Genome-wide Study of SBT Genes in Eight Plant Species: Evolution and Expression Outlines for Development and Stress

Sulin Wen

Shanghai Jiao Tong University <https://orcid.org/0000-0002-5994-3439>

Guanqun Chen

Shanghai Jiao Tong University

Tingting Huang

Shanghai Jiao Tong University

Xiaohui Shen (✉ ShenXH62@sjtu.edu.cn)

School of Design, Shanghai Jiao Tong University, Shanghai 200240, China.

Research article

Keywords: Subtilisin-like proteases, Genome-wide study, Evolution, Expression pattern, Arabidopsis, Rice

Posted Date: October 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-77460/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Subtilisin-like proteases (or subtilases, SBT), a specific family of serine peptidases, are a very big family in plants. Previous studies showed association of SBTs with environment response and development. However, it is still unclear about the function and features of SBTs. Also, there is little information about SBTs classification or protein function research.

Results: The characters of SBTs in eight plant species including *Arabidopsis*, rice, wheat etc. were analyzed. The analysis mainly contained evolutionary relationship, gene structure, chromosomal location, gene synteny and expression pattern. In short, the evolutionary trend of SBTs in the eight plant species was revealed. The expression pattern showed a scattered distribution of SBTs in diverse plant organs.

Conclusions: Our results reveal classification, feature and expression pattern of SBT. Then the results might offer clues for future studies about SBTs and additionally related to their possible function in plants against stress and development.

Background

Protease is a kind of important enzyme in most living things, which breakdown of proteins into smaller polypeptides or single amino acids. Proteases are involved in many biological functions, including digestion of ingested proteins, protein catabolism, and cell signaling (King et al 2014, Shen et al 2006). Subtilisin-like protease (SBT), which was first observed from *Bacillus subtilis* in 1947 (Lindstrom-Lang & Ottesen 1947), had been found in archaea, bacteria, eukaryote, fungi and yeast with its features like sequence, classification, function for the past decades (Coffeen & Wolpert 2004, Rawlings & Barrett 2000, Yamagata et al 1994), it was indicated an important protease in plants as well (Siezen & Leunissen 1997). The first plant SBT gene family was discovered in *Lycopersicon esculentum* (tomato) and was thought of 15 members (Meichtry et al 1999). Since then, in *Arabidopsis thaliana*, 56 known genes of this family was found (Rautengarten et al 2005), and 63 genes of this family in *Oryza sativa* (rice) was also indicated (Tripathi & Sowdhamini 2006). Besides, in the years that followed, the SBT family genes were explored in *Physcomitrella patens* (23 gene members), *Populus trichocarpa* (90 gene members), *Vitis vinifera* (82 gene members) and *Solanum tuberosum* (82 gene members), respectively (Figueiredo et al 2016, Schaller et al 2012, Sigrid Norero et al 2016).

The function of SBT showed widely spread in plants. It was reported involved in all aspects of plant life cycle, such as development of seeds and fruits, cell wall modification, processing of peptide growth factors and epidermal development and make response to biotic and abiotic stress (Schaller et al 2012). With abiotic stimulus, SBT showed be involved in stress like drought and salt resistance mechanisms (Budic et al 2013, Che et al 2010, Liu & Howell 2010a, Liu & Howell 2010b, Liu et al 2007). A representative example is that two SBTs in common bean (*Phaseolus vulgaris*), *PvSLP1* and *PvSLP2*, were responsible to drought (Budic et al 2013).

The SBT is constructed with three main domains, ie. Peptidase S8 domain(S8), protease-associated (PA) domain and inhibitor I9 domain (Figueiredo et al 2018). The peptidase S8 domain, also known as catalytic domain of SBT, is present in all SBTs with a highly conserved catalytic triad composed by aspartate (Asp), histidine (His) and serine (Ser) residues (Dodson & Wlodawer 1998). The PA domain is another main representative domain for SBT, consisting of an 120 to 160 amino acids insertion between the His and Ser active site residues (Luo & Hofmann 2001). Additionally, for the I9 domain(I9), its main function is keeping the inactive state of the zymogen and preventing the access of the substrate to the active site (Bryan 2002). Interaction between the PA and the I9 domains result in the cleavage of the N-terminus and it will open a access of substrates to the catalytic site, promoting the protease activity as well (Bergeron et al 2000).

The SBTs were widespread in plants (Figueiredo et al 2018). Although some researches have revealed gene members, features and function in some plant species like tomato, rice, potato (Rautengarten et al 2005, Sigrid Norero et al 2016, Tripathi & Sowdhamini 2006), there are still many questions about this enzyme. For discovering SBTs in more detail, we performed bioinformatics tools to process a more complete analysis for SBT family in eight plant species, including *Arabidopsis thaliana* and 6 monocotsie. three main crops (*Oryza sativa*, *Zea mays* and *Triticum aestivum*) and three our interest species (*Asparagus officinalis*, *Dendrobium catenatum*, *Phalaenopsis equestris*) belonging to Asparagales order of flower plants and another woody ornamental plant with many existed studies, *Rosa chinensis*. We used HMMER search method to identify more SBT family members in these species, and additionally performed phylogenetic analysis to uncover the evolutionary relationship of SBT in these eight species. Besides, we also analyze gene structure, chromosomal distribution, orthologous and paralogous, duplication, and expression pattern at the genome level for SBT. This study will lay the foundation for exploring the evolutionary relationship and biological function of SBT family members in plants.

Results

Identification of SBTs in eight plant species

We used HMMER search (based on Peptidase S8 domain) to identified a total of 702 genes in the eight plant species, *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, *Triticum aestivum*, *Asparagus officinalis*, *Dendrobium catenatum*, *Phalaenopsis equestris* and *Rosa chinensis*. The detailed information about selected SBT genes is available in Additional file 1. And the identified numbers were 57 in *Arabidopsis thaliana*, 63 in *Oryza sativa*, 66 in *Zea mays*, 255 in *Triticum aestivum*, 75 in *Asparagus officinalis*, 71 in *Dendrobium catenatum*, 51 in *Phalaenopsis equestris* and 64 in *Rosa chinensis*, respectively. Notably, through our used method, the numbers of identified genes in *Arabidopsis* and rice are similar same to the past two researches, 56 in *Arabidopsis* (Rautengarten et al 2005) and 63 in rice (Tripathi & Sowdhamini 2006), respectively.

Multiple sequence alignment of SBTs in eight plant species

The total of 702 genes and protein sequence were compared and the multiple sequence alignment result was shown as Fig1. All the SBT protein sequences were listed in the Additional file 2. And the similarity comparison matrix was shown in Additional file 3, which was used to draw the alignment heatmap. Fig1, the alignment heatmap, showed that many SBTs shared their sequences with high similarity. However, there were still some sequences showed less similarity with others. This indicated some SBT genes could have specific function in life activity. In addition, we selected the Peptidase S8 domain from Arabidopsis and rice to compare sequence similarity (Fig S1). The result showed that the Peptidase S8 domain displayed high similarity in some sequence areas, but in some other areas, the similarity exposed very low. This signposted conserved feature of Peptidase S8 domain but it was also variable in Arabidopsis and rice.

Phylogenetic analysis and classification of SBT genes in eight plant species

Additionally, the phylogenetic analysis indicated that SBTs can be divided into 9 clades among the eight plants (Fig 2). The subtrees with gene names are shown in Fig. S2. Clade 1 mainly contains the SBTs from *Asparagus officinalis*, *Dendrobium catenatum*, *Phalaenopsis equestris*, these three species belonging to Asparagales. Clade 2 is mainly constructed with subfamilies of SBT in wheat and maize. Clade 3 contains less subfamily members of SBT with some genes in wheat, rice, maize, *Asparagus officinalis*, *Dendrobium catenatum*, *Phalaenopsis equestris* and Arabidopsis AtSBT 1.2. Clade 4 contains SBT genes from the 7 species including rice OsSub19 gene and Arabidopsis Acetyl-CoA Carboxylase 2 (ACC2) gene (This gene has a similar protein domain as peptidase S8 domain). Clade 5 is a bigger clade with more SBT genes mainly from monocotyledon, with Arabidopsis SBT6.1 as well. Clade 6 is constructed mainly with Arabidopsis and wheat SBT genes. In this Clade, it has Arabidopsis SBT genes like AtSBT3.5, AtSBT 3.11, AtSBT5.1 and wheat TraesCS7B02G015300. Clade 7 is formed mainly with Arabidopsis "AtSBT4" genes (like AtSBT 4.1 and AtSBT4.9), wheat and *Rosa chinensis* SBT genes. Clade 8 contains SBT genes from the 7 plants, and the Arabidopsis SBT gene AtSBT2.5 is noticeable. Clade 9 mainly contains SBT genes from wheat. We compared the SBTs' sequences from Arabidopsis and rice (Fig S2). The sequences showed the several conserved domains in the front and the latter part of SBTs. That indicated SBTs function were conserved in some way.

Motif and gene structure analysis of SBTs gene family in Arabidopsis and rice

Because of abundant research resources for Arabidopsis and rice (Two model plants), to analyze the gene sequence structure of SBT, we selected SBT genes in Arabidopsis and rice for further analysis. We used MEME software to analyze the gene structure and motif of SBTs genes of Arabidopsis and rice (Fig 3). 10 motifs were found, and the detailed motif information was exposed in Additional file 4. Most of SBTs have conserved motif region with motif 6, 3, 7, 4, 8, and located in the front part of SBTs. On the other hand, the motif 2, 5, 1, 9, 10 are alternatively arranged in the latter part of SBTs. Many Arabidopsis SBT genes shared similar motifs such as AT1G322940, AT4G10550 and AT4G10540 with motif 6, 3, 7, 4, 8, 1, 5. Likewise, in rice, some SBT genes also were very similar because of same motifs like

Os01g0795100, Os01g0794800 and Os01g0795000 with motif 6, 3, 7, 4, 8, 1, 5, which suggested they probably shared similar function in biological process.

Gene Structure analysis showed that SBTs in rice and Arabidopsis displayed different gene structure (Fig 3). Rice SBTs' gene structure totally showed less introns for many isoforms. For example, Os03g0430500, Os04g0558900, Os05g0435800, these three genes have no introns. However, some Arabidopsis SBT genes showed much more introns in the gene structure, like AT1G20160, AT4G10550 and AT1G32960. which indicated a different translation pattern existed in Arabidopsis and rice.

Chromosomal location of SBTs in Arabidopsis and rice

We further analyzed chromosomal location and collinearity of SBTs in Arabidopsis and rice. The results showed that 14, 4, 4, 17 and 18 SBT genes located in Arabidopsis chromosome 1, 2, 3, 4, 5, respectively (Fig 4). On the other hand, in rice, 11, 14, 7, 15, 2, 3, 2, 2, 4, 2, 1, 1 SBT genes located in chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, respectively. It is noticeable that many SBT genes showed "gene cluster arrangement" in the chromosome. For example, AT1G32940, AT1G32950, AT1G32960, AT1G32970, AT1G32980 are closely arranged in Arabidopsis chromosome1. And Os01g0794800, Os01g0795000, Os01g0795100, Os01g0795200, Os01g07954000 are closely arranged in rice chromosome 1, too. This suggested that these genes might had happened tandem duplication with similar sequences and biological functions.

Collinearity analysis of SBTs genes in Arabidopsis and rice

We additionally analyzed collinearity of SBTs in Arabidopsis and rice by MCscanX software (Fig 5). All of the genome collinearity genes were shown in Additional file 5 and marked with gray lines in fig 4. On the other hand, the analysis showed that there existed 4 and three SBT collinearities in Arabidopsis and rice, respectively. In Arabidopsis, AT1G32950/AT4G10520, AT2G19170/AT4G30020, AT3G46840/AT5G59090, AT4G20430/AT5G44530, these 4 collinearity pairs were found. On the other hand, in rice, Os01g0868800/Os05g0435800, Os02g0665300/Os04g0558900, Os03g0119300/Os10g0524600, these three collinearity pairs were then found. The detail gene information was showed in Additional file 5. These results indicated that these genes might replicated through segmental duplication process.

Synten analysis of SBT genes among Arabidopsis, rice, wheat, and maize

To further conclude the phylogenetic mechanisms of SBT genes, we created comparative syntenic maps of SBT genes among the 4 representative plant species, including Arabidopsis (dicot) and three monocots (wheat, rice and maize). Through our procedure, the total gene pairs of Arabidopsis/rice, Arabidopsis/wheat, Arabidopsis/maize, rice/wheat, rice/maize were 2959, 6740, 2088, 44263, 8873, respectively (Fig 6). It was obvious that the monocot species (rice, wheat and maize) showed more genome similarity because of more gene pairs. And the synteny draft of rice/wheat and rice/maize were more likely the results of chromosome replication (Fig S4). Additionally, the numbers of SBTs' orthologous pairs of Arabidopsis/rice, Arabidopsis/wheat and Arabidopsis/maize were 3, 10, 3,

respectively. Besides, the numbers of orthologous pairs of rice/wheat and rice/maize were 92 and 21, respectively. The more orthologous pairs existing in rice/wheat and rice/maize indicated that SBTs had processed more duplication procedures in monocot class.

SBTs genes expression status analysis in Arabidopsis with different organs and development stages

The expression patterns of SBT genes in the typical model plant, Arabidopsis, was explored through Arabidopsis eFP Browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi). As a result, different SBTs showed different expression form for different development stage and part of Arabidopsis. The detailed data were displayed in Additional file 6. Fig 7 showed that SBT genes displayed distributed expression pattern in Arabidopsis. For instance, AT5G58820, AT4G21630 and AT4G15040 mainly expressed in mature flowers and young siliques, indicating they might involve in these organs development. However, some SBT genes, like AT2G05920 and AT4G20430, showed high expression in hypocotyl, root and shoot, suggesting a possible diverse role in some biological process. Then again, some members also showed another expression pattern such as AT5G45640 showing high expression level in the most organs compared with other SBTs except flower. AT2G04160 and AT1g32960 exhibited high expression levels in the rosette part. In contrast, AT4G34980, AT5G51750, AT2G19170, AT5G59120, AT3G46850, AT5G59810, AT5G58830 and AT4G21323, these genes displayed less expression levels in the most organs except flowers, young siliques. These results indicated that different SBTs obviously showed different expression patterns for different organs and development stages.

SBT genes in Arabidopsis response to exogenous environments

Furthermore, to better understand the influence of stress conditions to SBTs, we downloaded the data of multiple exogenous conditions to SBTs expression pattern. As a result, these several stress conditions including salt, heat, osmotic stress, cold, oxidative, wound, drought were presented (Fig 8). The complete data were displayed in Additional file 7. Diverse SBT gene showed different responsible models to stress. Some genes, such as AT4G34980, AT5G51750 showed a continuous high expression level for all the conditions. Alternatively, some of genes showed changed expression in some certain conditions, like AT3G14087, AT2G39850, AT1G01990 in cold, drought, heat. And some others displayed comparatively low-grade expression level for all the conditions, such as AT1G32960, AT5G58840. These results signposted that SBTs exhibited diverse expression patterns in different conditions, and suggested they might perform different functions in altered life process.

Discussion

SBTs were a big family that first found in *Bacillus subtilis* (Lindstrom-Lang & Ottesen 1947) and still remains a lot of questions in the plant research. The main function and structure had been identified (Dodson & Wlodawer 1998, Figueiredo et al 2018, Luo & Hofmann 2001). A previous study had been processed a systematic phylogenetic analysis of SBTs in potato (Sigrid Norero et al 2016) but it still exists many questions about the evolution relationship of SBT family members. Here, for better understanding the features and property of SBTs, we did the genome-wide association analysis by identifying family

members of SBTs in the eight plants. And we also analyzed their evolution relationship, gene structure, Collinearity relationship, expression pattern, stress response. These results will further our consideration of SBTs' phylogenesis and features for the future research.

Protein sequence alignments and phylogenetic relationship of SBTs

We totally found 702 protein sequences from the eight kinds of plant species. And the alignment schematic diagram showed conserved Peptidase S8 domain, which is the representative feature of SBTs (Dodson & Wlodawer 1998, Figueiredo et al 2018). This domain is also used for our HMMER search. Thus, we found these 702 protein sequences (belonging to unique genes) and proceeded their phylogenetic analysis. The 9 clades were clustered from these monocot and dicot species. With relatively more clades, our cluster process might be more detailed because of more species were added in our research. Sigrid Norero et al. clustered SBT gene family members to 7 clades before in potato and Arabidopsis (Sigrid Norero et al 2016). We added the extra species (rice, wheat, maize, *Asparagus officinalis*, *Dendrobium catenatum*, *Phalaenopsis equestris* and *Rosa chinensis*) for further research and the extra 2 clades were found. We used the same maximum likelihood method to study the phylogenetic relationship of SBTs so this result should be receivable. 2 more clades should be added for SBTs phylogenetic clusters.

Gene structural diversity of SBTs

Except Peptidase S8 domain of SBTs, SBTs also have the other 2 representative domains, PA and I9 domain (Chichkova et al 2010, Figueiredo et al 2018, Luo & Hofmann 2001, Schaller et al 2012). We were interested about SBT genes' conservative property as a result we did the gene structure investigation by selecting the two characteristic model plants, Arabidopsis and rice. Accordingly, the motif examination revealed that many conserved motifs existed in SBTs of Arabidopsis and rice, indicating possibly preserved gene function of SBT gene family members. However, many SBTs showed different gene structure since many rice SBTs had little intron but Arabidopsis SBTs were discontinuous with exons and introns. This result revealed a different splicing pattern and translation process of SBTs between Arabidopsis and rice SBTs.

Gene expression pattern of SBTs

Gene expression pattern is a very important part to investigate the gene function and help molecular breeding. Therefore, we also explored the expression mode of SBTs mainly from Arabidopsis and rice. Past researches revealed SBTs' involvement in plant response to both abiotic and biotic stress (Figueiredo et al 2018). Subtilases were informed to be involved in drought and salt tolerance (Budic et al 2013, Che et al 2010, Liu & Howell 2010a, Liu & Howell 2010b, Liu et al 2007). On the past year, some studies have painted the involvement of subtilases in plant fighting to pathogens. For example, In cotton, *GbSBT1*, was identified, categorized to *Verticillium dahliae*-induced resistance (Duan et al 2016)

. *SISTB3* in tomato, were indicated involved in resistance against *Manduca sexta* larvae (Meyer et al 2016). On considered the possibly important function in SBTs against stress and development, we analyzed SBTs expression in Arabidopsis and rice with different development stages and stress conditions.

Diverse SBTs displayed altered expression level in diverse organs and development stages might expose their different role in these situations. The Arabidopsis SBT gene family members' expression pattern showed a discrete distribution form in such organs—which suggested SBTs could perform different functions in these organs, not specific for one certain part. The stress-related expression pattern analysis also exposed SBTs in Arabidopsis really showed a variegated expression mode. Some SBTs presented dramatically responded to exogenous condition like heat, cold and oxidative but some others only exhibited very less expression level or little response to the environment stimulus. These results could offer clues to explore the essential SBT members in SBT gene families.

Conclusions

In conclusion, we identified 702 protein sequences from SBT genes in eight plant species, and the evolutionary pattern, gene structure, chromosomal location, gene synteny, gene expression and transcripts expression in multiple conditions and development stages were comprehensively analyzed. The identification of these SBT genes will help to clarify the molecular genetic basis of SBTs function and provide functional gene resources for transgenic research in plants. So far, few genes representing this gene family have been identified in the most species. Therefore, the SBT gene family of plants was analyzed comprehensively and systematically for the first time. This study provides useful resources for further study on the structure and function of SBT gene in plants. In addition, our analysis showed comprehensive evolutionary relationship of SBTs between different plants and complicated expression pattern for Arabidopsis and rice in different biotic and abiotic conditions, which may indicate the crucial role of SBTs against stress. This study will also help us understand the possibly essential function of SBTs in plants development and stress tolerance to environments.

Methods

Identification of SBT genes in eight plant species

The protein sequences of the SBT family were from *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, *Triticum aestivum*, *Asparagus officinalis*, *Dendrobium catenatum*, *Phalaenopsis equestris* and *Rosa chinensis*. The source was from the EnsemblPlants database (Available online: <http://plants.ensembl.org/info/website/ftp/index.html>). The hidden Markov model (HMM) file of the Peptidase S8 domain (PF00082) was used from Pfam database (<http://pfam.sanger.ac.uk/>). HMMER 3.0 was used to search the SBT genes from the eight plants' genome database. The cutoff value was set to 0.01. All candidate genes were further examined to approve the existence of the Peptidase S8 core sequences through PFAM and SMART software.

Phylogenetic and protein motif analyses in eight plant species

Amino acid sequences encoded by 702 SBT genes from eight plant species were aligned by using ClustalW in MEGA-X (Kumar et al 2018). Phylogenetic analysis of these 702 amino acid sequences was carried by the neighbor-joining method through Fast-tree software (Price et al 2010). The conserved protein sequences of 702 proteins encoded by the SBT gene are identified by MEME software (available online: <http://meme-suite.org/tools/meme>) based on the full-length protein sequence of each putative member in the SBT gene family.

Gene structure of SBT genes in eight plant species

Gene Structure Display Server (GSDS, available online: <http://gsds.cbi.pku.edu.cn>) was used to analyze the exon intron structure in the coding sequence and the genomic sequence of each predicted SBT gene from the EnsemblPlants database (Hu et al 2015).

Chromosome distribution and gene replication in Arabidopsis and rice

All SBT genes were mapped on Arabidopsis and rice chromosomes through Circos (Kryzwiniski et al 2009) according to the physical location in Arabidopsis and rice genome database. Considering probable research value and abundant genomic resources for the model plants, Arabidopsis and rice, we used the Multiple Collinearity Scan toolkit (MCScanX) (Wang et al 2012) with default parameters to analyze gene duplication events in Arabidopsis and rice. The collinearity of SBT gene between the 4 species (Arabidopsis, rice, maize, wheat) was established by using two system mapping software (<https://github.com/CJ-Chen/TBtools>) (Liu et al 2017).

Analysis of SBT gene family expression patterns in Arabidopsis and rice

The expression pattern of SBT gene in organs of Arabidopsis was received from The Bio-Analytic Resource database (Toufighi et al 2005). Affymetrix microarray data of Arabidopsis SBTs was downloaded from ePlant (Waese et al 2017), including various abiotic stresses like heat, cold and salt. The RNA-seq raw data for diverse SBTs transcripts expression level analysis from Arabidopsis and rice was downloaded from The European Nucleotide Archive (ENA). We referred the research in Arabidopsis (Schlaen et al 2015) and rice (Buti et al 2018). Additionally, RSEM (<https://github.com/deweylab/RSEM>) was used to calculate expression level of transcripts (Li & Dewey 2011). This dataset contained three replicates, and the FPKM value was converted by log2. The expression heat map was drawing by MeV (Multiple Experiment Viewer) (Saeed et al 2003).

Declarations

Acknowledgments

We thank all participants in the laboratory for their beneficial discussions and practical assistance. We are appreciative to the editors and reviewers for their critical evaluation of the manuscript and for their

constructive suggestions for the enhancement of the manuscript.

Authors' contributions

WS designed the experiments; WS conducted experiments. WS analyzed the data; WS, HT and CG interpreted the data; WS and SX wrote the article. All authors have read and approved the manuscript.

Funding

The research was funded by The National Natural Science Foundation of China (31670694). Funds were used for the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and material

The datasets supporting the conclusions of this article are included within the article and its additional files. The genome sequences of related plant used for identifying the SBT genes in this study were located in the EnsemblPlants database (Available online: <http://plants.ensembl.org/info/website/ftp/index.html>). And the expression datasets were from Arabidopsis eFP Browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi).

Ethics approval and consent to participate

Not applicable. The authors declared that experimental research works on the plants described in this paper comply with institutional, national and international guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

References

- Bergeron F, Leduc R, Day R. 2000. Subtilase-like pro-protein convertases: from molecular specificity to therapeutic applications. *Journal Of Molecular Endocrinology* 24: 1-22
- Bryan PN. 2002. Prodomains and protein folding catalysis. *Chemical Reviews* 102: 4805-15

- Budic M, Sabotic J, Meglic V, Kos J, Kidric M. 2013. Characterization of two novel subtilases from common bean (*Phaseolus vulgaris* L.) and their responses to drought. *Plant Physiology And Biochemistry* 62: 79-87
- Buti M, Pasquariello M, Ronga D, Milc JA, Pecchioni N, et al. 2018. Transcriptome profiling of short-term response to chilling stress in tolerant and sensitive *Oryza sativa* ssp. Japonica seedlings. *Functional & Integrative Genomics* 18: 627-44
- Che P, Bussell JD, Zhou W, Estavillo GM, Pogson BJ, Smith SM. 2010. Signaling from the Endoplasmic Reticulum Activates Brassinosteroid Signaling and Promotes Acclimation to Stress in Arabidopsis. *Science Signaling* 3: 12
- Chichkova NV, Shaw J, Galiullina RA, Drury GE, Tuzhikov AI, et al. 2010. Phytaspase, a relocalisable cell death promoting plant protease with caspase specificity. *Embo Journal* 29: 1149-61
- Coffeen WC, Wolpert TJ. 2004. Purification and characterization of serine proteases that exhibit caspase-like activity and are associated with programmed cell death in *Avena sativa*. *The Plant cell* 16: 857-73
- Dodson G, Wlodawer A. 1998. Catalytic triads and their relatives. *Trends In Biochemical Sciences* 23: 347-52
- Duan X, Zhang Z, Wang J, Zuo K. 2016. Characterization of a Novel Cotton Subtilase Gene GbSBT1 in Response to Extracellular Stimulations and Its Role in Verticillium Resistance. *Plos One* 11: e0153988
- Figueiredo J, Costa GJ, Maia M, Paulo OS, Malho R, et al. 2016. Revisiting Vitis vinifera Subtilase Gene Family : A Possible Role in Grapevine Resistance against Plasmopara viticola. *Frontiers in plant science* 7: 1783
- Figueiredo J, Sousa Silva M, Figueiredo A. 2018. Subtilisin-like proteases in plant defence: the past, the present and beyond. *Molecular plant pathology* 19: 1017-28
- Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31: 1296-97
- King JV, Liang WG, Scherpelz KP, Schilling AB, Meredith SC, Tang WJ. 2014. Molecular basis of substrate recognition and degradation by human presequence protease. *Structure (London, England : 1993)* 22: 996-1007
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* 19: 1639-45
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology And Evolution* 35: 1547-49

- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* 12: 16
- Lindstrom-Lang K, Ottesen M. 1947. A New Protein from Ovalbumin. *Nature* 159: 807-08
- Liu C, Xie T, Chen C, Luan A, Long J, et al. 2017. Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (*Ananas comosus*). *BMC genomics* 18: 503
- Liu J-X, Howell SH. 2010a. bZIP28 and NF-Y Transcription Factors Are Activated by ER Stress and Assemble into a Transcriptional Complex to Regulate Stress Response Genes in Arabidopsis. *The Plant cell* 22: 782-96
- Liu J-X, Howell SH. 2010b. Endoplasmic Reticulum Protein Quality Control and Its Relationship to Environmental Stress Responses in Plants. *The Plant cell* 22: 2930-42
- Liu J-X, Srivastava R, Che P, Howell SH. 2007. Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant Journal* 51: 897-909
- Luo XM, Hofmann K. 2001. The protease-associated domain: a homology domain associated with multiple classes of proteases. *Trends in biochemical sciences* 26: 147-48
- Meichtry J, Amrhein N, Schaller A. 1999. Characterization of the subtilase gene family in tomato (*Lycopersicon esculentum* Mill.). *Plant Molecular Biology* 39: 749-60
- Meyer M, Huttenlocher F, Cedzich A, Procopio S, Stroeder J, et al. 2016. The subtilisin-like protease SBT3 contributes to insect resistance in tomato. *Journal of experimental botany* 67: 4325-38
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* 5: e9490
- Rautengarten C, Steinhauser D, Altmann T. 2005. Inferring hypotheses on functional relationships of genes: Analysis of the Arabidopsis thaliana subtilase gene family. *Plant Biology (Rockville)* 2005: 264-64
- Rawlings ND, Barrett AJ. 2000. MEROPS: the peptidase database. *Nucleic acids research* 28: 323-25
- Saeed AI, Sharov V, White J, Li J, Liang W, et al. 2003. TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34: 374-+
- Schaller A, Stintzi A, Graff L. 2012. Subtilases - versatile tools for protein turnover, plant development, and interactions with the environment. *Physiologia Plantarum* 145: 52-66
- Schlaen RG, Mancini E, Sanchez SE, Perez-Santángelo S, Rugnone ML, et al. 2015. The spliceosome assembly factor GEMIN2 attenuates the effects of temperature on alternative splicing and circadian rhythms. *Proceedings of the National Academy of Sciences of the United States of America* 112: 9382-87

- Shen Y, Joachimiak A, Rosner MR, Tang WJ. 2006. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature* 443: 870-4
- Siezen RJ, Leunissen JAM. 1997. Subtilases: The superfamily of subtilisin-like serine proteases. *Protein Science* 6: 501-23
- Sigrid Norero N, Alfredo Castellote M, de la Canal L, Enrique Feingold S. 2016. Genome-Wide Analyses of Subtilisin-Like Serine Proteases on *Solanum tuberosum*. *American Journal Of Potato Research* 93: 485-96
- Toufighi K, Brady SM, Austin R, Ly E, Provart NJ. 2005. The Botany Array Resource: e-Northern, Expression Angling, and Promoter analyses. *Plant Journal* 43: 153-63
- Tripathi LP, Sowdhamini R. 2006. Cross genome comparisons of serine proteases in *Arabidopsis* and rice. *BMC genomics* 7: 200
- Waese J, Fan J, Pasha A, Yu H, Fucile G, et al. 2017. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. *The Plant cell* 29: 1806-21
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40: e49
- Yamagata H, Masuzawa T, Nagaoka Y, Ohnishi T, Iwasaki T. 1994. Cucumisin, a serine protease from melon fruits, shares structural homology with subtilisin and is generated from a large precursor. *Journal Of Biological Chemistry* 269: 32725-31

Figures

Figure 1

The heatmap of 702 gene sequences similarity in eight plant species.

Figure 2

Phylogenetic relationship of all SBT genes from the eight plants. Maximum likelihood method was used to construct this phylogenetic tree. Different colors indicated different clades.

Figure 3

Phylogenetic relationships, gene structure and conserved protein motifs of SBT genes from Arabidopsis and rice. The motifs information is provided in Additional file 4. The length of protein can be assessed using the scale at bottom.

Figure 4

Chromosomal location of SBT genes in Arabidopsis and rice. Arabidopsis chromosomes were marked with the prefix “At” and rice chromosomes were marked with the prefix “Os”. The length of the chromosomes could be assessed using the scale at left.

Figure 5

Schematic of SBT genes synteny in Arabidopsis and rice. Gray lines specify all synteny blocks in the Arabidopsis and rice genome, and the green lines indicate SBT gene pairs which have synteny for each other. Chromosome numbers were marked at the side of the chromosomes.

Figure 6

Synteny analysis of SBT genes among Arabidopsis, rice, maize and wheat. Gray lines indicate the collinear blocks within these plant species genomes, while the red lines direct the syntenic SBT gene pairs. Chromosome numbers were marked at the side of the chromosomes.

Figure 7

The expression hierarchical clustering of SBT genes in Arabidopsis. The gene expression data was obtained from Arabidopsis eFP Browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi), and the expression data values were transformed to a heatmap as shown. The original data is showed in Additional file 6.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile4.html](#)
- [Additionalfile1.xlsx](#)
- [Additionalfile6.xlsx](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile8.xlsx](#)
- [Additionalfile7.xlsx](#)
- [Additionalfile3.xlsx](#)
- [FigS2.jpg](#)
- [FigS3.jpg](#)
- [FigS4.jpg](#)
- [FigS1.jpg](#)