HAYATI
Journal of Biosciences

Vol. 32 No. 4, July 2025 940-949
DOI:10.4308/hjb.32.4.940-949
ISSN: 1978-3019
EISSN: 2086-4094

Research Article

Check for updates

OPEN ACCESS

The Trehalose-6-Phosphate Synthase and Trehalose-6-Phosphate Phosphatase in Cocoa (*Theobroma cacao* L.): Genome-Wide Identification and Expression Analysis

Quy Phuong Nguyen[1], Phi Bang Cao[1], Ngoc Thi Bich Chu[1], Lan Thi Mai Tran[1], Man Thi Le[1], Hang Thi Luong[1,3], Quynh Thi Ngoc Le[3], Gioi Huy Dong[4], Ha Duc Chu[5]*

[1]*Faculty of Natural Sciences, Hung Vuong University, Phu Tho Province 35000, Vietnam*
[2]*Tien Phong Junior High School, Doan Hung, Phu Tho Province 35300, Vietnam*
[3]*Department of Biotechnology, Thuyloi University, Hanoi City 122300, Vietnam*
[4]*Faculty of Biotechnology, Vietnam National University, Hanoi Hanoi City 122300, Vietnam*
[5]*Faculty of Agricultural Technology, University of Engineering and Technology, Hanoi City 122300, Vietnam*

ARTICLE INFO

Article history:
Received October 9, 2024
Received in revised form January 28, 2025
Accepted March 5, 2025

KEYWORDS:
Cocoa,
expression pattern,
investigation,
trehalose-6-phosphate phosphatase,
trehalose-6-phosphate synthase,

ABSTRACT

Cocoa (*Theobroma cacao* L.), a vital industrial crop renowned for its economic and nutritional significance, faces increasing challenges due to climate change-induced stresses. To enhance the understanding of cocoa's adaptive mechanisms, a comprehensive analysis was conducted on the trehalose-6-phosphate phosphatase (*TPP*) and trehalose-6-phosphate synthase (*TPS*) gene families, which play crucial roles in plant stress responses and development. Five *TcTPP* and eight *TcTPS* genes were identified using the latest cocoa genome assembly, distributed unevenly across nine of the ten chromosomes. Detailed physicochemical characterization revealed significant variability in amino acid length, molecular weight, isoelectric point, and hydrophilicity among these proteins, suggesting functional diversity. Phylogenetic analyses, performed using the maximum likelihood method, classified the *TcTPP* family into three distinct clades and the *TcTPS* family into two main groups. Gene structure examination uncovered variations in exon-intron organization, with *TcTPP* genes containing nine to twelve exons and *TcTPS* genes ranging from three to eighteen exons, indicating structural diversity within these families. Based on publicly available datasets, expression profiling demonstrated differential expression patterns of *TcTPP* and *TcTPS* genes during embryo development and under biotic stress conditions, such as pathogen infection by *Phytophthora megakarya*. Certain genes exhibited significant upregulation or downregulation in response to stress, implicating them in cocoa's defense mechanisms. Taken together, this study provides valuable insights into the *TPP* and *TPS* gene families in cocoa. It lays a foundation for developing strategies to enhance stress tolerance and sustainability in cocoa cultivation amidst changing climatic conditions.

1. Introduction

Cocoa (*Theobroma cacao* L.), a prominent industrial crop within the Malvaceae family, holds considerable global economic and cultural significance (Diaz-Valderrama *et al*. 2020). Native to the tropical regions of Central and South America, cocoa is now cultivated in over fifty countries situated in humid tropical zones around the world (Diaz-Valderrama *et al*. 2020; Jaimez *et al*. 2022). The cocoa beans produced by this

* Corresponding Author
E-mail Address: cd.ha@vnu.edu.vn

plant are a rich source of essential nutrients, minerals, and antioxidants, making them highly valuable for a variety of applications, including the production of chocolate, confectionery goods, and cosmetic products (Tan *et al*. 2021; Samanta *et al*. 2022). The demand for cocoa-based products continues to grow internationally, underscoring its importance in global trade and industry. However, the cultivation of cocoa is increasingly threatened by the adverse effects of climate change. Adverse environmental conditions, including prolonged droughts, irregular rainfall patterns, heat stress, and an increase in pest and disease outbreaks, are significantly affecting cocoa production (Schroth *et al*. 2016; Gateau-Rey *et al*. 2018). These stressors not only reduce yields and compromise bean quality but also endanger the livelihoods of millions of smallholder farmers globally. Addressing these challenges requires a deeper understanding of the molecular mechanisms underlying cocoa's ability to adapt and respond to these stress conditions. Identifying and characterizing genes associated with these mechanisms can provide critical insights for breeding more resilient cocoa varieties (Adeniyi 2019; Pokou *et al*. 2019).

Trehalose is a non-reducing disaccharide found widely across higher plants, where it plays a crucial role in protecting cellular membranes and proteins from adverse environmental stress (Iordachescu & Imai 2008; Paul *et al*. 2018). Additionally, trehalose is involved in plant growth and development via various physiological and biochemical mechanisms (Hassan *et al*. 2023). The biosynthesis of trehalose in plants is primarily facilitated by two key enzymes: trehalose-6-phosphate synthase (*TPS*) and trehalose-6-phosphate phosphatase (*TPP*). *TPS* catalyzes the conversion of uridine diphosphate-glucose and glucose-6-phosphate into trehalose-6-phosphate, which *TPP* subsequently dephosphorylates to produce trehalose (Iordachescu & Imai 2008; Fichtner & Lunn 2021). In plants, *TPS* is composed of two domains: an N-terminal domain that binds the nucleotide cofactor UDP-glucose and a C-terminal domain that binds glucose-6-phosphate (Ponnu *et al*. 2011). The *TPP* enzyme, on the other hand, contains a catalytic domain with a central beta-sheet surrounded by alpha-helices and a regulatory domain that modulates its activity (Paul *et al*. 2018). Recently, these enzyme families have been studied in various higher plant species, including the model plant *Arabidopsis thaliana* (Vandesteene *et al*. 2012; Yang *et al*. 2012), *Populus trichocarpa* (Gao *et al*. 2021; Yang *et al*. 2012), tomato (*Solanum lycopersicum*) (Mollavali & Börnke 2022), and peanut (*Arachis hypogea*) (Chu *et al*. 2024; La *et al*. 2022). However, information on the *TPS* and *TPP* families in cocoa is currently unavailable.

In this study, bioinformatics tools were employed to identify the *TPS* and *TPP* gene families within the peanut genome. Analyses were performed to investigate the evolutionary dynamics of these gene families, including gene duplication, gene structure, and phylogenetic relationships. Additionally, the protein characteristics and subcellular localization of *TPS* and *TPP* proteins in peanut plants were analyzed. The expression profiles of the *TPS* and *TPP* genes were also examined across major stages of embryo development in cocoa.

2. Materials and Methods

2.1. Data Assemblies

The newest cocoa genome (NCBI RefSeq assembly: GCF_000208745.1) (Argout *et al*. 2011) available in the Phytozome and NCBI was utilized for the *in silico* analyses.

2.2. Physical and Chemical Characterization

To analyze the physicochemical characteristics of proteins, we utilized the ExPASy ProtParam tool (Gasteiger *et al*. 2003; Gasteiger *et al*. 2005) as previously described (La *et al*. 2022; Chu *et al*. 2024). Amino acid sequences of each protein were retrieved from the cocoa proteome database and individually submitted to the ProtParam server (https://web.expasy.org/protparam/). This tool calculated various properties based on the input protein sequences, including molecular weight (kDa), amino acid length, theoretical isoelectric point (pI), grand average of hydropathy (GRAVY) and aliphatic index (AI).

2.3. Construction of Phylogenetic Tree

To construct the phylogenetic tree of proteins, we employed the Maximum Likelihood algorithm as previously described (La *et al*. 2022; Chu *et al*. 2024). Initially, *TPP* and *TPS* protein sequences from cocoa and well-characterized *TPP* and *TPS* protein sequences from *Arabidopsis* and *Populus* species (Vandesteene *et al*. 2012; Gao *et al*. 2021). Multiple sequence alignment was performed using ClustalW (Thompson *et al*. 2002; Larkin *et al*. 2007) integrated within MEGA X software (Kumar *et al*.

2018), utilizing default alignment parameters to ensure consistency. The aligned sequences were then subjected to phylogenetic analysis using the Maximum Likelihood approach implemented in MEGA X software (Kumar *et al*. 2018). To evaluate the strength of the phylogenetic relationships, a bootstrap analysis with 1,000 replicates was conducted.

2.4. Exon and Intron Analysis

To analyze the exon and intron organization of the identified genes, we employed the Gene Structure Display Server web-based tool (Hu *et al*. 2015) as previously described (La *et al*. 2022; Chu *et al*. 2024). Initially, both the coding sequences and the corresponding full-length genomic DNA sequences of each gene were retrieved from the genomic database. These sequences were then uploaded to the Gene Structure Display Server platform (Hu *et al*. 2015), which aligns the coding sequences with the genomic sequences to generate a graphical representation of the gene structure. Adobe Illustrator software was then utilized to illustrate the distribution, length, and number of exons and introns within each gene.

2.5. Expression Analysis

To investigate the gene expressions during different developmental stages of somatic and zygotic embryos of cocoa plants, we analyzed data from a public repository (accession number: GSE55476) in a recent study (Maximova *et al*. 2014), accessible via the NCBI Gene Expression Omnibus database (Barrett *et al*. 2013; Clough & Barrett 2016). Additionally, the expression profiles of genes under pathogen infection were examined by analyzing a previously published microarray dataset (accession number: GSE116041) (Pokou *et al*. 2019). In this study, "inoculation" refers to the deliberate introduction of the fungal pathogen *Phytophthora megakarya* onto cocoa plants to simulate infection conditions. Relative expression levels were calculated using Actin 11 as a reference gene, recognized for its stable expression across various cocoa tissues (Pinheiro *et al*. 2011), following the approach described recently (Cao 2022). Genes were classified as upregulated or downregulated based on an absolute fold-change threshold of ≥1.5 when comparing time points at 6, 24, and 72 hours after inoculation to the baseline at 0 hour after inoculation (Cao 2022).

3. Results

3.1. Survey of the *TPP* and *TPS* Families in Cocoa

In order to identify and annotate the *TPP* and *TPS* families in cocoa, the conserved domains of these enzymes were screened in the newest genome. Available in the Phytozome and NCBI. A total of five *TPP* genes (*TcTPP*) and eight *TPS* genes (*TcTPS*) were identified in the cocoa genome, distributed across nine out of the ten chromosomes. The gene names were renamed from *TcTPP1* to *TcTPP5* and *TcTPS1* to *TcTPS5* based on their locations and starting position on the chromosomes, beginning with chromosome 1. Out of the five *TcTPP* genes, three members were located on chromosome 3, one each on chromosomes 1 and 8. Meanwhile, two members of the *TcTPS* family were detected on chromosomes 2, 3, and 6 each. Only one *TcTPS* gene was distributed on chromosomes 7 and 9. No *TcTPP* and *TcTPS* genes were detected on chromosomes 4, 5, and 10, indicating an unequal distribution of *TcTPP* and *TcTPS* genes throughout the cocoa genome (Figure 1). The absence of *TcTPP* and *TcTPS* genes on certain chromosomes may be attributed to several factors, including chromosomal evolution, gene loss, or duplication events during the evolution

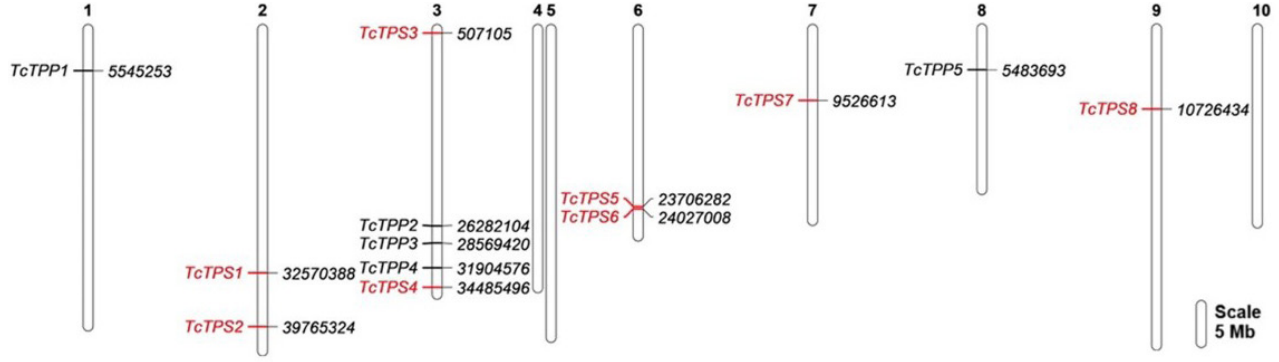

Figure 1. Physical locations of the *TcTPP* and *TcTPS* gene families in the cacao genome

of the cocoa genome. Chromosomes 4, 5, and 10 might have undergone evolutionary pressures that led to the loss of these gene families in these regions, potentially due to non-essential roles of *TPP* and *TPS* genes in these chromosomes or functional redundancy provided by genes on other chromosomes. Additionally, certain chromosomes may lack the necessary regulatory sequences or chromatin structure conducive to the retention of these gene families. Further comparative genomic analyses with closely related species could provide insights into whether this absence is a common evolutionary pattern or a unique feature of cocoa.

3.2. Physical and Chemical Characterization of The *TPP* and *TPS* Families in Cocoa

In order to calculate the physicochemical characteristics of the *TcTPP* and *TcTPS* proteins in cocoa, the ExPASy ProtParam tool was utilized to analyze the individual full-length protein sequence. As a result, the physical and chemical characterization of the *TcTPP* and *TcTPS* families was provided in Table 1.

The *TcTPP* and *TcTPS* enzymes in cocoa displayed amino acid lengths ranging from 341 residues (*TcTPP3*) to 406 residues (*TcTPP4*) and from 838 residues (*TcTPS8*) to 948 residues (*TcTPS7*), with mW varying between 38.49 kDa in *TcTPP3* and 45.48 kDa in *TcTPP4*, and between 95.43 kDa in *TcTPS8* and 106.88 kDa in *TcTPS7*. The pI values, which indicated the optimal pH for protein activity, spanned from 5.98 (*TcTPP4*) to 9.45 (*TcTPP5*) and from 5.80 (*TcTPS6*) to 6.45 (*TcTPS3*) in the *TcTPP* and *TcTPS* families in cocoa, respectively. Interestingly, a total of nine *TcTPP* and *TcTPS* proteins have predicted pI values below 6.50, indicating that most of these proteins are acidic, with potential roles in interacting with alkaline cellular environments or facilitating enzymatic activity under specific pH conditions. Next, the GRAVY scores of all members of the *TcTPP* and *TcTPS* families were negative, ranging from -0.33 (*TcTPP2*) to -0.47 (*TcTPP4*) and from -0.17 (*TcTPS1*) to -0.39 (*TcTPS7*), suggesting that these proteins are hydrophilic. This hydrophilic nature is likely essential for their solubility and functionality in aqueous environments, where they may mediate key metabolic and stress-response processes. Finally, the AI values of these *TcTPP* and *TcTPS* proteins in cocoa varied from 79.01 (*TcTPP1*) to 92.24 (*TcTPS4*), suggesting that these proteins possess a high degree of thermal stability, which could be critical for maintaining their structure and activity under stress conditions such as elevated temperatures. Collectively, these physicochemical characteristics highlight the adaptability of *TcTPP* and *TcTPS* proteins in responding to the dynamic cellular environment of cocoa plants, particularly under stress conditions.

3.3. Classification of The *TPP* and *TPS* Families in Cocoa

As a result, two phylogenetic trees of the *TPP* and *TPS* families in cocoa, *Arabidopsis* and *Populus* were described in Figures 2A and 2B. As expected, a Maximum-Likelihood phylogenetic analysis clustered 25 *TPP* sequences from three species into three main clades (Figure 2A). These sequences included five from cocoa, 10 from *Arabidopsis*, and 10 from Populus. Clades A and B were further subdivided into three and two subgroups, respectively. The *TcTPP* proteins from cocoa were distributed across all three clades. Specifically, clade A contained two *TcTPP* proteins (*TcTPP1* and *TcTPP5*), clade B comprised two TcTPP proteins (*TcTPP2* and *TcTPP4*), with only one member of the *TcTPP* family in clade C. Next, Figure 2B revealed that the *TPS* families in

Table 1. Survey of the *TcTPP* and *TcTPS* families in cacao

Gene name	Locus name	Lenght (aa)	mW (kDa)	pI	Gravy	AI
TcTPP1	Thecc.01G104100	375	42.53	9.05	-0.41	79.01
TcTPP2	Thecc.03G156600	384	43.12	8.66	-0.33	80.96
TcTPP3	Thecc.03G194100	341	38.49	9.04	-0.36	84.90
TcTPP4	Thecc.03G256900	406	45.48	5.98	-0.47	80.89
TcTPP5	Thecc.08G109500	364	41.03	9.45	-0.34	84.62
TcTPS1	Thecc.02G245900	861	96.91	5.99	-0.17	91.10
TcTPS2	Thecc.02G308700	862	97.58	5.83	-0.20	90.08
TcTPS3	Thecc.03G008600	862	98.00	6.45	-0.20	87.71
TcTPS4	Thecc.03G306600	862	97.38	5.48	-0.19	92.24
TcTPS5	Thecc.06G160200	934	105.06	6.19	-0.31	90.69
TcTPS6	Thecc.06G165100	857	96.71	5.80	-0.25	84.70
TcTPS7	Thecc.07G134200	948	106.88	6.34	-0.39	84.01
TcTPS8	Thecc.09G163100	838	95.43	5.79	-0.32	86.40

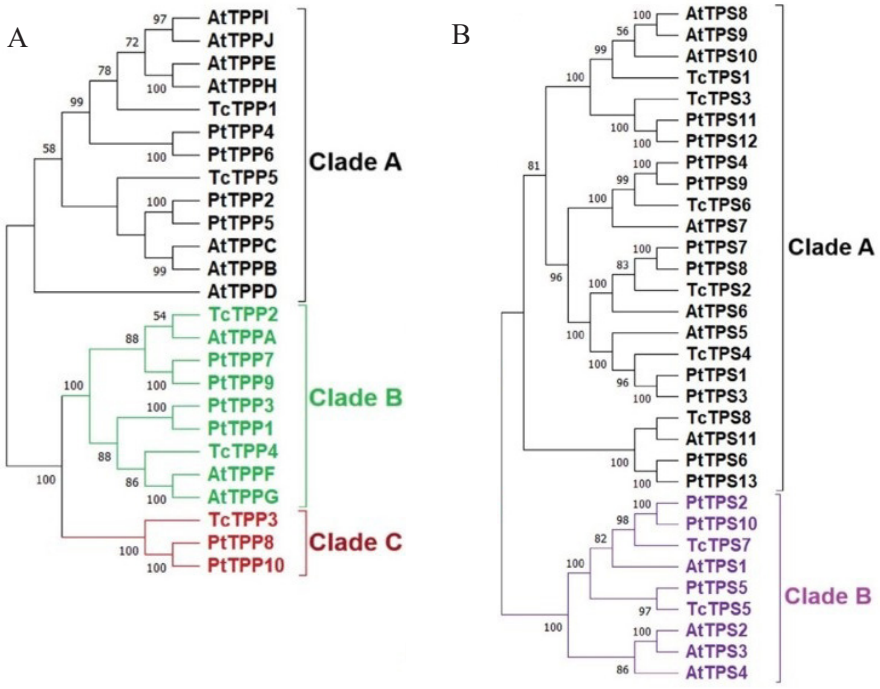

Figure 2. Classification of the (A) *TcTPP* and (B) *TcTPS* families in cacao based on well-characterized *TPP* and *TPS* protein in *Arabidopsis* and *Populus* species

cocoa, *Arabidopsis* and *Populus* could be classified into two main clades. Clade A included six *TcTPS* proteins and clade B encompassed two members of the *TcTPP* family.

3.4. Gene organization of The *TPP* and *TPS* Gene Families in Cocoa

In order to better understand the structural diversity of cocoa *TcTPP* and *TcTPS* genes, we examined their exon-intron architectures by using the Gene Structure Display Server website. As expected, Figures 3A and 3B illustrated the gene organization of the *TcTPP* and *TcTPS* gene families in cocoa, respectively. We found that the members of the *TcTPP* and *TcTPS* gene families in cocoa exhibited variations in the number of exons and the size of genomic sequences. Particularly, the exon counts among these *TcTPP* genes ranged from nine to 12, with *TcTPP1* having the highest number at 12 exons, noticeably more than any other *TcTPP* gene. Two members of the *TcTPP* gene family, including *TcTPP2* and *TcTPP5*, contained 11 exons, whereas *TcTPP4* had 10 exons. Additionally, the *TcTPP* genes' genomic sequences varied from 2029 bp (*TcTPP3*) to 2451 bp (*TcTPP1*). Next, the exon amounts of the TcTPS gene family in cocoa ranged from three to 18. Among them, a majority of members of the *TcTPS* gene family, particularly six out of eight (*TcTPS1*, *TcTPS2*, *TcTPS3*, *TcTPS4*, *TcTPS6*, and *TcTPS8*), had only three exons. One gene, namely *TcTPS7* contained 17 exons, while *TcTPS5* had 18 exons. Meanwhile, the genomic sequences of the *TcTPS* gene family in cocoa varied from 2782 bp (*TcTPS8*) to 10172 bp (*TcTPS7*).

3.5. Expression Patterns of The *TPP* and *TPS* Gene Families in Cocoa

In order to analyze the expression profiles of the *TcTPP* and *TcTPS* gene families in various organs, two recent datasets, including GSE55476 and GSE116041, were utilized. As a result, the expression patterns of the *TcTPP* and *TcTPS* gene families were provided in Figures 4 and 5. In particular, all members of the *TcTPP* gene family showed variable expression levels in various embryo samples during the development of fruits (Figure 4A). Only one gene, namely *TcTPP4*, exhibited a higher expression pattern in the mature stage of zygotic embryo samples than in other examined samples. Three genes, including *TcTPP1*, *TcPP3*, and *TcTPP5*, had lower expression patterns in at least one developmental stage of zygotic and somatic embryo samples. Under the biotic stress condition, the *TcTPP* genes in cocoa also had differential expression levels (Figure 4B). For example, *TcTPP1* was noted to reduce in Nanay genotypes at 24 hours after inoculation of *Phytophthora megakarya* (~ -2.20-fold) but up-regulate in Scavina genotypes at 24 and 72 hours after treatment by ~ 3.40-fold and ~ 2.14-fold, respectively. We also found that TcTPP3 was up-

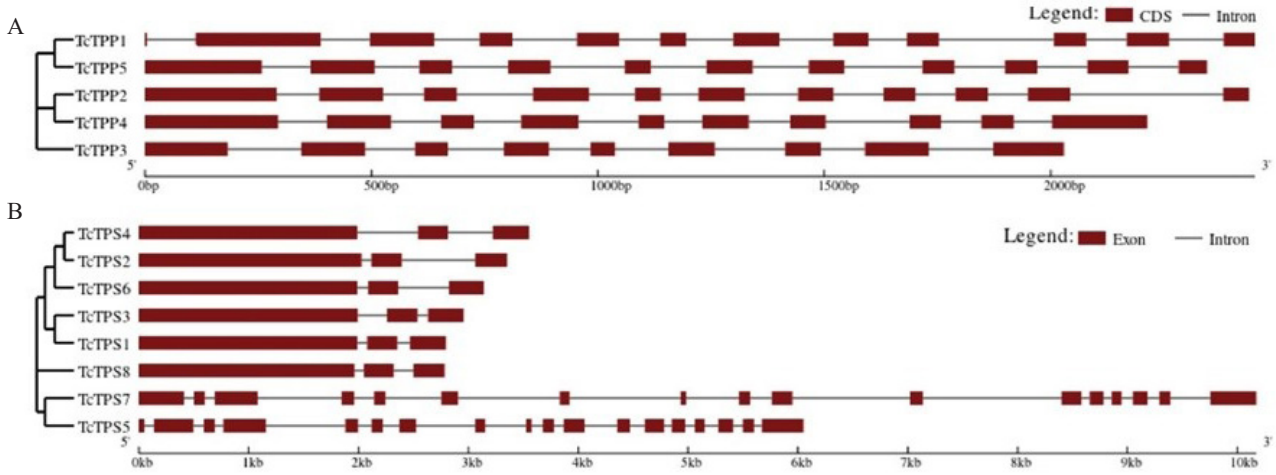

Figure 3. Gene organization of the (A) *TcTPP* and (B) *TcTPS* gene families in cacao

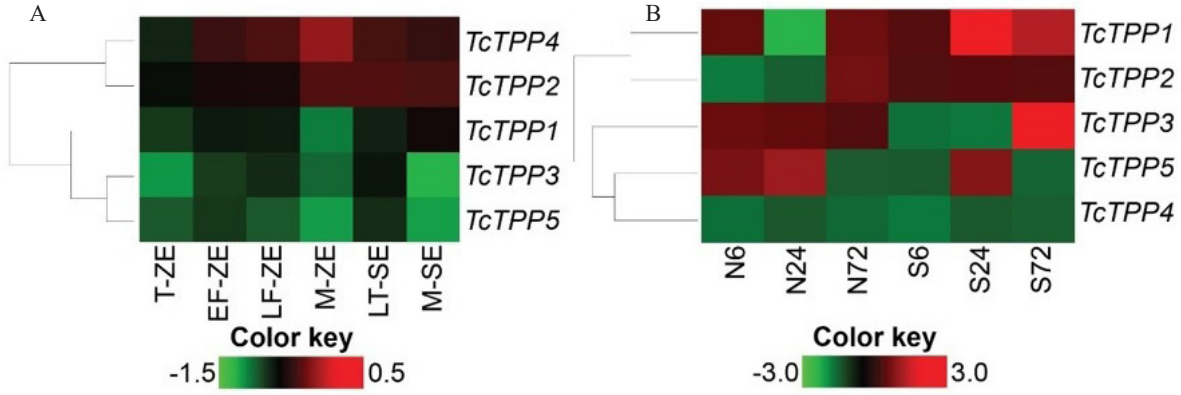

Figure 4. Expression profiles of the *TcTPP* gene family in (A) embryo samples, including zygotic embryos at 14 (T-ZE), 16 (EF-ZE), 18 (LF-ZE), 20 (M-ZE) weeks after pollination and somatic embryos at whole late torpedo (LT-SE), mature (M-SE) stage, and (B) Nanay genotypes at 6 (N6), 24 (N24), 72 (N72) hours after inoculation with *Phytophthora megakarya* treatment and Scavina genotypes at 6 (S6), 24 (S24) 72 (S72) hours after treatment

regulated in Scavina genotypes at 72 hours after treatment (~ 3.33-fold), while *TcTPP5* was induced in both Nanay and Scavina genotypes at 24 hours after treatment by ~ 1.83-fold and ~ 1.57-fold, respectively.

Similarly, the expression patterns of the *TcTPS* gene family in cocoa were also variable. During the developmental stage of zygotic and somatic embryos, the expression level of *TcTPS5* was not specific in any embryo samples, whereas *TcTPS2* showed a higher expression in zygotic embryos at 20 weeks after pollination (Figure 5A). Additionally, the expression pattern of *TcTPS8* was low in three developmental stages of zygotic embryos. Under the biotic stress treatment, *TcTPS3* and *TcTPS5* were up-regulated and down-regulated in both two genotypes, respectively (Figure 5B). Interestingly, *TcTPS1* was reduced in Nanay genotypes at 24 hours after treatment (~ -1.65-fold) but induced in Scavina genotypes at 72 hours after treatment (~ 1.70-fold). Another member of the *TcTPS* gene family, namely *TcTPS8*, was up-regulated in Nanay genotypes at 6 (~ 1.80-fold) and 72 (~ 1.72-fold) hours after treatment, but down-regulated at 24 hours after treatment (~ -1.63-fold). This gene was also induced in Scavina genotypes at 6 hours after treatment (~ 2.15-fold).

4. Discussion

4.1. Comparative Genomic Analysis of *TPP* and *TPS* Gene Families Across Plant Species

Recently, enzymes involved in trehalose biosynthesis have been identified in various higher plant species (Zang *et al*. 2011; Gao *et al*. 2021; Song *et al*. 2021; Du *et al*. 2022; Hu *et al*. 2022; Liu & Zhou 2022; Wang *et al*. 2022). Specifically, 12 and 10 members of the

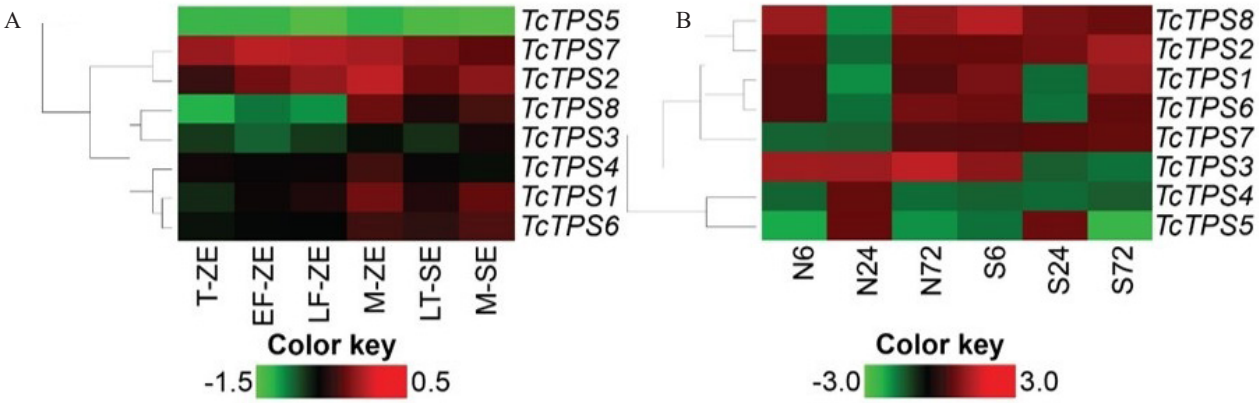

Figure 5. Expression profiles of the *TcTPS* gene family in (A) embryo samples, including zygotic embryos at 14 (T-ZE), 16 (EF-ZE), 18 (LF-ZE) and 20 (M-ZE) weeks after pollination and somatic embryos at whole late torpedo (LT-SE), mature (M-SE) stage, and (B) Nanay genotypes at 6 (N6), 24 (N24), 72 (N72) hours after inoculation with *Phytophthora megakarya* treatment and Scavina genotypes at 6 (S6), 24 (S24) 72 (S72) hours after treatment

TPP family were discovered in rice and *A. thaliana*, respectively (Vandesteene *et al*. 2012; Rahman *et al*. 2020). In the *Populus* genome, 10 *TPP* and 13 *TPS* genes were identified (Gao *et al*. 2021). In wheat, 31 *TPP* family genes were recently reported (Du *et al*. 2022). For cotton, the *TPP* family consists of 12, 17, 24, and 26 members in *G. raimondii*, *G. arboreum*, *G. hirsutum*, and *G. barbadense*, respectively (Wang *et al*. 2022). Additionally, eight *TPS* genes were identified as being distributed across sweet orange chromosomes (Liu & Zhou 2022), while 11 genes encoding *TPS* proteins were documented in rice (Zang *et al*. 2011). Furthermore, 11 and 26 *TPS* genes have been found in *M. truncatula* and *B. napus* (Song *et al*. 2021; Hu *et al*. 2022). Overall, our comparisons suggested that the *TPS* and *TPP* gene families in cocoa, and potentially in other plant species, are multi-gene families.

4.2. Structural and Biochemical Diversity of *TPP* and *TPS* Gene Families in Higher Plants

Previous studies supported our categorization of the *TcTPP* and *TcTPS* families in cocoa. For example, an unrooted phylogenetic analysis of 86 *TPP* family members from wheat, maize, rice, Brachypodium, *Arabidopsis*, and *Populus* indicated that these proteins can be classified into three subfamilies (Du *et al*. 2022). Similarly, recent studies have constructed unrooted phylogenetic trees for *TPP* families in *Arabidopsis* and cotton, dividing these proteins into three main clades (Wang *et al*. 2022). This classification pattern was also observed in the phylogenetic analysis of *TPP* proteins in *Populus* species (Gao *et al*. 2021). Additionally, an unrooted phylogenetic tree encompassing all *TPS* proteins from Populus, rice, *Arabidopsis*, soybean, and *M. domestica* revealed a categorization similar to that of the cocoa *TcTPS* family (Gao *et al*. 2021). These comparisons suggested that the *TcTPP* and *TcTPS* families in cocoa, and possibly in other higher plant species, can be classified into three and two typical groups, respectively.

Recent studies have extensively analyzed the physicochemical properties of *TPP* and *TPS* families in various higher plant species, supporting our characterization of the *TcTPP* and *TcTPS* families in cocoa. Previous studies have demonstrated that the sizes of *TPP* family proteins in four cotton species vary between 134 and 422 amino acid residues, with an average molecular weight of 39.54 kDa and an average pI value of 8.41 (Wang *et al*. 2022). In wheat, *TPP* proteins have been reported to range from 249 to 584 amino acids in length, averaging 386 residues, and their calculated mW values spanned from 28.66 to 96.02 kDa (Du *et al*. 2022). The theoretical pI values of wheat TPP proteins also exhibited significant variation, ranging from acidic (pI = 5.53) to alkaline (pI = 9.26) (Du *et al*. 2022). In the case of Populus, *TPS* proteins have been found to be between 846 and 922 amino acid residues long, with mW values ranging from 84.55 to 90.46 kDa (Gao *et al*. 2021). The *TPP* proteins in P. tomentosa are shorter than their *TPS* proteins, ranging from 235 to 387 amino acids, and possess mW values between 75.10 and 83.08 kDa (Gao *et al*. 2021). Notably, all *TPS* and *TPP* proteins in *P. tomentosa* have been characterized as relatively hydrophilic (Gao *et al*. 2021). Taken together, our comparisons strongly suggested that the *TPP* and *TPS* families in higher

plant species exhibited significant diversity in protein size, mW, pI, and hydrophilicity.

Comparative studies on gene structure further reinforce our findings in cocoa. Recent investigations have increasingly examined the gene structures of *TPP* and *TPS* gene families in higher plant species, including rice, wheat, cotton, and rapeseed. For instance, rice *TPP* genes commonly feature motifs comprising 10 or 11 exons (Rahman *et al*. 2021). In wheat, the exon count within the *TPP* gene family ranges from five to 13, with the majority of genes containing nine exons (14 out of 31 members) or ten exons (10 out of 31 members) (Du *et al*. 2022). The *TPP* gene family in four cotton species has been characterized as having complex gene structures; notably, 53 out of 79 *TPP* genes in these species possess at least ten exons (Wang *et al*. 2022). Similarly, the number of exons in the *TPS* genes of rapeseed varies significantly, spanning from three to 18 (Xia *et al*. 2021). Our study showed that cocoa's *TcTPP* and *TcTPS* genes often share similar exon-intron organizations within the same phylogenetic branch. Specifically, three exons were most commonly observed in the *TcTPP* gene family, while the number of exons in the *TcTPP* gene family was variable.

In conclusion, our comprehensive analysis of the *TcTPP* and *TcTPS* gene families in cocoa revealed significant genetic diversity and their potential roles in development and stress responses. Five *TcTPP* and eight *TcTPS* genes were identified with uneven chromosomal distribution. Physicochemical characterization revealed that these genes encode proteins with varying amino acid lengths, mW, pI, AI, and GRAVY scores. Phylogenetic classification grouped the *TcTPP* proteins into three main clades and the *TcTPS* proteins into two. Examination of exon-intron structures showed considerable variation in exon numbers and genomic sequence lengths among the genes. Expression profiling across different developmental stages and under biotic stress conditions demonstrated differential expression patterns, with certain genes being upregulated or downregulated, implying their involvement in embryo development and stress responses. These findings provide a foundation for future research to improve cocoa's resilience and productivity under environmental challenges.

Acknowledgements

This work was funded by the fundamental research program of Hung Vuong University under the project grant No. 06/2024/KHCN (HV06.2024).

References

Adeniyi, D., 2019. Diversity of cacao pathogens and impact on yield and global production. in: Aikpokpodion, P.O., (Eds.), *Theobroma Cacao-Deploying Science for Sustainability of Global Cocoa Economy*. IntechOpen., London, pp. 1-20.

Argout, X., Salse, J., Aury, J.M., Guiltinan, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J.F., Sabot, F., Kudrna, D., Ammiraju, J.S., Schuster, S.C., Carlson, J.E., Sallet, E., Schiex, T., Dievart, A., Kramer, M., Gelley, L., Shi, Z., Berard, A., Viot, C., Boccara, M., Risterucci, A.M., Guignon, V., Sabau, X., Axtell, M.J., Ma, Z., Zhang, Y., Brown, S., Bourge, M., Golser, W., Song, X., Clement, D., Rivallan, R., Tahi, M., Akaza, J.M., Pitollat, B., Gramacho, K., D'Hont, A., Brunel, D., Infante, D., Kebe, I., Costet, P., Wing, R., McCombie, W.R., Guiderdoni, E., Quetier, F., Panaud, O., Wincker, P., Bocs, S., Lanaud, C., 2011. The genome of Theobroma cacao. *Nat Genet*, 43, 101-108. doi:10.1038/ng.736

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res*. 41, D991-995. doi:10.1093/nar/gks1193

Cao, P.B., 2022. In silico structural, evolutionary, and expression analysis of small heat shock protein (shsp) encoding genes in cocoa (*Theobroma cacao* L.). *Journal of Animal and Plant Sciences*. 32, 1394-1402. doi:https://doi.org/10.36899/JAPS.2022.5.0546

Chu, H.D., Yen, T.T.H., Pham, C.T., Quynh, L.T.N., Huyen, T.T.T., Trung, N.Q., Gioi, D.H., Tien, T.V., Cao, P.B., 2024. Identification of two enzymes for trehalose synthesis and their potential function in growth and development in peanut (*Arachis hypogaea*). *Journal of Tropical Life Science*. 14, 83-94. doi:10.11594/jtls.14.01.10

Clough, E., Barrett, T., 2016. The gene expression omnibus database. *Methods Mol. Biol*. 1418, 93-110. doi:10.1007/978-1-4939-3578-9_5

Diaz-Valderrama, J.R., Leiva-Espinoza, S.T., Aime, M.C., 2020. The History of Cacao and Its Diseases in the Americas. *Phytopathology*. 110, 1604-1619. doi:10.1094/PHYTO-05-20-0178-RVW

Du, L., Li, S., Ding, L., Cheng, X., Kang, Z., Mao, H., 2022. Genome-wide analysis of trehalose-6-phosphate phosphatases (TPP) gene family in wheat indicates their roles in plant development and stress response. *BMC Plant Biol*. 22, 120. doi:10.1186/s12870-022-03504-0

Fichtner, F., Lunn, J.E., 2021. The role of trehalose 6-phosphate (Tre6P) in plant metabolism and development. *Annual Review of Plant Biology*. 72, 737-760. https://doi.org/10.1146/annurev-arplant-050718-095929

Gao, Y., Yang, X., Yang, X., Zhao, T., An, X., Chen, Z., 2021. Characterization and expression pattern of the trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase gene families in Populus. *Int J Biol Macromol*. 187, 9-23. doi:10.1016/j.ijbiomac.2021.07.096

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 Res.* 31, 3784-3788. doi:10.1093/nar/gkg563

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. in: Walker, J.M., (Eds.), *The Proteomics Protocols Handbook.* Humana Press., UK, pp. 571-607.

Gateau-Rey, L., Tanner, E.V.J., Rapidel, B., Marelli, J.P., Royaert, S., 2018. Climate change could threaten cocoa production: Effects of 2015-16 El Nino-related drought on cocoa agroforests in Bahia, Brazil. *PLoS One*. 13, e0200454. doi:10.1371/journal.pone.0200454

Hassan, M.U., Nawaz, M., Shah, A.N., Raza, A., Barbanti, L., Skalicky, M., Hashem, M., Brestic, M., Pandey, S., Alamri, S., Mostafa, Y.S., Sabagh, A.E.L., Qari, S.H., 2023. Trehalose: A key player in plant growth regulation and tolerance to abiotic stresses. *Journal of Plant Growth Regulation*. 42, 4935-4957. doi:10.1007/s00344-022-10851-7

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, 1296-1297. doi:10.1093/bioinformatics/btu817

Hu, M., Xie, M., Cui, X., Huang, J., Cheng, X., Liu, L., Liu, S., Tong, C., 2022. Genome-wide characterization of trehalose-6-phosphate synthase gene family of *Brassica napus* and potential links with agronomic traits. *Int J Mol Sci*. 23. https://mdpi-res.com/d_attachment/ijms/ijms-23-15714/article_deploy/ijms-23-15714-v2.pdf?version=1671520935

Iordachescu, M., Imai, R., 2008. Trehalose biosynthesis in response to abiotic stresses. *J Integr Plant Biol*. 50, 1223-1229. doi:10.1111/j.1744-7909.2008.00736.x

Jaimez, R.E., Barragan, L., Fernandez-Nino, M., Wessjohann, L.A., Cedeno-Garcia, G., Sotomayor Cantos, I., Arteaga, F., 2022. *Theobroma cacao* L. cultivar CCN 51: a comprehensive review on origin, genetics, sensory properties, production dynamics, and physiological aspects. *PeerJ*. 10, e12676. doi:10.7717/peerj.12676

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35, 1547-1549. doi:10.1093/molbev/msy096

La, H.V., Chu, H.D., Tran, C.D., Nguyen, K.H., Le, Q.T.N., Hoang, C.M., Cao, B.P., Pham, A.T.C., Nguyen, B.D., Nguyen, T.Q., Van Nguyen, L., Ha, C.V., Le, H.T., Le, H.H., Le, T.D., Tran, L.P., 2022. 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. doi:10.1016/j.gene.2022.146210

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and clustal X version 2.0. *Bioinformatics*. 23, 2947-2948. doi:10.1093/bioinformatics/btm404

Liu, K., Zhou, Y., 2022. Genome-wide identification of the trehalose-6-phosphate synthase gene family in sweet orange (*Citrus sinensis*) and expression analysis in response to phytohormones and abiotic stresses. *PeerJ*. 10, e13934. doi:10.7717/peerj.13934

Maximova, S.N., Florez, S., Shen, X., Niemenak, N., Zhang, Y., Curtis, W., Guiltinan, M.J., 2014. Genome-wide analysis reveals divergent patterns of gene expression during zygotic and somatic embryo maturation of *Theobroma cacao* L., the chocolate tree. *BMC Plant Biol*. 14, 185. doi:10.1186/1471-2229-14-185

Mollavali, M., Börnke, F., 2022. Characterization of trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase genes of tomato (*Solanum lycopersicum* L.) and analysis of their differential expression in response to temperature. *Int J Mol Sci*. 23, 11436. doi:10.3390/ijms231911436

Paul, M.J., Gonzalez-Uriarte, A., Griffiths, C.A., Hassani-Pak, K., 2018. The role of trehalose 6-phosphate in crop yield and resilience. *Plant Physiol*. 177, 12-23. doi:10.1104/pp.17.01634

Pinheiro, T.T., Litholdo, C.G., Jr., Sereno, M.L., Leal, G.A., Jr., Albuquerque, P.S., Figueira, A., 2011. Establishing references for gene expression analyses by RT-qPCR in *Theobroma cacao* tissues. *Genet Mol Res*. 10, 3291-3305. doi:10.4238/2011.November.17.4

Pokou, D.N., Fister, A.S., Winters, N., Tahi, M., Klotioloma, C., Sebastian, A., Marden, J.H., Maximova, S.N., Guiltinan, M.J., 2019. Resistant and susceptible cacao genotypes exhibit defense gene polymorphism and unique early responses to *Phytophthora megakarya* inoculation. *Plant Mol Biol*. 99, 499-516. doi:10.1007/s11103-019-00832-y

Ponnu, J., Wahl, V., Schmid, M., 2011. Trehalose-6-phosphate: connecting plant metabolism and development. *Front Plant Sci*. 2, 70. doi:10.3389/fpls.2011.00070

Rahman, M.M., Rahman, M.M., Eom, J.-S., Jeon, J.-S., 2020. Genome-wide identification, expression profiling and promoter analysis of trehalose-6-phosphate phosphatase gene family in rice. *J Plant Biol*. 64, 55-71. doi:10.1007/s12374-020-09279-x

Rahman, M.M., Rahman, M.M., Eom, J.-S., Jeon, J.-S., 2021. Genome-wide identification, expression profiling and promoter analysis of trehalose-6-phosphate phosphatase gene family in rice. *Journal of Plant Biology*. 64, 55-71. doi:10.1007/s12374-020-09279-x

Samanta, S., Sarkar, T., Chakraborty, R., Rebezov, M., Shariati, M.A., Thiruvengadam, M., Rengasamy, K.R.R., 2022. Dark chocolate: An overview of its biological activity, processing, and fortification approaches. *Curr Res Food Sci*. 5, 1916-1943. doi:10.1016/j.crfs.2022.10.017

Schroth, G., Laderach, P., Martinez-Valle, A.I., Bunn, C., Jassogne, L., 2016. Vulnerability to climate change of cocoa in West Africa: Patterns, opportunities and limits to adaptation. *Sci Total Environ*. 556, 231-241. doi:10.1016/j.scitotenv.2016.03.024

Song, J., Mao, H., Cheng, J., Zhou, Y., Chen, R., Zeng, L., Li, H., Wang, Y., 2021. Identification of the trehalose-6-phosphate synthase gene family in *Medicago truncatula* and expression analysis under abiotic stresses. *Gene*. 787, 145641. doi:https://doi.org/10.1016/j.gene.2021.145641

Tan, T.Y.C., Lim, X.Y., Yeo, J.H.H., Lee, S.W.H., Lai, N.M., 2021. The health effects of chocolate and cocoa: A Systematic Review. *Nutrients*. 13. doi:10.3390/nu13092909

Thompson, J.D., Gibson, T.J., Higgins, D.G., 2002. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. Chapter 2, Unit 2.3. doi:10.1002/0471250953.bi0203s00

Vandesteene, L., López-Galvis, L., Vanneste, K., Feil, R., Maere, S., Lammens, W., Rolland, F., Lunn, J.E., Avonce, N., Beeckman, T., Van Dijck, P., 2012. Expansive evolution of the trehalose-6-phosphate phosphatase gene family in *Arabidopsis*. *Plant Physiol*. 160, 884-896. doi:10.1104/pp.112.201400

Wang, W., Cui, H., Xiao, X., Wu, B., Sun, J., Zhang, Y., Yang, Q., Zhao, Y., Liu, G., Qin, T., 2022. Genome-wide identification of cotton (*Gossypium* spp.) trehalose-6-phosphate phosphatase (TPP) gene family members and the role of GhTPP22 in the response to drought stress. *Plants*. 11, 1079. doi:10.3390/plants11081079

Xia, J., Wang, D., Peng, Y., Wang, W., Wang, Q., Xu, Y., Li, T., Zhang, K., Li, J., Xu, X., 2021. Genome-wide analysis of the YABBY transcription factor family in rapeseed (*Brassica napus* L.). *Genes*. 12, 981. doi:10.3390/genes12070981

Yang, H.-L., Liu, Y.-J., Wang, C.-L., Zeng, Q.-Y., 2012. Molecular evolution of trehalose-6-phosphate synthase (TPS) gene family in Populus, *Arabidopsis* and Rice. *PLoS One*. 7, e42438. doi:10.1371/journal.pone.0042438

Zang, B., Li, H., Li, W., Deng, X.W., Wang, X., 2011. Analysis of trehalose-6-phosphate synthase (TPS) gene family suggests the formation of TPS complexes in rice. *Plant Mol Biol*. 76, 507-522. doi:10.1007/s11103-011-9781-1