

# Bioinformatic Analyses to Uncover Genes Involved in Trehalose Metabolism in the Polyploid Sugarcane

**Lauana Oliveira**

Universidade de São Paulo

**Bruno Navarro**

Universidade de São Paulo

**João Pedro Pereira**

Universidade de São Paulo

**Adriana Lopes**

Instituto Butantan

**Marina Martins**

Universidade de São Paulo

**Diego Riaño-Pachón**

Universidade de São Paulo

**Marcos Buckeridge** (✉ [msbuck@usp.br](mailto:msbuck@usp.br))

Universidade de São Paulo

---

## Research Article

**Keywords:** Bioethanol, Orthologous Groups, Sucrose, Sugarcane, Sugar Signaling, Trehalose-6-phosphate

**Posted Date:** January 11th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1226136/v1>

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

[Read Full License](#)

---

# Abstract

Trehalose-6-phosphate (T6P) is an intermediate of trehalose biosynthesis that plays an essential role in plant metabolism and development. Here, we comprehensively analyzed sequences from enzymes of trehalose metabolism in sugarcane, one of the main crops used for bioenergy production. We identified protein domains, phylogeny, and *in silico* expression levels for all classes of enzymes. However, post-translational modifications and residues involved in catalysis and substrate binding were analyzed only in trehalose-6-phosphate synthase (TPS) sequences. We retrieved 71 putative full-length TPS, 93 trehalose-6-phosphate phosphatase (TPP), and 3 trehalase (TRE) of sugarcane, showing all their conserved domains, respectively. Putative TPS (Classes I and II) and TPP sugarcane sequences were categorized into well-known groups reported in the literature. We measured the expression levels of the sequences from one sugarcane leaf transcriptomic dataset. Furthermore, TPS Class I has specific *N*-glycosylation sites inserted in conserved motifs and carries catalytic and binding residues in its TPS domain. Some of these residues are mutated in TPS Class II members, which implies loss of enzyme activity. Our approach retrieved many homo(eo)logous sequences for genes involved in trehalose metabolism, paving the way to discover the role of T6P signaling in sugarcane.

## Introduction

Sugars, and mainly sucrose, lie at the heart of plant metabolism. During photosynthesis, plants synthesize sucrose that is transiently stored in vacuoles, used for cellular activities, and exported from source to sink tissues to sustain metabolism and growth [1,2]. Simultaneously, starch accumulates in the leaf chloroplasts as short-term storage degraded during the night to meet the continuous carbon demand. Thus, plants regulate their sugar levels in temporal and spatial scales.

The appropriate balance among carbon assimilation, partitioning, and use is critical for plant development, survival, and reproductive success. Sugars function as substrates for growth and affect sugar-sensing systems that regulate how, when, and where sugars are utilized. Although both abundance and depletion of sugars significantly affect gene expression [3,4], resolving the mechanisms and physiological significance of sugar signaling in plants has proved to be challenging. This depends on multiple pathways that respond to different sorts of sugar, which interact with each other and in conjunction with additional nutrients (e.g., nitrogen and phosphorus) and the environmental and phytohormone responses [5–9]. As a signaling molecule, trehalose-6-phosphate (T6P), the intermediate of trehalose biosynthesis, is a sucrose-specific signal in plants and has a far-reaching influence on metabolism, growth, and development [10]. Therefore, T6P is a potential target for improving model and crop plants. Despite multiple trehalose biosynthetic routes, the only one found in plants involves the enzymes trehalose-6-phosphate synthase (TPS) and trehalose-6-phosphate phosphatase (TPP) [11]. TPS catalyzes the transfer of glucose from UDP-glucose to glucose-6-phosphate, producing T6P and uridine diphosphate. TPP dephosphorylates T6P to form trehalose and inorganic phosphate. Trehalose is a nonreducing disaccharide involved in osmoregulation and stress protection and its breakdown in several

organisms occurs via the hydrolytic enzyme trehalase (TRE), resulting in the formation of two glucose molecules.

TPS and TPP enzymes are encoded by multigenic families divided into distinct subfamilies based on their sequence similarity to the homologs *Saccharomyces cerevisiae* TPS (*ScTps1*) or TPP (*ScTps2*) [12–14]. *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays* TPS genes are divided into Class I (*AtTPS1–4*, *OsTPS1*, and *ZmTPSI.1.1-1.2*) and Class II (*AtTPS5-11*, *OsTPS2-11*, and *ZmTPSII.2.1-5.4*), both having a glycosyltransferase family 20 domain (TPS domain) and a trehalose-phosphatase domain (TPP domain) [15,16]. Class I's TPS domain is closely related to the *ScTps1* and *Escherichia coli* TPS (*otsA*), while Class II encodes proteins with higher similarity to *ScTps2*. Except for *AtTPS3*, only the Class I isoforms have catalytic activity confirmed by the complementation of yeast mutants [17–20]. Several Class II TPS genes respond to sugar availability [21–24] and hormones [18,25,26], besides displaying diurnal cycles of expression [27]. However, the function of most Class II genes remains enigmatic, although it has been suggested that they play regulatory roles. *Arabidopsis* has ten TPP genes (*AtTPPA-AtTPPJ*) containing a TPP domain, and three HAD motifs also found in Class II TPS proteins and a highly variable N-terminal region of unknown function [14]. All *AtTPP* genes originated from whole-genome duplication, and their encoded proteins are catalytically functional [28]. Poaceae species present similar quantities of *TPP* genes: 10 in rice and 11 in maize [15,16]. Different from *TPS* and *TPP*, *AtTRE* is encoded by a single gene. Based on the distribution of TPS, TPP, and TRE genes, trehalose metabolism appears to be universal in the plant kingdom and has ancient origins [12–14].

T6P levels are positively correlated with sucrose in several tissues, developmental stages, and species [29,30]. T6P was proposed as a signal of sucrose availability that also exerts a negative feedback regulation on sucrose levels to maintain them within a proper range according to the cellular metabolic status and the plant developmental stage [31,32]. However, this function is unclear in species such as sugarcane (*Saccharum* spp), accumulating a large amount of sucrose in its stems. This species is an important crop worldwide used as feedstock for sugar and bioethanol production. Brazil was the first country to introduce bioethanol as an efficient renewable fuel for transportation and stands out as one of the largest bioethanol exporters [33]. Improving sugar production per unit area and/or sucrose concentration in the stems has been an important goal of breeding programs [34,35]. Nevertheless, gains in sucrose content in commercial sugarcane varieties are about 1.0–1.5% per year and are believed to be near their limit [36–38].

The sugarcane genome is complex, interspecific, polyploid, and displays extensive aneuploidy [39], but research in sugarcane genomics has advanced [40–51]. One monoploid mosaic reference genome for the sugarcane hybrid R570 was released [52] along with a high-quality chromosome level genome for *S. spontaneum*, one of the parents of sugarcane hybrids [53]. Nevertheless, the incomplete coverage of the whole genome still hampers sugarcane biotechnological improvement. We identified the sequences and evolutionary relationships among *A. thaliana* and Panicoideae members of gene families encoding enzymes involved in trehalose metabolism. We combined search of orthologous genes in Viridiplantae, phylogenetic analysis, identification of functional protein domains and residues involved in catalysis, and

binding. These analyses were further combined with three-dimensional structure prediction, post-translational modifications, and *in silico* expression profiles. This approach should establish a foundation for further functional studies to uncover the physiological roles of T6P signaling in sugarcane. Such knowledge will help decipher the regulation of carbon partitioning and allocation, essential for the more efficient conversion of sugarcane biomass into bioproducts.

## Results

### Identification of sugarcane sequences involved in trehalose metabolism

Protein sequences of TPS, TPP, and TRE from *A. thaliana*, *Z. mays*, and *O. sativa* were used as queries to identify the orthologous groups (OGs) that they have been assigned to in the EggNOG database [54]. These OGs made it possible to retrieve homologous sequences from other Viridiplantae species (Supplementary Table S1). Subsequently, publicly available transcriptomic and genomic datasets from sugarcane (Table 1) were annotated with EggNOG.

The OGs obtained from model species and sugarcane were joined, recovering 15 OGs: nine for TPS, five for TPP, and one for TRE (Supplementary Table S1). The search on the sugarcane databases retrieved 444 sequences from all homo(eo)logous targets (Table 2). This is conceivable because the datasets used represent distinct sugarcane cultivars or genotypes, including the *S. spontaneum* genome (Figure 1b). Three OGs were excluded from further analyses: OG 1EKVF presented WD40 domains, which did not correspond to TPS proteins; OG 1EDRK and 1EDRH (Supplementary Table S1) contained only Arabidopsis sequences. These findings left us with 12 OGs and 430 sequences. To further assess the phylogenetic relationships among *Saccharum* spp homo(eo)logous sequences, 12 amino acid-based phylogenies were constructed, one for each OG (Supplementary Table S1 and Supplementary Figure S1: a-l).

### The protein domains are conserved in a few retrieved sugarcane sequences

The presence of conserved protein domains in sequences of sugarcane, *Sorghum bicolor*, and *A. thaliana* was verified with HMMER scan [57]. Both glycosyltransferase family 20 (TPS domain - PF00982) and trehalose-phosphatase (TPP domain - PF02358) domains were found for all Class I and II TPS sequences. For TPP family and TRE single domains, a TPP and a trehalase were obtained, respectively (Supplementary Figure S1: a-l). Only sugarcane sequences having  $\geq 80\%$  of their respective model domains (71 of TPS, 93 of TPP, and three of TRE) (Figure 1a and b) had their proteins shown in the phylogenetic trees (Supplementary Table S1 and Supplementary Figure S1: a-l). Most of the retrieved sequences from the sugarcane databases used in this study do not harbor their respective protein domains or have them incomplete and thus are most likely transcript or protein fragments. These results reinforce the importance of using different datasets when working with polyploid species that still lack well-annotated genomes.

### Classification of TPS and TPP sequences from sugarcane

To classify the filtered sugarcane sequences based on previously established clades [15,16], the TPS and TPP amino acid-based phylogenetic trees were rebuilt. TPS Class I (clade B) and Class II (clade A) could be distinguished (Figure 2). Clade B is divided into two (B1 and B2). As observed in other studies, gene sequences belonging to subclade B2 were unique to *A. thaliana* [58]. Subclades A1 to A5 can also be identified as some topological disagreements arose, although not well endorsed by the computed low branch support values. Most of the sequences already classified in previous studies (*A. thaliana*, maize, and rice) [15,16] remained in their clades, but AtTPS11 was not well resolved, making it difficult to distinguish between A4 and A5 subclades.

Similarly, as for the TPS family, our analysis of TPPs recovered most of the previously identified clades but with some topological disagreements. The TPP family is displayed in two clades (A and B) divided into three (A1 to A3) and two (B1 and B2) subclades. In this study, subclades A2 and A3 were unique to monocots, and AtTPPD was grouped in clade B2 instead of B1 (Figure 3). Besides, the classification in OGs by EggNOG does not necessarily reflect the different subclades in the phylogenetic analyses for both TPS and TPP.

### Identity analysis of TPS, TPP, and TRE in sugarcane

Because most of the trehalose metabolizing enzymes have isoforms with high similarity in some cases [84], and we retrieved the corresponding sugarcane sequences from different datasets (Figure 1b), the TPS, TPP, and TRE filtered sequences were submitted to identity analysis. A global pairwise alignment for multiple sequences was performed for each of the three targets. From all filtered sequences, 14 TPS and 13 TPP were identified as redundant sequences with 100% of identity (Figure 4). Likewise, 26 TPS, 60 TPP, and three TRE sequences showed an identity of  $\leq 97\%$  (Figure 4 and Supplementary Table S2). Our findings indicate that different datasets (from different cultivars/different sequencing approaches) could recover some identical sequences. Additionally, highly similar sequences could represent alternative splice variants or recent paralogs.

### *In silico* expression in sugarcane leaves

To provide evidence of how *TPS*, *TPP*, and *TRE* are expressed in sugarcane leaves (var. SP80-3280), we analyzed published RNA-seq data [56]. Except for *TRE*, we recovered the expression levels (Transcripts Per Million - TPM) of sequences from most databases used in this analysis (please see materials and methods for details), including 31 *TPS* and 46 *TPP* from distinct clades (Figure 5a-c). *TPP* clade A3 contained the biggest number (15) of expressed sequences. For *TPS* clades with regulatory functions (A2, A3, and, A4/A5) *SCA3\_SP803280\_c117830\_g1\_i1\_m.154722*, *SCA2\_1\_c95057f1p13337.p1*, and *SCA2\_1\_c94760f1p43082.p1* had 2-fold higher expression than sequences that had no expression value. Interestingly, the sequence with the highest expression levels (2,5 -fold higher) (*SCA1\_2\_Sh\_239M11\_p000040*) among all *TPS* belongs to clade B1 (Class I), characterized by sequences with catalytic function (Figure 5a). The *TPP* transcripts with the maximum expression values were *SCA5\_SCEZRZ3019C12.g.p1*, *SCA3\_SP803280\_c104339\_g1\_i3\_m.83754*, and *SCA2\_1\_c72120f1p11566.p1*, had 1,8, 1,9, and 2,3, respectively, fold increased expression than sequences

that had no expression value. These sequences belong to clades A1, A3, and B1, respectively (Figure 5b). *TPP* sequences in clades A2 and B2 were not expressed. *TRE* showed a unique sequence with a low expression value (1.5) compared with the above-mentioned *TPS* and *TPP* sequences (Figure 5c).

### **Class I TPS harbors specific predicted N-glycosylation sites in conserved motifs**

We evaluated the putative glycosylation sites (Supplementary Table S3) and conserved motifs (Figure 6) among filtered TPS Class I and II sequences. Sequences of *Ostreococcus tauri* that have two TPS, one Class I, and one Class II enzyme [14] were included in this analysis.

Many sequences (73.8%) harbors predicted N-glycosylation sites (Asn), in which 17 and 41 were exclusive to Class I and II, respectively, whereas three appeared in both classes (Supplementary Table S3). NDTV, NITE, and NSTL sites were detected in Class I sequences for most species, including *A. thaliana*, *Z. mays*, *O. sativa*, and sugarcane. NITE is localized in motif 3 and overlapped with the TPS domain in AtTPS1 and monocot species (Figure 6). Otherwise, the NDTV and NSTL sites were predicted to be located between the two domains (Figure 6), with NDTV present in the same sequences highlighted above (except rice) and NSTL only in monocot species (Supplementary Table S3). For OtTPS1 (*O. tauri*), AtTPS1 (*Arabidopsis*), OsTPS1 (rice), ZmTPSI.1.1 (maize), and two sugarcane sequences (SCA3\_SP803280\_c107577\_g2\_i1\_m.97794 and SCA1\_2\_Sh\_239M11\_p000040) of the Class I, the putative motifs 1, 3, 4, 5, 6, 7, and 10 were conserved and together constitute the TPS domain. Interestingly, in other isoforms of *A. thaliana* TPS Class I (AtTPS2, AtTPS3, and AtTPS4) motif 7 was not present. The putative motifs 2 and 11 constitute the TPP domain in all Class I TPS except for *O. tauri*. All sequences of TPS Class II contained the same motifs present in the TPS and TPP domains. Additionally, motif 9 from the TPS domain and motif 8 from the TPP domain were not presented in TPS Class I.

### **TPS Class I and II catalytic and binding residues show mutations at TPS and TPP domains**

Sugarcane and *A. thaliana* protein sequences that had  $\geq 80\%$  similarity with the respective model domains were used to construct three-dimensional (3D) structures. Crystallized proteins of *Aspergillus fumigatus* (Model 5hvm.1.A - TPS), *Mycobacterium tuberculosis* (Model 5gvx.1.A - TPP), and *E. coli* (Model 2jjb.1.A - TRE) were defined as templates in SWISS-MODEL [62]. The templates had at least 30% sequence identity for TPS, 31% for TPP, and 35% for TRE (Supplementary Table S1 and Supplementary Figure S2). When the protein folding was analyzed, TPS proteins shared similar structures independently from their catalytic (Class I) or regulatory (Class II) function based on the phylogenetic classification (data not shown). Alternatively, TPP 3D structures seem different when comparing the monocot (sugarcane) and eudicot (*A. thaliana*). The latter is closest to the template (Supplementary Figure S2).

To distinguish sugarcane catalytic TPS Class I from the regulatory Class II, we aimed at identifying all residues involved in catalysis and binding from each domain and their putative mutations. For that, the two sugarcane sequences with the highest expression levels from each clade (Figure 5a), as well as those from *A. thaliana*, *O. sativa*, *Z. mays*, and *O. tauri* (Supplementary Figure S3), were aligned to the TPS

from *E. coli* (OtsA) [63], *Candida albicans* (Tps1) [64], and *C. albicans* TPP (Tps2) [65]. This alignment (Supplementary Figure S3) showed that for TPS Class I, all species listed above presented catalytic and binding residues already described for OtsA at the TPS domain. The TPS Class II presented mutations at the TPS domain, implying in loss of enzyme activity (Table 3). Moreover, the TPP domain of all Class I and Class II sugarcane TPS have remarkable similarities with the TPP domain of *C. albicans* Tps2 [65], displaying mutations that allow the differentiation between catalytic and regulatory classes (Supplementary Figure S3 and Table 3).

The deduced proteins from the most highly expressed sugarcane TPS Class I and II enzyme-coding transcripts (SCA1\_2Sh\_23M11\_p000040 and SCA3\_SP803280\_c117830\_g1\_il\_m.154722, respectively) were submitted to 3D structure analysis using SWISS-MODEL (Figure 7). For sugarcane TPS Class I, two distinct templates, 5hut (Qmean of -1.7562, sequence ID 50.1%) and 5hvm (Qmean of -1.6795, sequence ID 50.75%), appeared as best hits, and therefore 5hvm was chosen. The same template was used for the sugarcane TPS Class II.

Both sugarcane TPS Class I (Figure 7a) and II (Figure 7b) displayed similar folding structures (Supplementary Figure S2). However, sugarcane TPS Class I presented mutations at D25G (the change of D in *C. albicans* to G in sugarcane at position 25 of the *C. albicans*) catalytic residue of TPP domain and at the binding residues D27N, R67S, K133R, H140N, R142K, K176S, and N178S. TPS Class II displayed the replacement of residues involved in binding at the TPP domain (Table 3), mainly at R67K. Besides, the division of TPS Class II subclades is related to specific amino acid replacements involved in substrate binding or catalysis residues on both domains (Figure 7). For instance, residue R9 at the TPS domain, in which Q/T replaces at A3 and A2 subclade, and F/M at subclade A5 (Figure 2).

## Discussion

Sugarcane plays a key role in the Brazilian bioenergy sector regarding economic and societal aspects [66], such as environmental sustainability [67]. Gains in sugarcane yield have the potential to increase not only bioethanol production, an effective alternative to mitigate CO<sub>2</sub> emissions and climate change, but also other bioproducts [33]. However, some efforts remain necessary to achieve essential improvements in productivity, for instance, dealing with the complexity of its genome.

Due to its high ploidy levels, presence of aneuploidy, high rates of polymorphism, and repeat content, the sugarcane genome is still a challenge for genome sequencing, contributing to the lack of information about molecular function and structure [51]. However, recent advances in sequencing technologies and computational strategies for genome assembly are opening the way to deciphering the sugarcane genome [49,51–53]. Combining different sequencing strategies to mine datasets makes it possible to retrieve more accurate information about homo(eo)logous sequences. Therefore, the present work resorted to bioinformatics methods to identify the sugarcane trehalose pathway-related targets, accessing distinct sugarcane datasets [55,56,49,47,52,53]. Altogether, 430 sequences related to the

trehalose pathway were classified into 12 OGs from the EggNOG database. Only 36% (167 sequences) displayed all the predicted domains with high similarity ( $\geq 80\%$ ) to the established domain templates.

Identification of trehalose pathway-related target sequences in sugarcane is of particular interest to understand the mechanisms involved in sucrose accumulation since T6P is a specific sucrose sensor [31,32]. Information about trehalose metabolism in sugarcane is limited to the characterization of the TRE enzyme [68–70] and transcriptomic studies that have identified changes in the expression of putative genes encoding TPS and TPP [56,71–75]. These findings are consistent with the hypothesis that T6P could be a master key sensor in this species. Most of these studies have inferred the role of trehalose metabolism on abiotic stress tolerance and regulation of photosynthesis. However, the precise identification of the isoforms was unfeasible. Trehalose levels in sugarcane culms were five orders of magnitude lower than sucrose, ranging from less than 0.3 to 3.9 nmol g<sup>-1</sup>FW, although these sugars did not correlate linearly [76]. In contrast, [77] showed that trehalose positively correlated with sucrose. Thus, the correlation between these two disaccharides remains unclear for sugarcane. Transgenic sugarcane plants overexpressing *TPS* and *TPP* genes showed increased TRE activity, whereas no changes were observed in transformants containing an RNAi transgene specific for *TRE* [78]. Nevertheless, there is essentially no information about how T6P signaling operates in sugarcane or its potential impact on sucrose accumulation.

All genomes analyzed so far contain genes coding for those enzymes, indicating that trehalose metabolism has ancient origins [14]. When filtered sequences were used for rebuilding the phylogenetic trees of TPS and TPP [15,16], sugarcane TPS proteins were classified into Class I and Class II clades and their respective subclades (Figure 2 and Figure 3). Most diploid plants have only one TPS Class I gene, except for paleopolyploid species such as *Z. mays* and poplar (*Populus trichocarpa*), which have two [14]. Four sugarcane sequences from three datasets were grouped in TPS Class I, suggesting a similar pattern observed for the other paleopolyploid species (Figure 2). These sequences are present in the subclades B1 and B2 that contain all the catalytically active proteins from *A. thaliana* (AtTPS1, AtTPS2, and AtTPS4), *O. sativa* (OsTPS1), and maize (ZmTPSI.1.1). An evaluation of the catalytic residues showed that all amino acids involved in catalysis and binding are present at the sugarcane sequences allocated at the B1 subclade (Class I), indicating that these enzymes are likely to be active and physiologically relevant (Table 3 and Supplementary Figure S3).

Differently from AtTPS1, the predicted AtTPS2-4 proteins lack the N-terminal auto-inhibitory domain and appear to be restricted to the Brassicaceae [14], in which these sequences constitute the subclade B2 (Table 2 and Figure 2). Most of the filtered sugarcane TPS sequences (~94%) had high similarity with Class II proteins, reflecting more involvement in regulatory rather than catalytic function as found for most plant species studied to date. Similar results have been recently reported by [79], who used phylogenetic trees to classify one sugarcane sequence as possibly catalytic and eight as regulatory.

Unlike TPS, all TPP encode active enzymes [28], classified in clades A and B and their respective subclades (Figure 3) [15,16]. In the subclades A1 and B1, various eudicot and monocot species were

present, and A2, A3, and B2 were exclusive for monocots [16,58]. Sugarcane sequences were grouped in all subclades, but B1 displayed the highest number of sequences (~37%) (Figure 3). Differently from TPS and TPP, TRE is encoded by a single gene in Arabidopsis and rice, and among 29 sequences retrieved from sugarcane databases, three displayed high similarity with the protein domain template (Supplementary Table S1 and Supplementary Figure S2).

Multigenic families encode plant TPS and TPP through duplication events which started earlier for TPS than for TPP genes [15]. Regarding the similarity percentage among sequences belonging to a specific multigenic family, the highest identities for TPS and TPP sequences in maize are 97% and 90%, respectively (data not shown). For filtered sugarcane sequences, 26 TPS, 60 TPP, and three TRE showed less than 97% of identity (Figure 4), reflecting the existence of distinct predicted homo(eo)logous sequences, including alternative allelic versions of gene products, alternatively spliced variants, and possible paralogs. T6P is an essential regulator of sucrose in plants [80], and changes in their quantity can modify gene expression and metabolism, maintaining sucrose levels within an optimal range [80]. The *in silico* expression analysis indicates a variable TPS and TPP gene expression profile in leaves of the sugarcane cultivar SP80-3280 (Figure 5a and b).

For TPP, subcellular localization was used to identify all *AtTPP* cell- and tissue-specific expression patterns, suggesting neofunctionalization after gene duplications [28]. Our results showed that the sugarcane TPP sequence with the highest expression value belongs to the clade B1 (Figure 5b). *AtTPPD* and *AtTPPI*, belonging to this subclade, have been associated with abiotic features such as salt and oxidative stress resistance and responses to drought, respectively [81,82].

For TPS, the highest expression value belongs to a TPS Class I (Figure 5a), which might indicate that the TPS catalytic function in mature leaves is more relevant. It remains to be elucidated whether the high TPS expression would contribute to high T6P levels. Moreover, sugarcane TPS Class II sequences belonging to subclades A2 and A3 also had high expression values (Figure 5a). Subclades A2 and A3 also contain *AtTPS6* and *AtTPS7*, which regulate plant architecture, cell shape, and trichome branching [83]. Besides, they are thought to be related to signal transduction during stress resistance [84]. The functional characterization of these sequences in sugarcane might help to elucidate their molecular mechanisms.

The TPS Class I and II duplications led to the neofunctionalization of sequences in a determined clade [10], which is reflected by mutations at important residues. For TPS Class I, the residues involved in binding and catalysis were maintained in the TPS domain compared with well-characterized sequences. Otherwise, sugarcane TPS Class II sequences showed mutations at residues involved in catalysis, which could explain the acquired regulatory function (Figure 7, Table 3, and Supplementary Figure S3). The catalysis residues R9 and G22 were conserved in Class I and mutated in Class II sequences (Figure 2). These residues are important for binding with glucose-6-phosphate and UDP, respectively [65].

Likewise, the TPP domain of TPS sequences was also analyzed (Table 3). *C. albicans* Tps2 maintains the catalytic activity associated with the TPP domain [65]. Comparing sugarcane and *C. albicans* TPP domains led to identifying many residue mutations associated with catalysis and binding functions

(Table 3). Residue replacements at D25G and R67S indicate the complete loss of enzyme activity, which can explain why the TPP domain at sugarcane TPS Class I is probably inactive [65]. Moreover, TPS Class II presented replaced residues involved in binding at the TPP domain (Table 3), mainly at R67K. The replacement of this residue by alanine at *C. albicans* TPP inactivated this phosphatase [65].

Motif analysis in the TPS proteins was performed to identify specific protein regions associated with a regulatory or catalytic function (Figure 6). The *Z. mays* genome encodes two Class I TPS sequences (Figure 2). However, only one (ZmTPSI.1.1) is functional and has all conserved TPS motifs [16]. The second isoform (ZmTPSI.1.2) is truncated and does not have some of the residues necessary for substrate binding [16]. Some studies showed different amounts of TPS motifs, 6 in sugarcane [79], 20 in potato [84], and 12 in cotton [85]. Motif 9 (motif 2 in [79]) and motif 8 are present in the TPS domain of TPS Class II of sugarcane and cotton [85]. Together, these results associated with the residue mutations may point out differences in the TPS proteins that may justify the absence of catalytic activity in the regulatory sequences and the centralization of the catalysis in some members of this multigenic family (Table 3). However, further studies are required to validate this hypothesis.

Post-translational modifications have already been experimentally shown to influence TPS Class I sequences' activity and catalytic fidelity [86]. Phosphorylation at Ser827 and Ser941 and putative SUMOylation at Lys902 were identified in the TPP domain. The latter occurs inside a consensus sequence highly conserved in Class I TPS enzymes in all the major land plant groups and streptophyte algae [86]. *N*-glycosylation is one of the most common and chemically complex post-translational modifications in eukaryotes [87]. However, there is little information for TPS and TPP families members. Sugarcane TPS Class I and II showed 17 and 41 putative *N*-glycosylation sites, respectively (Supplementary Table S3). Although these sites do not provoke significant changes in protein structure, they might influence the dynamic properties, protein stability, and possibly the enzyme's catalytic activity [87].

Genetic manipulation of the trehalose pathway improves tolerance to different abiotic stresses [20,88]. Sugarcane is an annual crop cultivated in large geographical areas worldwide, facing constant environmental changes such as temperature and water availability. Water deficit differentially affects sugarcane during the distinct growth stages and is considered one of the main factors limiting its productivity [89–93]. However, water deficit is beneficial to enhance the influx of sucrose into the stems during the maturation phase [94]. Thus, a better understanding of the processes mediated by the trehalose pathway in sugarcane is also an alternative to mitigate current environmental pressures derived from climate change and boost sugarcane-derived products' production. As a large sugar and bioethanol producer, any modest gain in sugarcane productivity in Brazil represents significant profits for the bioenergy sector.

## Conclusions

The role of T6P as a sucrose sensor is well known. However, the involvement of trehalose metabolism reported so far for sugarcane recognized it as a putative mediator of osmoprotectant mechanism under stress. The high sucrose levels in sugarcane stems could indicate a role for T6P as a central regulator during sugarcane growth and development. This study uncovered a large number of sequences with high homology to the selected target genes. However, the exact numbers of TPS, TPP, and TRE sequences in sugarcane are not yet precise as even the most complete database was unable to cover the entire sugarcane genome. Apart from classifying TPS and TPP proteins from sugarcane into distinct clades, amino acid residue and motif analyses revealed specific alterations contributing to a catalytic or regulatory function. We managed to retrieve expression values from one sugarcane transcriptome dataset, but more information is needed to map under what conditions and in which tissues these genes are expressed. Our findings started to pave the way for functional studies to uncover the physiological roles of T6P signaling in sugarcane.

## Materials And Methods

### Identification of trehalose metabolizing enzymes in sugarcane

Gene sequences encoding trehalose metabolizing enzymes from *A. thaliana*, *Z. mays*, and *O. sativa* were first used as queries to identify the groups of orthologous genes they belonged to at the level of Viridiplantae, in the EggNOG v4.5.1 database [54]. The orthologous group (OG) IDs were then used to identify genes belonging to the same groups in species of the subfamily Panicoideae, whose genomes are publicly available. For sugarcane, a mix of genomics and transcriptomics datasets was used [55,56,49,47,52,53] (Table 2).

### Sequence alignment and phylogenetic analyses

To determine the percentage of identity among all filtered TPS and TPP protein sequences retrieved from different sugarcane databases, a global pairwise alignment among all sequence pairs was carried out with the program needle of the EMBOSS v6.0.0. suite [95]. Additionally, multiple sequence alignments for the protein sequences of each OG were generated with MAFFT [96], and dubious regions were removed from the alignments using TrimAl v1.2 [97].

Phylogenetic inference was performed by IQ-TREE v1.6.9 [59], with automatic evolutionary model selection and branch support values were computed as Shimodaira–Hasegawa approximate likelihood ratio test (SH-like aLRT) [60], and Ultrafast bootstrap (UFboot) [61]. Phylogenetic trees were rooted by reconciling them with the commonly accepted species tree with Notung v2.9 [98]. Lastly, the phylogenetic trees of filtered TPS and TPP sugarcane sequences were constructed based on [15] and [16], respectively, using the above settings.

### Domain characterization, three-dimensional protein structure analyses, and catalytic/binding residues

Protein sequences of *A. thaliana*, *S. bicolor*, and sugarcane were subjected to domain analysis using HMMER v2.41.2 [57,99]. Sequences that matched with a score above the gathering threshold and covered at least 80% of the domain model were considered for further analyses. This threshold reflected the domain coverage value, representing how much of the domains were detected in sugarcane sequences. Sequences harboring all the predicted conserved domains were illustrated using the Illustrator for biological sequences v1.0 [100]. The three-dimensional structures were modeled by SWISS-MODEL [62]. The best template of each sequence was selected, combining larger sequence coverage, global model quality estimation (GMQE), quaternary structure quality estimate (QSQE), and the sequence identity to the target. The obtained structures were processed using PyMol (TM) 2.4.2, and the catalytic and binding residues were identified at modeled structures by the alignment with TPS from *E. coli* [63,65], as well as TPS and TPP from *C. albicans* [64,65].

### ***In silico* transcript expression patterns**

TPS, TPP, and TRE expression levels were recovered from a published transcriptomics dataset of leaf development from the hybrid SP80-3280 [56]. As most of the databases were from hybrid cultivars, and to avoid errors in the analysis, the sequences of the *S. spontaneum* have been removed from this analysis. In [56] different developmental regions along the leaf +1 of two-month-old seedlings were evaluated but we focused on samples from the middle portion (4 biological replicates), as this is the most photosynthetically active region (NCBI Short Read Archive accession numbers: SRR1979669, SRR1979665, SRR1979662, and SRR1979660). Raw sequence reads were downloaded from NCBI's SRA, and cleaned with BBDuk2 [101], to remove remainders of rRNA and low-quality regions as well as adapters. Salmon v1.1.0 [102] was used to estimate transcript expression levels expressed as Transcripts per Million (TPM) transformed in (LOG10+1).

### **Prediction of *N*-glycosylation sites and conserved motifs**

The potential *N*-glycosylation (Asn) sites from TPS Class I and II filtered sequences were predicted with NetNGlyc 1.0 software [103]. As the software recommended, only *N*-glycosylation sites prediction with potential values >0.5. Subsequently, to identify the motif regions, the two sugarcane sequences with the highest expression values of each clade of the TPS Class I and II were submitted to MEME [104], using default parameters and the maximum number of motifs set to 12. For this analysis, were used sequences of *A. thaliana*, rice, maize, and a basal green alga, *O. tauri*. A workflow that summarizes all the steps followed in this article is in Supplementary Figure S4.

## **Declarations**

### **Acknowledgements**

This work was supported by the National Institute of Science and Technology of Bioethanol (INCT-Bioethanol) (São Paulo Research Foundation FAPESP 2014/50884-5 and National Council for Scientific and Technological Development CNPq 465319/2014-9). Lauana Pereira de Oliveira (CNPq 142090/2018-

2), Marina Câmara Mattos Martins (FAPESP 18/03764-5), Bruno Viana Navarro (CAPES PNPd 20131621-33002010156P0), Diego M. Riaño-Pachón (CNPq 310080/2018-5) and João Pedro de Jesus Pereira (CNPq 115313/2019-2) are grateful for the fellowships.

### Author Contributions

BVN, LPO, MCMM, and MSB conceived the study. BVN, DMRP, JPJP, ARL, LPO, and MCMM carried out analyzes and interpreted the data. BVN, LPO, MCMM, DMRP and MSB wrote and revised the manuscript.

### Competing Interest Statement

The authors declare no competing interests.

### Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files.

## References

1. Lunn, J. E. Sucrose Metabolism. in *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd. doi:10.1002/9780470015902.a0021259 (2008).
2. Ruan, Y. L. Sucrose metabolism: Gateway to diverse carbon use and sugar signaling. *Annual Review of Plant Biology* **65**, 33–67 (2014).
3. Smith, A. M. & Stitt, M. Coordination of carbon supply and plant growth. *Plant Cell Environ* **30**, 1126–1149 (2007).
4. Czedik-Eysenberg, A. *et al.* The interplay between carbon availability and growth in different zones of the growing maize leaf. *Plant Physiology* **172**, 943–967 (2016).
5. León, P. & Sheen, J. Sugar and hormone connections. *Trends in Plant Science* **8**, 110–116 (2003).
6. Nunes-Nesi, A., Fernie, A. R. & Stitt, M. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Molecular Plant* **3**, 973–996 (2010).
7. Li, L. & Sheen, J. Dynamic and diverse sugar signaling. *Current Opinion in Plant Biology* **33**, 116–125 (2016).
8. Forzani, C., Turqueto Duarte, G. & Meyer, C. The Plant Target of Rapamycin Kinase: A connectTOR between Sulfur and Growth. *Trends in Plant Science* **23**, 472–475 (2018).
9. Chen, Y. S. *et al.* Sugar starvation-regulated MYBS2 and 14-3-3 protein interactions enhance plant growth, stress tolerance, and grain weight in rice. *Proceedings of the National Academy of Sciences of the*

*United States of America* **116**, 21925–21935 (2019).

10. Fichtner, F. & Lunn, J. E. The Role of Trehalose 6-Phosphate (Tre6P) in Plant Metabolism and Development. *Annual Review of Plant Biology*. doi:10.1146/annurev-arplant-050718-095929 (2021).
11. Paul, M. J., Primavesi, L. F., Jhurrea, D. & Zhang, Y. Trehalose metabolism and signaling. *Annual Review of Plant Biology* **59**, 417–441 (2008).
12. Leyman, B., Van Dijck, P. & Thevelein, J. M. An unexpected plethora of trehalose biosynthesis genes in *Arabidopsis thaliana*. *Trends in Plant Science* **6**, 510–513 (2001).
13. Avonce, N., Mendoza-Vargas, A., Morett, E. & Iturriag, G. Insights on the evolution of trehalose biosynthesis. *BMC Evolutionary Biology* **6**, (2006).
14. Lunn, J. E. Gene families and evolution of trehalose metabolism in plants. *Funct Plant Biol* **34**, 550–563 (2007).
15. Yang, H.-L., Liu, Y.-J., Wang, C.-L. & Zeng, Q.-Y. Molecular Evolution of Trehalose-6-Phosphate Synthase (TPS) Gene Family in Populus, Arabidopsis and Rice. *PLOS ONE* **7**, e42438 (2012).
16. Henry, C. *et al.* The trehalose pathway in maize: conservation and gene regulation in response to the diurnal cycle and extended darkness. *Journal of experimental botany* **65**, 5959–5973 (2014).
17. Vogel, G. *et al.* Trehalose metabolism in Arabidopsis: Occurrence of trehalose and molecular cloning and characterization of trehalose-6-phosphate synthase homologues. *Journal of Experimental Botany* **52**, 1817–1826 (2001).
18. Ramon, M. *et al.* Extensive expression regulation and lack of heterologous enzymatic activity of the Class II trehalose metabolism proteins from *Arabidopsis thaliana*. *Plant, cell & environment* **32**, 1015–1032 (2009).
19. Vandesteene, L., Ramon, M., Le Roy, K., Van Dijck, P. & Rolland, F. A single active trehalose-6-P synthase (TPS) and a family of putative regulatory TPS-like proteins in Arabidopsis. *Molecular plant* **3**, 406–419 (2010).
20. Delorge, I., Figueroa, C. M., Feil, R., Lunn, J. E. & Van Dijck, P. Trehalose-6-phosphate synthase 1 is not the only active TPS in *Arabidopsis thaliana*. *Biochemical Journal* **466**, 283–290 (2015).
21. Usadel, B. *et al.* Global transcript levels respond to small changes of the carbon status during progressive exhaustion of carbohydrates in Arabidopsis rosettes. *Plant Physiology* **146**, 1834–1861 (2008).
22. Schlupepmann, H. *et al.* Trehalose mediated growth inhibition of Arabidopsis seedlings is due to trehalose-6-phosphate accumulation. *Plant Physiology* **135**, 879–890 (2004).

23. Osuna, D. *et al.* Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived *Arabidopsis* seedlings. *Plant Journal* **49**, 463–491 (2007).
24. Baena-González, E. & Sheen, J. Convergent energy and stress signaling. *Trends in Plant Science* **13**, 474–482 (2008).
25. Brenner, W. G., Romanov, G. A., Köllmer, I., Bürkle, L. & Schmölling, T. Immediate-early and delayed cytokinin response genes of *Arabidopsis thaliana* identified by genome-wide expression profiling reveal novel cytokinin-sensitive processes and suggest cytokinin action through transcriptional cascades. *Plant J* **44**, 314–333 (2005).
26. Tian, L. *et al.* The trehalose-6-phosphate synthase TPS5 negatively regulates ABA signaling in *Arabidopsis thaliana*. *Plant Cell Reports* **38**, 869–882 (2019).
27. Bläsing, O. E. *et al.* Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell* **17**, 3257–3281 (2005).
28. Vandesteene, L. *et al.* Expansive evolution of the trehalose-6-phosphate phosphatase gene family in *Arabidopsis*. *Plant physiology* **160**, 884–896 (2012).
29. Lunn, J. E. *et al.* Sugar-induced increases in trehalose 6-phosphate are correlated with redox activation of ADPglucose pyrophosphorylase and higher rates of starch synthesis in *Arabidopsis thaliana*. *Biochemical Journal* **397**, 139–148 (2006).
30. Fichtner, F. *et al.* Trehalose 6-phosphate is involved in triggering axillary bud outgrowth in garden pea (*Pisum sativum* L.). *Plant Journal* **92**, 611–623 (2017).
31. Lunn, J. E., Delorge, I., Figueroa, C. M., Van Dijck, P. & Stitt, M. Trehalose metabolism in plants. *The Plant Journal* **79**, 544–567 (2014).
32. Yadav, U. P. *et al.* The sucrose-trehalose 6-phosphate (Tre6P) nexus: specificity and mechanisms of sucrose signalling by Tre6P. *Journal of experimental botany* **65**, 1051–1068 (2014).
33. de Souza, A. P., Grandis, A., Leite, D. C. C. & Buckeridge, M. S. Sugarcane as a Bioenergy Source: History, Performance, and Perspectives for Second-Generation Bioethanol. *BioEnergy Research* **7**, 24–35 (2014).
34. Jackson, P. A. Breeding for improved sugar content in sugarcane. *Field Crops Research* **92**, 277–290 (2005).
35. Scortecci, K. C. *et al.* Challenges, opportunities and recent advances in sugarcane breeding. *Plant breeding* **1**, (2012).

36. Grof, C. P. L. & Campbell, J. A. Sugarcane sucrose metabolism: scope for molecular manipulation. *Functional Plant Biol.* **28**, 1–12 (2001).
37. Burnquist, W. L. Evaluating sugarcane R&D performance: evaluation of three breeding programs. *Proc. Int. Soc. Sugar Cane Technol.* **27**, 15 (2010).
38. Waclawovsky, A. J., Sato, P. M., Lembke, C. G., Moore, P. H. & Souza, G. M. Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnology Journal* **8**, 263–276 (2010).
39. Thirugnanasambandam, P. P., Hoang, N. V. & Henry, R. J. The Challenge of Analyzing the Sugarcane Genome. *Frontiers in Plant Science* **9**, 616 (2018).
40. Piperidis, G., Piperidis, N. & D’Hont, A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics* **284**, 65–73 (2010).
41. Heller-Uszynska, K., Uszynski, G. & Huttner, E. Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Mol Breeding* **28**, 37–55 (2011).
42. Garcia, A. A. F. *et al.* SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Scientific Reports* **3**, 1–10 (2013).
43. Grativol, C. *et al.* Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. *Plant Journal* **79**, 162–172 (2014).
44. de Setta, N. *et al.* Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* **15**, 540 (2014).
45. Okura, V. K., de Souza, R. S. C., de Siqueira Tada, S. F. & Arruda, P. BAC-pool sequencing and assembly of 19 Mb of the complex sugarcane genome. *Frontiers in Plant Science* **7**, (2016).
46. Racedo, J. *et al.* Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biology* **16**, 142 (2016).
47. Hoang, N. V. *et al.* A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* **18**, 395 (2017).
48. Miller, J. R. *et al.* Initial genome sequencing of the sugarcane CP 96-1252 complex hybrid. *F1000Research* **6**, 688 (2017).
49. Riaño-Pachón, D. M. & Mattiello, L. Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Research* **6**, 861 (2017).

50. Yang, X. *et al.* Mining sequence variations in representative polyploid sugarcane germplasm accessions. *BMC Genomics* **18**, 594 (2017).
51. Souza, G. M. *et al.* Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *GigaScience* **8**, 1–18 (2019).
52. Garsmeur, O. *et al.* A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications* **9**, (2018).
53. Zhang, J. *et al.* Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics* **50**, 1565–1573 (2018).
54. Huerta-Cepas, J. *et al.* EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **44**, D286–D293 (2016).
55. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Research* **46**, W200–W204 (2018).
56. Han, B. *et al.* Interspecies and intraspecies analysis of trehalose contents and the biosynthesis pathway gene family reveals crucial roles of trehalose in osmotic-stress tolerance in cassava. *International Journal of Molecular Sciences* **17**, 1077 (2016).
57. Biasini, M. *et al.* SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* **42**, (2014).
58. Kaasen, I., McDougall, J. & AR, S. Analysis of the *otsBA* operon for osmoregulatory trehalose synthesis in *Escherichia coli* and homology of the OtsA and OtsB proteins to the yeast trehalose-6-phosphate synthase/phosphatase complex. *Gene* **145**, 9–15 (1994).
59. Miao, Y. *et al.* Structural and In Vivo Studies on Trehalose-6-Phosphate Synthase from Pathogenic Fungi Provide Insights into Its Catalytic Mechanism, Biological Necessity, and Potential for Novel Antifungal Drug Design. *mBio* **8**, (2017).
60. Miao, Y. *et al.* Structures of trehalose-6-phosphate phosphatase from pathogenic fungi reveal the mechanisms of substrate recognition and catalysis. *Proceedings of the National Academy of Sciences* **113**, 7148 LP – 7153 (2016).
61. Bordonal, R. d. O. *et al.* Sustainability of sugarcane production in Brazil. A review. *Sustainability of sugarcane production in Brazil. A review* **38**, (2018).
62. Tammisola, J. Towards much more efficient biofuel crops - can sugarcane pave the way? *GM crops* **1**, 181–198 (2010).

63. Vettore, A. L. *et al.* Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome research* **13**, 2725–2735 (2003).
64. Mattiello, L. *et al.* Physiological and transcriptional analyses of developmental stages along sugarcane leaf. *BMC Plant Biology* **15**, (2015).
65. Glasziou, K. T. & Gayler, K. R. Sugar transport: Occurrence of trehalase activity in sugar cane. *Planta* **85**, 299–302 (1969).
66. Alexander, A. G. Studies on trehalase in *Saccharum* spp. leaf and storage tissues. *Plant and Cell Physiology* **14**, 157–168 (1973).
67. Fleischmacher, O. L., Vattuone, M. A., Prado, F. E. & Sampietro, A. R. Specificity of sugar cane trehalase. *Phytochemistry* **19**, 37–41 (1980).
68. Carson, D. L., Hockett, B. I. & Botha, F. C. Sugarcane ESTs differentially expressed in immature and maturing internodal tissue. *Plant Science* **162**, 289–300 (2002).
69. Casu, R. E. *et al.* Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Mol Biol* **52**, 371–386 (2003).
70. McCormick, A., Cramer, M. & Watt, D. Differential Expression of Genes in the Leaves of Sugarcane in Response to Sugar Accumulation. *Tropical Plant Biology* **1**, 142–158 (2008).
71. Junior, N. N., Nicolau, M. S. P., Mantovanini, L. J. & Zingaretti, S. M. Expression Analysis of Two Genes Coding for Trehalose-6-Phosphate Synthase (TPS), in Sugarcane (*Saccharum* spp.) under Water Stress. *American Journal of Plant Sciences* **2013**, (2013).
72. Vicentini, R. *et al.* Large-Scale Transcriptome Analysis of Two Sugarcane Genotypes Contrasting for Lignin Content. *PLOS ONE* **10**, e0134909 (2015).
73. Bosch, S., Botha, F. C. & Rohwer, J. M. Trehalose and carbon partitioning in sugarcane. *SUNScholar Research Repository* <https://scholar.sun.ac.za:443/handle/10019.1/1433> (2005).
74. Glassop, D., Roessner, U., Bacic, A. & Bonnett, G. D. Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation? *Plant and Cell Physiology* **48**, 573–584 (2007).
75. O'Neill, B. P., Purnell, M. P., Nielsen, L. K. & Brumbley, S. M. RNAi-mediated abrogation of trehalase expression does not affect trehalase activity in sugarcane. *SpringerPlus* **1**, 1–6 (2012).
76. Hu, X. *et al.* Genome-Wide Analysis of the Trehalose-6-Phosphate Synthase (TPS) Gene Family and Expression Profiling of ScTPS Genes in Sugarcane. *Agronomy* **10**, 969 (2020).

77. Figueroa, C. M. & Lunn, J. E. A Tale of Two Sugars: Trehalose 6-Phosphate and Sucrose. *Plant physiology* **172**, 7–27 (2016).
78. Krasensky, J., Broyart, C., Rabanal, F. A. & Jonak, C. The redox-sensitive chloroplast trehalose-6-phosphate phosphatase AtTPPD regulates salt stress tolerance. *Antioxidants and Redox Signaling* **21**, 1289–1304 (2014).
79. Lin, Q., Wang, S., Dao, Y., Wang, J. & Wang, K. *Arabidopsis thaliana* trehalose-6-phosphate phosphatase gene TPPI enhances drought tolerance by regulating stomatal apertures. *Journal of Experimental Botany* **71**, 4285–4297 (2020).
80. Chary, S. N., Hicks, G. R., Choi, Y. G., Carter, D. & Raikhel, N. V. Trehalose-6-phosphate synthase/phosphatase regulates cell shape and plant architecture in Arabidopsis. *Plant physiology* **146**, 97–107 (2008).
81. Xu, Y., Wang, Y., Mattson, N., Yang, L. & Jin, Q. Genome-wide analysis of the *Solanum tuberosum* (potato) trehalose-6-phosphate synthase (TPS) gene family: evolution and differential expression during development and stress. *BMC Genomics* **18**, 926 (2017).
82. Mu, M. *et al.* Genome-wide Identification and analysis of the stress-resistance function of the TPS (Trehalose-6-Phosphate Synthase) gene family in cotton. *BMC Genetics* **17**, 54 (2016).
83. Fichtner, F. *et al.* Functional features of TREHALOSE-6-PHOSPHATE SYNTHASE1, an essential enzyme in Arabidopsis. *Plant Cell* **32**, 1949–1972 (2020).
84. Lee, H. S., Qi, Y. & Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Scientific Reports* **5**, 1–7 (2015).
85. Paul, M. J., Gonzalez-Uriarte, A., Griffiths, C. A. & Hassani-Pak, K. The Role of Trehalose 6-Phosphate in Crop Yield and Resilience. *Plant physiology* **177**, 12–23 (2018).
86. Inman-Bamber, N. G. & Smith, D. M. Water relations in sugarcane and response to water deficits. *Field Crops Research* vol. 92 185–202 (2005).
87. Basnayake, J., Jackson, P. A., Inman-Bamber, N. G. & Lakshmanan, P. Sugarcane for water-limited environments. Variation in stomatal conductance and its genetic correlation with crop productivity. *J Exp Bot* **66**, 3945–3958 (2015).
88. Lakshmanan, P., & Robinson, N. Stress Physiology: Abiotic Stresses. *In* Sugarcane: Physiology, Biochemistry, and Functional Biology. *Wiley Online Library*, 411-434 (2013).
89. Ferreira, T. H. S. *et al.* Sugarcane Water Stress Tolerance Mechanisms and Its Implications on Developing Biotechnology Solutions. *Frontiers in Plant Science* **8**, 1077 (2017).

90. Marchiori, P. E. R. *et al.* Physiological Plasticity Is Important for Maintaining Sugarcane Growth under Water Deficit. *Frontiers in Plant Science* **8**, 2148 (2017).
91. De Souza, A. P., Grandis, A., Arenque-Musa, B. C. & Buckeridge, M. S. Diurnal variation in gas exchange and nonstructural carbohydrates throughout sugarcane development. *Functional Plant Biology* **45**, 865–876 (2018).
92. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276–277 (2000).
93. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).
94. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
95. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).
96. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).
97. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: A program for dating gene duplications and optimizing gene family trees. in *Journal of Computational Biology* **7**, 429–447 (2000).
98. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
99. Liu, W. *et al.* IBS: An illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **31**, 3359–3361 (2015).
100. Bushnell, B., Rood, J. & Singer, E. BBMerge – Accurate paired shotgun read merging via overlap. *PLoS One* **26**, (2017).
101. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
102. Gupta, R., Jung, E. & Brunak, S. Prediction of *N*-glycosylation sites in human proteins. <http://www.cbs.dtu.dk/services/NetNGlyc/> (2004).
103. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research* **37**, W202 (2009).
104. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307–321 (2010).

# Tables

**Table 1.** Publicly available datasets used to look for genes encoding trehalose metabolizing enzymes in sugarcane.

Dataset	Type of dataset	Variety - Species	Tissue	Sequencing technology	BioProject
SCA5 [63]	Expressed Sequence Tags (EST)	>3 varieties (Including SP80-3280).	Different tissues	ABI sequencer	sucest-fun.org*
SCA3 [64]	Transcriptomics	SP80-3280	Leaf	Illumina Hi-Seq2500	PRJNA244522
SCA4 [49]	Genomics	SP80-3280	Leaf	Illumina HiSeq2000	PRJNA272769
SAC2_1 and SAC2_2 [47]	Transcriptomics	10 varieties	Internode	Illumina HiSeq4000	PRJNA356226
	Transcriptomics	22 varieties	Leaf, internode and root tissues	PacBio	
SCA1_1 and SCA1_2 [52]	Transcriptomics	R570	leaves, roots, and stems	Illumina Hi-Seq2500	ERZ654945**
	Genomics	R570, <i>S. spontaneum</i> and <i>S. officinarum</i>		BAC, PacBio and Illumina WGS	
SSP [53]	Genomics	<i>S. spontaneum</i>	Leaf	BAC, Illumina HiSeq 2500, PacBio and Illumina HiSeq X Ten platform	PRJNA483885

\*Data available in <http://sucest-fun.org/>

\*\*Data deposited in the EMBL-European Bioinformatics Institute

**Table 2.** Number of putative trehalose pathway protein sequences.

Species	TPS Class I		TPS Class II	TPP	TRE
	B1	B2	A1 - A5	(A1, A2, A3, B1 and B2)	
<i>Arabidopsis thaliana</i>	4	5	19	22	1
<i>Oryza sativa</i>	2	0	21	20	1
<i>Triticum aestivum</i>	18	0	24	37	3
<i>Setaria</i> spp	3	0	20	27	2
<i>Sorghum bicolor</i>	2	0	25	39	2
<i>Miscanthus sinensis</i>	5	0	24	31	9
<i>Zea mays</i>	55	0	120	26	6
<i>Saccharum</i> spp	39	0	211	165	29

Numbers of putative protein sequences of trehalose-6-phosphate synthase (TPS) - Class I (catalytic) and II (regulatory), trehalose-6-phosphate phosphatase (TPP), and trehalase (TRE) in different plant species. The clades distribution follows the nomenclature established by [16].

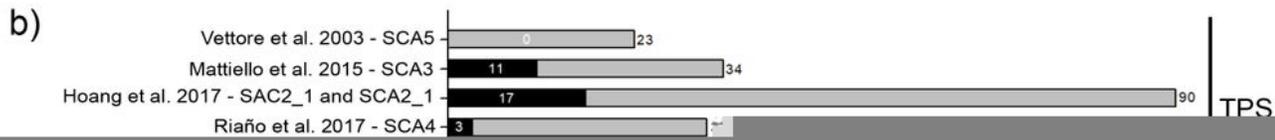
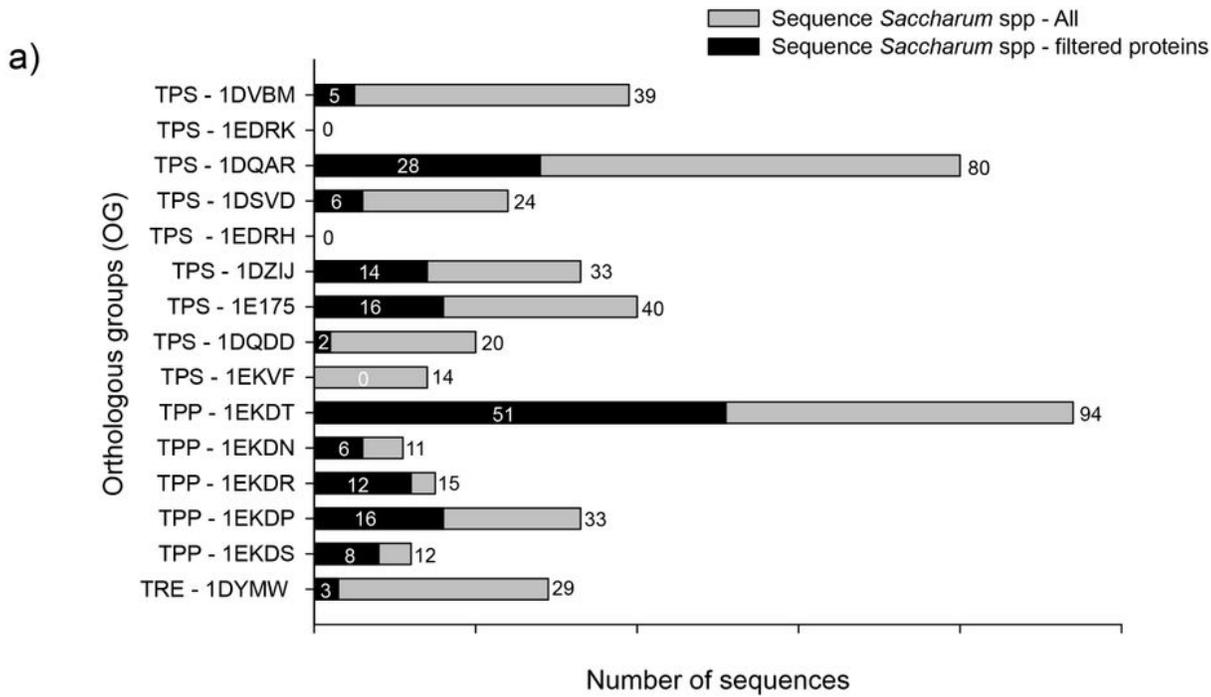
**Table 3.** Identification of catalytic, binding residues, and replacement at TPS Class I and II from *Saccharum* spp.

	TPS Class I		TPS Class II	
	Catalytic	Binding	Catalytic	Binding
TPS domain	R <sub>9</sub> -R	E <sub>171</sub> -E	<b>R-Q</b>	E-E
	G <sub>22</sub> -G	R <sub>370</sub> -R	<b>G-S</b>	R-R
	W <sub>41</sub> -W		<b>W-C</b>	
	Y <sub>76</sub> -Y		Y-Y	
	W <sub>85</sub> -W		W-W	
	D <sub>130</sub> -D		D-D	
	H <sub>154</sub> -H		H-H	
	R <sub>162</sub> -R		<b>R-D</b>	
	D <sub>371</sub> -D		D-D	
	E <sub>379</sub> -E		E-E	
TPP domain	<b>D<sub>25</sub>-G</b>	<b>D<sub>27</sub>-N</b>	D-D	D-D
	S <sub>65</sub> -S	<b>P<sub>32</sub>-E</b>	S-S	<b>P-/</b>
	K <sub>188</sub> -K	V <sub>34</sub> -V	K-K	<b>V-Q</b>
	<b>D<sub>230</sub>-H</b>	<b>R<sub>67</sub>-S</b>	D-D	<b>R-K</b>
	D <sub>234</sub> -D	E <sub>131</sub> -E	D-D	E-E
		<b>K<sub>133</sub>-R</b>		K-K
		<b>H<sub>140</sub>-N</b>		<b>H-C</b>
		<b>R<sub>142</sub>-K</b>		<b>R-E</b>
		<b>K<sub>176</sub>-S</b>		<b>K-P</b>
		<b>N<sub>178</sub>-S</b>		<b>N-S</b>
		E <sub>180</sub> -E		E-E

TPS domain catalytic and binding residues were identified based on *E. coli* TPS (Kaasen et al., 1994, Miao et al., 2016). TPP domain catalytic and binding residues were identified based on *Candida albicans* TPP (Miao et al., 2016). TPS Class I and TPS II from *Saccharum* spp. residues are based on the analysis

of two sequences more expressed SCA1\_2Sh\_23M11\_p000040 (catalytic) and SCA3\_SP803280\_c117830\_g1\_il\_m.154722 (regulatory). Bolded residues indicate residue replacement and / indicates a deletion. The numeration in residues are based on (Miao et al., 2016), for example mutations at X25Z (the change of X in *C. albicans* to Z in sugarcane at position 25). Other residues replacements are described at additional file 6 at the phylogenetic tree analysis (Fig. 2).

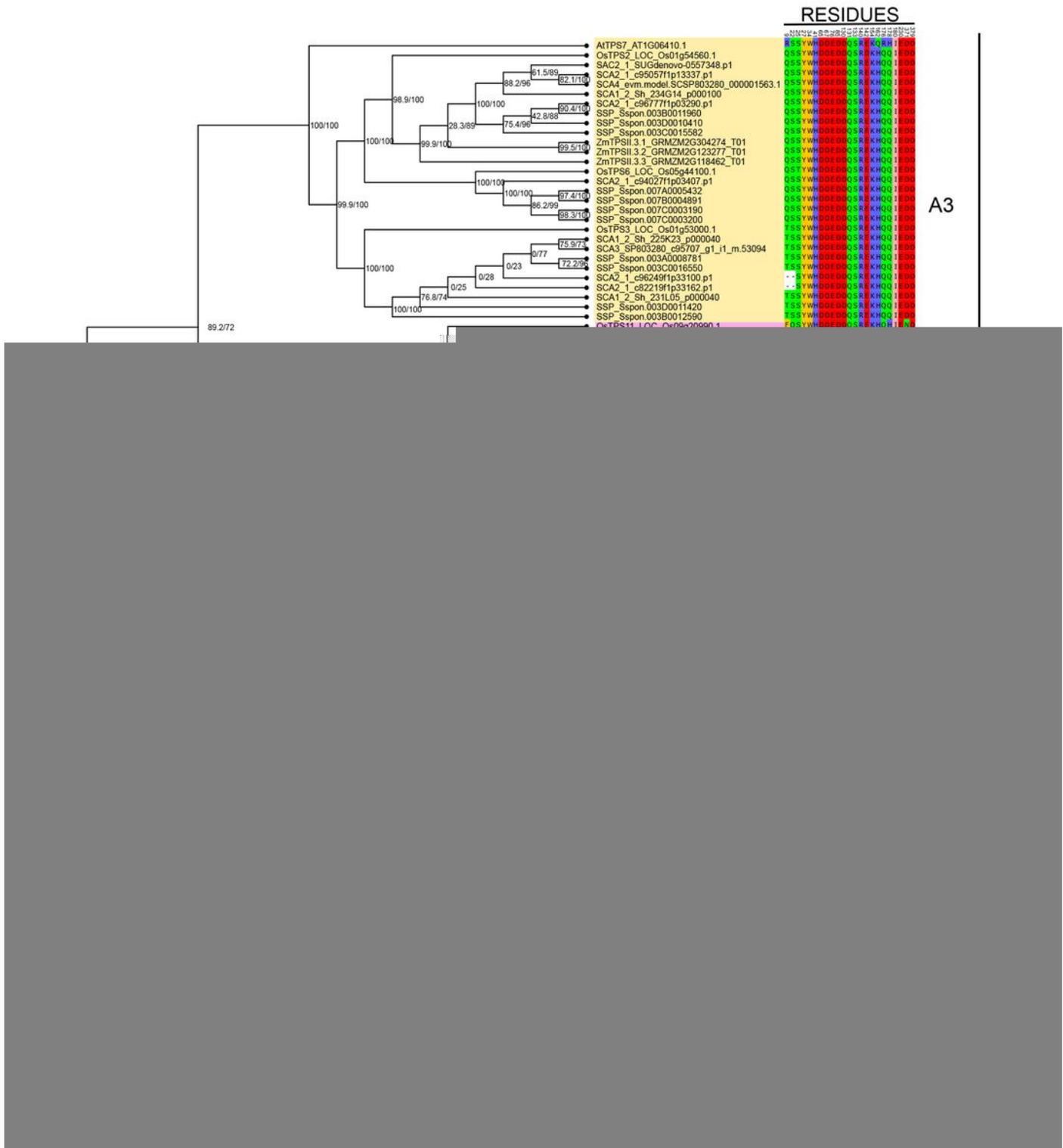
## Figures



**Figure 1**

The number of sequences related to the trehalose metabolism pathway in sugarcane. (a) Dispersion of total (grey) and filtered (black) sequences associated with orthologous groups (OG) from TPS, TPP, and TRE. The sequences related to each OG are described in Supplementary Table S1. (b) Dispersion of total (grey) and filtered (black) sequences among sugarcane publicly available datasets. The datasets

(sequence IDs) analyzed were described in Table 1: SCA5 [55]; SCA3 [56]; SCA4 [49]; SAC2\_2 and SCA2\_1 [47]; SCA1\_1 and SCA1\_2 [52]; SSP [53].



**Figure 2**

Phylogenetic tree of trehalose-6-phosphate synthase (TPS) from sugarcane, *Zea mays*, *Oryza sativa*, and *Arabidopsis thaliana*. Filtered sugarcane sequences were used to construct a phylogenetic tree based on

previous protein sequences obtained from <sup>15,16</sup>. The tree was built with IQ-TREE [59] using automatic evolutionary model selection, branch support values are shown as SH-like aLRT (%) [60] and ultrafast bootstrap (UFboot) (%) [61]. Branches with SH-like aLRT > 80% and UFboot > 95% are confident. Previous established clades are shown, although they are not always supported by the topology <sup>15,16</sup>. Residues related with the division of subclades of TPS Class I and II are highlighted. Databases and accession numbers are listed in Supplementary Table S1.

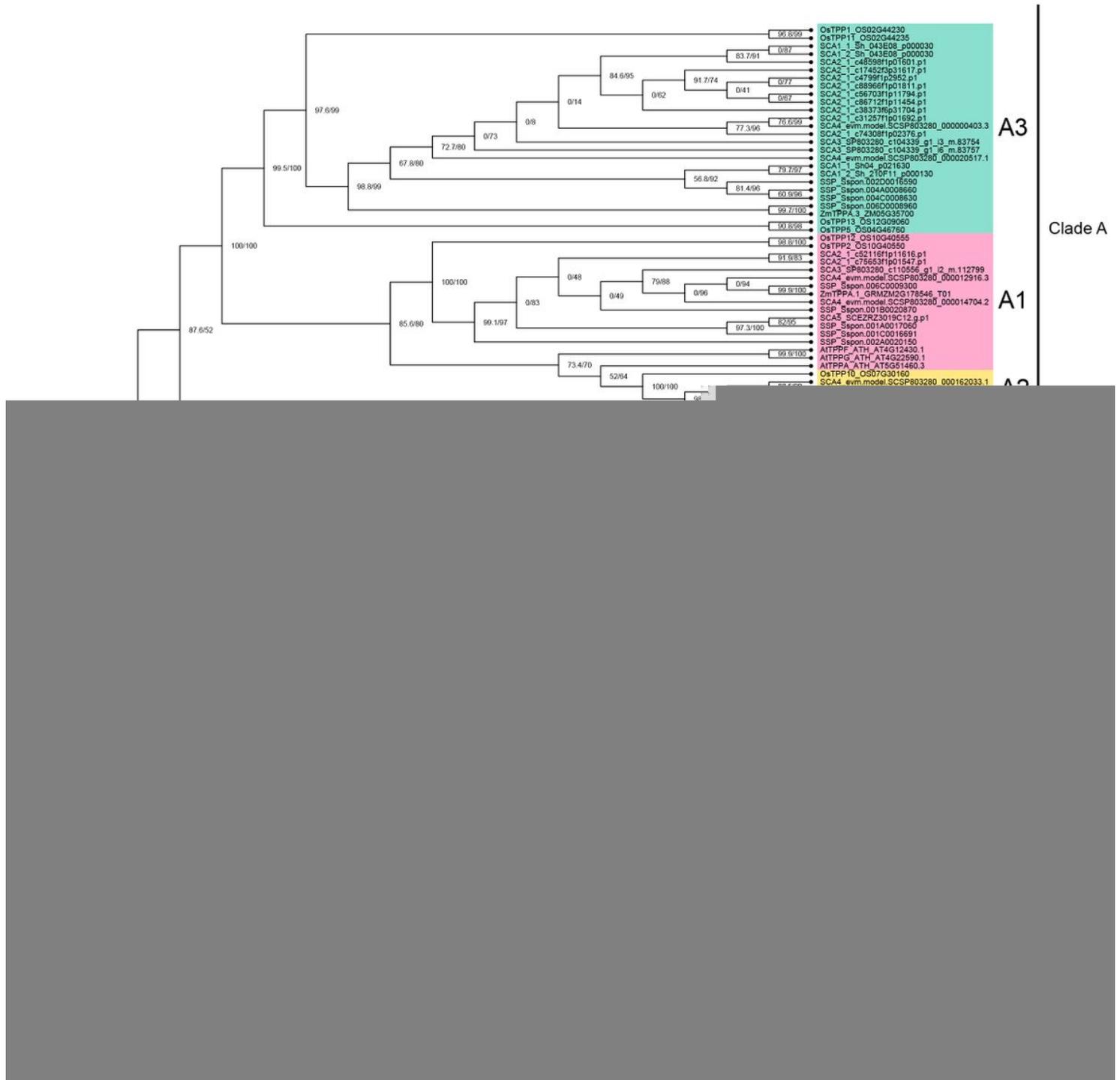
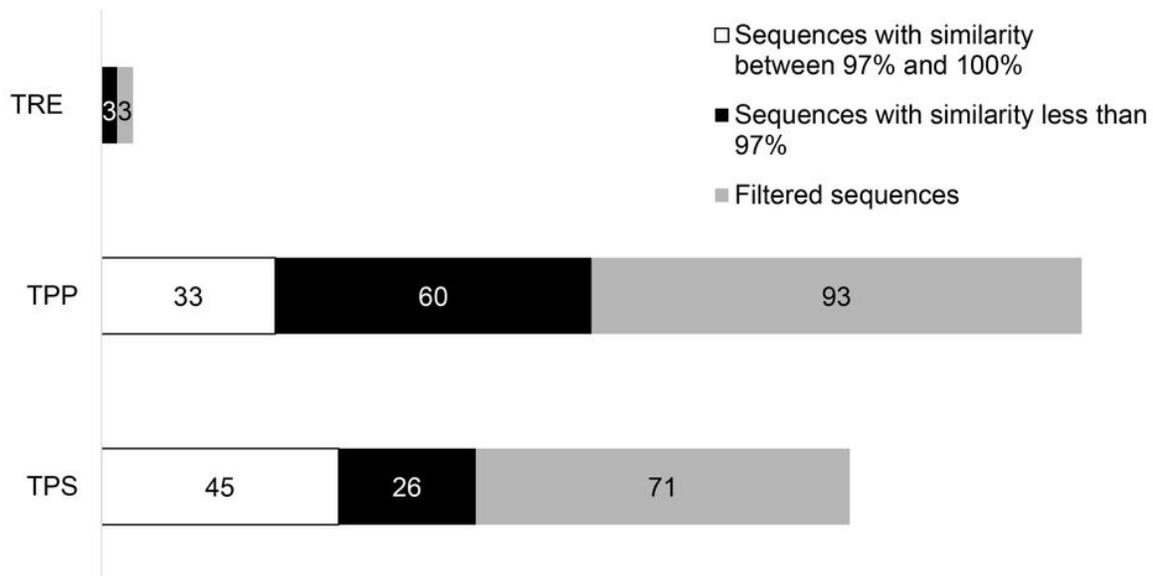


Figure 3

Phylogenetic tree of trehalose-6-phosphate phosphatase (TPP) from sugarcane, *Zea mays*, *Oryza sativa*, and *Arabidopsis thaliana*. Filtered sugarcane sequences were used to construct a phylogenetic tree based on previous protein sequences obtained from [16]. The tree was built with IQ-TREE [59] using automatic evolutionary model selection, branch support values are shown as SH-like aLRT (%) [60] and ultrafast bootstrap (UFboot) (%) [61]. Branches with SH-like aLRT > 80% and UFboot > 95% are confident. Previous established clades are shown, although they are not always supported by the topology<sup>15,16</sup>. Databases and accession numbers are listed in Supplementary Table S1.

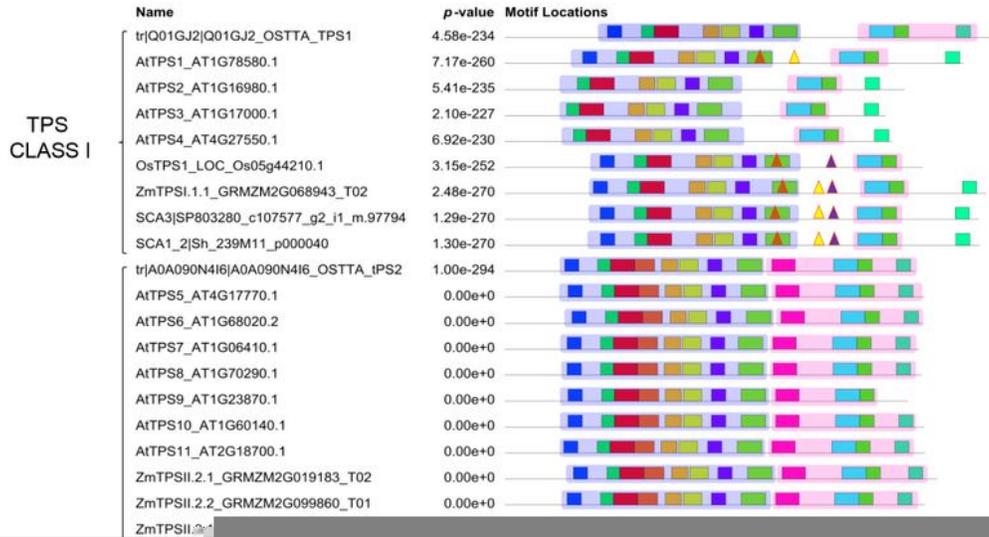


**Figure 4**

Identity of global pairwise alignment for multiple filtered sequences. Pairwise sequence alignments between all pairs of sequences were computed with the software needle from the EMBOSS v6.0.0. suite, using default settings.

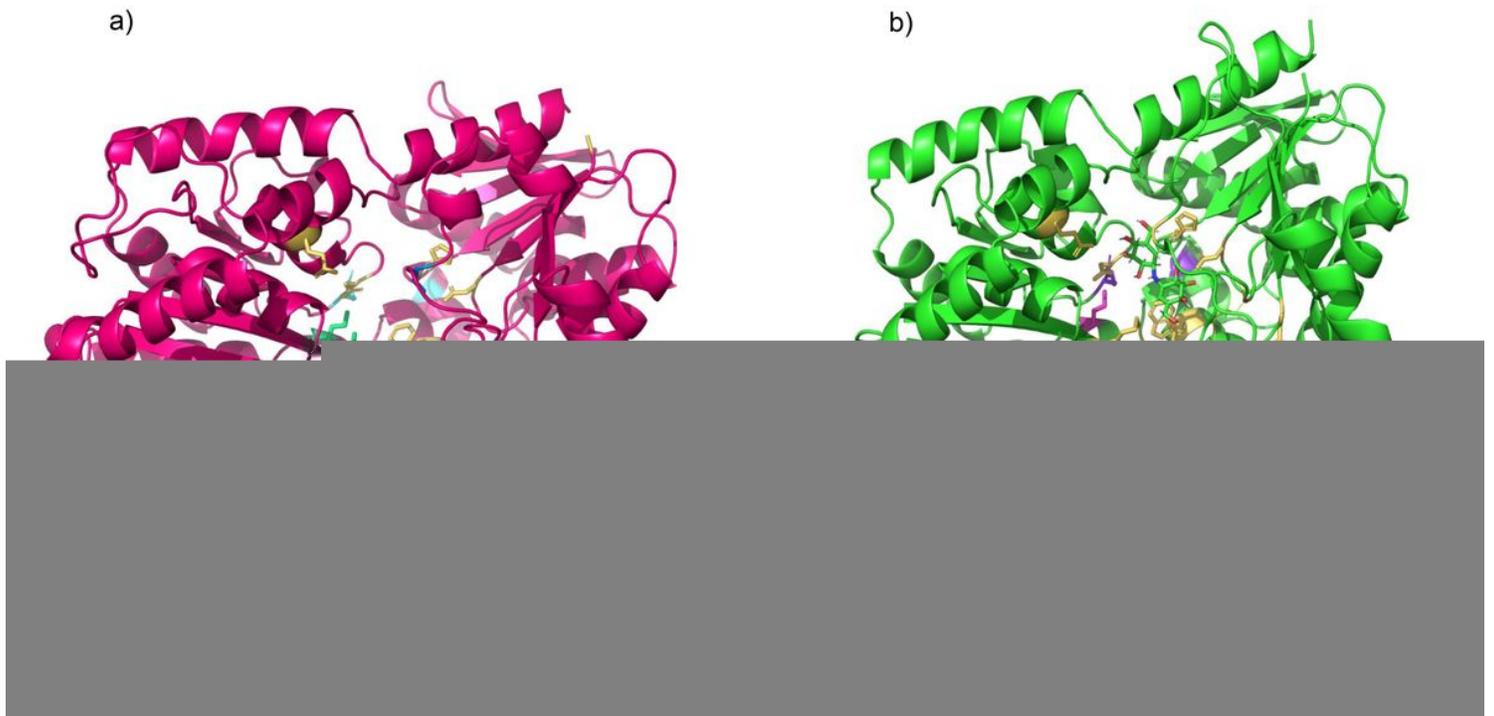
**Figure 5**

Expression values of trehalose-6-phosphate synthase (TPS) (a), trehalose-6-phosphate phosphatase (TPP) (b), and trehalase (TRE) (c). The heatmaps show the transformed TPM ( $\text{LOG}_{10}+1$ ), using a subset of the transcriptomics dataset from [56]. The sugarcane cultivar SP80-3280 was grown in a greenhouse for 60 days and gene expression in the portion of the leaf with the highest photosynthetic activity (middle part of the leaf +1) was assessed. Red and yellow indicate high and low expression values, respectively. Sequence IDs followed by an asterisk indicate sequences with higher expression value in each clade and underlined sequence IDs specify clustered sequences.



## Figure 6

Conserved motifs of trehalose-6-phosphate synthase (TPS) protein sequences showing *N*-glycosylation sites. Motif analysis was performed using the MEME online program. A total of 12 putative conserved motifs of sugarcane TPS Class I and II proteins was identified and classified as overlapped with the glycosyltransferase (transparent blue) or trehalose-phosphatase (transparent red) domains. Prediction of *N*-glycosylation sites were investigated for TPS sequences of *O. tauri*, *A. thaliana*, *Z. mays*, *O. sativa*, and *Saccharum* spp by NetNGlyc 1.0. Three *N*-glycosylation sites in Class I TPS are demonstrated (NITE, NDTV, and NSTL). All *N*-glycosylation sites are listed in Supplementary Table S3.



## Figure 7

Three-dimensional (3D) structures of trehalose-6-phosphate synthase (TPS) Class I and II. 3D template structures of TPS from *A. fumigatus* (ID model 5hvm.1.A) were used as models to predict the sugarcane TPS Class I and II structure. (a) Sugarcane TPS Class I structure (SCA1\_2\_Sh239M11\_p000040): yellow marked residues comprise the catalytic residues, purple-blue residues are involved in oligomer interaction, and light magenta highlights the K residue at the R/K pair. This analysis indicates that this enzyme is active and contains all catalytic residues. (b) Sugarcane TPS Class II structure (SCA3\_SP803280\_c117830\_g1\_i1\_m154722): marine blue indicates modified catalytic residues; yellow-orange catalytic conserved residues, ciane residues are involved in oligomer interaction, and lime green highlights K residue at the R/K pair.

## Supplementary Files

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

- [SupplementarytableS1.xlsx](#)
- [SupplementarytableS2.xlsx](#)
- [SupplementarytableS3.xlsx](#)
- [SupplementaryfigureS1.docx](#)
- [SupplementaryfigureS2.pdf](#)
- [SupplementaryfigureS3.pdf](#)
- [SupplementaryfigureS4.pdf](#)