



OPEN

Bioinformatic analyses to uncover genes involved in trehalose metabolism in the polyploid sugarcane

Lauana Pereira de Oliveira^{1,4,5}, Bruno Viana Navarro^{1,4,5}, João Pedro de Jesus Pereira^{1,4}, Adriana Rios Lopes^{1,2}, Marina C. M. Martins^{1,4}, Diego Mauricio Riaño-Pachón^{3,4}✉ & Marcos Silveira Buckeridge^{1,4}✉

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.

Sugars, 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 development¹⁰. 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

¹Laboratório de Fisiologia Ecológica de Plantas, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. ²Laboratório de Bioquímica, Instituto Butantan, São Paulo, Brazil. ³Laboratório de Biologia Computacional, Centro de Energia Nuclear na Agricultura, Evolutiva e de Sistemas, Universidade de São Paulo, São Paulo, Brazil. ⁴Instituto Nacional de Ciência e Tecnologia do Bioetanol, São Paulo, Brazil. ⁵These authors contributed equally: Lauana Pereira de Oliveira and Bruno Viana Navarro. ✉email: diego.riano@cena.usp.br; msbuck@usp.br

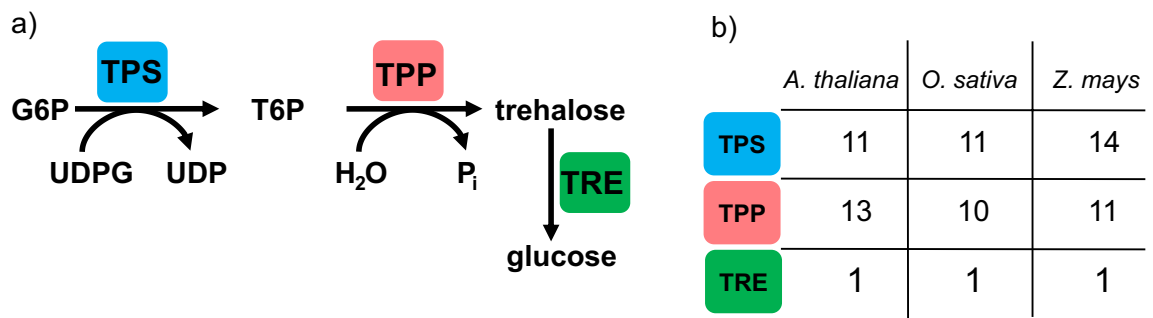


Figure 1. Plant trehalose metabolism. (a) Enzymes involved in trehalose synthesis and degradation. (b) The number of genes encoding trehalose-6-phosphate synthase (TPS), trehalose-6-phosphate phosphatase (TPP), and trehalase (TRE) in *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays*. Trehalose-6-phosphate (T6P) is a central sugar sensor in plant metabolism. Its biosynthesis occurs from UDP-glucose (UDPG) and glucose-6-phosphate (G6P) by TPS activity. Subsequently, the T6P is converted to trehalose by TPP. Trehalose is hydrolyzed into two molecules of glucose by TRE. Multigenic families encode TPS and TPP.

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 (Fig. 1).

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 expression²⁷. 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, three HAD motifs also found in Class II TPS proteins, and a highly variable N-terminal region of unknown function¹⁴. All *AtTPP* genes originated from whole-genome duplication and their encoded proteins are catalytically functional²⁸. 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 exporters³³. 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 aneuploidy³⁹, but research in sugarcane genomics has advanced^{40–51}. One monoloid mosaic reference genome for the sugarcane hybrid R570 was released⁵² along with a high-quality chromosome level genome for *S. spontaneum*, one of the parentals of sugarcane hybrids⁵³. 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 database⁵⁴. 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.

Dataset/Abbreviation	Type of dataset	Variety—Species	Tissue	Sequencing technology	BioProject
SCA5 ⁵⁵	Expressed Sequence Tag (EST)	> 3 varieties (Including SP80-3280)	Different tissues	ABI sequencer	sucest-fun.org*
SCA3 ⁵⁶	Transcriptomic	SP80-3280	Leaf	Illumina Hi-Seq2500	PRJNA244522
SCA4 ⁴⁹	Genomic	SP80-3280	Leaf	Illumina HiSeq2000	PRJNA272769
SAC2_1 and SAC2_2 ⁴⁷	Transcriptomic	10 varieties	Internode	Illumina HiSeq4000	PRJNA356226
	Transcriptomic	22 varieties	Leaf, internode and root	PacBio	
SCA1_1 and SCA1_2 ⁵²	Transcriptomic	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

Table 1. Publicly available datasets used to identify genes encoding trehalose metabolizing enzymes in sugarcane. *Data available in <http://sucest-fun.org/>. **Data deposited in the EMBL-European Bioinformatics Institute.

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

Table 2. Number of putative trehalose pathway protein sequences found in the species under study, using the EggNOG database. The defined OGs for sequences from model species and sugarcane were joined, recovering 15 OGs. 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 clade distribution follows the nomenclature established by^{15,16}.

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 (Fig. 2b). 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 Fig. 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 scan⁵⁷. 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 Fig. S1: a-l). Only sugarcane sequences having $\geq 80\%$ of their respective model domains (71 TPS, 93 TPP, and three TRE) (Fig. 2a and b) were considered filtered and had their proteins shown in the phylogenetic trees (Supplementary Table S1 and Supplementary Fig. 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 (Fig. 3). 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

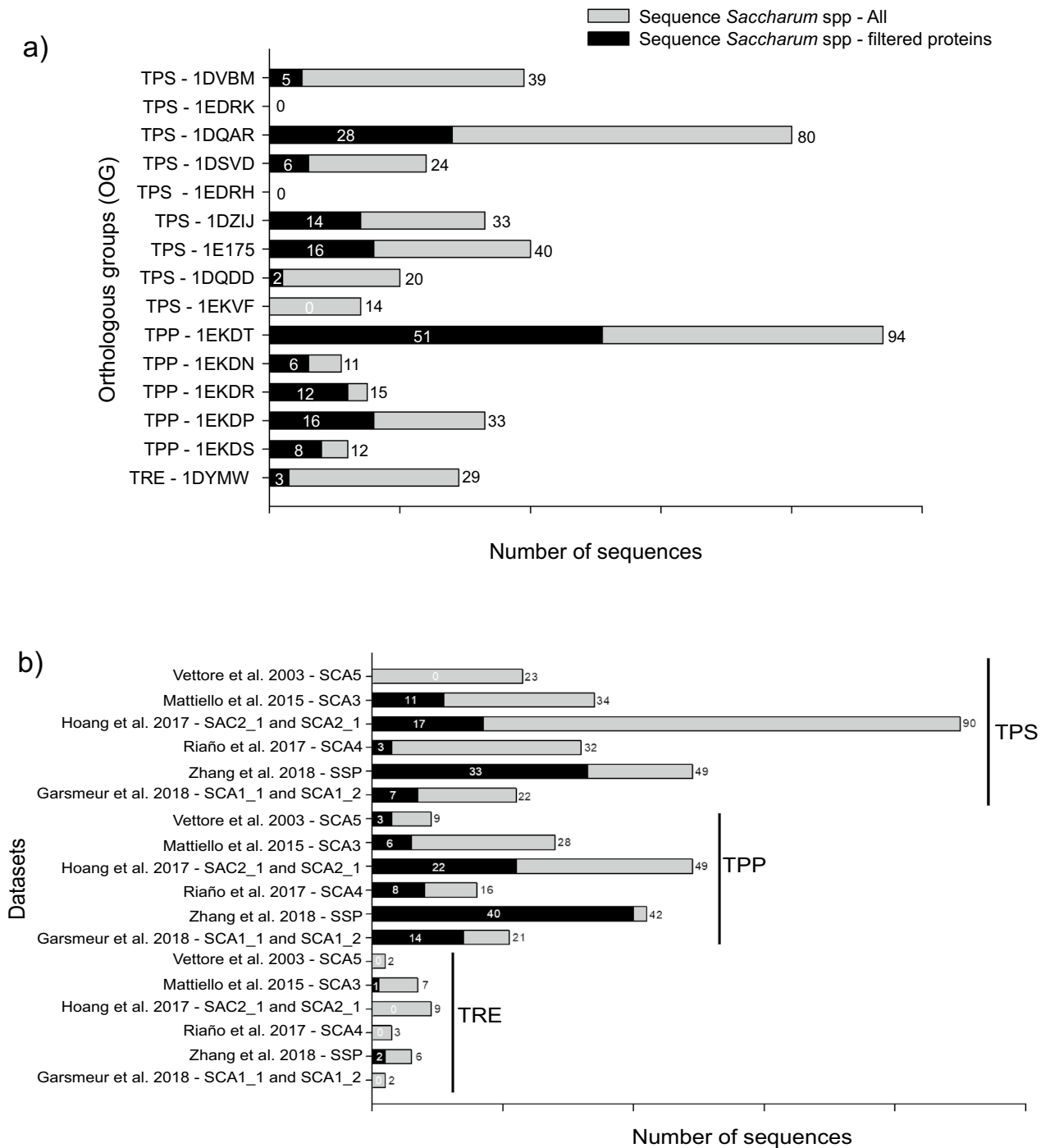


Figure 2. The number of sequences related to the trehalose metabolism pathway in sugarcane. **(a)** Dispersion of total (grey) and filtered (black) sequences associated with orthologous groups (OG) from TPS, TPP, and TRE. The sequences related to each OG are described in Supplementary Table S1. **(b)** Dispersion of total (grey) and filtered (black) sequences among sugarcane publicly available datasets. The datasets (sequence IDs) analyzed were described in Table 1: SCA5⁵⁵, SCA3⁵⁶, SCA4⁴⁹, SAC2_1 and SCA2_1⁴⁷, SCA1_1 and SCA1_2⁵²; SSP⁵³.

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 (Fig. 4). Besides, the classification in OGs by EggNOG does not necessarily reflect the different subclades in the phylogenetic analyses for both TPS and TPP.

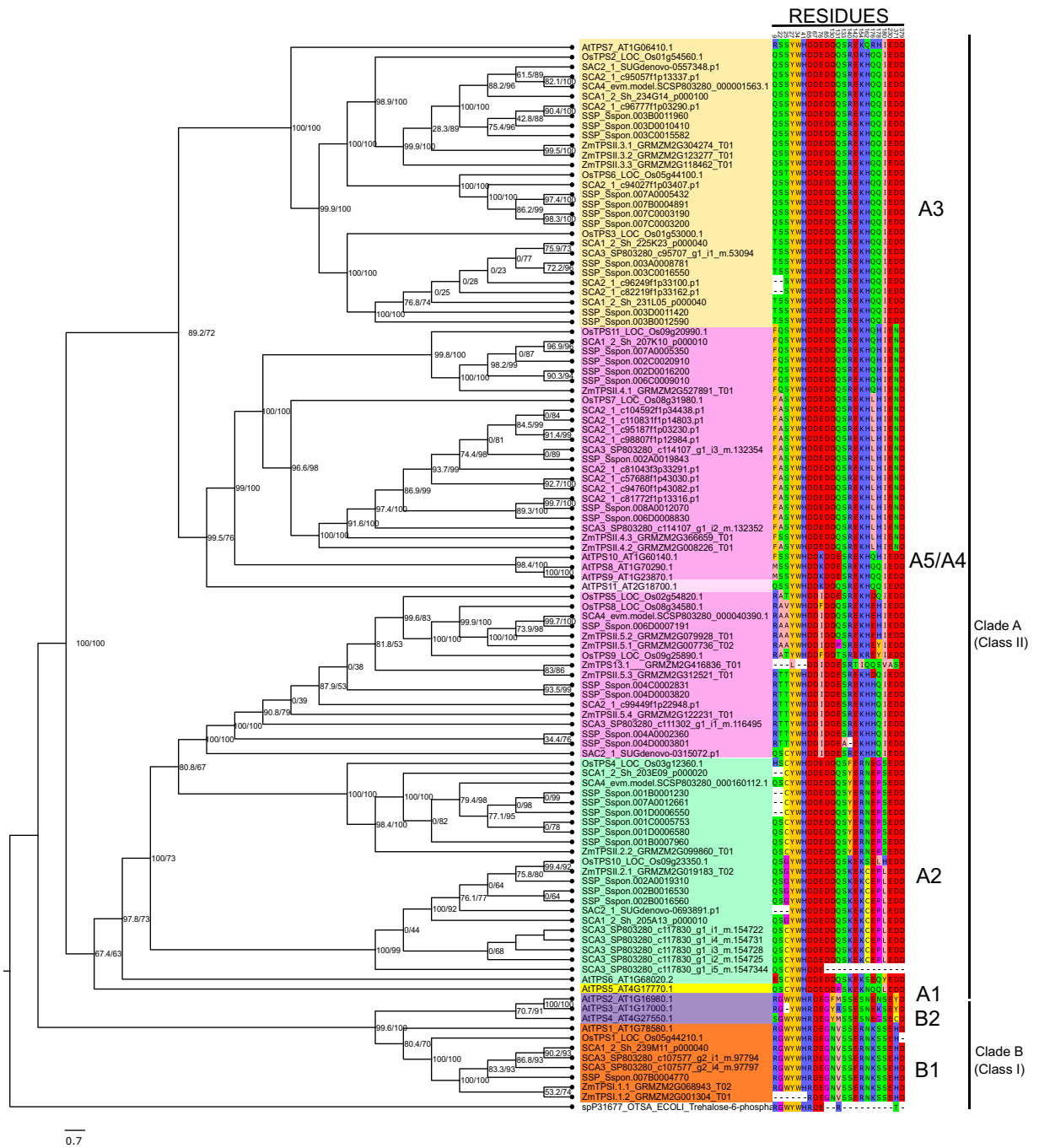


Figure 3. Phylogenetic tree of trehalose-6-phosphate synthase (TPS) from sugarcane, *Zea mays*, *Oryza sativa*, and *Arabidopsis thaliana*. Filtered sugarcane sequences were used to construct a phylogenetic tree based on previous protein sequences obtained from^{15,16}. The tree was built with IQ-TREE⁵⁹ 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^{15,16}. Residues related with the division of subclasses of TPS Class I and II are highlighted. Databases and accession numbers are listed in Supplementary Table S1.

Identity analysis of TPS, TPP, and TRE in sugarcane. Because most of the trehalose metabolizing enzymes have isoforms with high similarity in some cases⁶² and we retrieved the corresponding sugarcane sequences from different datasets (Fig. 2b), 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 (Fig. 5). Likewise, 26 TPS, 60 TPP, and three TRE sequences showed an identity of ≤ 97% (Fig. 5 and Supplementary

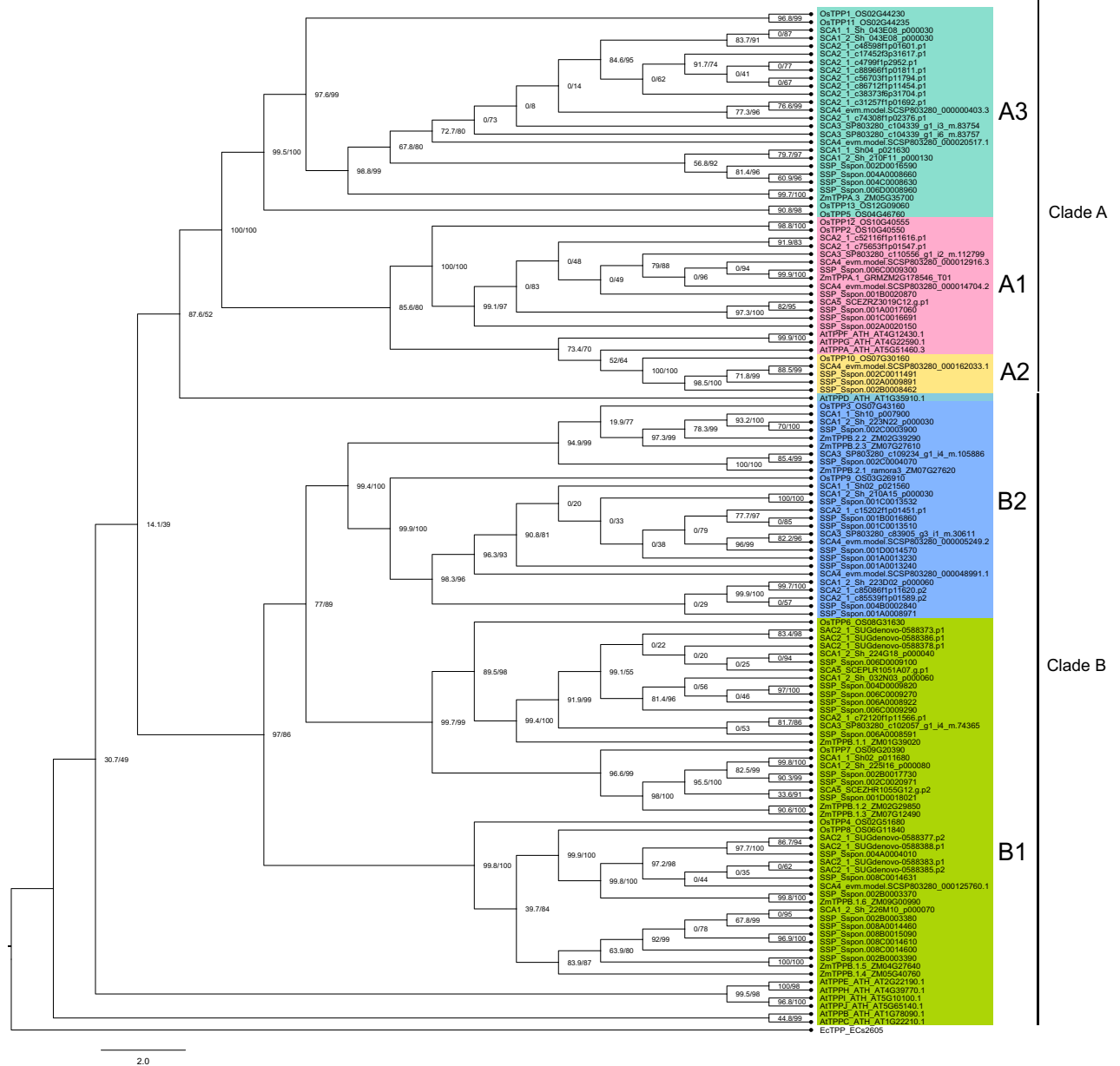


Figure 4. Phylogenetic tree of trehalose-6-phosphate phosphatase (TPP) from sugarcane, *Zea mays*, *Oryza sativa*, and *Arabidopsis thaliana*. Filtered sugarcane sequences were used to construct a phylogenetic tree based on previous protein sequences obtained from¹⁶. The tree was built with IQ-TREE⁵⁹ 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 topology¹⁶. Databases and accession numbers are listed in Supplementary Table S1.

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 a published RNA-seq data⁵⁶. 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 (Fig. 6a–c). *TPP* clade A3 contained the largest 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_*

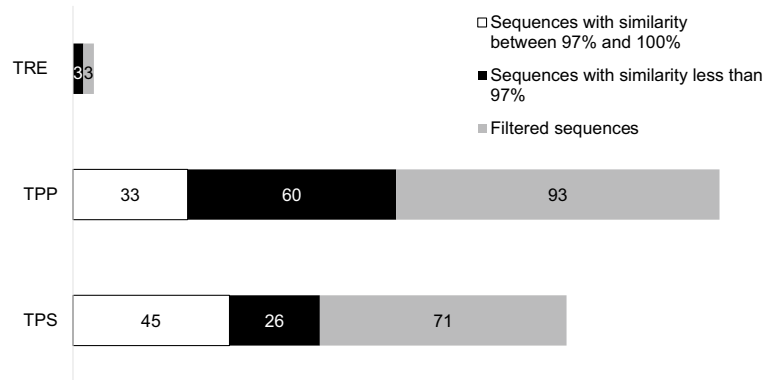


Figure 5. Identity of global pairwise alignment for multiple filtered sequences. Pairwise sequence alignments between all pairs of sequences were computed with the software needle from the EMBOSS v6.0.0. suite, using default settings.

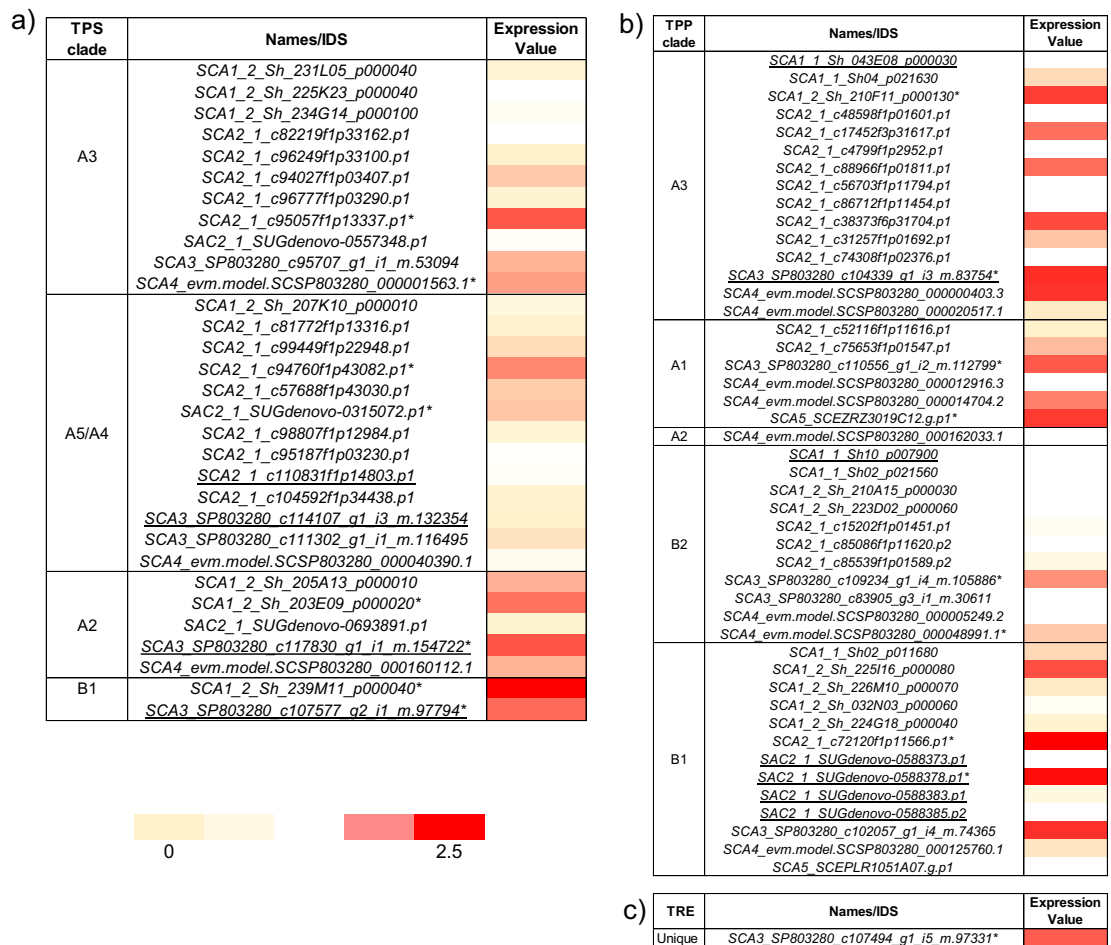


Figure 6. Expression values of trehalose-6-phosphate synthase (TPS) (a), trehalose-6-phosphate phosphatase (TPP) (b), and trehalase (TRE) (c). The heatmaps show the transformed TPM [log₁₀(TPM + 1)], using a subset of the transcriptomics dataset from⁵⁶. The sugarcane cultivar SP80-3280 was grown in a 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. Each clade has the sequences obtained from the selected datasets (Table 1), which were related to the expression values accessed in⁵⁶. Sequence IDs followed by an asterisk indicate sequences with higher expression value in each clade and underlined sequence IDs specify clustered sequences.

c94760f1p43082.p1 had higher expression values. 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 (Fig. 6a). 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*. These sequences belong to clades A1, A3, and B1, respectively (Fig. 6b). 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 (Fig. 6c).

Class I TPS harbors specific predicted N-glycosylation sites in conserved motifs. We evaluated the putative glycosylation sites (Supplementary Table S3) and conserved motifs (Fig. 7) among filtered TPS Class I and II sequences. Sequences of *Ostreococcus tauri* that have two TPS, one Class I and one Class II enzyme¹⁴, 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 (Fig. 7). Otherwise, the NDTV and NSTL sites were predicted to be located between the two domains (Fig. 7), 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 2jbb.1.A—TRE) were defined as templates in SWISS-MODEL⁶³. The templates had at least 30% sequence identity for TPS, 31% for TPP, and 35% for TRE (Supplementary Table S1 and Supplementary Fig. 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 Fig. 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 (Fig. 6a), as well as those from *A. thaliana*, *O. sativa*, *Z. mays*, and *O. tauri* (Supplementary Fig. S3), were aligned to the TPS from *E. coli* (OtsA)⁶⁴, *Candida albicans* (Tps1)⁶⁵, and *C. albicans* TPP (Tps2)⁶⁶. This alignment (Supplementary Fig. 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* Tps2⁶⁶, displaying mutations that allow the differentiation between catalytic and regulatory classes (Supplementary Fig. S3 and Table 3).

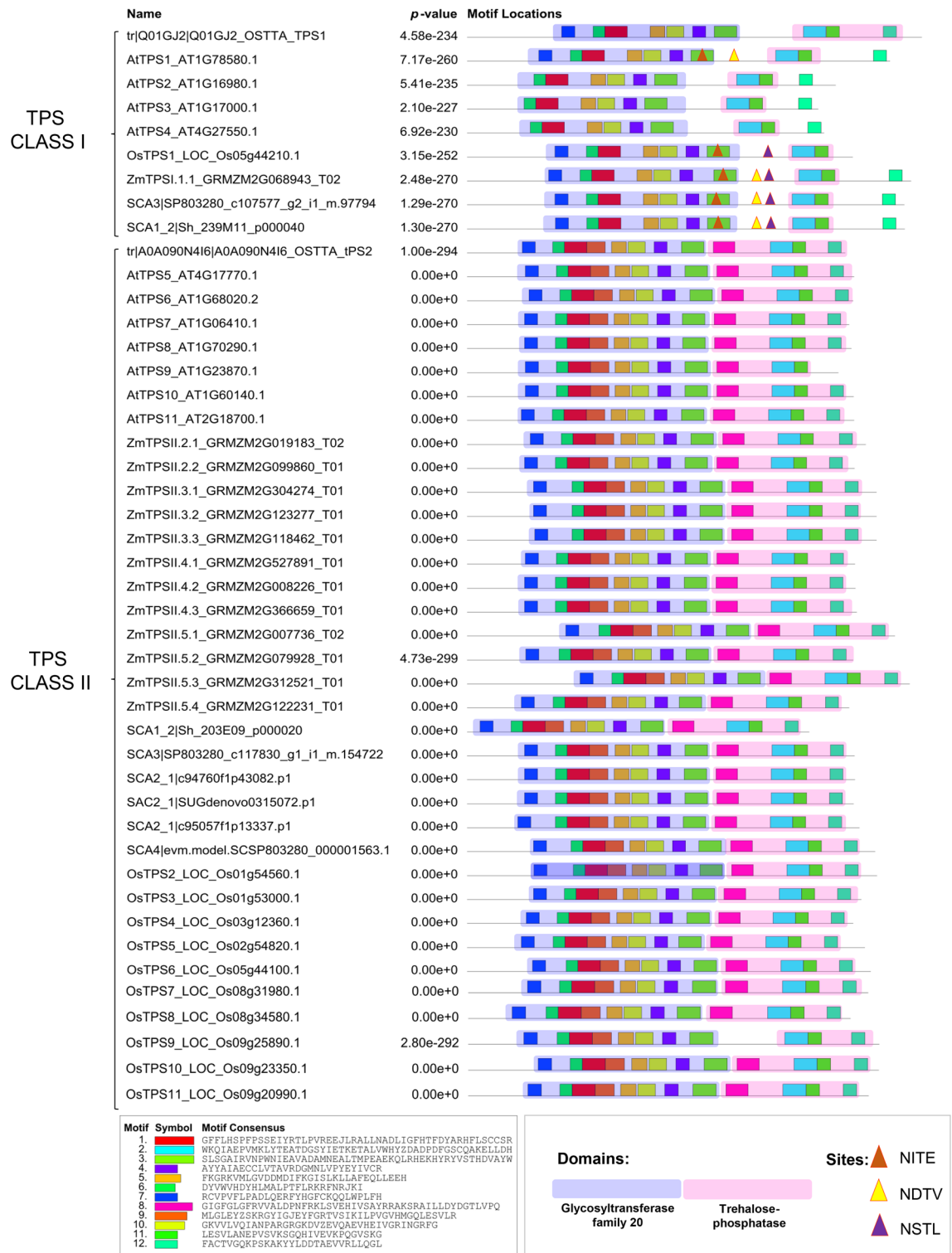
The deduced proteins from the most highly expressed sugarcane TPS Class I and II enzyme-coding transcripts (*SCA1_2Sh_23M11_p000040* and *SCA3_SP803280_c117830_g1_il_m.154722*, respectively) were submitted to 3D structure analysis using SWISS-MODEL (Fig. 8). For sugarcane TPS Class I, two distinct templates, 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 (Fig. 8a) and II (Fig. 8b) displayed similar folding structures (Supplementary Fig. 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 subclasses is related to specific amino acid replacements involved in substrate binding or catalysis residues on both domains (Fig. 8). For instance, residue R9 at the TPS domain, in which Q/T replaces at A3 and A2 subclass, and F/M at subclass A5 (Fig. 3).

Discussion

Sugarcane plays a key role in the Brazilian bioenergy sector regarding economic and societal aspects⁶⁷, such as environmental sustainability⁶⁸. Gains in sugarcane yield have the potential to increase not only bioethanol production, an effective alternative to mitigate CO₂ emissions and climate change, but also other bioproducts³³. 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



	TPS class I		TPS class II	
	Catalytic	Binding	Catalytic	Binding
TPS domain	R ₉ -R	E ₁₇₁ -E	R-Q	E-E
	G ₂₂ -G	R ₃₇₀ -R	G-S	R-R
	W ₄₁ -W		W-C	
	Y ₇₆ -Y		Y-Y	
	W ₈₅ -W		W-W	
	D ₁₃₀ -D		D-D	
	H ₁₅₄ -H		H-H	
	R ₁₆₂ -R		R-D	
	D ₃₇₁ -D		D-D	
E ₃₇₉ -E		E-E		
TPP domain	D₂₅-G	D₂₇-N	D-D	D-D
	S ₆₅ -S	P₃₂-E	S-S	P-/
	K ₁₈₈ -K	V ₃₄ -V	K-K	V-Q
	D₂₃₀-H	R₆₇-S	D-D	R-K
	D ₂₃₄ -D	E ₁₃₁ -E	D-D	E-E
		K₁₃₃-R		K-K
		H₁₄₀-N		H-C
		R₁₄₂-K		R-E
		K₁₇₆-S		K-P
		N₁₇₈-S		N-S
	E ₁₈₀ -E		E-E	

Table 3. Identification of catalytic, binding residues, and substitution at TPS Class I and II from *Saccharum* spp. TPS domain catalytic and binding residues were identified based on *E. coli* TPS^{64,66}. TPP domain catalytic and binding residues were identified based on *Candida albicans* TPP⁶⁶. 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 on⁶⁶, for example, mutations at X25Z. Other residues replacements are described in Supplementary Figure S3 at the phylogenetic tree analysis (Fig. 3).

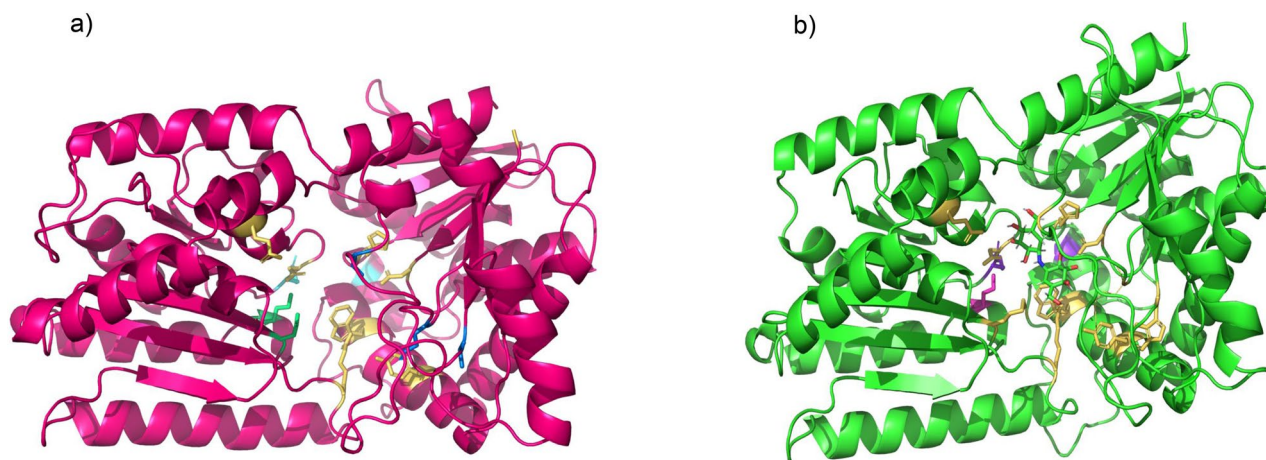


Figure 8. Three-dimensional (3D) structures of trehalose-6-phosphate synthase (TPS) Class I and II. 3D template structures of TPS from *A. fumigatus* (ID model 5hvm.1.A) were used as models to predict the sugarcane TPS Class I and II structure. (a) Sugar cane TPS Class I structure (SCA1_2_Sh239M11_p000040): 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) Sugar cane TPS Class II structure (SCA3_SP803280_c117830_g1_il_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.

function and structure⁵². However, recent advances in sequencing technologies and computational strategies for genome assembly are opening the way to deciphering the sugarcane genome^{49,52,53}. 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^{47,49,52,53,55,56}. 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^{69–71} and transcriptomic studies that have identified changes in the expression of putative genes encoding TPS and TPP^{56,72–76}. 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⁻¹FW, although these sugars did not correlate linearly⁷⁷. In contrast⁷⁸, 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*⁷⁹. 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 origins¹⁴. 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 (Figs. 3 and 4). Most diploid plants have only one TPS Class I gene, except for paleopolyploid species such as *Z. mays* and poplar (*Populus trichocarpa*), which have two¹⁴. Four sugarcane sequences from three datasets were grouped in TPS Class I, suggesting a similar pattern observed for the other paleopolyploid species (Fig. 3). 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 Fig. S3).

Differently from AtTPS1, the predicted AtTPS2–4 proteins lack the N-terminal auto-inhibitory domain and appear to be restricted to the Brassicaceae¹⁴, in which these sequences constitute the subclade B2 (Table 2 and Fig. 3). Most of the filtered sugarcane TPS sequences ($\sim 94\%$) 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 by⁸⁰, who used phylogenetic trees to classify one sugarcane sequence as possibly catalytic and eight as regulatory.

Unlike TPS, all TPP encode active enzymes²⁸, classified in clades A and B and their respective subclades (Fig. 4)^{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 ($\sim 37\%$) (Fig. 4). 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 Fig. S2).

Multigenic families encode plant TPS and TPP through duplication events which started earlier for TPS than for TPP genes¹⁵. 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 (Fig. 5), 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 plants⁸⁴, and changes in their quantity can modify gene expression and metabolism, maintaining sucrose levels within an optimal range⁸¹. The in silico expression analysis indicates a variable TPS and TPP gene expression profile in leaves of the sugarcane cultivar SP80-3280 (Fig. 6a and b).

For TPP, subcellular localization was used to identify all *AtTPP* cell- and tissue-specific expression patterns, suggesting neofunctionalization after gene duplications²⁸. Our results showed that the sugarcane TPP sequence with the highest expression value belongs to the clade B1 (Fig. 6b). *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^{82,83}.

For TPS, the highest expression value belongs to a TPS Class I (Fig. 6a), 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 (Fig. 6a). Subclades A2 and A3 also contain AtTPS6 and AtTPS7, which regulate plant architecture, cell shape, and trichome branching⁸⁴. Besides, they are thought to be related to signal transduction during stress resistance⁶². 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 clade¹⁰, 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 (Fig. 8, Table 3, and Supplementary Figure S3). The catalysis residues R9 and G22 were

conserved in Class I and mutated in Class II sequences (Fig. 3). These residues are important for binding with glucose-6-phosphate and UDP, respectively⁶⁶.

Likewise, the TPP domain of TPS sequences was also analyzed (Table 3). *C. albicans* Tps2 maintains the catalytic activity associated with the TPP domain⁶⁶. 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 inactive⁶⁶. 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 phosphatase⁶⁶.

Motif analysis in the TPS proteins was performed to identify specific protein regions associated with a regulatory or catalytic function (Fig. 7). The *Z. mays* genome encodes two Class I TPS sequences (Fig. 3). However, only one (ZmTPSI.1.1) is functional and has all conserved TPS motifs¹⁶. The second isoform (ZmTPSI.1.2) is truncated and does not have some of the residues necessary for substrate binding¹⁶. Some studies showed different amounts of TPS motifs, 6 in sugarcane⁸⁰, 20 in potato⁶², and 12 in cotton⁸⁵. Motif 9 (motif 2 in⁸⁰) and motif 8 are present in the TPS domain of TPS Class II of sugarcane and cotton⁸⁵. 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 fidelity⁸⁶. 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 algae⁸⁶. *N*-glycosylation is one of the most common and chemically complex post-translational modifications in eukaryotes⁸⁷. 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 activity⁸⁷.

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 phase⁹⁴. 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.

Material 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 database⁵⁴. 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^{47,49,52,53,55,56} (Table 1).

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. suite⁹⁵. Additionally, multiple sequence alignments for the protein sequences of each OG were generated with MAFFT⁹⁶, and dubious regions were removed from the alignments using TrimAI v1.2⁹⁷.

Phylogenetic inference was performed by IQ-TREE v1.6.9⁵⁹, 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 v2.9⁹⁸. Lastly, the phylogenetic trees of filtered TPS and TPP sugarcane sequences were constructed based on¹⁵ and¹⁶, 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 v1.0¹⁰⁰. The three-dimensional structures were modeled by SWISS-MODEL⁶³. 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*^{64,66}, as well as TPS and TPP from *C. albicans*^{65,66}.

In silico transcript expression patterns. TPS, TPP, and TRE expression levels were recovered from a published transcriptomics dataset of leaf development from the hybrid SP80-3280⁵⁶. 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. In⁵⁶ 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 BBDuk2¹⁰¹, to remove remainders of rRNA and low-quality regions as well as adapters. Salmon v1.1.0¹⁰² was used to estimate transcript expression levels expressed as Transcripts per Million (TPM) transformed as $\log_{10}(\text{TPM} + 1)$.

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 software¹⁰³. 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 MEME¹⁰⁴, 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 Fig. S4.

Data availability

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

Received: 3 January 2022; Accepted: 22 March 2022

Published online: 07 May 2022

References

- Lunn, J. E. Sucrose Metabolism. in *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0021259> (2008).
- Ruan, Y. L. Sucrose metabolism: Gateway to diverse carbon use and sugar signaling. *Annu. Rev. Plant Biol.* **65**, 33–67 (2014).
- Smith, A. M. & Stitt, M. Coordination of carbon supply and plant growth. *Plant Cell Environ* **30**, 1126–1149 (2007).
- Czedik-Eysenberg, A. *et al.* The interplay between carbon availability and growth in different zones of the growing maize leaf. *Plant Physiol.* **172**, 943–967 (2016).
- León, P. & Sheen, J. Sugar and hormone connections. *Trends Plant Sci.* **8**, 110–116 (2003).
- Nunes-Nesi, A., Fernie, A. R. & Stitt, M. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol. Plant* **3**, 973–996 (2010).
- Li, L. & Sheen, J. Dynamic and diverse sugar signaling. *Curr. Opin. Plant Biol.* **33**, 116–125 (2016).
- Forzani, C., Turqueto Duarte, G. & Meyer, C. The plant target of rapamycin kinase: A connecTOR between sulfur and growth. *Trends Plant Sci.* **23**, 472–475 (2018).
- 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. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21925–21935 (2019).
- Fichtner, F. & Lunn, J. E. The role of trehalose 6-phosphate (Tre6P) in plant metabolism and development. *Annu. Rev. Plant Biol.* <https://doi.org/10.1146/annurev-arplant-050718-095929> (2021).
- Paul, M. J., Primavesi, L. F., Jhurreea, D. & Zhang, Y. Trehalose metabolism and signaling. *Annu. Rev. Plant Biol.* **59**, 417–441 (2008).
- Leyman, B., Van Dijck, P. & Thevelein, J. M. An unexpected plethora of trehalose biosynthesis genes in *Arabidopsis thaliana*. *Trends Plant Sci.* **6**, 510–513 (2001).
- Avonce, N., Mendoza-Vargas, A., Morett, E. & Iturriag, G. Insights on the evolution of trehalose biosynthesis. *BMC Evol. Biol.* <https://doi.org/10.1186/1471-2148-6-109> (2006).
- Lunn, J. E. Gene families and evolution of trehalose metabolism in plants. *Funct. Plant Biol.* **34**, 550–563 (2007).
- 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).
- Henry, C. *et al.* The trehalose pathway in maize: conservation and gene regulation in response to the diurnal cycle and extended darkness. *J. Exp. Bot.* **65**, 5959–5973 (2014).
- Vogel, G. *et al.* Trehalose metabolism in *Arabidopsis*: Occurrence of trehalose and molecular cloning and characterization of trehalose-6-phosphate synthase homologues. *J. Exp. Bot.* **52**, 1817–1826 (2001).
- 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 Environ.* **32**, 1015–1032 (2009).
- 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*. *Mol. Plant* **3**, 406–419 (2010).
- 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*. *Biochem. J.* **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 Physiol.* **146**, 1834–1861 (2008).
22. Schluepmann, H. *et al.* Trehalose mediated growth inhibition of *Arabidopsis* seedlings is due to trehalose-6-phosphate accumulation. *Plant Physiol.* **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 J.* **49**, 463–491 (2007).
24. Baena-González, E. & Sheen, J. Convergent energy and stress signaling. *Trends Plant Sci.* **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 Rep.* **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 Physiol.* **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*. *Biochem. J.* **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 J.* **92**, 611–623 (2017).
31. Lunn, J. E., Delorge, I., Figueroa, C. M., Van Dijck, P. & Stitt, M. Trehalose metabolism in plants. *Plant J.* **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. *J. Exp. Bot.* **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 Res.* **7**, 24–35 (2014).
34. Jackson, P. A. Breeding for improved sugar content in sugarcane. *Field Crop Res.* **92**, 277–290 (2005).
35. Scortecci, K. C. *et al.* Challenges, opportunities and recent advances in sugarcane breeding. *Plant Breed.* **1**, 267–296 (2012).
36. Grof, C. P. L. & Campbell, J. A. Sugarcane sucrose metabolism: scope for molecular manipulation. *Funct. Plant Biol.* **28**, 1–12 (2001).
37. Burnquist, W. L. Evaluating sugarcane R&D performance: Evaluation of three breeding programs. *Proc. Int. Soc. Sugarcane 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 Biotechnol. J.* **8**, 263–276 (2010).
39. Thirugnanasambandam, P. P., Hoang, N. V. & Henry, R. J. The challenge of analyzing the sugarcane genome. *Front. Plant Sci.* **9**, 616 (2018).
40. Piperidis, G., Piperidis, N. & D'Hont, A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol. Genet. 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. Breed.* **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. *Sci. Rep.* **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 J.* **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. *Front. Plant Sci.* **7**, 342 (2016).
46. Racedo, J. *et al.* Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol.* **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. *Nat. Commun.* **9**, 1–10 (2018).
53. Zhang, J. *et al.* Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **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 Res.* **44**, D286–D293 (2016).
55. Vettore, A. L. *et al.* Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res.* **13**, 2725–2735 (2003).
56. Mattiello, L. *et al.* Physiological and transcriptional analyses of developmental stages along sugarcane leaf. *BMC Plant Biol.* **15**, 1–21 (2015).
57. Potter, S. C. *et al.* HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
58. 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. *Int. J. Mol. Sci.* **17**, 1077 (2016).
59. 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).
60. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
61. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
62. 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).
63. Biasini, M. *et al.* SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).

64. Kaasen, I., McDougall, J. & Strøm, A. R. 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).
65. 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**, e00643 (2017).
66. Miao, Y. *et al.* Structures of trehalose-6-phosphate phosphatase from pathogenic fungi reveal the mechanisms of substrate recognition and catalysis. *Proc. Natl. Acad. Sci.* **113**, 7148–7153 (2016).
67. Bordonal, R. D. O. *et al.* Sustainability of sugarcane production in Brazil. A review. *Sustain. Sugarcane Prod. Braz. Rev.* **38**, 1–23 (2018).
68. Tammissola, J. Towards much more efficient biofuel crops—can sugarcane pave the way?. *GM Crops* **1**, 181–198 (2010).
69. Glasziou, K. T. & Gayler, K. R. Sugar transport: Occurrence of trehalase activity in sugar cane. *Planta* **85**, 299–302 (1969).
70. Alexander, A. G. Studies on trehalase in *Saccharum* spp. leaf and storage tissues. *Plant Cell Physiol.* **14**, 157–168 (1973).
71. Fleischmacher, O. L., Vattuone, M. A., Prado, F. E. & Sampietro, A. R. Specificity of sugar cane trehalase. *Phytochemistry* **19**, 37–41 (1980).
72. Carson, D. L., Huckett, B. I. & Botha, F. C. Sugarcane ESTs differentially expressed in immature and maturing internodal tissue. *Plant Sci.* **162**, 289–300 (2002).
73. 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).
74. McCormick, A., Cramer, M. & Watt, D. Differential expression of genes in the leaves of sugarcane in response to sugar accumulation. *Trop. Plant Biol.* **1**, 142–158 (2008).
75. 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. *Am. J. Plant Sci.* <https://doi.org/10.4236/ajps.2013.412A3011> (2013).
76. Vicentini, R. *et al.* Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. *PLoS ONE* **10**, e0134909 (2015).
77. 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).
78. Glassop, D., Roessner, U., Bacic, A. & Bonnett, G. D. Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation?. *Plant Cell Physiol.* **48**, 573–584 (2007).
79. 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).
80. 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).
81. Figueroa, C. M. & Lunn, J. E. A tale of two sugars: Trehalose 6-phosphate and sucrose. *Plant Physiol.* **172**, 7–27 (2016).
82. Krasensky, J., Broyart, C., Rabanal, F. A. & Jonak, C. The redox-sensitive chloroplast trehalose-6-phosphate phosphatase AtTPPD regulates salt stress tolerance. *Antioxid. Redox Signal.* **21**, 1289–1304 (2014).
83. 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. *J. Exp. Bot.* **71**, 4285–4297 (2020).
84. 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 Physiol.* **146**, 97–107 (2008).
85. 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 Genet.* **17**, 54 (2016).
86. Fichtner, F. *et al.* Functional features of TREHALOSE-6-PHOSPHATE SYNTHASE1, an essential enzyme in *Arabidopsis*. *Plant Cell* **32**, 1949–1972 (2020).
87. 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. *Sci. Rep.* **5**, 1–7 (2015).
88. Paul, M. J., Gonzalez-Urriarte, A., Griffiths, C. A. & Hassani-Pak, K. The role of trehalose 6-phosphate in crop yield and resilience. *Plant Physiol.* **177**, 12–23 (2018).
89. Inman-Bamber, N. G. & Smith, D. M. Water relations in sugarcane and response to water deficits. *Field Crop Res.* **92**, 185–202 (2005).
90. 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).
91. Lakshmanan, P., & Robinson, N. Stress physiology: Abiotic stresses. *Sugarcane Physiol. Biochem. Funct. Biol.* 411–434 (2013).
92. Ferreira, T. H. S. *et al.* Sugarcane water stress tolerance mechanisms and its implications on developing biotechnology solutions. *Front. Plant Sci.* **8**, 1077 (2017).
93. Marchiori, P. E. R. *et al.* Physiological plasticity is important for maintaining sugarcane growth under water deficit. *Front. Plant Sci.* **8**, 2148 (2017).
94. De Souza, A. P., Grandis, A., Arenque-Musa, B. C. & Buckeridge, M. S. Diurnal variation in gas exchange and nonstructural carbohydrates throughout sugarcane development. *Funct. Plant Biol.* **45**, 865–876 (2018).
95. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
96. 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 Res.* **30**, 3059–3066 (2002).
97. 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).
98. Chen, K., Durand, D. & Farach-Colton, M. NOTUNG: A program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **7**, 429–447 (2000).
99. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
100. Liu, W. *et al.* IBS: An illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **31**, 3359–3361 (2015).
101. Bushnell, B., Rood, J. & Singer, E. BBMerge—Accurate paired shotgun read merging via overlap. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0185056> (2017).
102. 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).
103. Gupta, R., Jung, E. & Brunak, S. Prediction of N-glycosylation sites in human proteins. <http://www.cbs.dtu.dk/services/NetNGlyc/> (2004).
104. Bailey, T. L. *et al.* MEME suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202 (2009).

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 Matos 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

B.V.N., L.P.O., M.C.M.M., and M.S.B. conceived the study. B.V.N., D.M.R.P., J.P.J.P., A.R.L., L.P.O., and M.C.M.M. carried out analyzes and interpreted the data. B.V.N., L.P.O., M.C.M.M., D.M.R.P. and M.S.B. wrote and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-11508-x>.

Correspondence and requests for materials should be addressed to D.R.-P. or M.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022