Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genome-wide identification and analysis of subtilisin-like proteases in cassava (Manihot esculenta) reveals developmental and stress-responsive expression patterns

Huong Lan Thi Do
Hanoi Pedagogical University 2
Hong Viet La
Hanoi Pedagogical University 2
Quynh Thi Ngoc Le
Thuyloi University
Gioi Huy Dong
Vietnam National University of Agriculture
Bach Duc Nguyen
Vietnam National University of Agriculture
Ha Duc Chu
cd.ha@vnu.edu.vn

University of Engineering and Technology, Vietnam National University Hanoi

Research Article

Keywords: Manihot esculenta, subtilisin-like protease, gene family evolution, drought response, transcriptome analysis

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11086067/v1

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

Additional Declarations: No competing interests reported.

Genome-wide identification and analysis of subtilisin-like proteases in cassava (*Manihot esculenta*) reveals developmental and stress-responsive expression patterns

Huong Lan Thi Do[1] · Hong Viet La[2] · Quynh Thi Ngoc Le[3] · Gioi Huy Dong[4] · Bach Duc Nguyen[4] · Ha Duc Chu[5,*]

[1]Faculty of Biology, Hanoi Pedagogical University 2, Phu Tho Province 35000, Vietnam

[2]Institute of Scientific Research and Applications, Hanoi Pedagogical University 2, Phu Tho Province 35000, Vietnam

[3]Department of Biotechnology, Faculty of Chemistry and Environment, Thuyloi University, Hanoi City 100000, Vietnam

[4]Faculty of Biotechnology, Vietnam National University of Agriculture, Hanoi City 100000, Vietnam

[5]Faculty of Agricultural Technology, University of Engineering and Technology, Vietnam National University Hanoi, Hanoi City 100000, Vietnam

Corresponding author: Ha Duc Chu (Email address: cd.ha@vnu.edu.vn; Tel.: +84 983 766 070)

Statements and Declarations

Funding Declaration: This work is funded by the Foundation for Science and Technology Development, Hanoi Pedagogical University 2 via the grant number HPU2.2023-UT-02.

Competing Interests: The authors declare that they have no conflict of interest.

Ethical issues: None.

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work, the authors used Grammarly Premium and QuillBot Premium to edit grammar. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Abstract

Subtilisin-like proteases (SBTs) participate in peptide maturation, plant development, and stress signaling. Despite the agronomic importance of cassava, the composition, evolutionary history, organ expression, and drought responsiveness of its SBT family remain unresolved. A genome-wide search based on the Peptidase_S8 profile identified 75 cassava SBT genes. Protein properties, chromosomal distribution, exon-intron structure, phylogenetic relationships, and duplication patterns were examined. Public RNA-seq datasets from 11 organs and drought-treated leaves were subsequently re-analyzed. Eight candidates were further assessed by qRT-PCR in detached leaves after 0, 6, and 12 h of dehydration. The 75 *MeSBT* genes were distributed across 17 chromosomes and separated into six phylogenetic clades. Twenty genes were intronless, and 27 duplication events comprised 11 tandem, one segmental, and 15 whole-genome duplications. Twenty-five genes reached an FPKM value of at least 10 in one or more organs. *MeSBT21* showed high transcript abundance in seven organs, whereas *MeSBT62* and *MeSBT75* exceeded 100 FPKM in the petiole, stem, and storage root. Drought RNA-seq analysis classified 43 genes as upregulated, 21 as downregulated, and 11 as unchanged. Detached-leaf qRT-PCR detected concordant response directions for the induced genes *MeSBT29*, *MeSBT59*, *MeSBT21*, and *MeSBT62* and the repressed genes *MeSBT03*, *MeSBT75*, *MeSBT65*, and *MeSBT36*. This study addresses the cassava-specific knowledge gap in SBT family organization and identifies experimentally supported, drought-responsive candidates. These genes provide defined targets for whole-plant genetic and biochemical studies of cassava adaptation to water deficit.

Keywords: *Manihot esculenta* · subtilisin-like protease · gene family evolution · drought response · transcriptome analysis

Introduction

Cassava (*Manihot esculenta*), native to South America, has emerged as a fundamental staple food for millions worldwide, particularly in tropical and subtropical regions (Malik et al., 2020). Esteemed for its robustness and adaptability, cassava thrives in less fertile soils and can withstand variable climatic conditions, making it a crucial crop in challenging environments (Brown et al., 2016; Chavarriaga-Aguirre et al., 2016; Ray et al., 2019). It is a significant source of carbohydrates and provides essential nutrients such as vitamin C, calcium, and potassium, though its protein content is relatively low (Sugiharto et al., 2018). The cultivation of cassava is characterized by its minimal maintenance requirements, yet it is susceptible to various biotic and abiotic stresses (Tomlinson et al., 2018). The exploration of the molecular mechanisms governing cassava's responses to environmental stimuli is vital for enhancing its resilience to stress, improving yields, and ensuring food security (Shan et al., 2018). Moreover, understanding these mechanisms can facilitate the development of cassava varieties with improved nutritional profiles, further bolstering its role in addressing nutritional deficiencies in vulnerable populations.

Genome-wide gene-family surveys coupled with transcript profiling have become effective approaches for identifying genes associated with plant development and abiotic-stress responses. Conserved-domain searches, gene-structure analysis, phylogenetic classification, and expression profiling have defined stress-related gene families in lentil and pigeonpea (Shivadarshan et al., 2025; Shruti Sinha 2025). Genomic and transcriptomic integration has also identified pathway-associated genes in wheat, while comparative RNA-Seq has prioritized stress-responsive candidates across mungbean tissues (Kothari et al., 2026; Mona Yadav et al., 2025). Subtilisin-like proteases (SBTs), also termed subtilases, form a major class of plant serine proteases characterized by a catalytic triad of serine, histidine, and aspartate residues (Figueiredo et al., 2014; Figueiredo et al., 2018). These enzymes are defined by the peptidase S8 domain (Berger and Altmann 2000; Figueiredo et al., 2014), ensuring broad substrate specificity and important proteolytic functionality (Figueiredo et al., 2014). Their extensive involvement in plant physiology and development includes roles in senescence, stress response (Figueiredo et al., 2014; Figueiredo et al., 2018), growth, and developmental regulation (Ali et al., 2022), notably through the processing of hormone precursors to modulate environmental response mechanisms (Figueiredo et al., 2018; Pearce et al., 2010). Given their pivotal functions, SBTs have been subjects of significant interest, aimed at unlocking their potential for crop improvement. Up till now, the SBT families have been identified and characterized in various species, including melon fruits (*Cucumis melo*) (Yamagata et al., 1994), tomato (*Solanum lycopersicum*) (Tornero et al., 1996), Arabidopsis (*Arabidopsis thaliana*) (Rautengarten et al., 2005), grass endophytic fungus (*Epichloë festucae*) (Bryant et al., 2009), grape (*Vitis vinifera*) (Cao et al., 2014), potato (*Solanum tuberosum*) (Norero et al., 2016), rubber (*Hevea brasiliensis*) (Ekchaweng et al., 2017), pineapple (*Ananas comosus*) (Jin et al., 2021), rice (*Oryza sativa*) (Zheng et al., 2022), four cotton species (*Gossypium hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii*) (Dai et al., 2022) and wheat (*Triticum aestivum*) (Zhao et al., 2023). Although a cassava reference genome is available (Bredeson et al., 2016), no genome-wide study has defined the *MeSBT* repertoire, its evolutionary expansion, or its transcriptional patterns across cassava organs and drought-stressed leaves. This knowledge gap has limited the selection of *MeSBT* candidates for functional testing and drought-resilience breeding.

The objective of the current study was to comprehensively identify, annotate, and characterize the SBT family in cassava. Initially, potential members of the SBT family were screened and validated within the cassava genome assembly. Subsequently, the general characteristics of the identified proteins and genes were examined using web-based analytical tools. An unrooted phylogenetic tree of the SBT proteins was constructed to elucidate their evolutionary relationships and to infer duplication events within the cassava *SBT* gene family. Lastly, a re-analysis of an existing transcriptome database was conducted to assess the expression patterns of the *SBT* genes across different cassava tissues.

Materials and Methods

Identification of the SBT proteins in cassava

To screen the members of the SBT family within cassava genomes, comprehensive genomic and proteomic data for cassava (Bredeson et al., 2016) were acquired from the National Center for Biotechnology Information database. Utilizing the hidden Markov model profile corresponding to the conserved functional domain of SBT, namely peptidase_S8 (Pfam accession: PF00082) (Figueiredo et al., 2014) sourced from the Pfam database (Mistry et al., 2021), a screening process was undertaken against the cassava proteome (Bredeson et al., 2016) to isolate potential SBT members. For each identified member, the full-length protein sequences, genomic DNA sequences, and coding DNA sequences were extracted, laying the groundwork for further *in silico* analyses.

Analysis of parameters of the SBT proteins in cassava

Utilizing the full-length amino acid sequences of SBT proteins from cassava as seed sequences, a targeted search was conducted in the ExPASy ProtParam tool (Gasteiger et al., 2003), following previous guidelines (La 2022; La et al., 2022a; La et al., 2022b; Le et al., 2022; Tran et al., 2018). This search was specifically aimed at determining common characteristics of the SBT proteins, such as protein length, isoelectric point (pI), molecular weight (mW), aliphatic index (AI), and the grand average of hydropathicity (GRAVY).

Generation of the phylogenetic tree of the SBT proteins in cassava

Utilizing the full-length protein sequences of SBT proteins from cassava and well-characterized AtSBT proteins from Arabidopsis (Rautengarten et al., 2005), an unrooted phylogenetic tree was constructed following the previously described method (Gonzalez and Galli 2021). Particularly, the process began with the use of ClustalW software (Larkin et al., 2007) for the multiple sequence alignment of SBT proteins from cassava and well-characterized AtSBT proteins from Arabidopsis (Rautengarten et al., 2005) to enrich comparative analysis. The alignments were transferred to the Molecular Evolutionary Genetics Analysis software (Tamura et al., 2021) to construct the unrooted phylogenetic tree. The Maximum Likelihood estimation approach, adhering to default settings for model selection, was used as previously

described (La 2022; La et al., 2022a; La et al., 2022b; Le et al., 2022; Tran et al., 2018). Subsequent to the phylogenetic tree's generation, Adobe Illustrator software was applied for the purposes of enhancing the visualization of the tree.

Prediction of gene duplication of the SBT genes in cassava

The prediction of duplication events within the *SBT* gene family in cassava was based on the similarity level of each gene pair as previously detailed (La et al., 2022b). In particular, the *SBT* members that shared a similarity > 70% were hypothesized to constitute duplicated pairs (La et al., 2022b). A tandem duplication event was identified when the gene pairs were found in immediate proximity on the same chromosome (Panchy et al., 2016). In contrast, segmental duplication events were characterized by the duplication of DNA segments varying from 1 to 200 kb, also located within the same chromosome (Panchy et al., 2016). Furthermore, gene pairs stemming from whole genome duplication events were distinguished by their distribution across different chromosomes, reflecting a broader genomic rearrangement (Panchy et al., 2016).

Exon/intron structural analysis of the SBT genes in cassava

The exon-intron structure characteristics of genes encoding SBT proteins in cassava were analyzed following established protocols (La 2022; La et al., 2022a; La et al., 2022b; Le et al., 2022; Tran et al., 2018). For this analysis, both the genomic DNA sequence and the coding DNA sequence of each SBT-encoding gene in cassava were aligned using the Gene Structure Display Server (Hu et al., 2015). The arrangement of the SBT proteins in cassava, as derived from the phylogenetic tree, was subsequently employed to facilitate the visualization of their gene structures.

Transcriptome analysis of the SBT genes in cassava

The exploration of expression profiles for genes encoding SBT proteins in cassava utilized microarray databases, specifically those accessible via the Gene Expression Omnibus portal (Barrett et al., 2013). Particularly, expression patterns across a variety of 11 major organs and tissues, including leaf blade, leaf mid vein, petiole, stem, lateral bud, shoot apical meristem, storage root, fibrous root, root apical meristem, organized embryogenic structure, and friable embryogenic callus, were analyzed by querying the GSE82279 dataset (Wilson et al., 2017). Expression levels were quantified using the fragments per kilobase of transcript per million mapped reads metric (Wilson et al., 2017) as previously reported . Additionally, the GSE98537 dataset (Zhu et al., 2020) was re-examined to assess responses under drought stress. These heatmaps were systematically clustered utilizing an R script and illustrated by Adobe Illustrator software.

Plant materials and dehydration treatments

Cassava cultivar KM94 (international name: KU50) (Utsumi et al., 2022), obtained from the Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences (Hanoi City, Vietnam), was used as the plant material in this study. Healthy and uniform stem cuttings were propagated hydroponically in half-strength Hoagland nutrient solution for 40 days before dehydration treatment. During cultivation, the nutrient solution was renewed every 3 days. The plants were maintained under greenhouse conditions at 28 ± 2 ºC, with a 16 hours light/8 hours dark photoperiod, relative humidity of 65 - 75%, and a light intensity of 250 - 300 µmol m^{-2} s^{-1}. Fully expanded leaves were collected from plants of similar growth status and subjected to dehydration treatment under laboratory ambient conditions. Briefly, detached leaves were placed on dry sterile filter paper on the laboratory bench at 25 ± 2 ºC for 0 (control), 6, and 12 hours. At each time point, leaf samples were immediately frozen in liquid nitrogen and stored at -80 ºC for further steps.

RNA isolation and quantitative reverse transcription-PCR analysis

Total RNA was extracted from control and drought-treated cassava leaves using the RNAprep Pure Plant Plus Kit (Tiangen Biotech, China) according to the manufacturer's instructions. Residual genomic DNA was removed using gDNA Eraser (Takara Bio Inc., Japan). RNA concentration and purity were assessed with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) based on the A_{260}/A_{280} and A_{260}/A_{230} ratios. RNA integrity was examined by electrophoresis on a 1.0% (w/v) agarose gel. First-strand cDNA was synthesized from 1 µg of total RNA in a final volume of 20 µL using the PrimeScript RT Reagent Kit (Takara Bio Inc., Japan). Eight drought-responsive *MeSBT* genes with detectable leaf expression were selected for qRT-PCR validation. The selected genes comprised *MeSBT29* (Manes.05G175400), *MeSBT59* (Manes.13G059200), *MeSBT21* (Manes.04G074500), *MeSBT62* (Manes.14G157400), *MeSBT03* (Manes.01G070400), *MeSBT75* (Manes.18G044300), *MeSBT65* (Manes.16G094900), and *MeSBT36* (Manes.06G013400). Gene-specific primers were designed using Primer3 (Untergasser et al., 2012). Cassava UBQ10 (POLYUBIQUITIN 10, accession no. DV441403) was used as the internal reference gene under drought treatment (Hu et al., 2016). The primer sequences for UBQ10 were 5′-TGCATCTCGTTCTCCGATTG-3′ and 5′-GCGAAGATCAGTCGTTGTTGG-3′ for the forward and reverse primers, respectively. qRT-PCR was performed on a Stratagene Mx3000P Real-Time PCR System (Agilent Technologies, USA) using Thunderbird SYBR qPCR Mix (Toyobo, Japan) according to the manufacturer's protocol. The thermal program consisted of initial denaturation at 95 ºC for 30 s, followed by 40 cycles at 95 ºC for 5 s and 60 ºC for 30 s. Three independent biological replicates were analyzed, with three technical replicates for each biological sample. Relative transcript abundance was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). Differences between the control and drought treatments were evaluated using Student's *t*-test at $P < 0.05$. Genes were classified as upregulated when their relative expression increased by at least two-fold and as downregulated when their relative expression decreased to 0.50-fold or less of the control.

Results

Identification and annotation of the SBT family in cassava

The Peptidase_S8 hidden Markov model profile identified 75 putative SBT proteins in the cassava proteome. Their corresponding genes were designated *MeSBT01* to *MeSBT75*, and the locus identifiers are provided in Table 1. The *MeSBT* genes were distributed unevenly across 17 of the 18 cassava chromosomes

(Fig. 1). Chromosome 1 contained the largest number, with 13 genes, followed by chromosomes 5, 6, and 18 with eight, seven, and seven genes, respectively. Chromosome 7 contained only one *MeSBT* gene, while no family member was detected on chromosome 15. Each of the remaining represented chromosomes contained between two and six *MeSBT* genes.

Analysis of the general features of SBT family in cassava

The 75 MeSBT proteins varied in their physicochemical properties (Table 1). Protein length ranged from 506 amino acids in MeSBT31 to 1,424 amino acids in MeSBT39, and the predicted molecular weight ranged from 55.51 to 156.73 kDa. The theoretical isoelectric point ranged from 4.84 for MeSBT72 to 9.32 for MeSBT56. Forty proteins had basic pI values above 7.0, whereas 35 had acidic pI values below 7.0. The aliphatic index ranged from 75.13 in MeSBT47 to 92.91 in MeSBT69. Negative GRAVY values from -0.339 to -0.002 were obtained for 65 proteins, which were classified as hydrophilic. Positive GRAVY values were recorded for the other 10 proteins, namely MeSBT05, MeSBT06, MeSBT08, MeSBT10, MeSBT21, MeSBT22, MeSBT27, MeSBT42, MeSBT64, and MeSBT69, which were classified as hydrophobic.

Construction of the phylogenetic tree of the SBT family in cassava

The maximum-likelihood phylogenetic tree constructed from 75 MeSBT and 56 Arabidopsis AtSBT protein sequences classified the cassava proteins into six clades (Fig. 2). Clade I was the largest and contained 28 MeSBT proteins, followed by clades IV and V with 18 and 16 members, respectively. Clade II contained seven proteins, namely MeSBT06, MeSBT17, MeSBT18, MeSBT37, MeSBT62, MeSBT64, and MeSBT69. Clade III comprised MeSBT07, MeSBT19, MeSBT20, and MeSBT42, whereas clade VI included only MeSBT39 and MeSBT68.

Gene structure and gene duplication of the SBT family in cassava

The exon-intron structures varied among the 75 *MeSBT* genes (Fig. 3). Twenty genes were intronless, five contained two exons, and eight contained eight exons. Nineteen and 16 genes contained 10 and 11 exons, respectively. *MeSBT74*, *MeSBT36*, *MeSBT64*, *MeSBT15*, *MeSBT08*, *MeSBT44*, and *MeSBT39* contained 3, 4, 5, 7, 9, 12, and 35 exons, respectively. Genomic sequence length ranged from 2,265 bp for *MeSBT12* to 25,171 bp for *MeSBT39*. Duplication analysis identified 27 events within the *MeSBT* family. These comprised 11 tandem duplications, one segmental duplication, and 15 whole-genome duplications (Table 2).

Analysis of the *SBT* genes expression profiles during the growth and development of cassava

The FPKM-based expression profiles of the 75 *MeSBT* genes differed among 11 cassava organs and tissues (Fig. 4). Twenty-nine genes had no detectable transcripts in any sample. Forty-six genes were

expressed in at least one sample, of which 23 reached an FPKM value of at least 10. *MeSBT21* showed the most extensive expression profile, with high transcript abundance in all examined samples and the highest levels in fibrous root, lateral bud, midvein, petiole, shoot apical meristem, and stem. *MeSBT36* and *MeSBT67* also showed high expression across most samples. *MeSBT03*, *MeSBT23*, *MeSBT59*, *MeSBT62*, and *MeSBT75* were abundant in several vegetative and meristematic tissues. Tissue-preferential expression was detected for *MeSBT51* and *MeSBT70* in fibrous root, *MeSBT29* and *MeSBT65* in leaf and midvein, *MeSBT55* in organized embryogenic structures and root apical meristem, *MeSBT68* in root apical meristem, and *MeSBT74* in fibrous and storage roots. *MeSBT75* showed particularly high expression in petiole, stem, and storage root and had the highest transcript abundance in leaf and storage root.

Relative expression level of *MeSBT* genes in leaf samples under drought treatment

RNA-seq analysis revealed extensive changes in *MeSBT* transcript levels in cassava leaves under drought stress (Fig. 5). Genes with fold-change values of at least 2 were classified as upregulated, while genes with values of -2 or lower were classified as downregulated. Of the 75 *MeSBT* genes, 43 were upregulated, 21 were downregulated, and 11 were classified as unchanged. *MeSBT28* and *MeSBT29* showed the strongest induction, with fold-change values of 642.10, followed by *MeSBT26* and *MeSBT27* at 526.15, *MeSBT02* at 512.41, and *MeSBT06* at 507.45. High induction was also detected for *MeSBT51*, *MeSBT30*, *MeSBT31*, *MeSBT32*, *MeSBT10*, and *MeSBT15*. The greatest reductions were recorded for *MeSBT03*, *MeSBT04*, and *MeSBT05*, each with a fold-change value of -120.50, followed by *MeSBT66*, *MeSBT71*, and *MeSBT75*. The fold-change values of *MeSBT16*, *MeSBT17*, *MeSBT18*, *MeSBT19*, *MeSBT20*, *MeSBT25*, *MeSBT39*, *MeSBT40*, *MeSBT44*, *MeSBT53*, and *MeSBT61* remained between -2 and 2.

Eight *MeSBT* genes were selected to examine whether the expression directions identified by RNA-seq re-analysis under drought stress were reproduced during detached-leaf dehydration (Fig. 6). qRT-PCR showed progressive induction of *MeSBT29*, *MeSBT59*, *MeSBT21*, and *MeSBT62* across the 6- and 12-h treatments. Their relative transcript levels ranged from 2.50-fold to 84.60-fold at 6 h and from 6.20-fold to 342.00-fold at 12 h. *MeSBT29* showed the strongest induction at both time points, followed by *MeSBT59*. In contrast, *MeSBT03*, *MeSBT75*, *MeSBT65*, and *MeSBT36* showed progressive repression. Their relative transcript levels decreased to 0.041-0.428-fold at 6 h and 0.0011-0.115-fold at 12 h. *MeSBT03* showed the greatest reduction, while *MeSBT65* and *MeSBT36* showed more gradual decreases. The response directions of all eight genes agreed with the RNA-seq re-analysis, with four genes induced and four genes repressed in both analyses. The time-course profiles also showed stronger responses after prolonged dehydration, although the response magnitudes differed between the drought RNA-seq dataset and the detached-leaf dehydration experiment.

Discussion

The SBT gene family, initially discovered in melon fruits (Yamagata et al., 1994), has subsequently been identified across a diverse array of species, including tomato (Tornero et al., 1996), Arabidopsis (Rautengarten et al., 2005), grass endophytic fungus (Bryant et al., 2009), grape (Cao et al.,

2014), potato (Norero et al., 2016), rubber (Ekchaweng et al., 2017), pineapple (Jin et al., 2021), rice (Zheng et al., 2022), four cotton species (Dai et al., 2022) and wheat (Zhao et al., 2023). Specifically, investigations have reported 80 and 255 SBT family members in grape and wheat, respectively (Cao et al., 2014; Zhao et al., 2023). Further studies have identified 56 SBT proteins in Arabidopsis (Rautengarten et al., 2005) and 63 in rice (Zheng et al., 2022), 15 in tomato (Tornero et al., 1996), 74 in potato (Norero et al., 2016). In more extensive screenings, *G. hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii* have been found to possess 120, 112, 67 and 69 putative SBT proteins, respectively (Dai et al., 2022). These findings underscore the notion that SBT families constitute large and highly variable groups within plant species. Consequently, our study has identified 75 MeSBT proteins in cassava. This amount is larger than recorded for other higher plant species, including 15 in tomato (Tornero et al., 1996), 56 in Arabidopsis (Rautengarten et al., 2005), 63 in rice (Zheng et al., 2022), 67 in *G. arboreum* (Dai et al., 2022), 69 in *G. raimondii* (Dai et al., 2022) and 74 in potato (Norero et al., 2016). This relatively high number of members of the MeSBT family in cassava likely indicates the substantial physiological roles that the subtilase gene family fulfills in this species.

Previously, the common parameters of the SBT proteins in other species have been reported. For example, the size of the TaSBT proteins in wheat exhibited great variation, with lengths ranging from 220 residues (TaSBT253) to 1491 residues (TaSBT235). This variation in length corresponded to a mW range from 24.22 kDa to 163.20 kDa (Zhao et al., 2023). The pI scores of the TaSBT proteins in wheat spanned from 4.89 (TaSBT99) to 9.65 (TaSBT103), indicating a balance between acidic and basic amino acids (Zhao et al., 2023). Additionally, the sizes of the 54 identified AcoSBT proteins in pineapple were greatly variable, from 228 residues (AcoSBT5.9) to 2759 residues (AcoSBT2.5) (Jin et al., 2021). This variation in protein length corresponds to a broad mW value, extending from 23.66 kDa to 297.43 kDa (Jin et al., 2021). Furthermore, the pI values of these AcoSBT proteins demonstrated a diverse range, from 5.29 (AcoSBT1.11) to 9.77 (AcoSBT1.15), indicating a wide variance in their charge at physiological pH levels (Jin et al., 2021). In cotton, the identified SBT genes were found to encode proteins with amino acid lengths varying from 136 to 2111 residues (Dai et al., 2022). These proteins exhibited a wide range of pI values, from 4.387 to 10.369, with an average pI of 7.382 (Dai et al., 2022).

To elucidate the evolutionary dynamics of the SBT family across wheat, Arabidopsis, rice and maize, a comprehensive phylogenetic analysis was undertaken, incorporating a total of 435 SBT protein sequences (Zhao et al., 2023). This dataset comprised 255 TaSBT proteins from wheat (Zhao et al., 2023), 56 AtSBT proteins from Arabidopsis (Rautengarten et al., 2005), 66 OsSBT proteins from rice (Zheng et al., 2022) and 61 ZmSBT proteins from maize. The phylogenetic tree delineated these members into seven distinct groups, labeled clade I through clade VII (Zhao et al., 2023). Notably, clade VII emerged as the largest group, containing 182 members, distributed among TaSBT (90 members), AtSBT (39 members), OsSBT (28 members) and ZmSBT (22 members) proteins (Zhao et al., 2023). Additionally, alignment of the full-length protein sequences from 56 AtSBT proteins from Arabidopsis (Rautengarten et al., 2005) and 54 AcoSBT proteins from pineapple (Jin et al., 2021) facilitated the construction of a Neighbor-Joining phylogenetic tree (Jin et al., 2021). This analysis categorized the SBT proteins into six sub-families (Jin et al., 2021). Notably, group I emerged as the largest sub-family, comprising nine and 30 AtSBT and AcoSBT proteins, respectively (Jin et al., 2021). Conversely, group VI was identified as the smallest group, harboring

only two SBT proteins from both pineapple and Arabidopsis (Jin et al., 2021). Next, phylogenetic analyses of the SBT proteins in grape enabled the classification of this family into eight distinct groups, sequentially labeled from group 1 through group 8. Notably, group 5 emerged as the largest clade, comprising 29 members (Cao et al., 2014). In Arabidopsis, 56 AtSBT proteins were delineated six unique sub-families (Rautengarten et al., 2005). Gene duplication and structural divergence appear to have contributed to the expansion of the cassava SBT family. Comparable genome-wide surveys have associated duplication history, intron distribution, conserved motifs, and phylogenetic clustering with diversification of stress-related gene families in lentil and pigeonpea (Shivadarshan et al., 2025; Shruti Sinha 2025). In cassava, 11 tandem, one segmental, and 15 whole-genome duplication events provide genomic evidence for *MeSBT* expansion. Uneven exon-intron structures and clade distributions also indicate possible divergence among paralogs, although their biochemical and physiological functions require experimental testing. In Arabidopsis, *SBT1.4* gene, which is expressed in all aboveground organs, negatively regulates the production of siliques and the branching of inflorescences throughout the reproductive developmental stages (Wang et al., 2018). Conversely, in legumes, such as *Medicago truncatula* and *Pisum sativum*, proteins encoded by *SBT1.1* gene are localized within the endosperm and play a crucial role in modulating seed size variation by regulating the division of embryo cells during reproductive development (D'Erfurth et al., 2012). Throughout the growth and development phases of wheat, a total of 140 *TaSBT* genes exhibited differential expression (Zhao et al., 2023). Briefly, differential expression was observed during eight growth stages, including the seedling stage (77 genes), trilophyll stage (59 genes), tiller stage (73 genes), flag leaf stage (63 genes), flowering stage (110 genes), milk ripening stage (42 genes), grain filling stage (29 genes), and maturity stage (28 genes) (Zhao et al., 2023). Furthermore, an analysis of *TaSBT* genes' differential expression across various wheat tissues revealed that out of 142, 93, 54, 110 and 30 *TaSBT* genes were differentially expressed in the root, leaf, ear, and seed, respectively (Zhao et al., 2023). This indicates a significant role for *TaSBT* genes in the expression profiles of different wheat tissues (Zhao et al., 2023). To sum up, our re-analysis suggested that these *MeSBT* genes could be instrumental in the regulation of tissue development within cassava plants. Stress-responsive transcription differs across genes, genotypes, and treatment duration. Abscisic acid-independent candidate genes differed between drought-tolerant and drought-susceptible *Saccharum* and *Erianthus* clones, while drought-induced *NAC14* and *NAC19* expression varied among *Brassica juncea* and *Sinapis alba* genotypes (Dharitree Phukan et al., 2016; Priji and Hemaprabha 2014). Chickpea *P5CS* also showed time-dependent transcription under dehydration, with strong induction at 3 h (Kondawar et al., 2025). These studies support time-course expression analysis for prioritization of drought-responsive candidates. However, transcript abundance does not establish a causal role in tolerance. In cassava, the RNA-seq re-analysis identified 43 upregulated and 21 downregulated *MeSBT* genes, whereas 11 genes remained within the -2 to 2 threshold. The opposing responses indicate that individual *MeSBT* members may act at different points in the dehydration-response network. Functional tests are required to determine whether these transcript changes alter drought tolerance.

A principal limitation of the present study is that differential transcript abundance does not establish a functional role for any *MeSBT* protein in drought tolerance. The public RNA-seq data and detached-leaf qRT-PCR assay also represent different water-deficit systems. Leaf excision prevents separation of dehydration responses from wound responses and excludes root-to-shoot signaling. Therefore,

agreement between the two datasets provides experimental support for drought responsiveness but does not constitute direct validation of the RNA-seq results.

Future work should first repeat RNA-seq and qRT-PCR with the same KM94 biological samples and matched sampling times under a controlled whole-plant soil-drying and rewatering treatment. Expression data should be paired with soil water content, leaf water potential, relative water content, stomatal conductance, photosynthetic rate, survival, biomass, and storage-root recovery. Drought studies in wheat show the value of combining physiological traits with molecular markers across diverse accessions (Choudhary et al., 2026). The induced genes *MeSBT29*, *MeSBT59*, *MeSBT21*, and *MeSBT62* and the repressed genes *MeSBT03*, *MeSBT75*, *MeSBT65*, and *MeSBT36* should then be tested through CRISPR/Cas9 knockout, overexpression, and native-promoter complementation. Multiplex editing is warranted for close paralogs that may have overlapping functions. Protease activity assays, active-site substitution, subcellular localization, and N-terminomic substrate discovery would test whether drought phenotypes depend on MeSBT catalytic activity and identify downstream targets. Single and multiplex CRISPR/Cas9 knockouts should be compared with overexpression and native-promoter complementation lines, since paralogous *MeSBT* genes may mask single-gene phenotypes. Cassava genome editing has been demonstrated experimentally (Odipio et al., 2017). Complementation with an active-site serine substitution would determine whether an observed phenotype depends on proteolytic activity. Abscisic acid sensitivity and reactive oxygen species assays would allow comparison with genetic analyses of rice *OsSub14* (Wang et al., 2025). Mechanistic studies should examine MeSBT localization, prodomain removal, catalytic activity, and substrate identity through MeSBT-Green fluorescent protein imaging, immunoblot analysis, fluorophosphonate-based activity profiling (Kaschani et al., 2009), apoplastic N-terminomics, and recombinant cleavage assays. These approaches could determine whether cassava SBTs target abscisic acid-signaling proteins or peptide precursors, as reported for the ARABIDOPSIS SENESCENCE-ASSOCIATED SUBTILISIN PROTEASE-OPEN STOMATA 1 (Wang et al., 2018) and SBT3.8-PHYTOSULFOKINE 1 PRECURSOR pathways (Stührwohldt et al., 2021).

Conclusion

This study provides a genome-wide characterization of the subtilisin-like protease family in cassava. Seventy-five *MeSBT* genes were identified across 17 chromosomes and assigned to six phylogenetic clades. Variation in protein properties and exon-intron organization indicated considerable structural diversity within the family. The duplication analysis detected 11 tandem, one segmental, and 15 whole-genome duplication events. Tissue transcript profiles identified 25 genes with FPKM values of at least 10 in one or more organs, with *MeSBT21*, *MeSBT62*, and *MeSBT75* among the most prominent members. RNA-seq re-analysis identified 43 upregulated and 21 downregulated genes in drought-treated leaves, whereas 11 genes remained within the defined fold-change threshold. The detached-leaf qRT-PCR assay supported the RNA-seq response directions for the induced genes *MeSBT29*, *MeSBT59*, *MeSBT21*, and *MeSBT62* and the repressed genes *MeSBT03*, *MeSBT75*, *MeSBT65*, and *MeSBT36*. These transcriptional patterns identify candidate *MeSBT* genes for drought-related research but do not establish their direct functions in drought tolerance. Whole-plant drought experiments, genetic manipulation,

protease activity measurements, and substrate identification are required to determine their physiological and molecular roles in cassava.

Author Contribution: H.L.T.D., H.V.L., and H.D.C. conceived and designed the study. H.L.T.D., H.V.L., Q.T.N.L., G.H.D. and B.D.N. conducted the literature search and data analysis. H.L.T.D. and H.D.C. prepared the original draft. All authors reviewed and edited the manuscript. Q.T.N.L. prepared the visualizations. H.V.L. and H.D.C. supervised the study. H.D.C. administered the project and acquired funding. All authors read and approved the final manuscript.

Funding: This work is funded by the Foundation for Science and Technology Development, Hanoi Pedagogical University 2 via the grant number HPU2.2023-UT-02.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The cassava (*Manihot esculenta*) reference genome and proteome sequences used for *MeSBT* identification are publicly available through the NCBI RefSeq assembly M.esculenta_v8 (accession GCF_001659605.2). The Peptidase_S8 hidden Markov model is available from the Pfam database under accession PF00082. The RNA-seq dataset used for tissue-specific expression analysis is available from the NCBI Gene Expression Omnibus under accession GSE82279 (BioProject PRJNA324539). The drought-responsive RNA-seq dataset is available under accession GSE98537 (BioProject PRJNA385393). The genomic, coding, and protein sequences of the identified *MeSBT* genes are contained within the cassava reference assembly, and their locus identifiers are provided in Table 1. The qRT-PCR primer sequences are provided in the manuscript. No new high-throughput sequencing data were generated in this study. Raw qRT-PCR data are available from the corresponding author upon reasonable request.

References

Ali, A. et al. (2022). A putative SUBTILISIN-LIKE SERINE PROTEASE 1 (SUBSrP1) regulates anther cuticle biosynthesis and panicle development in rice. *J Adv Res*, 42(1), 273-287. 10.1016/j.jare.2022.01.003

Barrett, T. et al. (2013). NCBI GEO: archive for functional genomics data sets - Update. *Nucleic acids research*, 41(1), D991-D995. 10.1093/nar/gks1193

Berger, D., Altmann, T. (2000). A subtilisin-like serine protease involved in the regulation of stomatal density and distribution in *Arabidopsis thaliana*. *Genes & development*, 14(9), 1119-1131. 10.1101/gad.14.9.1119

Bredeson, J. V. et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature biotechnology*, 34(5), 562-570. 10.1038/nbt.3535

Brown, A. L., Cavagnaro, T. R., Gleadow, R., Miller, R. E. (2016). Interactive effects of temperature and drought on cassava growth and toxicity: implications for food security? *Global change biology*, 22(10), 3461-3473. 10.1111/gcb.13380

Bryant, M. K., Schardl, C. L., Hesse, U., Scott, B. (2009). Evolution of a subtilisin-like protease gene family in the grass endophytic fungus *Epichloë festucae*. *BMC Evol Biol*, 9(1), 168. 10.1186/1471-2148-9-168

Cao, J., Han, X., Zhang, T., Yang, Y., Huang, J., Hu, X. (2014). Genome-wide and molecular evolution analysis of the subtilase gene family in *Vitis vinifera*. *BMC Genom*, 15(1), 1116. 10.1186/1471-2164-15-1116

Chavarriaga-Aguirre, P. et al. (2016). The potential of using biotechnology to improve cassava: A review. *In vitro cellular & developmental biology Plant : journal of the Tissue Culture Association*, 52(5), 461-478. 10.1007/s11627-016-9776-3

Choudhary, D. et al. (2026). Assessing genetic diversity and drought adaptive potential in diverse hexaploid wheat accessions. *Indian J Genet Plant Breed*, 86(3), 317-332. 10.1007/s44489-026-00027-y

D'Erfurth, I. et al. (2012). A role for an endosperm-localized subtilase in the control of seed size in legumes. *The New phytologist*, 196(3), 738-751. 10.1111/j.1469-8137.2012.04296.x

Dai, M. et al. (2022). Genome-wide analysis of the SBT gene family involved in drought tolerance in cotton. *Frontiers in plant science*, 13(1097732. 10.3389/fpls.2022.1097732

Dharitree Phukan, P. Bhupendra Singh, Indu Ravi, Amitha S. V. Mithra, Devendra K. Yadava, Viswanathan Chinnusamy, Mohapatra, T. (2016). Cloning and characterization of drought stress-induced NAC transcription factors from *Brassica juncea* and *Sinapis alba*. *Indian J Genet Plant Breed*, 76(3), 233-245. 10.5958/0975-6906.2016.00038.9

Ekchaweng, K., Khunjan, U., Churngchow, N. (2017). Molecular cloning and characterization of three novel subtilisin-like serine protease genes from *Hevea brasiliensis*. *Physiol Mol Plant Pathol*, 97(79-95. https://doi.org/10.1016/j.pmpp.2016.12.007

Figueiredo, A., Monteiro, F., Sebastiana, M. (2014). Subtilisin-like proteases in plant-pathogen recognition and immune priming: A perspective. *Frontiers in plant science*, 5(739. 10.3389/fpls.2014.00739

Figueiredo, J., Sousa Silva, M., Figueiredo, A. (2018). Subtilisin-like proteases in plant defence: The past, the present and beyond. *Molecular plant pathology*, 19(4), 1017-1028. 10.1111/mpp.12567

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, 31(13), 3784-3788. 10.1093/nar/gkg563

Gonzalez, H. C. B., Galli, V. (2021). The CYP707A gene family in strawberry (*Fragaria* × *ananassa*). *Braz Arch Biol Technol*, 64(e21200133. 10.1590/1678-4324-2021200133

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., Gao, G. (2015). GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*, 31(8), 1296-1297. 10.1093/bioinformatics/btu817

Hu, M., Hu, W., Xia, Z., Zhou, X., Wang, W. (2016). Validation of reference genes for relative quantitative gene expression studies in cassava (*Manihot esculenta* Crantz) by using quantitative real-time PCR. *Frontiers in plant science*, 7(1), 680. https://doi.org/10.3389/fpls.2016.00680

Jin, X. et al. (2021). Genome-wide investigation of SBT family genes in pineapple and functional analysis of AcoSBT1.12 in floral transition. *Front Genet*, 12(730821. 10.3389/fgene.2021.730821

Kaschani, F., Gu, C., Niessen, S., Hoover, H., Cravatt, B. F., van der Hoorn, R. A. (2009). Diversity of serine hydrolase activities of unchallenged and botrytis-infected *Arabidopsis thaliana*. *Mol Cell Proteom*, 8(5), 1082-1093. 10.1074/mcp.M800494-MCP200

Kondawar, V., Deokar, A. A., Kohli, D., Yadav, S., Srinivasan, R., Jain, P. K. (2025). Promoter characterization of chickpea delta-1-pyrroline-5-carboxylate synthetase (P5CS) gene provides

novel insights into its stress responsiveness. *Indian J Genet Plant Breed*, 85(04), 572-577. 10.31742/ISGPB.85.4.5

Kothari, D. et al. (2026). Preliminary investigation of phosphorus-use efficiency in mungbean (*Vigna radiata*): A comparative RNA-Seq study. *Indian J Genet Plant Breed*, 86(3), 358-368. 10.1007/s44489-026-00031-2

La, H. V. (2022). Genome-wide identification and analysis of Heat shock Protein 70 family in *Theobroma cacao*. *Pak J Biol Sci*, 25(7), 608-618. 10.3923/pjbs.2022.608.618

La, H. V. et al. (2022a). SWEET gene family in sugar beet (*Beta vulgaris*): Genome-wide survey, phylogeny and expression analysis. *Pak J Biol Sci*, 25(5), 387-395.

La, H. V. et al. (2022b). Insights into the gene and protein structures of the CaSWEET family members in chickpea (Cicer arietinum), and their gene expression patterns in different organs under various stress and abscisic acid treatments. *Gene*, 819(146210. 10.1016/j.gene.2022.146210

Larkin, M. A. et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947-2948. 10.1093/bioinformatics/btm404

Le, M. T. et al. (2022). Genome-wide identification and analysis of genes encoding putative Heat shock protein 70 in papaya (*Carica papaya*). *Pakistan journal of biological sciences : PJBS*, 25(6), 468-475. 10.3923/pjbs.2022.468.475

Livak, K. J., Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\text{Delta Delta C(T)})}$ Method. *Methods (San Diego, Calif)*, 25(4), 402-408. 10.1006/meth.2001.1262

Malik, A. I. et al. (2020). Cassava breeding and agronomy in Asia: 50 years of history and future directions. *Breeding science*, 70(2), 145-166. 10.1270/jsbbs.18180

Mistry, J. et al. (2021). Pfam: The protein families database in 2021. *Nucleic acids research*, 49(D1), D412-D419. 10.1093/nar/gkaa913

Mona Yadav, A. S., Apoorv Tiwari, Bhawna Sheoran, Era Chaudhary, Ritu Nain, Sanjay Kumar Singh, Garg, M. (2025). Identification and characterization of genes involved in tocochromanol biosynthesis pathway through genomics and transcriptomics in durum and bread wheat. *Indian J Genet Plant Breed*, 85(3), 407-417. 10.31742/ISGPB.85.3.6

Norero, N. S., Castellote, M. A., de la Canal, L., Feingold, S. E. (2016). Genome-wide analyses of subtilisin-like serine proteases on *Solanum tuberosum*. *American Journal of Potato Research*, 93(5), 485-496. 10.1007/s12230-016-9525-5

Odipio, J., Alicai, T., Ingelbrecht, I., Nusinow, D. A., Bart, R., Taylor, N. J. (2017). Efficient CRISPR/Cas9 genome editing of *Phytoene desaturase* in cassava. *Frontiers in plant science*, 8(01780. 10.3389/fpls.2017.01780

Panchy, N., Lehti-Shiu, M., Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol*, 171(4), 2294-2316. 10.1104/pp.16.00523

Pearce, G., Yamaguchi, Y., Barona, G., Ryan, C. A. (2010). A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. *Proceedings of the National Academy of Sciences*, 107(33), 14921-14925. 10.1073/pnas.1007568107

Priji, P. J., Hemaprabha, G. (2014). Sugarcane-specific drought responsive candidate genes belonging to ABA-independent pathway identified from tolerant and susceptible clones of *Saccharum* and *Erianthus* species. *Indian J Genet Plant Breed*, 74(01), 64-72. .

Rautengarten, C., Steinhauser, D., Büssis, D., Stintzi, A., Schaller, A., Kopka, J., Altmann, T. (2005). Inferring Hypotheses on Functional Relationships of Genes: Analysis of the Arabidopsis thaliana Subtilase Gene Family. *PLOS Computational Biology*, 1(4), e40. 10.1371/journal.pcbi.0010040

Ray, D. K., West, P. C., Clark, M., Gerber, J. S., Prishchepov, A. V., Chatterjee, S. (2019). Climate change has likely already affected global food production. *PLoS One*, 14(5), e0217148-e0217148. 10.1371/journal.pone.0217148

Shan, Z., Luo, X., Wei, M., Huang, T., Khan, A., Zhu, Y. (2018). Physiological and proteomic analysis on long-term drought resistance of cassava (Manihot esculenta Crantz). *Scientific Reports*, 8(1), 17982. 10.1038/s41598-018-35711-x

Shivadarshan, S. J. et al. (2025). Genome-wide identification, evolution and expression analysis of the HSP20 gene family in lentil [*Lens culinaris* (L.) Medikus]. *Indian J Genet Plant Breed*, 85(04), 619-628. 10.31742/ISGPB.85.4.10

Shruti Sinha, K. K., Palak Gupta, Harsha Srivastava, Sandhya Sharma, Rishu Jain, Kumar Durgesh, Kishor U. Tribhuvan, Kishor Gaikwad (2025). Identification and characterization of cold shock domain proteins (CSPs) for their response to cold stress in pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Indian J Genet Plant Breed*, 85(2), 262-270. 10.31742/isgpb.85.2.11

Stührwohldt, N., Bühler, E., Sauter, M., Schaller, A. (2021). Phytosulfokine (PSK) precursor processing by subtilase SBT3.8 and PSK signaling improve drought stress tolerance in *Arabidopsis*. *Journal of experimental botany*, 72(9), 3427-3440. 10.1093/jxb/erab017

Sugiharto, S., Yudiarti, T., Isroli, I., Widiastuti, E. (2018). The potential of tropical agro-industrial by-products as a functional feed for poultry. *Iranian Journal of Applied Animal Science*, 8(375-385.

Tamura, K., Stecher, G., Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular biology and evolution*, 38(7), 3022-3027. 10.1093/molbev/msab120

Tomlinson, K. R., Bailey, A. M., Alicai, T., Seal, S., Foster, G. D. (2018). Cassava brown streak disease: historical timeline, current knowledge and future prospects. *Molecular plant pathology*, 19(5), 1282-1294. 10.1111/mpp.12613

Tornero, P., Conejero, V., Vera, P. (1996). Primary structure and expression of a pathogen-induced protease (PR-P69) in tomato plants: Similarity of functional domains to subtilisin-like endoproteases. *Proceedings of the National Academy of Sciences of the United States of America*, 93(13), 6332-6337. 10.1073/pnas.93.13.6332

Tran, C. D., Chu, H. D., Nguyen, K. H., Watanabe, Y., La, H. V., Tran, K. D., Tran, L.-S. P. (2018). Genome-wide identification of the TCP transcription factor family in chickpea (*Cicer arietinum* L.) and their transcriptional responses to dehydration and exogenous abscisic acid treatments. *J Plant Growth Regul*, 37(4), 1286-1299. 10.1007/s00344-018-9859-y

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., Rozen, S. G. (2012). Primer3 - New capabilities and interfaces. *Nucleic acids research*, 40(15), e115. 10.1093/nar/gks596

Utsumi, Y. et al. (2022). Agrobacterium-mediated cassava transformation for the Asian elite variety KU50. *Plant molecular biology*, 109(3), 271-282. 10.1007/s11103-021-01212-1

Wang, Q. et al. (2018). Arabidopsis subtilase SASP is involved in the regulation of ABA signaling and drought tolerance by interacting with OPEN STOMATA 1. *Journal of experimental botany*, 69(18), 4403-4417. 10.1093/jxb/ery205

Wang, Y. et al. (2025). Characterization of drought-responsive subtilisin genes in upland rice: The role of *OsSub14* in ABA signaling and ROS accumulation. *BMC Plant Biol*, 25(1), 1462. 10.1186/s12870-025-07528-0

Wilson, M. C. et al. (2017). Gene expression atlas for the food security crop cassava. *The New phytologist*, 213(4), 1632-1641. 10.1111/nph.14443

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. *The Journal of biological chemistry*, 269(52), 32725-32731.

Zhao, X., Goher, F., Chen, L., Song, J., Zhao, J. (2023). Genome-wide identification, phylogeny and expression analysis of subtilisin (SBT) gene family under wheat biotic and abiotic stress. *Plants*, 12(17), 3065. 10.3390/plants12173065

Zheng, K., Pang, L., Xue, X., Gao, P., Zhao, H., Wang, Y., Han, S. (2022). Genome-wide comprehensive survey of the subtilisin-like proteases gene family associated with rice caryopsis development. *Frontiers in plant science*, 13(1), 943184. 10.3389/fpls.2022.943184

Zhu, Y. et al. (2020). Antioxidant enzymatic activity and its related genes expression in cassava leaves at different growth stages play key roles in sustaining yield and drought tolerance under moisture stress. *J Plant Growth Regul*, 39(2), 594-607. 10.1007/s00344-019-10003-4

Figure and table legends

Figure 1. Chromosomal distribution and predicted duplication relationships of the cassava *MeSBT* gene family. The 75 *MeSBT* genes were mapped to 17 of the 18 cassava chromosomes. Chromosome numbers are shown above or below the chromosome diagrams, and black ticks indicate gene positions. Blue brackets indicate tandem duplicate groups, the red bracket on chromosome 18 indicates the segmental duplicate pair *MeSBT71-MeSBT72*, and colored curves between chromosomes indicate relationships attributed to whole-genome duplication.

Figure 2. Phylogenetic relationships of SBT proteins from cassava and *Arabidopsis thaliana*. The unrooted maximum-likelihood tree was constructed from the full-length amino acid sequences of 75 MeSBT and 56 AtSBT proteins. Branch and label colors delineate the six phylogenetic clades, designated I-VI. Numbers beside internal branches indicate bootstrap support values.

Figure 3. Exon-intron organization of the cassava *MeSBT* genes. The 75 genes are ordered according to their phylogenetic relationships, as shown by the dendrogram on the left. Dark red blocks represent exons, and black connecting lines represent introns. Gene structures are drawn to scale in kilobases.

Figure 4. Transcript profiles of *MeSBT* genes across cassava organs and tissues. The heatmap was generated from RNA-seq-derived FPKM values obtained from dataset GSE82279. The analysis comprised friable embryogenic callus, fibrous root, lateral bud, leaf blade, leaf midvein, organized embryogenic structure, petiole, root apical meristem, shoot apical meristem, stem, and storage root. Genes were grouped by hierarchical clustering. Green indicates low or undetectable transcript abundance, black indicates detectable or intermediate abundance, and red indicates high abundance. FPKM: fragments per kilobase of transcript per million mapped reads.

Figure 5. RNA-seq-derived fold changes of *MeSBT* genes in cassava leaves under drought treatment. Fold-change scores were calculated for drought-treated leaves relative to the control from dataset GSE98537. Genes are arranged from the strongest reduction on the left to the strongest induction on the right. Green, black, and red indicate decreased, unchanged, and increased transcript abundance, respectively. Genes with fold-change scores of -2.0 or lower were classified as downregulated, whereas scores of 2.0 or higher indicated upregulation. Scores between -2.0 and 2.0 were classified as unchanged.

Figure 6. Relative transcript levels of eight *MeSBT* genes in cassava leaves during detached-leaf dehydration. Transcript levels were measured by qRT-PCR at 0, 6, and 12 h after treatment. Expression was normalized against cassava *UBQ10* and calculated with the $2^{-\Delta\Delta Ct}$ method, with the 0-h value set to 1. Values represent means ± SD from three independent biological replicates, each assessed with three technical replicates. The dashed line indicates the 0-h reference level. Red bars indicate induction of at least twofold, blue bars indicate expression of 0.5-fold or lower, and black bars indicate the reference level or a nonsignificant response. Asterisks indicate significant differences from 0 h according to Student's *t*-test at $p < 0.05$.

Table 1. Physicochemical properties of the 75 predicted MeSBT proteins and their encoding genes in cassava. Gene size is reported in base pairs, and protein length is reported in amino acids. MW, molecular weight. pI, theoretical isoelectric point. AI, aliphatic index. GRAVY, grand average of hydropathicity. Positive and negative GRAVY values indicate relative hydrophobicity and hydrophilicity, respectively. A plus sign between Phytozome locus identifiers indicates that one *MeSBT* entry is associated with two locus identifiers.

Table 2. Predicted duplication events among *MeSBT* genes in cassava. Duplication events were classified as tandem, segmental, or whole-genome duplication according to the criteria described in the Materials and Methods. A slash separates members of a duplicated pair or group. Parentheses identify a tandem duplicate group assigned to one side of a whole-genome duplication relationship.

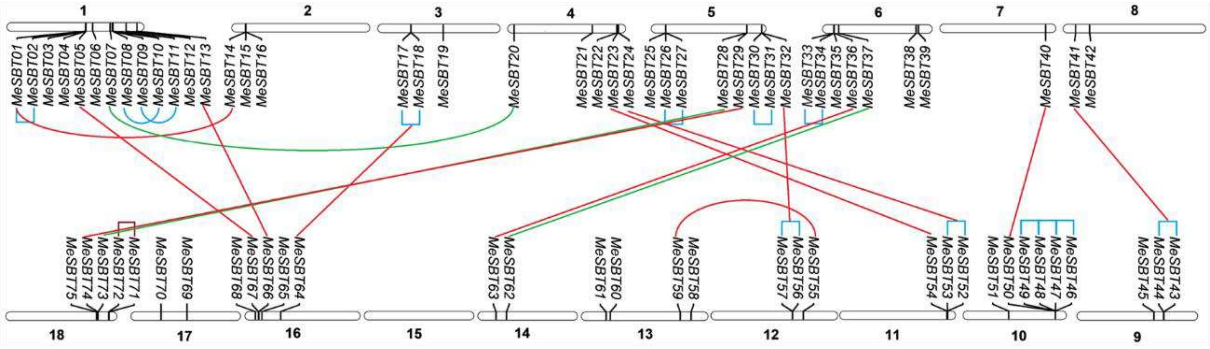


Figure 1. Chromosomal distribution and predicted duplication relationships of the cassava *MeSBT* gene family. The 75 *MeSBT* genes were mapped to 17 of the 18 cassava chromosomes. Chromosome numbers are shown above or below the chromosome diagrams, and black ticks indicate gene positions. Blue brackets indicate tandem duplicate groups, the red bracket on chromosome 18 indicates the segmental duplicate pair *MeSBT71-MeSBT72*, and colored curves between chromosomes indicate relationships attributed to whole-genome duplication.

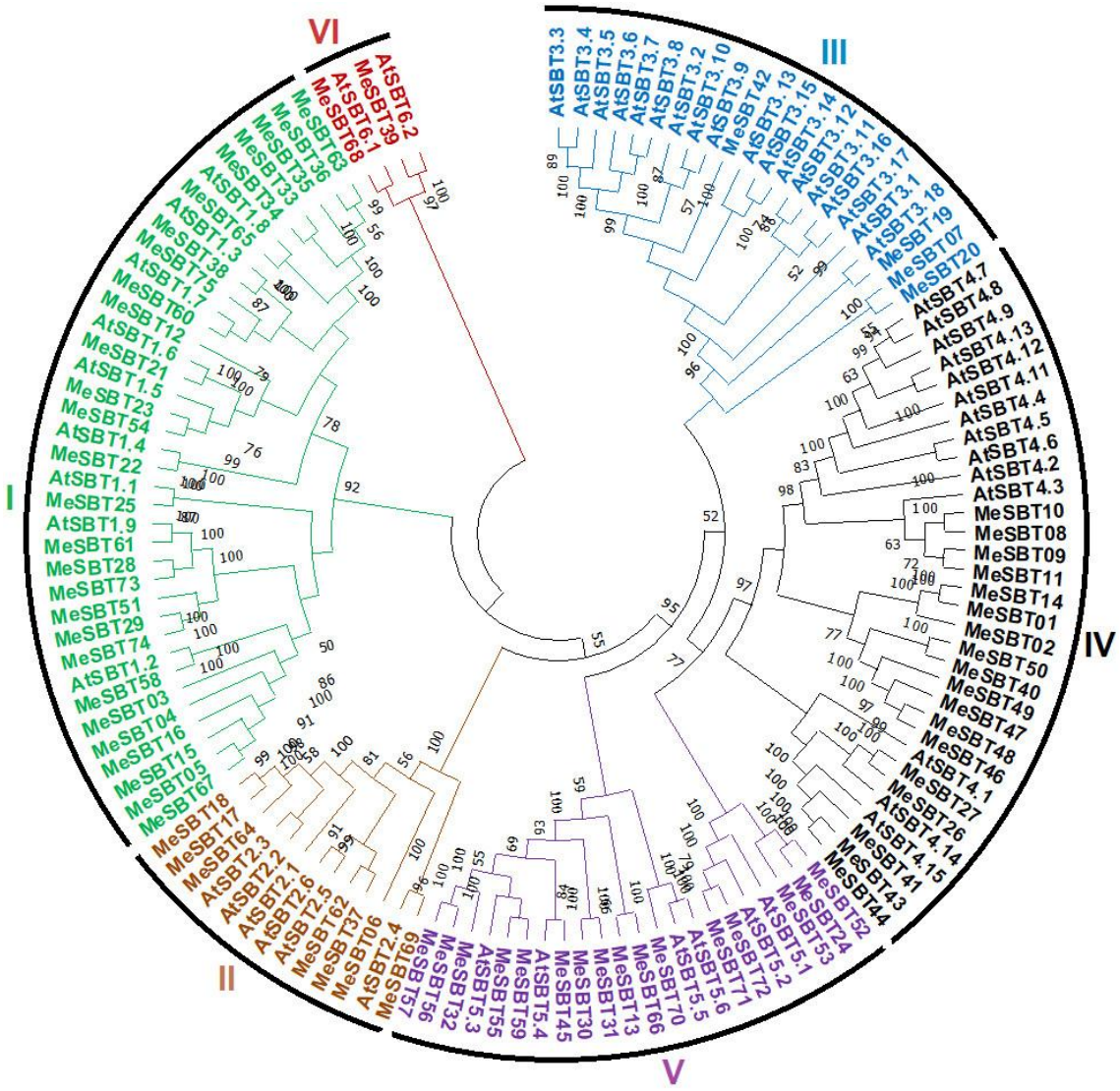


Figure 2. Phylogenetic relationships of SBT proteins from cassava and *Arabidopsis thaliana*. The unrooted maximum-likelihood tree was constructed from the full-length amino acid sequences of 75 MeSBT and 56 AtSBT proteins. Branch and label colors delineate the six phylogenetic clades, designated I-VI. Numbers beside internal branches indicate bootstrap support values.

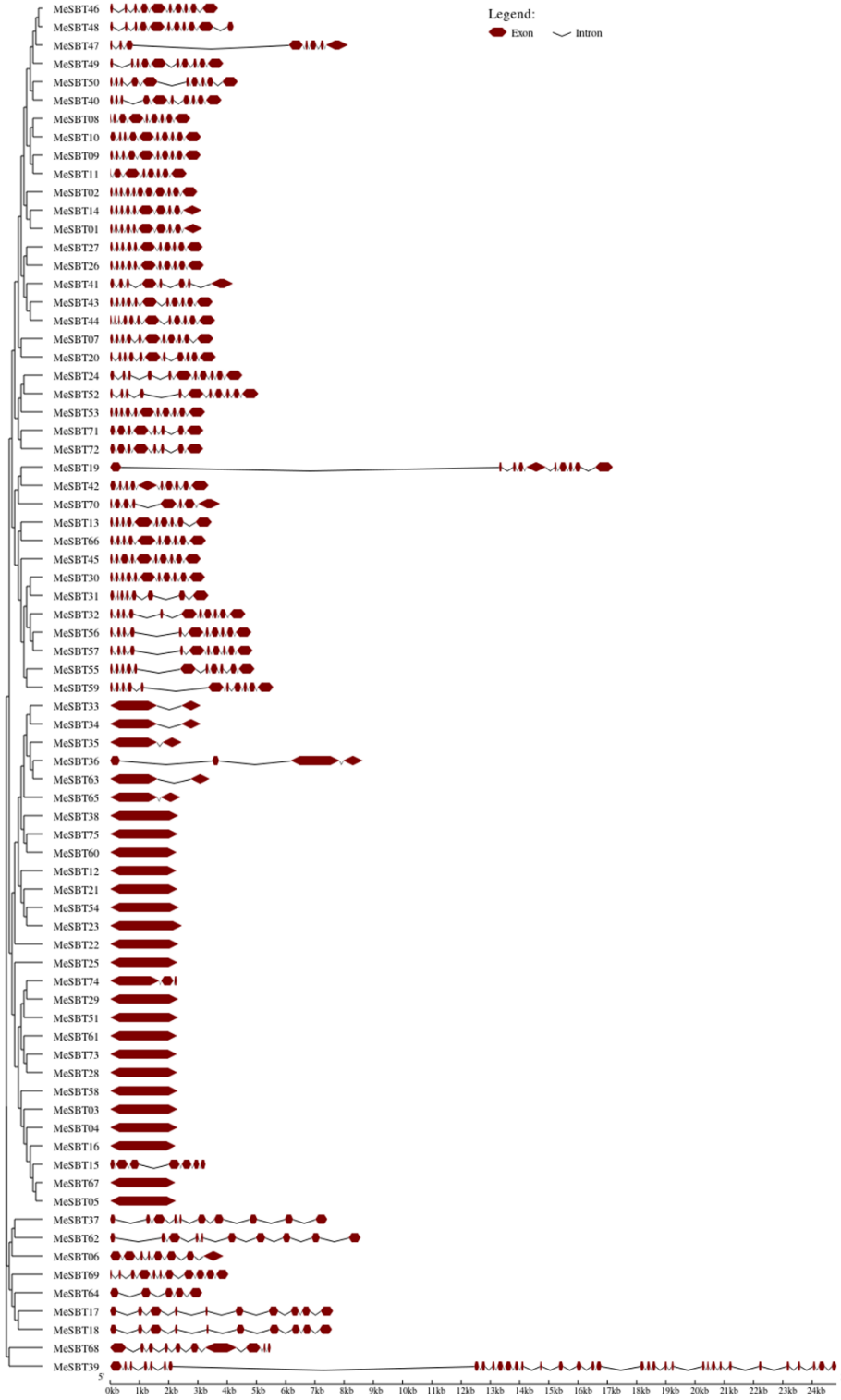


Figure 3. Exon-intron organization of the cassava *MeSBT* genes. The 75 genes are ordered according to their phylogenetic relationships, as shown by the dendrogram on the left. Dark red blocks represent exons, and black connecting lines represent introns. Gene structures are drawn to scale in kilobases.

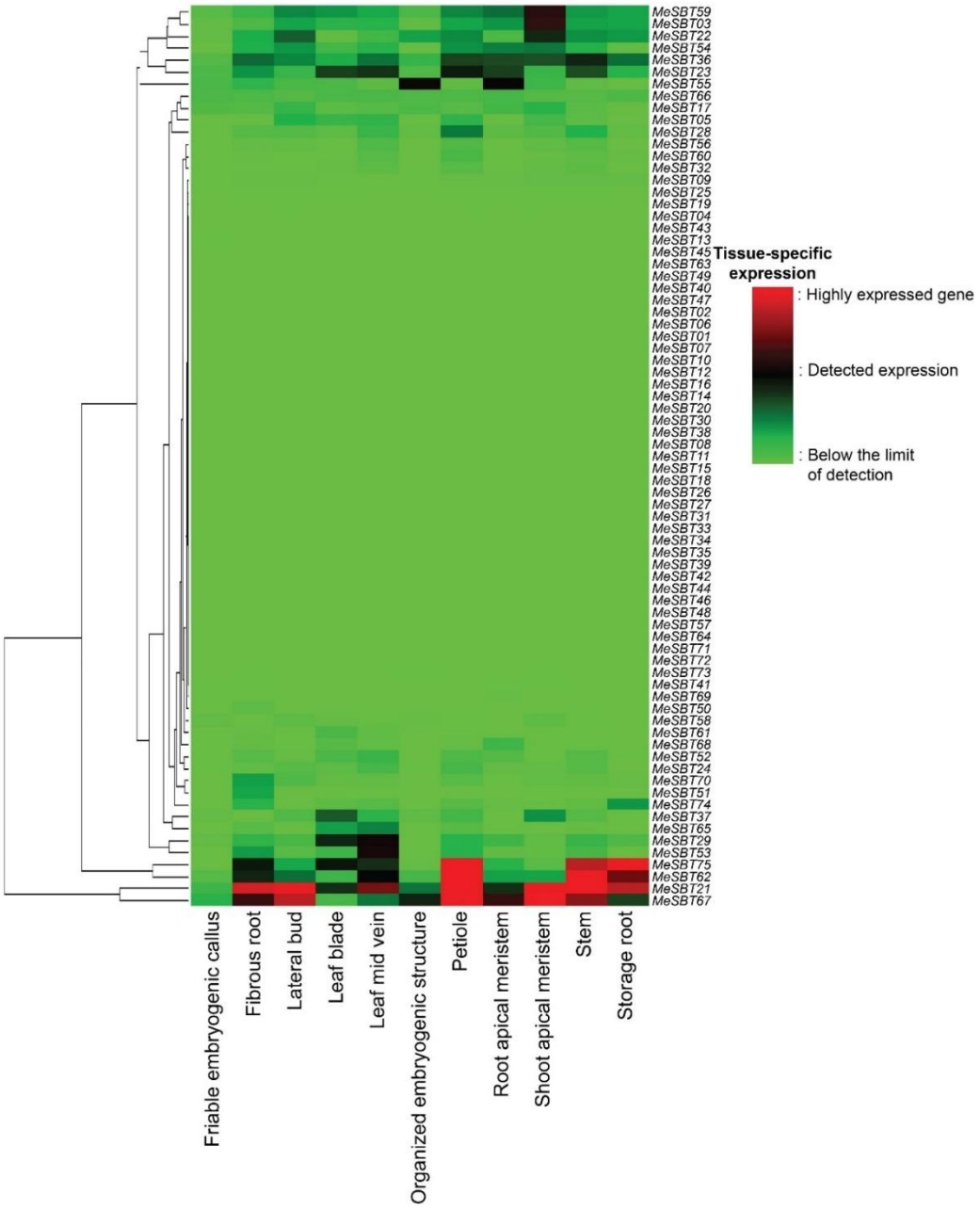
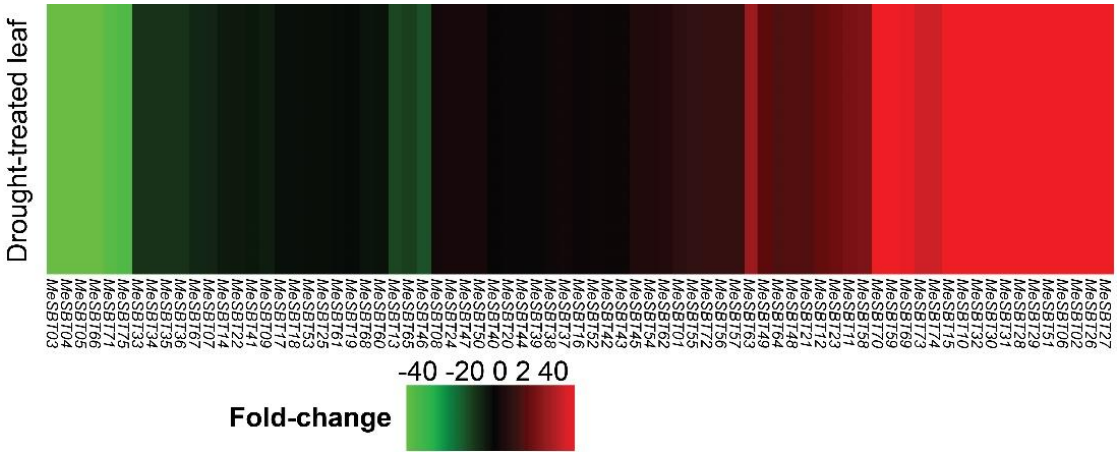


Figure 4. Transcript profiles of *MeSBT* genes across cassava organs and tissues. The heatmap was generated from RNA-seq-derived FPKM values obtained from dataset GSE82279. The analysis comprised friable embryogenic callus, fibrous root, lateral bud, leaf blade, leaf midvein, organized embryogenic structure, petiole, root apical meristem, shoot apical meristem, stem, and storage root. Genes were grouped by hierarchical clustering. Green indicates low or undetectable transcript abundance, black indicates detectable or intermediate abundance, and red indicates high abundance. FPKM: fragments per kilobase of transcript per million mapped reads.



674
675

676 **Figure 5. RNA-seq-derived fold changes of *MeSBT* genes in cassava leaves under drought treatment.**
677 Fold-change scores were calculated for drought-treated leaves relative to the control from dataset
678 GSE98537. Genes are arranged from the strongest reduction on the left to the strongest induction on the
679 right. Green, black, and red indicate decreased, unchanged, and increased transcript abundance,
680 respectively. Genes with fold-change scores of -2.0 or lower were classified as downregulated, whereas
681 scores of 2.0 or higher indicated upregulation. Scores between -2.0 and 2.0 were classified as unchanged.

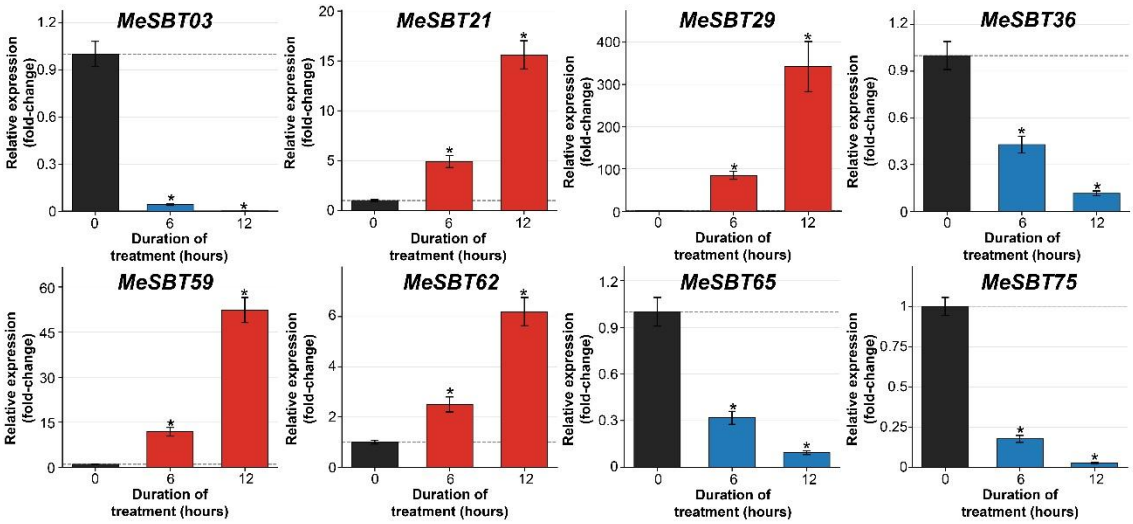


Figure 6. Relative transcript levels of eight *MeSBT* genes in cassava leaves during detached-leaf dehydration. Transcript levels were measured by qRT-PCR at 0, 6, and 12 h after treatment. Expression was normalized against cassava *UBQ10* and calculated with the $2^{-\Delta\Delta Ct}$ method, with the 0-h value set to 1. Values represent means ± SD from three independent biological replicates, each assessed with three technical replicates. The dashed line indicates the 0-h reference level. Red bars indicate induction of at least twofold, blue bars indicate expression of 0.5-fold or lower, and black bars indicate the reference level or a nonsignificant response. Asterisks indicate significant differences from 0 h according to Student's *t*-test at $p < 0.05$.

Table 1. Physicochemical properties of the 75 predicted MeSBT proteins and their encoding genes in cassava. Gene size is reported in base pairs, and protein length is reported in amino acids. MW, molecular weight. pI, theoretical isoelectric point. AI, aliphatic index. GRAVY, grand average of hydropathicity. Positive and negative GRAVY values indicate relative hydrophobicity and hydrophilicity, respectively. A plus sign between Phytozome locus identifiers indicates that one *MeSBT* entry is associated with two locus identifiers.

Gene name	Phytozome Locus ID	Gene size (bp)	Protein length (aa)	mW (kDa)	pI	AI	GRAVY
MeSBT01	Manes.01G069100	3342	771	82.99	5.61	80.99	-0.037
MeSBT02	Manes.01G069200	2978	707	75.89	5.44	77.69	-0.129
MeSBT03	Manes.01G070400	2462	767	82.09	5.76	77.65	-0.108
MeSBT04	Manes.01G070500	2539	767	81.39	5.68	87.21	-0.049
MeSBT05	Manes.01G070600	2417	749	78.75	5.85	83.50	0.002
MeSBT06	Manes.01G088700	4141	834	89.94	8.76	87.51	0.049
MeSBT07	Manes.01G158500	3782	765	82.04	8.96	83.95	-0.037
MeSBT08	Manes.01G167600	2746	710	75.63	5.55	87.23	0.037
MeSBT09	Manes.01G167700	3096	732	77.71	9.01	83.17	-0.070
MeSBT10	Manes.01G167600	3104	764	81.94	6.06	90.38	0.078
MeSBT11	Manes.01G167600	2614	686	72.83	9.16	81.91	-0.115
MeSBT12	Manes.01G222700	2265	754	81.04	5.28	86.25	-0.095
MeSBT13	Manes.01G267200	3644	769	82.89	7.89	82.98	-0.134
MeSBT14	Manes.02G030800	3417	774	82.79	5.29	80.92	-0.002
MeSBT15	Manes.02G029500	5515	604	64.45	6.47	84.80	-0.080
MeSBT16	Manes.02G028900	2329	744	79.31	5.96	85.51	-0.009
MeSBT17	Manes.03G066700	7724	844	90.74	6.89	83.58	-0.037
MeSBT18	Manes.03G066700	7850	844	90.63	7.33	83.47	-0.031
MeSBT19	Manes.03G086500	17289	881	95.70	8.38	87.76	-0.040
MeSBT20	Manes.04G005200	3643	730	78.16	6.97	78.15	-0.127
MeSBT21	Manes.04G074500	2544	767	81.34	8.81	84.95	0.023
MeSBT22	Manes.04G146100	2580	777	82.49	6.34	88.13	0.034
MeSBT23	Manes.04G150200	2714	817	86.25	9.02	81.40	-0.039
MeSBT24	Manes.04G153200	5721	793	85.88	8.92	85.73	-0.104
MeSBT25	Manes.05G041400	2612	766	81.84	8.78	81.79	-0.102
MeSBT26	Manes.05G041500	3363	742	79.45	9.14	84.95	-0.039
MeSBT27	Manes.05G041600	3343	741	78.64	6.20	87.41	0.040
MeSBT28	Manes.05G175300	2661	763	82.06	5.46	84.15	-0.037
MeSBT29	Manes.05G175400	2328	787	85.96	6.78	85.50	-0.040
MeSBT30	Manes.05G206600	6019	767	81.61	7.61	84.38	-0.135

MeSBT31	Manes.05G206600	8458	506	55.51	8.63	87.29	-0.158
MeSBT32	Manes.05G206800	5090	771	83.30	9.17	77.69	-0.216
MeSBT33	Manes.06G013200	3317	754	81.22	9.20	88.47	-0.017
MeSBT34	Manes.06G013250	3333	754	81.22	9.20	88.47	-0.017
MeSBT35	Manes.06G013300	2518	753	82.10	9.29	84.52	-0.085
MeSBT36	Manes.06G013400	8629	758	82.31	9.15	85.37	-0.082
MeSBT37	Manes.06G020400	7679	817	87.99	8.20	84.65	-0.072
MeSBT38	Manes.06G139800	2858	776	83.38	8.41	83.71	-0.117
MeSBT39	Manes.06G148500	25171	1424	156.73	6.34	85.50	-0.339
MeSBT40	Manes.07G122300	3955	736	79.48	5.70	79.42	-0.144
MeSBT41	Manes.08G023600	4407	742	79.47	8.75	83.89	-0.128
MeSBT42	Manes.08G057700	3462	811	86.46	8.53	88.75	0.035
MeSBT43	Manes.09G057300	3506	746	80.49	9.24	79.12	-0.197
MeSBT44	Manes.09G059400	4660	729	78.80	9.30	78.02	-0.206
MeSBT45	Manes.09G072100	3099	768	82.47	8.09	81.02	-0.128
MeSBT46	Manes.10G021500	3783	740	79.07	5.29	79.64	-0.090
MeSBT47	Manes.10G021400	8130	679	71.77	4.91	75.13	-0.090
MeSBT48	Manes.10G021600	4219	797	85.77	5.64	79.31	-0.094
MeSBT49	Manes.10G021700	6377	737	78.43	6.03	80.85	-0.045
MeSBT50	Manes.10G021800	4497	731	78.44	6.13	82.23	-0.039
MeSBT51	Manes.10G079600	2615	774	84.20	5.18	79.29	-0.144
MeSBT52	Manes.11G010100	5321	771	83.29	8.92	82.62	-0.122
MeSBT53	Manes.11G010200	3432	766	83.33	8.88	82.28	-0.156
MeSBT54	Manes.11G013600	2672	783	82.65	8.64	84.44	-0.008
MeSBT55	Manes.12G067800	5100	770	82.57	8.87	82.49	-0.115
MeSBT56	Manes.12G081800	4969	770	83.20	9.32	78.70	-0.186
MeSBT57	Manes.12G081800	5074	760	82.18	9.29	77.43	-0.214
MeSBT58	Manes.13G039400	2310	769	83.88	9.10	84.45	-0.086
MeSBT59	Manes.13G059200	5891	770	82.25	7.62	83.16	-0.088
MeSBT60	Manes.13G113900	2762	756	79.86	6.13	79.25	-0.074
MeSBT61	Manes.13G123200	2503	761	81.85	7.62	86.96	-0.042
MeSBT62	Manes.14G157400	8980	816	87.54	6.55	83.21	-0.075
MeSBT63	Manes.14G164700	3480	755	82.14	9.24	80.21	-0.158
MeSBT64	Manes.16G063000	9539	535	56.88	6.95	87.55	0.084
MeSBT65	Manes.16G094900	5512	762	80.92	8.15	83.36	-0.053
MeSBT66	Manes.16G104900	3549	778	83.63	8.75	80.51	-0.139
MeSBT67	Manes.16G106100	2599	742	79.91	8.83	82.84	-0.139
MeSBT68	Manes.16G115100	6451	1039	115.53	6.54	85.77	-0.172

MeSBT69	Manes.17G027600	4261	827	88.07	6.58	92.91	0.150
MeSBT70	Manes.17G062600	4029	778	84.15	8.88	81.53	-0.188
MeSBT71	Manes.18G010600	3189	691	73.89	5.42	83.13	-0.162
MeSBT72	Manes.18G012100	3182	685	73.05	4.84	79.17	-0.134
MeSBT73	Manes.18G039000	2280	759	81.78	5.82	88.33	-0.027
MeSBT74	Manes.18G039300	2279	727	79.36	6.19	82.54	-0.139
MeSBT75	Manes.18G044300	2604	771	80.95	6.04	82.35	-0.006

697

Table 2. Predicted duplication events among *MeSBT* genes in cassava. Duplication events were classified as tandem, segmental, or whole-genome duplication according to the criteria described in the Materials and Methods. A slash separates members of a duplicated pair or group. Parentheses identify a tandem duplicate group assigned to one side of a whole-genome duplication relationship.

Duplication event	Duplicated genes (gene 1/gene 2)
Tandem duplication	*MeSBT01/MeSBT02, MeSBT08/MeSBT10, MeSBT09/MeSBT11, MeSBT17/MeSBT18, MeSBT26/MeSBT27, MeSBT30/MeSBT31. MeSBT33/MeSBT34, MeSBT43/MeSBT44, MeSBT46/MeSBT47/MeSBT48/MeSBT49, MeSBT52/MeSBT53, MeSBT56/MeSBT57*
Segmental duplication	*MeSBT71/MeSBT72*
Whole genome duplication	*MeSBT01/MeSBT14, MeSBT05/MeSBT67, MeSBT07/MeSBT20, MeSBT13/MeSBT66, MeSBT24/(MeSBT52/MeSBT53), (MeSBT17/MeSBT18)/MeSBT64, MeSBT23/MeSBT54, MeSBT28/MeSBT73, MeSBT29/MeSBT74, MeSBT32/(MeSBT56, MeSBT57), MeSBT36/MeSBT63, MeSBT37/MeSBT62, MeSBT40/MeSBT50, MeSBT41/MeSBT43, MeSBT55/MeSBT59*