 2. Genes with no detectable expression in 32 DM and 16 RH RNA-Seq libraries (http://solanaceae.plantbiology.msu.edu/) were removed from the analyses

functions later identified in sugar hydrolysis in subtilases from other organisms such as fungi (Muszewska et al. 2011).

Sub-Cellular Localization

Primary structure studies of proteins allow predicting its cellular and subcellular localizations generating hypothesis about its function (Emanuelsson et al. 2000; Rautengarten et al. 2005). The S8A subtilases appear mostly to be directed towards the secretory pathway; however, in *Arabidopsis thaliana*, locations to chloroplasts, mitochondria and other organelles with no evident signal sequences were also reported (Beers et al. 2004). In this work we have found 60 *St*SBT predicted to have extracellular localization, five *St*SBTs were predicted to be targeted in mitochondria, one in the chloroplast, two in the cytoplasm and thirteen *St*SBTs in plasmatic,

cloroplastic or vacuolar membrane. These results are based on motif predictions, homology with known proteins and neural networks of the predicted programs used (ESM 6: Table 2).

Considering that the extracellular space is where the first host-pathogen interaction, recognition and signaling events take place (Dixon and Lamb 1990) the accumulation of subtilases here may account for an important role during pathogenesis that could be related to protein turnover, defense or/and signaling (Tornero et al. 1996; Jordá et al. 1999; Tian et al. 2004). We found that the majority of *St*SBT1 and *St*SBT4 and *St*SBT5 subtilases were predicted to be in the extracellular space, while most *St*SBT3 subtilases were predicted to be localized in organelles. Coincidently with members of *St*SBT4, P69B and P69C tomato subtilases are secreted to the appoplast where they interact with *P. infestans* effectors (EPI1 and EPI10) (Tian et al. 2004, 2005).

Subtilases present in vacuolar membrane were proposed to have a degradative function that could also be related to protein turnover or defense (Rautengarten et al. 2005) while subtilases directed to mitochondria and chloroplasts suggest they could play a specific role in these organelles.

Expression Profile

Expression of potato subtilases was analyzed on different plant organs and biotic stress conditions in the light of the doubled monoploid *Solanum tuberosum* group Phureja RNA-Seq expression data. From 74 subtilases genes, 10 genes did not show detectable expression in neither 32 DM nor 16 RH expression libraries (http://solanaceae.plantbiology.msu.edu/). Although the tissues and conditions of theses libraries are very comprehensive, it can occur that these genes are expressed under different conditions, phenological stages or tissues considered, or a combination of them. Most probably, this result can be explained by the occurrence of pseudogenes. These genes were omitted from the expression analyses depicted in Fig. 4.

Potato subtilases are generally expressed in most plant organs but only 22 % of potato subtilases are ubiquitous (Fig. 4). These are preferentially subtilases from group *St*SBT3 and could be related to general protein turnover functions (Rautengarten et al. 2005; Golldack et al. 2003). Most subtilases, especially those from groups *St*SBT1 and *St*SBT4, evidenced tissue specific expression patterns, suggesting that they could have a more restricted function such as plant-specific proprotein convertases (Fig. 4). Major expression levels of subtilases were found first in immature fruit, and second in leaves, flowers and stems. In the case of group *St*SBT4, we found major leaf expression levels of PGSC0003DMG400003939 and PGSC0003DMG 400003913, homologs to p69C and p69A tomato subtilases. Also we found that PGSC0003DMG400003938 and PGSC0003DMG401003937, homologs to P69B, as well as genes related with this, PGSC0003DMG400034790 and PGSC0003DMG400045235 are expressed mainly in leaves including petioles (data not shown) (Fig. 4).

We are specifically interested in the expression profile upon biotic stress as subtilases have been related to pathogen defense (Jordá et al. 1999; Tian et al. 2004; Tornero et al. 1997; Fernández et al. 2012). The analyses of the expression profile of detached potato leaves after *P. infestans* infection or after BABA or BTH treatment showed 39 subtilases genes with changes in their expression compared with the control. From these, twenty genes were up regulated after *P. infestans* infection and were located on chromosome I, II, III, IV, VII, VIII and X (ESM 6: Table 3). Most of the up regulated subtilase genes were included in groups *St*SBT1, *St*SBT3 and *St*SBT4. Moreover, PGSC003\ DMP400018521 subtilase coded by PGSC003D\ MG400010470 (from *St*SBT1) have been found as a major protein in apoplast of detached potato leaves after *P. infestans* infection (Fernández et al. 2012). This subtilase evidenced DEVDasa activity and was related to programmed cell death functions (Fernández et al. 2014, 2015). BABA treatment have shown a protector effect on potato cultivars exposed to pathogen infection by inducing genes related with defense (PRs) (Jakab et al. 2001; Altamiranda et al. 2008). We observed the up regulation of twelve potato subtilases in BABA treatments which could be involved in induced resistance (IR). Thus, it can be explored if the effect of induced response triggered by BABA on potato cultivars could be also mediated by these subtilases genes. Figure 4 also showed that ten subtilases are up regulated after BTH treatment -an analog to salicylic acid- suggesting that some subtilases may be involved also in systemic acquired resistance and/or basal defense against pathogens (Halim et al. 2007).

*St*SBT Genome Localization and Meta-QTLs of Late Blight Resistance Relationship

Additional evidence of potato subtilases gene candidates related with *P. infestans* defense was searched by analyzing their co-localization with metaQTLs of late blight resistance (Danan et al. 2011). Thus, we constructed an integrated potato map that links a potato genetic map with a potato physical map based on sequenced molecular markers described by Sharma et al. (2013) and included in the consensus potato map of Danan et al. (2011) (Fig. 5). Localization and chromosome positions of these molecular markers closest to the 74 *St*SBT genes are indicated in ESM 7: Table 3.

We found 29 *St*SBT genes included in six out of twenty four metaQTLs of late blight resistance described by Danan et al. (2011) (Fig. 5). Fourteen of these genes were up regulated by *P. infestans* infections (Figs. 4 and 5) which includes: Nine genes from *St*SBT4 that co-localized with metaQTL2 of late blight resistance in chromosome VIII (abbreviated MQTL_2_late_blight_8); three genes from *St*SBT1 in MQTL_1_late_blight_1; one gene from *St*SBT1 in MQTL_1_late_blight_10 and one gene from *St*SBT3 in

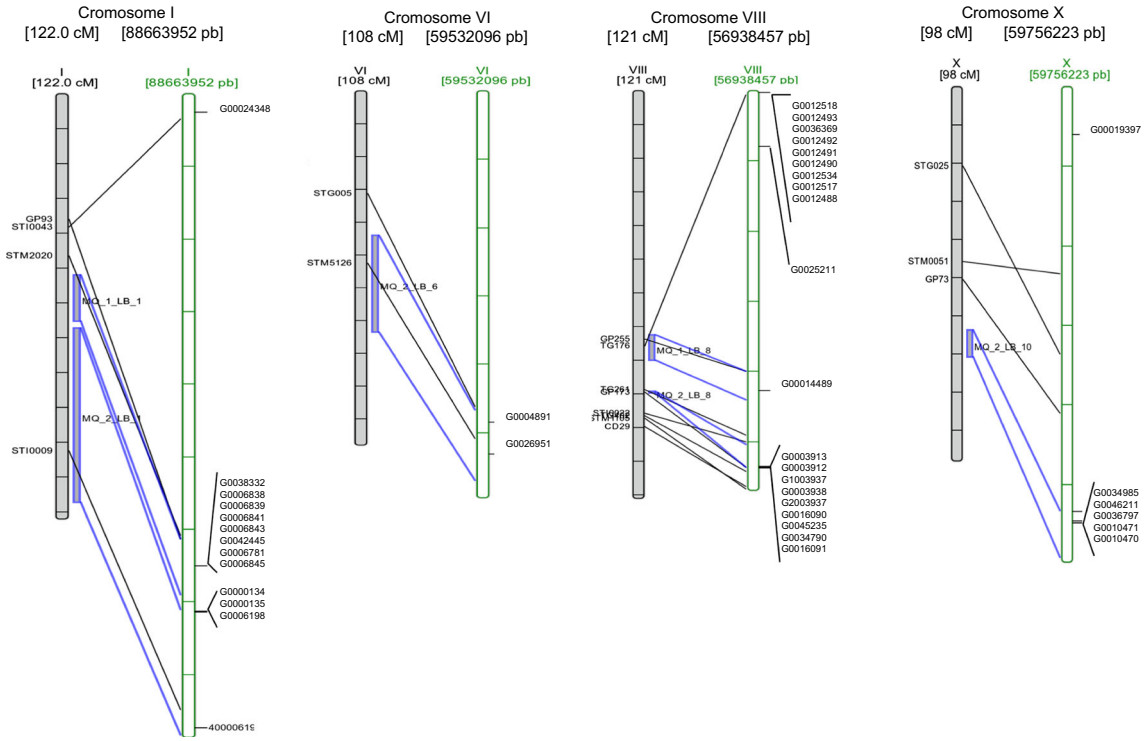

Fig. 5 Co-localization of *St*SBT genes with metaQTL of resistance to *Phythophthora infestans*. The figure shows and integrates a genetic map (in cM) and a physical map (in bp) based on the potato metaQTL described by Danan et al. (2011) and markers from the consensus map (https://solgenomics.net) and Sharma et al. (2013)

MQTL_1_late_blight_6 (Fig. 5). The co-localization described above remains to be further experimentally examined. Preliminary results in our laboratory with resistant and susceptible potato cultivars plants infected with *P. infestans* evidence an induction of gene expression of homologs to p69b genes in the resistant cultivar (data not shown). The significance of these will be validated with functional genomics studies.

Conclusions

This work presents the first identification and characterization of potato subtilisin-like serine protease family, comprising 74 genes and 82 putative *St*SBT proteins (S8A, S08.001, MEROPS). Potato subtilases constitute a multigenic family coded by 47 intronless genes of 74 studied. Many of these genes were clustered in tandem arrays in the potato genome suggesting gene expansion, a mechanism reported in the evolution of defense gene families. Phylogenetic analyses show that potato subtilases can be classified in five groups (*St*SBT1 to *St*SBT5) based on amino acid similarity rather than gene intron-exon or protein domain structures.

Potato subtilases phylogeny, gene annotation, expression pattern and subcellular localization analyses relate subtilases with diverse functions. In particular, we found fourteen subtilases genes related with pathogen defense as they are up regulated during *P. infestans* infection and co-localize with 6 metaQTLs related with *P. infestans* resistance. These subtilases genes are present mainly in *St*SBT1 and *St*SBT4 groups that include tomato homologues to p69b and p69c genes and *Le*SBTs related with pathogen defense as well as a potato subtilase reported to be related also with programmed cell death. The fact that 60 % of subtilases were predicted to be located in the extracellular space reinforce its putative role as part of first barriers on pathogen defense.

The findings reported here set the bases for future studies of candidate genes to be analyzed on expression studies as well as in reverse genetics experiments such as gene editing or gene silencing experiments aimed to determine the functions of potato subtilases and their specific role of subtilases on plant-pathogen interactions.

Acknowledgments This paper is submitted in partial fulfillment of Natalia Norero Doctoral degree at Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata, Argentina. We thanks Gabriela Massa for fruitful discussions while working in the paper. This work was financed by the Instituto Nacional de Tecnología Agropecuaria (INTA), Ministerio de Tecnología e Innovación Productiva, Agencia Nacional de Promoción Científica y Tecnológica, PICT-2010-2037 and NN doctoral fellowship from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

References

Altamiranda, E.A.G., A.B. Andreu, G.R. Daleo, and F.P. Olivieri. 2008. Effect of β-aminobutyric acid (BABA) on protection against *Phytophthora infestans* throughout the potato crop cycle. *Australasian Plant Pathology* 37(4): 421–427.

Antão, C.M., and F.X. Malcata. 2005. Plant serine proteases: biochemical, physiological and molecular features. *Plant Physiology and Biochemestry* 4: 637–50.

Arya, P., G. Kumar, and A.K. Acharya. 2014. Genome-wide identification and expression analysis of NBS-encoding genes in *Malus x domestica* and expansion of NBS genes family in Rosaceae. *PloS One* 9(9), e107987. doi:10.1371/journal.pone.0107987.

Beers, E., A.M. Jones, and A.W. Dickerman. 2004. The S8 serine, CIA cysteine and AI aspartic protease families in Arabidopsis. *Phytochemistry* 65: 43–58.

Berger, D., and T. Altmann. 2000. A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana*. *Genes Development* 14: 1119–1131.

Cao, J., X. Han, T. Zhang, Y. Yang, J. Huang, and X. Hu. 2014. Genome-wide and molecular evolution analysis of the subtilase gene family in *Vitis vinifera*. *BMC Genomics* 15: 1116.

Coffeen, W.C., and T.J. Wolpert. 2004. Purification and characterization of seríne proteases that exhibit caspase-like activity and are associated with programmed cell death in *Avena sativa*. *The Plant Cell* 16: 857–873.

Danan, S., J.B. Veyrieras, and V. Lefebvre. 2011. Construction of a potato consensus map and QTL meta-analysis offer new insights into the genetic architecture of late blight resistance and plant maturity traits. *BMC Plant Biology* 11: 16.

Dixon, R.A., and C.J. Lamb. 1990. Molecular communication in interactions between plants and microbial pathogens. *Annual Review of Plant Physiology and Plant Molecular Biology* 41: 339–367.

Dong, S., R. Stam, L.M. Cano, J. Song, J. Sklenar, K. Yoshida, T.O. Bozkurt, R. Oliva, Z. Liu, M. Tian, J. Win, M.J. Banfield, A.M.E. Jones, R.A.L. van der Hoorn, and S. Kamoun. 2014. Effector specialization in a lineage of the Irish potato famine pathogen. *Science* 343: 552–555.

Eddy, S. 1998. Profile hidden markov models. *Bioinformatics* 14(9): 755–763.

Emanuelsson, O., N. Nielsen, S. Brunak, and G. von Heijne. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300: 1005–1016.

Emanuelsson, O., S. Brunak, G. von Heijne, and H. Nielsen. 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. *Nature Protocols* 2: 953–971.

Fernández, M.B., G.R. Daleo, and M.G. Guevara. 2012. DEVDase activity is induced in potato leaves during *Phytophtthora inestans* infection. *Plant Physiology and Biochemestry* 61: 197–203.

Fernández, M.B., G.R. Daleo, and M.G. Guevara. 2015. Isolation and characterization of a *Solanum tuberosum* subtilisin-like protein with caspase-3 activity (*St*SBTc-3). *Plant Physiology and Biochemistry* 86: 137–146.

Golldack, D., P. Vera, and K.J. Dietz. 2003. Expression of subtilisin-like serine proteases in *Arabidopsis thaliana* is cell-specific and responds to jasmonic acid and heavy metals with developmental differences. *Physiologia Plantarum* 118(1): 64–73.

Haas, B.J., S. Kamoun, M.C. Zody, R.H. Jiang, R.E. Handsaker, L.M. Cano, M. Grabherr, C.D. Kodira, S. Raffaele, T. Torto-Alalibo, T.O. Bozkurt, A.M. Ah-Fong, L. Alvarado, V.L. Anderson, M.R. Armstrong, A. Avrova, L. Baxter, J. Beynon, P.C. Boevink, S.R. Bollmann, J.I. Bos, V. Bulone, G. Cai, C. Cakir, J.C. Carrington, M. Chawner, L. Conti, S. Costanzo, R. Ewan, N. Fahlgren, M.A. Fischbach, J. Fugelstad, E.M. Gilroy, S. Gnerre, P.J. Green, L.J. Grenville-Briggs, J. Griffith, N.J. Grünwald, K. Horn, N.R. Horner, C.H. Hu, E. Huitema, D.H. Jeong, A.M. Jones, J.D. Jones, R.W. Jones, E.K. Karlsson, S.G. Kunjeti, K. Lamour, Z. Liu, L. Ma, D. Maclean, M.C. Chibucos, H. McDonald, J. McWalters, H.J. Meijer, W. Morgan, P.F. Morris, C.A. Munro, K. O'Neill, M. Ospina-Giraldo, A. Pinzón, L. Pritchard, B. Ramsahoye, Q. Ren, S. Restrepo, S. Roy, A. Sadanandom, A. Savidor, S. Schornack, D.C. Schwartz, U.D. Schumann, B. Schwessinger, L. Seyer, T. Sharpe, C. Silvar, J. Song, D.J. Studholme, S. Sykes, M. Thines, P.J. van de Vondervoort, V. Phuntumart, S. Wawra, R. Weide, J. Win, C. Young, S. Zhou, W. Fry, B.C. Meyers, P. van West, J. Ristaino, F. Govers, P.R. Birch, S.C. Whisson, H.S. Judelson, and C. Nusbaum. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461(7262): 393–8.

Halim, V.A., L. Eschen-Lippold, S. Altmann, M. Birschwilks, D. Scheel, and S. Rosahl. 2007. Salicylic acid is important for basal defense of *Solanum tuberosum* against *Phytophthora infestans*. *Molecular Plant Microbe Interactions* 20(11): 1346–1352.

Jain, M., P. Khurana, A.K. Tyagi, and J.P. Khurana. 2008. Genome-wide analysis of intronless genes in rice and Arabidopsis. *Functional Integrative Genomics*. doi:10.1007/s10142-007-0052-9.

Jakab, G., V. Cottier, V. Toquin, G. Rigoli, L. Zimmerli, J.P. Metraux, and B. Mauch-Mani. 2001. β-Aminobutyric acid-induced resistance in plants. *European Journal of Plant Pathology* 107(1): 29–37.

Jones, P., D. Binns, H. Chang, M. Fraser, W. Li, C. McAnulla, M.W. Hamish, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A.F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.W. Yong, R. Lopez, and S. Hunter. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. doi:10.1093/bioinformatics/btu031.

Jordá, L., A. Coego, V. Conejero, and P. Vera. 1999. A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. *Journal of Biological Chemistry* 274(4): 2360–2365.

Jordá, L., V. Conejero, and P. Vera. 2000. Characterization of P69E and P69F, two differentially regulated genes encoding new members of the subtilisin-like proteinase family from tomato plants. *Plant Physiology* 122(1): 67–74.

Käll, L., A. Krogh, and E.L.L. Sonnhammer. 2007. Advantages of combined transmembrane topology and signal peptide prediction. The Phobius web server. *Nucleic Acids Research* 35: W429–32.

Letunic, I., and P. Bork. 2006. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1): 127–8.

Letunic, I., and P. Bork. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* 39: W475–8.

Luo, X., and K. Hofmann. 2001. The protease associated domain: a homology domain associated with multiple classes of proteases. *Trends Biochememical Science* 26: 147–148.

Mahon, P., and A. Bateman. 2000. The PA domain: a protease-associated domain. *Protein Science* 9: 1930–1934.

Mazin PV, Gelfand MS, Mironov AA, Rakhmaninova AB, Rubinov AR, Russell RB, Kalinina OV. 2010. An automated stochastic approach to the identification of the protein specificity determinants and functional subfamilies. *Algorithms for Molecular Biology*:5–29.

Meichtry, J., N. Amrhein, and A. Schaller. 1999. Characterization of the subtilase gene family in tomato (*Lycopersicon esculentum* Mill). *Plant Molecular Biology* 39(4): 749–760.

Muszewska, A., J.W. Taylor, P. Szczesny, and M. Grynberg. 2011. Independent subtilases expansions in fungi associated with animals. *Molecular Biology Evolution* 28(12): 3395–3404.

Neuteboom, L.W., J.M. Ng, M. Kuyper, O.R. Clijdesdale, P.J. Hooykaas, and B.J. van der Zaal. 1999. Isolation and characterization of cDNA clones corresponding with mRNAs that accumulate during auxin-induced lateral root formation. *Plant Molecular Biology* 39(2): 273–87.

Nowicki M, Foolad MJ, Nowakowska M, Kozik EU (2012). Potato and tomato late blight caused by *Phytophthora infestans*: an overview of pathology and resistance breeding. *Plant Disease* 96(1).

Page, R.D.M., and E.C. Holmes. 1998. *Molecular evolution: a phylogenetic approach. blackwell science*, 1st ed. Oxford: Utopia Press.

Potato Genome Sequence Consortium (PGSC). 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195.

Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R.R. Eddy, A.A. Bateman, and R.D. Finn. 2012. The Pfam protein families database: nucleic acids research. *Database Issue* 40: 290–301.

Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R.R. Eddy, A.A. Bateman, and R.D. Finn. 2014. The Pfam protein families database: nucleic acids research. *Database Issue* 42: 222–230.

Raffaele, S., R.A.L. Farrer, M.L. Cano, D.J. Studholme, D. MacLean, M. Thines, R.H.Y. Jiang, M.C. Zody, S.G. Kunjeti, M.N. Donofrio, B.C. Meyers, C. Nusbaum, and S. Kamoun. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330: 1540–1543.

Ramirez, V., A. López, B. Mauch-Mani, M.J. Gil, and P. Vera. 2013. An extracellular subtilases switch for immune priming in *Arabidopsis. PLoS Pathology* 9(6), e1003445. doi:10.1371/journal.ppat.1003445.

Rautengarten, C., D. Steinhauser, D. Büssis, A. Stintzi, A. Schaller, J. Kopka, and T. Altmann. 2005. Inferring hypotheses on functional relationships of genes: analysis of the *Arabidopsis thaliana* subtilase gene family. *PLoS Computational Biology* 1: e40. doi:10.1371/journal.pcbi.0010040.

Rawlings, N., F. Morton, and A. Barrett. 2006. MEROPS: the peptidase database. *Nucleic Acids Research* 34: D270–D272. doi:10.1093/nar/gkj089.

Ribeiro, A., A.D. Akkermans, A. van Kammen, T. Bisseling, and K. Pawlowski. 1995. A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of actinorhizal nodule development. *Plant Cell* 7(6): 785–94.

Riggs, C.D., K. Zeman, R. Deguzman, A. Rzepczyk, and A.A. Taylor. 2001. Antisense inhibition of a tomato meiotic proteinase suggests functional redundancy of proteinases during microsporogenesis. *Genome* 44: 644–650.

Sakharkar, M.K., and P. Kangueane. 2004. Genome SEGE: a database for 'intronless' genes in eukaryotic genomes. *BMC Bioinformatics* 5: 67.

Schaller, A., A. Stintzi, and L. Graff. 2012. Subtilases – versatile tools for protein turnover, plant development, and interactions with the environment. *Physiologia Plantarum* 145: 52–66.

Schuster-Böckler, B., J. Schultz, and S. Rahmann. 2004. HMM Logos for visualization of protein families. *BMC Bioinformatics* 5: 7.

Sharma, S.K., D. Bolser, J. de Boer, M. Sonderkær, W. Amoros, M.F. Carboni, J.M. D'Ambrosio, G. de la Cruz, A. Di Genova, D.S. Douches, M. Eguiluz, X. Guo, F. Guzman, C.A. Hackett, J.P. Hamilton, G. Li, Y. Li, R. Lozano, A. Maass, D. Marshall, D. Martinez, K. McLean, N. Mejía, L. Milne, S. Munive, I. Nagy, O. Ponce, M. Ramirez, R. Simon, S.J. Thomson, Y. Torres, R. Waugh, Z. Zhang, S. Huang, R.G.F. Visser, C.W.B. Bachem, B. Sagredo, S.E. Feingold, G. Orjeda, R.E. Veilleux, M. Bonierbale, J.M.E. Jacobs, D. Milbourne, D.M.A. Martin, and G.J. Bryan. 2013. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3: Genes, Genomes, Genetics* 3: 2031–2047.

Siezen, R.J. 1996. Subtilases: subtilisin-like serine protease. *Advances in Experimental Medicine and Biology* 379: 75–93.

Siezen, R.J., and J.A. Leunissen. 1997. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Science* 6(3): 501–23.

Siezen, R.J., B. Renckens, and J. Boekhorst. 2007. Evolution of prokaryotic subtilases: genome-wide analysis reveals novel subfamilies with different catalytic residues. *Proteins* 67(3): 681–994.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

Tanaka, H., H. Onouchi, M. Kondo, I. Hara-Nishimura, M. Nishimura, C. Machida, and Y. Machida. 2001. A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development* 128(23): 4681–9.

Tian, M., E. Huitema, L. Cunha, T. Torto-Alalibo, and S. Kamoun. 2004. A Kazal-like extracellular seríne protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. *The Journal of Biology Chemistry* 279: 26370–26377.

Tian, M., B. Benedetti, and S. Kamoun. 2005. A second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the appoplastic pathogenesis-related protease P69B of tomato. *Plant Physiology* 138: 1785–1793.

Tornero, P., V. Conejero, and P. Vera. 1996. Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: Similarity of functional domains to subtilisin-like endoproteases. *Proceedings National Academic of Science. Plant Biology. EEUU* 93: 6332–6337.

Tornero, P., V. Conejero, and P. Vera. 1997. Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants. *Journal of Biologycal Chemistry* 272(22): 14412–9.

Tripathi, L.P., and R. Sowdhamini. 2006. Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genomics* 7: 200.

Vartapetian, A.B., A.I. Tuzhikov, N.V. Chichkova, M. Taliansky, and T.J. Wolpert. 2011. A plant alternative to animal caspases: subtilisin-like proteases. *Cell Death Differentiation* 18: 1289–1297.

Von Groll, U., D. Berger, and T. Altmann. 2002. The subtilisin-like serine protease SDD1 mediates cell-to-cell signaling during Arabidopsis stomatal development. *Plant Cell* 14(7): 1527–39.

Zhang, R., F. Murat, C. Pont, T. Langin, and J. Salse. 2014. Paleo-evolutionary plasticity of plant disease resistance genes. *BMC Genomics* 15: 187. doi:10.1186/1471-2164-15-187.

Zhao, C., B.J. Johnson, B. Kositsup, and E.P. Beers. 2000. Exploiting secondary growth in Arabidopsis. *Plant Physiology* 123(3): 1185–1196.