Functional and comparative analysis of *THI1* gene in grasses with a focus on sugarcane

Henrique Moura Dias[1,*], Andreia Prata Vieira[1,*], Erika Maria de Jesus[1], Nathalia de Setta[1,2], Gesiele Barros[1] and Marie-Anne Van Sluys[1]

[1] Botanica/IB, Universidade de Sao Paulo, Sao Paulo, Sao Paulo, Brazil
[2] Universidade Federal do ABC, Sao Bernardo do Campo, Sao Paulo, Brazil
* These authors contributed equally to this work.

ABSTRACT

De novo synthesis of thiamine (vitamin B1) in plants depends on the action of thiamine thiazole synthase, which synthesizes the thiazole ring, and is encoded by the *THI1* gene. Here, we investigated the evolution and diversity of *THI1* in Poaceae, where C4 and C3 photosynthetic plants co-evolved. An ancestral duplication of THI1 is observed in Panicoideae that remains in many modern monocots, including sugarcane. In addition to the two sugarcane copies (*ScTHI1-1* and *ScTHI1-2*), we identified *ScTHI1-2* alleles showing differences in their sequence, indicating divergence between *ScTHI1-2a* and *ScTHI1-2b*. Such variations are observed only in the Saccharum complex, corroborating the phylogeny. At least five *THI1* genomic environments were found in Poaceae, two in sugarcane, *M. sinensis*, and *S. bicolor*. The *THI1* promoter in Poaceae is highly conserved at 300 bp upstream of the start codon ATG and has cis-regulatory elements that putatively bind to transcription factors associated with development, growth, development and biological rhythms. An experiment set to compare gene expression levels in different tissues across the sugarcane R570 life cycle showed that *ScTHI1-1* was expressed mainly in leaves regardless of age. Furthermore, *ScTHI1* displayed relatively high expression levels in meristem and culm, which varied with the plant age. Finally, yeast complementation studies with THI4-defective strain demonstrate that only *ScTHI1-1* and *ScTHI1-2b* isoforms can partially restore thiamine auxotrophy, albeit at a low frequency. Taken together, the present work supports the existence of multiple origins of THI1 harboring genomic regions in Poaceae with predicted functional redundancy. In addition, it questions the contribution of the levels of the thiazole ring in C4 photosynthetic plant tissues or potentially the relevance of the THI1 protein activity.

Subjects Agricultural Science, Evolutionary Studies, Genetics, Genomics, Plant Science
Keywords Thiazole biosynthesis, Evolutionary diversity, Genomic characterization, Plant development, Promoter analysis, Genetic complementation, Gene expression

INTRODUCTION

The activity of thiamine thiazole synthase (THI1) is key for the biosynthesis of thiamine, also known as a vitamin B1 (*Ribeiro et al., 1996*). In plants, thiamine is produced through the condensation of two heterocyclic phosphate molecules named hydroxymethyl-thiazole (HET-P) and hydroxymethyl-pyrimidine (HMP-P) (*Kong et al., 2008*; *Jurgenson, Begley &*

Ealick, 2009; *Goyer, 2010*; *Rapala-Kozik, 2011*). As a co-factor of several enzymes, thiamine participates in carbohydrate metabolisms (*Rapala-Kozik, 2011*; *Fitzpatrick & Noordally, 2021*) such as pyruvate dehydrogenases involved in the finetuning between glucose and fatty acid oxidation (*LeClere, Rampey & Bartel, 2004*; *Marsden, McMillan & Hanefeld, 2020*), transketolase associated with photosynthesis in the Calvin cycle (*Frank, Leeper & Luisi, 2007*; *Rapala-Kozik, 2011*; *Rocha et al., 2014*), and pyruvate decarboxylases that act in alcoholic fermentation (*Forlani, Mantelli & Nielsen, 1999*). All these pathways and enzymes are essential for maintaining homeostasis and cell functioning.

In addition, to the canonical pathway of the thiazole ring biosynthesis for thiamine production, cumulative evidence in the literature has raised the hypothesis that THI1 plays roles that are still unclear. Initial studies in bacteria and yeast associated *THI1* to tolerance to oxidative stress and DNA repair mechanisms (*Machado et al., 1996*, *1997*; *Medina-Silva et al., 2006*). In *Arabidopsis thaliana*, THI1 was suggested to have a protective function against salt and osmotic stress linked with the abscisic acid (ABA) hormone signaling (*Rapala-Kozik et al., 2012*). More recently, *Li et al. (2016)* have demonstrated that the AtTHI1 protein interacts with Ca^{2+}-dependent protein kinases and modulates stomata closure through ABA signaling during drought stress. A single gene encodes the AtTHI1 protein (*Machado et al., 1996*) which has two signal peptides at the N-terminus of the protein due to the presence of two initiation codons that targets each of the isoforms to either chloroplasts or mitochondria (*Chabregas et al., 2001*, *2003*).

As drought stress strongly constrains crop yield (*Dietz, Zörb & Geilfus, 2021*), we aimed to understand the evolution and diversity of *THI1* in Poaceae, where C4 and C3 photosynthetic plants co-evolved. Species of agricultural interest, such as rice, corn, wheat, sorghum, and sugarcane, are all members of this important family. Sugarcane (*Saccharum* spp.) attracts special attention due to its ability to accumulate higher levels of sucrose in the stalk internodes (*Whittaker & Botha, 1997*). Sucrose is the cheapest and one of the most accessible sources of carbon for human and livestock consumption, and sugarcane provides the best ratio of sugar yield per cultivated area worldwide (*Formann et al., 2020*). Sugarcane is a C4 high-yield photosynthetic crop with the potential for bioenergy. This energy source is explored in Latin America in Brazil, Mexico, and Colombia. Yet, other countries could also increase the contribution of sugarcane to their fuel and energy sectors (*Khan & Khan, 2019*).

The complexity of the sugarcane genome has hindered its full sequencing and accurate gene annotation. A steady and progressive effort worldwide has produced encouraged several initiatives starting with an EST database for sugarcane (*Vettore et al., 2003*) and subsequently, a sugarcane BAC collection (*de Setta et al., 2014*), a monoploid version (*Garsmeur et al., 2018*) and an allele resolved version of a commercial cultivar (*Souza et al., 2019*). In all these genomic resources, two *AtTHI1* homologs are identified. Our synteny studies revealed that C4 photosynthetic plants maintain multiple copies of the THI1 gene however with unequal distribution of genomic loci across closely related species. *THI1* gene sequence diversification using phylogeny in Poaceae and clustering in Saccharum complex is revealed. *Cis*-regulatory elements in the promoter region, as well as differential gene expression pattern of the gene copies along development support that all three

variants are expressed. Furthermore, *ScTHI1-1* and *ScTHI1-2b* were capable of genetically complementing yeast THI4 mutants. Taken together, our findings provide information on the evolution and divergence of the *THI1* gene in Poaceae and shed light on biological aspects of the *ScTHI1* genes in sugarcane.

MATERIALS AND METHODS

Screening of sugarcane *ScTHI1* homologs in the BAC library and in other grasses sequence collection of *THI1* genes in grasses

Based on a conserved region of the two SAS sequences from the SUCEST database (*Vettore et al., 2001*) identified as putative *THI1* homologs, a pair of primers (thi1_F: CAC CAT GGC CGA GAA CAG; thi1_R: CGT ACG AGC TCT CCA AGG AC) was used to screen for the presence of *ScTHI1* in the sugarcane BAC library (*de Setta et al., 2014*). As a result, nineteen BACs were selected for sequencing and assembly as described previously (*de Setta et al., 2014*).

Phytozome database v13 (*Goodstein et al., 2012*) and PLAZA Monocots 4.5 database (*Van Bel et al., 2018*) were screened using BLAST search to identify *THI1* homologs in Poaceae group plants with an *E*-value cutoff of 10^{-5} and coverage ≥70%. Only nucleotide and protein sequences from species with complete sequenced genomes were selected. Manual inspection was performed to verify all putative candidates using the online software InterProScan (*Blum et al., 2021*) and Pfam (*El-Gebali et al., 2019*) to check the presence of the full THI4 protein family domain (PF01946).

Phylogenetic and synteny analyses

Full-length protein sequences were aligned and inspected using MAFFT v7.450 (*Katoh & Standley, 2013*). The topology of phylogenetic tree was generated using the Bayesian analysis algorithm with MrBayes v3.2.6 (*Ronquist et al., 2012*). The node's significance was evaluated by one run of 1,000,000 generations with Metropolis-coupled Monte Carlo Markov chains (MCMC). The Bayesian model parameters were nucmodel = 4by4, nst = 2, and aamodel = mixed for amino acid alignments. Markov chains were sampled every 1,000,000 generations. The remaining trees were used to compute the majority rule consensus tree, the posterior probability of clades and branches lengths. Both analyses were performed in Geneious Prime (version 2021.0.3).

To evaluate synteny across species within *THI1* gene region, a 200 kb region was explored for conserved genes both upstream and downstream within the genome of 24 examined species. *A. thaliana* genome sequence was used as the reference, since it is currently the best genome assembly with highest quality and the completest genome annotation. tBLASTn and the best-fit results (E-value ≤ 2e−10 and identity ≥80%) were selected to explore these genome assemblies. An R package was used to design genes in a chromosome-scale (*Anand & Rodriguez Lopez, 2022*) with adjustments in Inkscape Illustrator.

Saccharum spp. THI1 sequence diversity analysis

A network analysis approach was used to address the THI1 homologs sequence diversity in the *Saccharum* complex. A pair of primers (thi1Conserved_F: CTC CTC AAG TCC TCC TTC GC and thi1Conseved_R: TCA TGC CGA TGT CCT GGA G) was used to generate sequences corresponding to a conserved region of *ScTHI1*. Genomic DNA from closely related species *S. spontaneum* (Mandalay), *S. spontaneum* (IN84-58), *Miscanthus sp.*, *S. officinarum*, Brazilian hybrids SP8032-80, SP7011-43, SP8132-50, RB835486, RB72454, RB867515, and other cultivars (POJ-2878 from Java; NA56-79 from Argentina; NCo-310 from South Africa; and Co-290 from India) were used as template in the PCR as well as the R570 nineteen BACs. Amplicons were cloned and sequenced using the ABI PRIS 3730 DNA ANALYZER (Applied Biosystems™, Waltham, MA, USA). Sequence quality and assembly was performed with Phred-Phrap-Consed package (*Ewing & Green, 1998*; *Gordon, Abajian & Green, 1998*; *Gordon, 2003*). Only bases with phred quality ≥20 were used, resulting in primary sequences of about 330 bp. Homologous regions from SbTHI1, ZmTHI1, MsTHI1 were retrieved by BLASTn from Phytozome v13 database (*Goodstein et al., 2012*).

As described in *Vieira (2018)*, ClustalOmega (*Sievers et al., 2011*) using default parameters was used to align all sequences. DnaSP5 (*Librado & Rozas, 2009*) and NETWORK 4.6.1.2 (*Bandelt, Forster & Röhl, 1999*) software with default parameters were used to generate the network graph. Representative sequences from three R570 BACs were selected as follows, 108_C04 BAC represents *ScTHI1-1*, the one found in the 017_B18 BAC represents *ScTHI1-2a* and the 094_O04 BAC represents *ScTHI1-2b*.

Promoter characterization and distribution of *cis*-regulatory elements in *THI1* homologs

We performed a comparative analysis of the promoter regions of the different occurrences of *THI1* in: *Zea mays*, *Miscanthus sinensis*, *Sorghum bicolor*, *Saccharum spontaneum* (monoploid), and *Saccharum* sp. var. R570. The *THI1* gene sequences were identified *via* tBLASTn on PLAZA database and in BACs of sugarcane (R570 variety), by using the homolog protein sequence of *A. thaliana*, as a query. Upstream sequences of 2,000 bp from the start codon were assessed for conserved features and motifs, by using MEME suite tools (*Bailey et al., 2009*). Twenty motifs were allowed to be from 5 to 25 bp in length, with an *E-value* less than 0.05, the default parameter for MEME (*Powell et al., 2019*). The retrieved motifs were run through TomTom (*Gupta et al., 2007*), *via* the JASPAR Core Plants database (*Khan et al., 2018*), and the respective Uniprot IDs results were collected, if *p-value* was equal or smaller than 0.01. The Uniprot IDs were then used to collect biological GO terms for functions assignment of each motif, and g:Profiler (*Raudvere et al., 2019*) was used for statistical analysis of GO terms overrepresentation.

Subcellular localization prediction

According to *Chabregas et al. (2003)*, we identified two start codons in the *THI1* sequences. The whole amino acids sequence of *THI1* starting from the first start codon (1st ATG) and second (2nd ATG) was used for prediction of signal peptide cleavage site with SignalP

(*Armenteros et al., 2019b*). To subcellular localization prediction was used TargetP (*Armenteros et al., 2019a*).

Plant growth conditions for evaluating the expression of *ScTHI1* homologs

Sugarcane (*Saccharum sp*. var. R570) was vegetatively propagated from the GaTE-Lab sugarcane collection (Instituto de Biociências, USP). The culms were disinfected with 1.5% hypochlorite, germinated on vermiculite, and maintained in a greenhouse for 15 days. The seedlings were then transferred to pots (50 L) with a mixture of substrate 3:1 (commercial substrate and vermiculite). Irrigation occurred systematically, with nutrient supplementation 15 days before harvest. Plants were harvested at 3, 6, and 9 months after sprouting, and separated into apical meristem, leaf, culm, and root.

Total RNA and cDNA synthesis

Total RNA was isolated from fine powder ground tissue with TRIzol (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's recommendations. RNA quantity and purity were measured in an ND-1000 NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The quality and integrity of the RNA were verified by electrophoresis on 1% agarose gels. Total RNA samples were treated with a Turbo DNA-*free* kit (Invitrogen, Waltham, MA, USA). Super Script First III Strand System for RT-PCR kit (Invitrogen, Waltham, MA, USA) was used for complementary DNA (cDNA) synthesis from 500 ng/μL RNA samples.

RT-qPCR assay and gene quantification

Each cDNA sample was subjected to reverse transcriptase quantitative PCR (RT-qPCR) reactions for all genes of interest in each using cDNA-specific TaqManGene Expression Assays on QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems, Waltham, MA, USA). Reaction mixture in a total 10 μl reaction is composed of 1 μl cDNA, 0.25 μL (0.25 μM) TaqMan Probe, 0.3 μL (0.3 μM) of Forward and Reverse primers, 5 μl TaqMan Fast Advanced Master Mix 2x (Applied Biosystems, Waltham, MA, USA) and 3.15 μL nuclease free water. This allowed for the consistent use of standardized thermal cycling conditions: 95 °C for 2 min, followed by 40 cycles of 95 °C for 1 s and 60 °C for 20 s. The standard curve was generated using synthetic genes cloned in BlueHeron pUC MinusMCS plasmid. The standard curve was generated using synthetic CDS of *ScTHI1* genes cloned in BlueHeron pUC MinusMCS vector, serially diluted (10×). The transcripts copy numbers were determined by interpolation of the standard curve. Each sample and standard curve was run in triplicate to ensure reproducibility. Absolute expression data (number of transcripts) were log-transformed to enable statistical analysis (ANOVA), assuming the log normality of the data.

Yeast complementation assay

As previously described in *Vieira (2018)*, three synthetic versions of *ScTHI1* were performed by Blue Heron® Biotech, LLC. Alternative versions of these CDS lacking the N-terminal chloroplast transit peptide (DelN) were produced through PCR amplification

on the synthetic CDSs, using a pair of primers (thi1.1DelN_F: GGA TCC ATG ACC CGC CGC TA for *ScTHI1-1*; and thi1.2DelN_F: GGA TCC ATG ACC CGG CGG TA for both *ScTHI1-2a* and *ScTHI1-2b*) and thi1DelN_R (GTC GAC TCA GGC GTC CAC) for all corresponding to a transcript that includes the second initiation codon (amino acid 76 in Fig. S2) up to the stop codon (amino acid 360).

A restriction enzyme BamH1-Sal1 fragment of the corresponding CDSs were inserted into the yeast expression vector pG-1 (*Schena, Picard & Yamamoto, 1991*) using T4 ligase (Promega, Madison, WI, USA) according to the manufacturer's protocol. Two other constructs were used as controls: the positive control A184V, which corresponds to the *A. thaliana tz-201* mutant CDS and is able to complement the thiamine auxotrophy in the *thi4* yeast mutant (*Papini-Terzi et al., 2003*); and DelN as negative control, which is the *A. thaliana* wild-type (WT) *AtTHI1* lacking the N-terminal chloroplast transit peptide (*Papini-Terzi et al., 2003*). All constructs were transformed into wild type strain W303 (mata, ade2-1, trp1-1, leu2-3-112, can1-100, ura3-1, his3-11-115) as a positive control (*Praekelt, Byrne & Meacock, 1994*) and the thiamine auxotroph KBY5 mutant strain (THI4::URA3) by the heat shock method (*Schiestl & Gietz, 1989*).

All transformant selection was on yeast nitrogen base (YNB) (BD Biosciences, Franklin Lakes, NJ, USA) medium lacking tryptophan. Thiamine auxotrophy was examined by growing cells overnight in liquid YNB without tryptophan, centrifuged, and re-suspended in 10 mM MgSO$_4$. From an OD = 1, three serial and tenfold dilutions were prepared. Yeast complementation was assayed on minimal medium (YNB lacking both thiamine-HCL (YNB-thia) and tryptophan, from USBiological) plates, either with or without thiamine and tryptophan, incubated for 28 days at 30 °C, and analyzed every 4 days.

RESULTS

Gene synteny of *THI1* homologs in sugarcane and Poaceae

As many Poaceae genomes are publicly available, we analyzed the *THI1* by a comparative genomic approach across a wide range of species. Sequences from *Arabidopsis thaliana* and *Joinvillea ascendens* were included as outgroups. For sugarcane, besides the genome from *S. spontaneum* (*Zhang et al., 2018*), additional 19 bacterial artificial chromosomes (BACs) (Table S1) from the R570 cultivar (*de Setta et al., 2014*) containing *THI1* homologs were sequenced and annotated. All sequences containing the putative protein were compared using PFAM and InterproScan databases to screen for the presence of the THI4 protein family domain (PF01946). After removing sequences with incomplete domains and redundant sequences retrieved from this search, 49 THI1 homologs were identified in 23 Poaceae species (Table S2).

A Bayesian phylogenetic tree based on multiple alignments from these 49 protein sequences and the two outgroups was built. The tree topology revealed that the THI1 homologs were separated into two distinct clades, Clade A and Clade B (Fig. 1A and Table S2). Clade A is composed by THI1-homologue single copy sequences from *B. distachyum, B. hybridum, B. mexicanum, B. stacei, B. sylvaticum, H. vulgare, T. aestivum, T. turgidum, T. intermedium*. It also displays a diversifying branch from *Hordeum vulgare, Triticum aestivum, Triticum turgidum,* and *Thinopyrum intermedium*
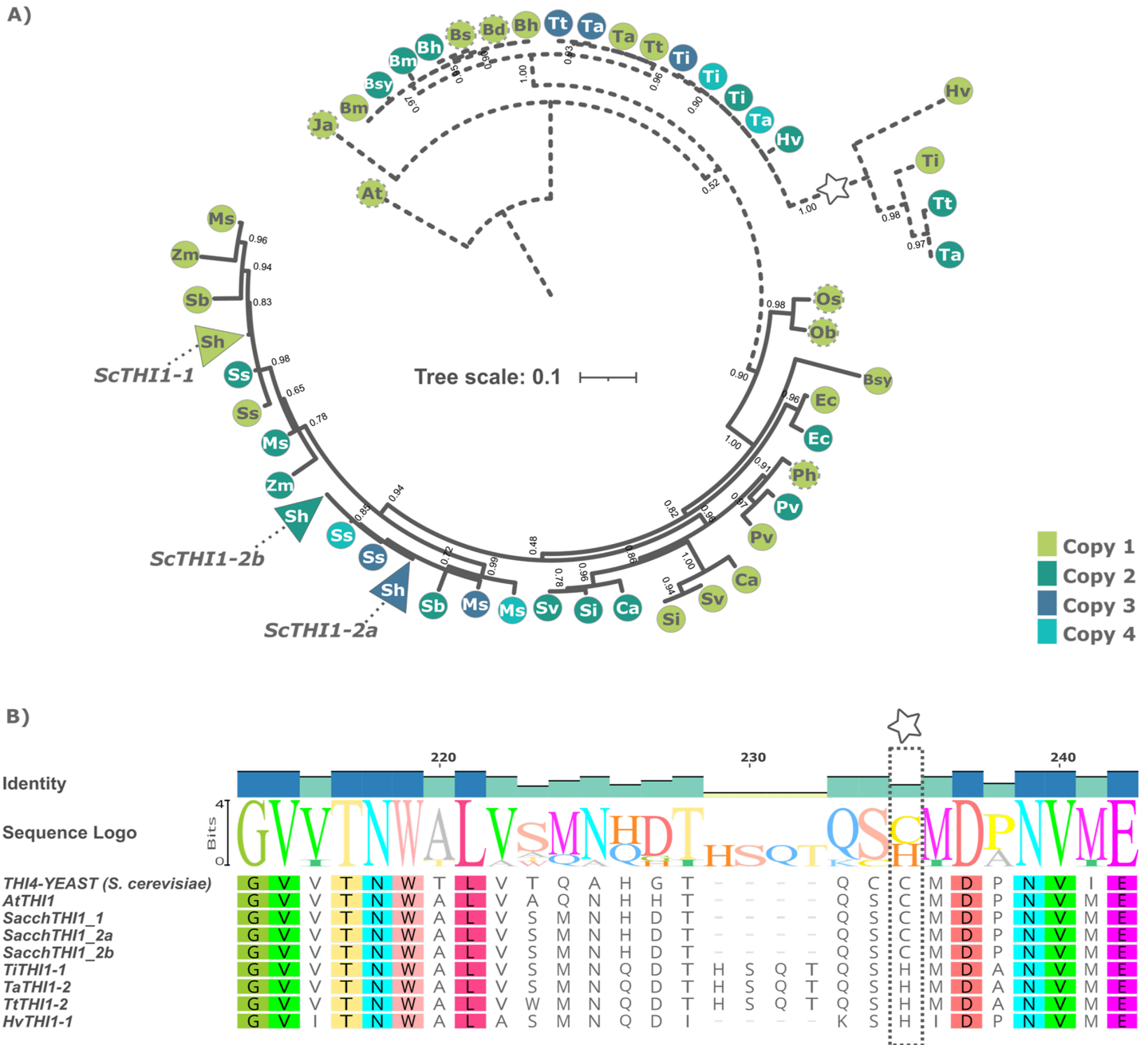
Figure 1 Phylogenetic relationships of THI1 (thiamine thiazole synthase) in Poaceae. (A) The tree containing 51 protein sequences is subdivided into two major clades. Clade A (dashed line) comprises the C3 grasses (*B. distachyum* (Bd), *B. hybridum* (Bh), *B. mexicanum* (Bd), *B. stacei* (Bs), *B. sylvaticum* (Bsy), *H. vulgare* (Hv), *T. aestivum* (Ta), *T. turgidum* (Tt), *T. intermedium* (Ti)) and Clade B (black line) the C4 grasses (*E. coracana* (Ec), *C. americanus* (Ca), *P. hallii* (Ph), *P. virgatum* (Pv), *M. sinensis* (Ms), *Saccharum sp.* var. R570 (ScTHI1), *S. spontaneum* (Ss), *S. italica* (Si), *S. viridis* (Sv), *S. bicolor* (Sb), and *Z. mays* (Zm)). The outgroups are represented by *A. thaliana* (At) and *J. ascendens* (Ja). The circle indicate species and are colored according to the numbers of gene copies. Circles with dashed border show a THI1 homolog single copy. The triangle shows collapsed BAC sequences. The scale shows the phylogenetic distance between protein sequences. For an expanded version with protein names and IDs see Table S2. (B) Multiple sequence alignment of THI1 homologs from *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Saccharum sp.* var. R570, *Hordeum vulgare*, *Triticum aestivum*, *Triticum turgidum*, and *Thinopyrum intermedium* showing diversity at position 238 (stars).

Full-size DOI: 10.7717/peerj.14973/fig-1

that has a previously described amino acid substitution at position 238 (Fig. 1B and Table S2). The *in-silico* analyses of the R570 sugarcane BAC sequences revealed that there are at least two gene variants of *ScTHI1* supporting the two SAS previously described. Those, alongside with *S. spontaneum* sequences, are distributed into two genomic subgroups with a common ancestral node. The amino acid sequence from ScTHI1 copies derived from the SAS1 probe (ScTHI1-1) clustered with SbTHI1-1, ZmTHI1-1, ZmTHI1-2, MsTHI1-1, MsTHI1-2, SsTHI1-1, and SsTHI1-2, while the sequences derived from SAS2 (ScTHI1-2a and ScTHI1-2b) clustered with SbTHI1-2, MsTHI1-3, MsTHI1-4, SsTHI1-3, and SsTHI1-4. Clade B (*Cenchrus americanus, Eleusine coracana, Miscanthus sinensis, Panicum hallii, Panicum virgatum, Saccharum sp., Setaria italica, Setaria viridis, Sorghum bicolor*, and *Zea mays*) present at least two THI1-homologues, with the exception of *Oryza sativa* and *Oryza brachyantha* that have a single copy THI1 homologue.

To investigate the genomic context of *ScTHI1* homologs, we selected their genomic vicinities and compared them with TH1 phylogenetic relatedness. We identified five different genomic regions (A–E) flanking *THI1* genes (Fig. 2 and Table S3). *S. bicolor THI1* genes were found on chromosomes 2 (*SbTHI-2*) and 3 (*SbTHI1-2*) while in *Z. mays* the genes located on chromosomes 3 (*ZmTHI1-2*) and 8 (*ZmTHI1-1*). One copy of *THI1* has been identified in *O. brachyantha, O. sativa*, (both in chromosome 7) *and P. hallii* (chromosome 4). Homologs of *E. coracana* (chromosome 7), *C. americanus* (chromosomes 5 and 7), *S. italica and S. viridis* (both in chromosomes 2 and 4), and *P. virgatum* (chromosome 4) are also included in the present analysis.

Sugarcane *THI1* was found in two of these genomic loci (D and E) that by synteny are grouped in different *ScTHI1* clades. Two well-supported sugarcane clades (bootstrap 100 for *ScTHI1-1* and 99 for *ScTHI1-2*) were found and the clade of *ScTHI1-2* was divided into two subclades. BAC clones 017_B18, 030_H05, 251_N23, and 092_F09 cluster with *ScTHI1-2a* while BACs clones 094_O04, 109_G13, 183_N05, 184_N05, 190_F02, and 222_C03 with *ScTHI1-2b*, also with higher bootstrap values (Fig. 2).

Molecular structure of THI1 homologues

Exon-intron boundaries was examined for each of these 51 genes and the 19 BACs. While *AtTHI1* has two introns, all Poaceae species harbor only one intron with varying sizes (63 to 149 bp) (Fig. S1), suggesting high gene structure conservation. The protein length of the THI1 homologs is also highly conserved (Fig. 3A), with amino acid sequences at least 67% identical, including the less conserved region containing the previously described organellar targeting to mitochondria and/or chloroplast (*Chabregas et al., 2001, 2003*). Despite these features, the amino acid pre-sequence encoded by first start codon is variable, in length and amino acid content among species. Although the sequence and length of signal peptides can vary substantially, computational analyzes depict the existence of a series of conserved aminoacids at given positions, probably resulting in secondary structure conservation (see Fig. 3B). The signal peptide prediction results show that all *ScTHI1* proteins have a chloroplast targeting peptide when transcribed by the first initiation codon (1st ATG), as seen in other Poaceae species. The differences among the
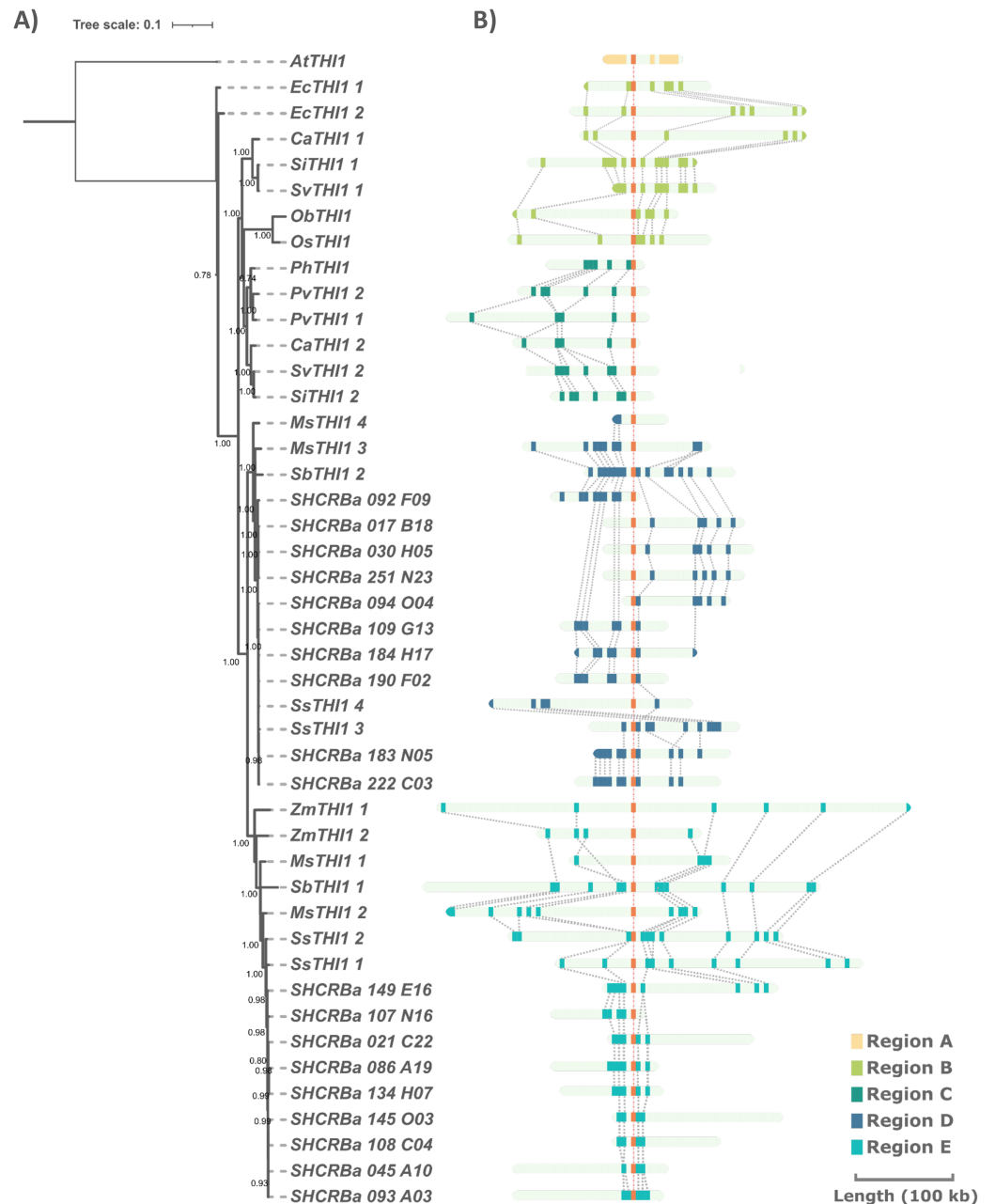
Figure 2 Comparison of *THI1* gene and its genomic flanking regions in sugarcane and other plants through phylogenetic and synteny analyses. (A) Numbers at nodes are branch support values estimated by the aLRT SHlike method implemented in PhyML3.0 (see Methods). The tree is based on the multiple alignments of the two exons of the single *THI1* copy from *A. thaliana* (AtTHI1), *E. coracana* (EcTHI1), *C. americanus* (CaTHI1), *O. sativa* (OsTHI1), *O. brachyantha* (ObTHI1), *S. italica* (SiTHI1), *S. viridis* (SvTHI1), *S. bicolor* (SbTHI1), *M. sinensis* (MsTHI1), *S. spontaneum* (SsTHI1) and BACs sequences of R570 sugarcane (BAC 108_C04, BAC 017_B18, and BAC 094_O04 represent *ScTHI1-1, ScTHI1-2a*, and *ScTHI1-2b*, respectively). (B) Synteny analysis was performed using blastx among sugarcane BAC sequences and sequences from the other genomes obtained from the plant's database (see Table S2). Rectangles indicate genes. Full-size ◩ DOI: 10.7717/peerj.14973/fig-2

A)

B)
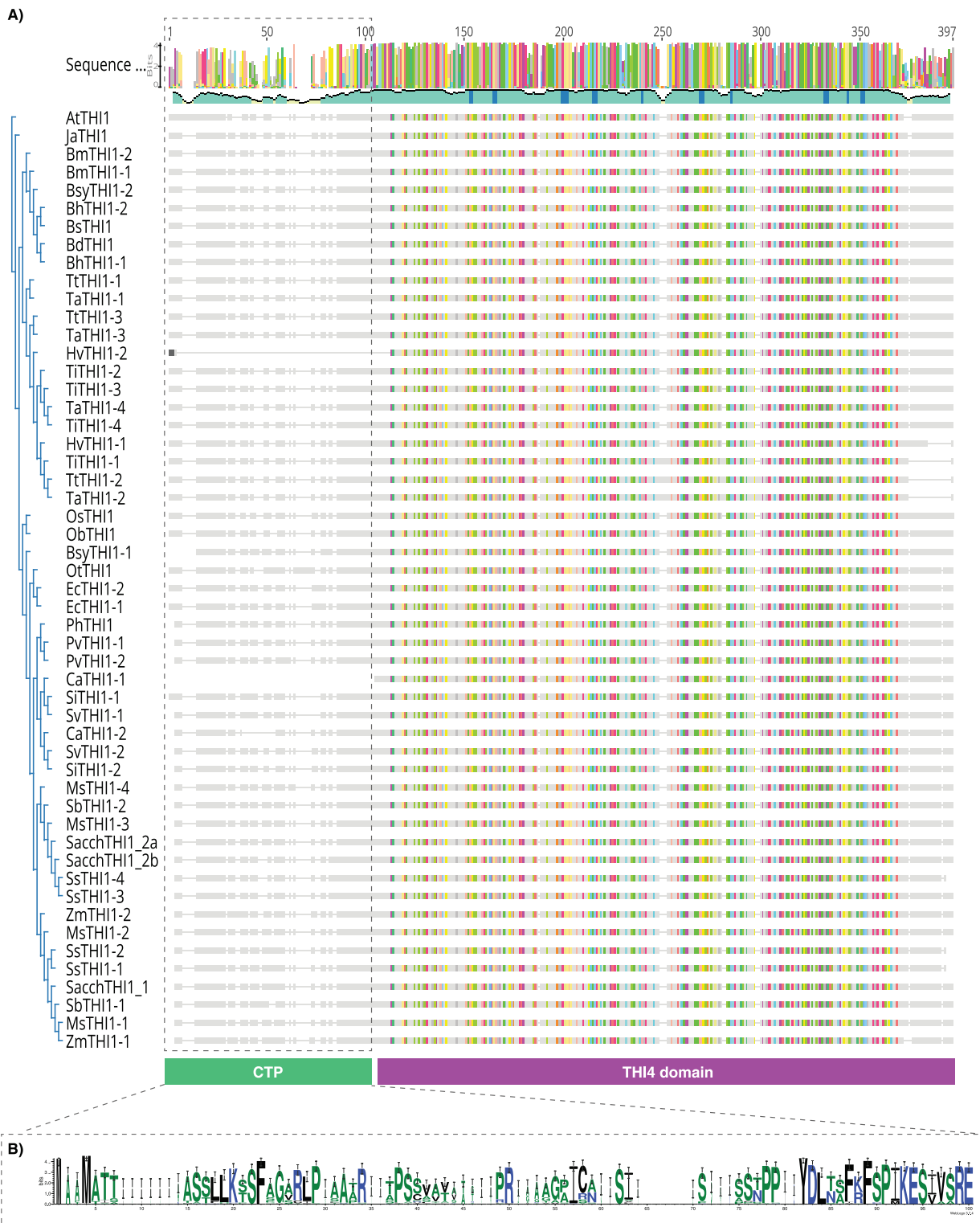
Figure 3 **Alignment of amino acid sequences of THI1.** (A) 51 THI1 sequences alignment across various species of Poaceae, using *J. ascendens* and *A. thaliana* as outgroups. The alignment was built with MAFFT v7. (B) WebLogo representation of multiple sequence alignment of N-terminal protein region indicating the relative frequency of amino acids at a given position (height). Abbreviation: CTP, Chloroplast Transient Peptide.

Full-size 🖼 DOI: 10.7717/peerj.14973/fig-3

two sugarcane sequence subgroups lies at the 5′ region, where the Chloroplast Transit Peptide (CTP) is predicted (*Machado et al., 1996*).

Due to their high similarity, only one version of each ScTHI1 was taken into consideration for further detailed analysis of their nucleotide genomic sequence as described, ScTHI1-1 sequence was retrieved from BAC 108_C04 and ScTHI1-2a, ScTHI1-2b from BAC 017_B18, BAC 094_O04 respectively (Figs. S2A and S2B). When their CDS sequences were compared, high nucleotide identity and similarity were found (>91%) (Fig. 3B). Outside of the N-terminal region, only minor differences were detected mostly resulting in synonymous amino acid substitutions (Fig. S2C). The ScTHI1-1 protein differs from both ScTHI1-2 subgroups in one residue of MPS (*Chabregas et al., 2003*), and other seven residues along the protein (Fig. S2C).

THI1 promoter region analysis

To understand the diversity of the *ScTHI1* promoter region in sugarcane, we performed an *in-silico* analysis of the 2 kb sequence preceding the start codon of the predicted genes (region D and E, Fig. 2). First, we identified 20 conserved motifs with sizes between 5 and 25 bp among the five species with the highest genome conservation (*i.e.*, *M. sinensis*, *S. bicolor*, sugarcane, *S. spontaneum*, and *Z. mays*). Our analyses revealed different degrees of conservation between the motifs (support statistical of *p*-value ≤ 0.05), in which a group of seven motifs formed a standard portion among the species (Fig. 4A). The promoter regions of these genes were highly conserved on the 300 bp next to the start codon, where at least seven motifs were common among 15 analyzed sequences. Therefore, this region was considered as the core promoter (Fig. 4B). The promoter regions of the three variants of the sugarcane var. R570, *ScTHI1-1*, *ScTHI1-2a* and *ScTHI1-2b*, shared most *cis*-regulatory elements (CREs). However, ScTHI1-1 presented a unique distribution and organization, while *ScTHI1-2a* and *ScTHI1-2b* shared as many CREs as distribution, supporting the presence of ScTHI1 in sugarcane along two genomic regions (Fig. 4A).

Next, transcription factors (TFs) with the potential to bind to CREs were identified in all 20 motifs (*p*-value ≤ 0.01), and GO terms association allowed us to understand the biological processes in which they are involved (see Fig. 4C). The GO terms of the core promoter (highlighted in blue in Fig. 4A) revealed enrichment on biological processes related to: 'Biological regulation' (GO:0065007), 'Developmental process' (GO:0032502), 'Post-Embryogenic development' (GO:0009791), 'Response to a stimulus' (GO:0050896), 'Response to a light stimulus' (GO:0009416), 'Rhythmic process' (GO:0048511), 'Circadian rhythm' (GO:0007623), and 'Regulation of circadian rhythm' (GO:0042752). *Fitzpatrick & Noordally (2021)* reported the relationship between the genes of the thiamine pathway and the influence of light/dark transitions, suggesting a daily transcription regulation of genes involved in the synthesis of vitamin B1.

When considering only the genes from sugarcane, TFs with potential binding in the CREs of the *ScTHI1-1* and *ScTHI1-2* promoters (highlighted in green in Fig. 4A), show functional redundancy because both GO terms are related to 'Response to stimulus' (GO: 0050896), 'Developmental process' (GO:0032502), and 'Biological regulation' (GO: 0065007). These results reinforce the putative biological identity of THI1 regulation in the

Figure 4 Promoter analysis and motif enrichment of *THI1* genes. (A) Motif identification in promoter regions (2,000 bp upstream of ATG) at *THI1* genes. The following parameters were used: site distribution = zero or one occurrence per sequence (zoops); minimum width = 5; minimum width = 25 bp; number of motifs = 20. The blue boxes showed conserved motifs among all grasses and the green boxes showed motifs present in *ScTHI1* paralogues. (B) Sequence logo of seven motifs (highly conserved) identified in the core promoter region of *THI1* of grasses. (C) GO enrichment analysis of motifs localized in the core promoter region of *THI1* of grasses (blue boxes in A). (D) GO enrichment analysis of motifs localized in the promoter region of *ScTHI1* (green boxes in A). Full-size ☒ DOI: 10.7717/peerj.14973/fig-4

thiamine pathway and its role in coordinating plant development processes. Additionally, the enrichment of terms related to plant-microorganism interaction was observed, such as 'Biological process involved in interspecies interaction between organisms' (GO:0044419), 'Immune system process' (GO:0002376), and 'Immune effector process' (GO:0002252), in as well as GO terms associated with 'Signaling' (GO:0023052) and 'Cell communication' (GO:0007154). Although these *in silico* observations still need to be confirmed experimentally, studies in *A. thaliana*, rice, and other crops (tobacco, cucumber, and tomato) have already demonstrated the participation of thiamine during the process of infection by pathogenic microorganisms (*Ahn, Kim & Lee, 2005*) and response to oxidative stress (*Tunc-Ozdemir et al., 2009*), Moreover, improvement in infection resistance was described for mutants of *A. thaliana* and *O. sativa* (*Dong, Stockwell & Goyer, 2015*; *Dong et al., 2016*).

THI1 sequence diversity in modern sugarcane cultivars

In order to address the haplotype sequence diversity of *THI1* in the *Saccharum* complex, we designed a primer pair in a conserved region between amino acids 49 to 161 (Fig. S2C). This conserved region was amplified from the closely related species *Miscanthus sp, S. officinarum*, *S. spontaneum* and 10 modern sugarcane hybrids cultivars. Table S4 describe how many sequences of the copies of *ScTHI1* were amplified for each of the genotypes/varieties.

Amplicons varying between 321–330 bp each were sub-cloned and 197 were sequenced, assembled and aligned against the 19 BAC sequences. A network analysis was performed (Fig. 5), where two main clusters emerge that are not species- or cultivar-specific, but *ScTHI1* copy-specific (Table S4) (*Vieira, 2018*).

Developmental and tissue-specific expression profiles of sugarcane *ScTHI1* genes

To gain insights into how *ScTHI1* genes are expressed in sugarcane we investigated changes in their expression patterns in different tissues (meristem, root, culm, and leaf) along the development of *Saccharum sp.* var R570. Our analysis revealed that all identified variants were expressed in all tissues and ages (Fig. 6). Furthermore, differential expression was detected when comparing different tissue (ANOVA p-value ≤ 0.01) and age, the latter more significantly evidenced in the meristematic tissue for all *ScTHI1* variants (Fig. 6). In culm, a change in expression level was seen in the transition from 3 to 6 months, corroborating previous findings (*Partida et al., 2021*).

Functional complementation assay

A *Saccharomyces cerevisiae* mutant strain, KBY5 (THI4::URA3), has a truncated *THI4* gene, which impairs growth on a minimal medium without thiamine supplementation. *Machado et al. (1996)* reported that *A. thaliana* homolog (*AtTHI1*) can restore the thiamine auxotrophy of this strain. An analogous complementation assay using *ScTHI1-1*, *ScTHI1-2a*, and *ScTHI1-2b* was performed. Two versions of each of the three CDSs were synthesized and transferred to the yeast expression vector pG-1 (*Vieira, 2018*). One version
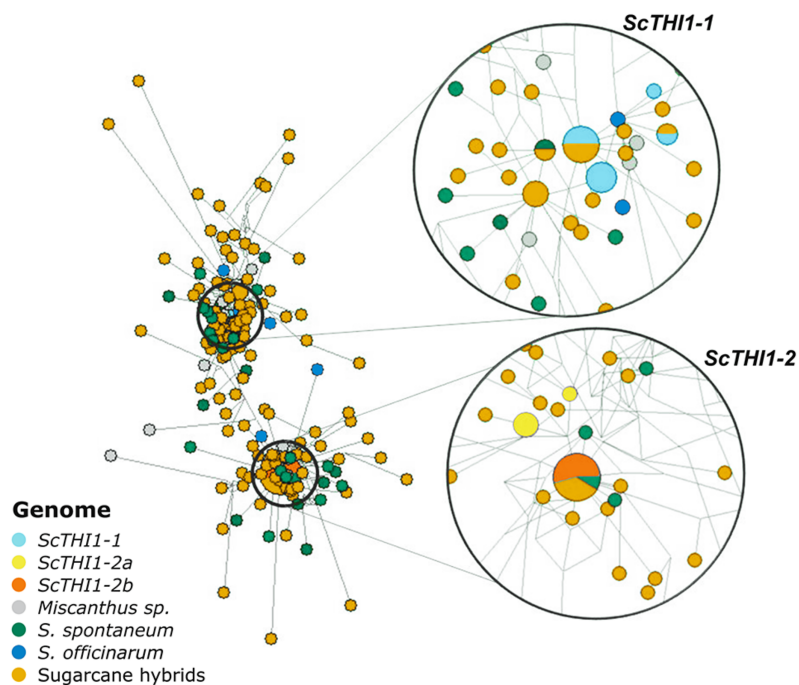
Figure 5 Network analysis of *THI1* gene in *Saccharum* complex. The network was built using the NETWORK 4.6.1.3 software (*Bandelt, Forster & Röhl, 1999*) with default parameters (Median-joining method). An alignment of a region of 539 bp of 210 sequences of varieties of sugarcane was used to construct the network. The right part of the figure is a close-up of the entire network shown on the left. The size of the circles is proportional to the number of sequences in the haplotype; the distance between clusters is proportional to the number of substitutions observed between sequences.

Full-size 🖼 DOI: 10.7717/peerj.14973/fig-5

represents the full *ScTHI1-1*, *ScTHI1-2a*, and *ScTHI1-2b* CDS and the second version is the CDS lacking the N-terminal chloroplast transit peptide sequences, named *ScTHI1-1/DelN*, *ScTHI1-2a/DelN*, and *ScTHI1-2b*/DelN. Two other constructs, A184V and DelN in the same expression vector were used as controls. A184V is the cDNA of the *THI1* gene from the *A. thaliana tz-201* mutant and was used as positive control. In fact, it served as a control for partial complementation (*Papini-Terzi et al., 2003*). The negative control DelN contains a cDNA of the wild-type *AtTHI1* gene without the N-terminal chloroplast transit peptide (*Papini-Terzi et al., 2003*). Growth was evaluated after 4 and 28 days of incubation at 30 °C.

KBY5 strain transformed independently with the three sugarcane constructs and plated on YNB medium supplemented with thiamine (thia) and tryptophan (W) could efficiently grow (Fig. 7A). The inability of the wild-type strain W303a and KBY5 to grow on selective [−W/+ thia] YNB medium shows that the transformation with the sugarcane constructs was efficient and complemented W auxotrophy in the recipient strain (Fig. 7B). Only wild type strain W303a and auxotrophic complemented strains by the recipient plasmid could grow on plates supplemented with W only after four (Fig. 7C) and 28 (Fig. 7E) days of incubation. None of the full-length *ScTHI1* CDSs complemented KBY5 thiamine auxotrophy, but both DelN versions of *ScTHI1-1* and *ScTHI1-2b* did. However, this
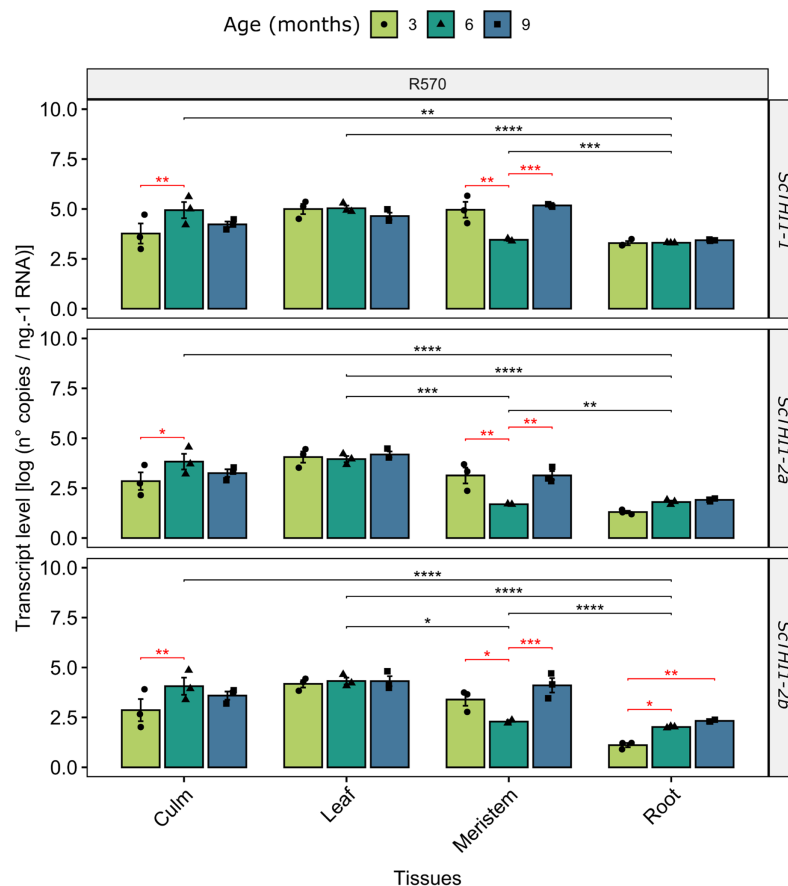
Figure 6 Expression analysis of *ScTHI1* genes in different tissues and ages in the sugarcane cultivar R570. Data are the means of three biological replicates; error bars indicate SD. Asterisks indicate significant differences among compared groups using the ANOVA and *p*-adjust (Bonferroni method) <0.05 (*), 0.01(**), 0.001 (***), and 0.0001(****). Black and red lines represent statistically significant differences between tissues and ages. Full-size 🖼 DOI: 10.7717/peerj.14973/fig-6

complementation was less efficient than the *A. thaliana* A184V allele (*Papini-Terzi et al., 2003*), with yeasts forming smaller colonies and taking more time to develop.

DISCUSSION

Studying precursor genes for coenzymes has been of increasing interest in biology, especially in plant crops (*Pourcel, Moulin & Fitzpatrick, 2013*; *Strobbe et al., 2021*). For many years, plant science has focused on genes directly responsible for the increase in biomass. More recently, several studies have looked at coenzymes and their impact on metabolic pathways, especially on the carbohydrate (see review of *Fitzpatrick & Noordally (2021)*). Here, we show an integrative view for understanding the biology of the *THI1* genes Poaceae, especially in sugarcane. We aimed at elucidating this gene's structure, organization, and distribution in the genomes and clarifying aspects of the regulation of *THI1* expression, targeting, and gene diversity.

Sugarcane is an important tropical crop, cultivated for the production of sucrose and bioethanol. Several efforts for sequencing its complex genome structure have been made
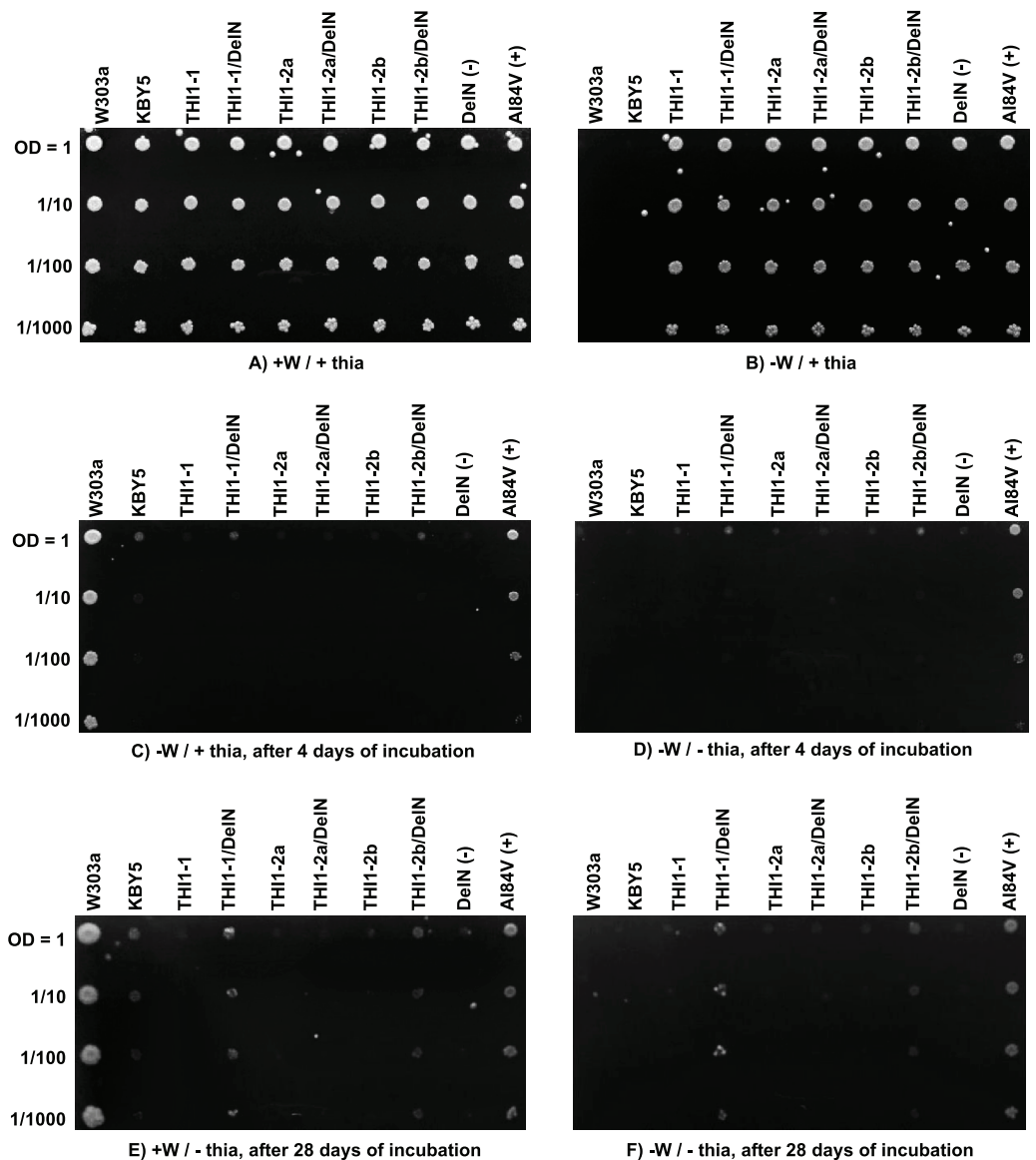
Figure 7 Functional yeast complementation assay. *S. cerevisiae* strain KBY5 was transformed with the three versions of the *ScTHI1* gene found in the sugarcane genome, with the positive control *A184V* and the negative control and DelN, and plated in YNB media with or without tryptophan and thiamine. W303a strain was used as positive control. W = tryptophan and "thia" = thiamine. Each column represents one transformant. Lines represent dilution series. (A and B) are the experimental controls to check if all strains can grow (A) and if all transformed ones can grow without tryptophan (B). (C and D) show growth after 4 days of incubation, while (E and F) after 28 days of incubation, at 30 °C.

Full-size 🖼 DOI: 10.7717/peerj.14973/fig-7

and the most complete managed to cover 373,869 genes (99.1% of the sugarcane genome) (*Souza et al., 2019*). The unavailability of the full genome poses a challenge in investigating the genomic context of genes in this species. However, our study provides the first complete overview of *THI1* in Poaceae, including protein characterization, phylogeny, gene structure, chromosome location, synteny, and gene expression patterns.

Features of *THI1* genes have been conserved during Poaceae speciation

Our genome-wide investigation of *THI1* genes include all Poaceae species sequenced to date, which have at least one *THI1* homologue gene. The aligned sequences displayed that with the exception of the variable N-terminal region, most of the nucleotide changes have resulted in synonymous substitution. The variability found at the N-terminal region could be explained by its targeting function that depends on the classes of amino acids (hydrophobic and positively charged) present rather than on specific amino acid sequences (*von Heijne & Gavel, 1988*; *Käll, Krogh & Sonnhammer, 2004*). As a result, this region could accommodate non-synonymous substitutions without loss of function.

In addition to the subgroups formed among ScTHI1, the homologous of the species *S. italica*, *S. viridis*, *C. americanus*, *P. virgatum*, *E. coracana* and *P. hallii* form a group (Paniceae) sister to the Andropogoneae tribe. Interestingly, members of Bambusoideae, Oryzoideae, and Pooideae subfamilies (B.O.P) grouped together in the clade comprised by *B. distachyon*, *B. hybridum*, *B. mexicanum*, *B. stacei*, *B. sylvaticum*, *T. intermedium*, *T. aestivum, and T. turgidum*. Despite the species *O. sativa* and *O. brachyantha* being part of this same clade (B.O.P), it is possible to see that the homologs of the genus *Oryza* are less related. The topology of the evolutionary tree of the predicted amino acid sequences from THI1 of Poaceae was similar to the one described by *Soreng et al. (2015*, *2017)*. Furthermore, homologs of the same species clustered together, indicating that these duplication events happened independently after speciation.

Despite the variation found in the N-terminal region, THI4 domains are reported to be highly conserved (*Hwang et al., 2014*). The exceptions were THI1 homologs non-Cys present in *T. aestivum, T. intermedium, T. turgidum*, and *H. vulgare*. The expression levels of non-Cys variants were found to be lower than those with this cysteine residue conserved (*Joshi et al., 2020*). Similar to Archaea species, such as *Methanococcus igneus* (*Zhang et al., 2016*) and *Methanocaldococcus jannashi* (*Eser et al., 2016*), the obtention of thiazole ring is accomplished by the use of the nicotinamide adenine dinucleotide, glycine and free sulfide. This suggests that a THI1 isoform non-Cys is only restricted to those cereal species, which possibly need a Cys containing THI1 isoform capable of donating a sulfur molecule to form the thiazole ring.

THI1 has a central role in thiamine biosynthesis that in turn is an essential cofactor for several metabolic pathways, such as amino acids metabolism (*Duggleby, 2006*; *Duggleby, McCourt & Guddat, 2008*) and carbohydrate (*Belanger et al., 1995*). According to the amino acid sequences analyzed, the residue (Cys205) required for the Fe^{2+}-binding (*Zhang et al., 2016*; *Eser et al., 2016*; *Joshi et al., 2020*) is fully conserved in THI1 from Poaceae. This Cys residue is known to be the sulfur donor in yeast (*Chatterjee et al., 2011*), plants (*Godoi et al., 2006*), and the Archaea species *H. volcanii* (*Hwang et al., 2014*). Our results indicate that THI1 sulfur donation function is present at least in one gene copy and is suggestive of its role in the thiazole ring formation in all Poaceae isoforms.

The molecular characterization of the *THI1* gene revealed that gene duplication has not only occurred in C4 plants and that *THI1* genes are also positioned in different genomic

regions in Poaceae, probably due to the number of genome duplications in those groups (*de Setta et al., 2012*; *Svačina et al., 2020*; *Lee et al., 2020*). Recent studies based on comparative genomics support the occurrence of whole-genome duplication in angiosperms (*Jiao et al., 2011*) and early polyploidy in monocots (*Tang et al., 2010*). Furthermore, our study shows that C3 species generally have a single copy of *THI1*. However, this is not the case for wheat and barley, representatives of the Triticeae group that are the most prominent example of duplication under the influence of the domestication event (*Qiao et al., 2019*).

The comparative analysis identified five non-syntenic genomic regions among the Poaceae genomes. Evidence of the *THI1* gene duplications and translocations as well as duplications of genomic regions containing this gene are presented. The common ancestor of the Andropogoneae tribe (sugarcane, sorghum, and maize) has two loci carrying *THI1* copies (*THI1-1*—present in region D and *THI1-2*—present in region E Fig. 2), which is observed in high collinearity between sugarcane and sorghum chromosomes (*Ming et al., 1998*; *Vieira, 2018*). *S. italica* and *S. viridis* have two *THI1* copies, one in a region syntenic to those of *C. americanus, O. sativa, O. brachyantha,* and *E. coracana* (region B). Another copy is located in a different area shared among *P. hallii, P. virgatum* and other copy of the *C. americanus* (region C). *Z. mays* duplicated the entire region E. Finally, the *A. thaliana THI1* region is non-syntenic to grasses genomic regions analyzed here (region A). These results provide an overview of Poaceae *THI1*, including their gene numbers, evolutionary relationship, and structural conservation locus.

ScTHI1 genes share conserved molecular features

We identified 19 BACs of sugarcane var. R570 containing *THI1*-like genes. Our genomic characterization revealed at least two groups of *ScTHI1* genes, nine alleles of *ScTHI1-1* and ten alleles of *ScTHI1-2* (*Vieira, 2018*). Phylogenetic (Fig. 1B) and network analyses (Fig. 5) of its sequences showed that, despite the similarity among *ScTHI1-2* genes, a diversification of *ScTHI1-2* has occurred. Two subgroups were identified, including four alleles of the *ScTHI1-2a* group and six alleles of the *ScTHI1-2b* group.

As shown in the synteny analysis, each paralogue of *ScTHI1* is present in a different genomic environment. In addition, the core promoter located 600 bp upstream the start codon (ATG) is highly conserved, among the sugarcane ScTHI1 paralogs and Poaceae. Further analysis of the promoter regions (2 kb upstream from ATG) revealed that *ScTHI1.1* and *ScTHI1.2* have distinct sets of of CRE and TF binding sites conserved across all *Saccharum* species, which contributed to the prediction of their potential function. Together, these data support the idea of a gene duplication occurring in a common ancestor of the Andropogoneae tribe, preserving the gene and its core promoter region along the evolution.

The diversification of two *ScTHI1-2* subgroups is supported by the Network analysis. *ScTHI1-2b* sequences fall into one haplotype along with one *S. spontaneum* and four modern cultivars sequences, whereas *ScTHI1-2a* sequences fall into two distinct haplotypes, one comprised of several BACs (017_B18, 030_H05 and 251_N23) and a second composed by 092_F09.

Our data revealed that the expression pattern between the variants is very similar. Looking only at the leaf tissue, where the three variants showed constantly higher expression levels over the development, a 10-fold difference in expression of *ScTHI1-1* compared to *ScTHI1-2a* and 5-fold compared to *ScTHI1-2b* is seen. Comparing the different tissues, the leaf presents the largest number of transcripts while the root has the lowest number, showing the relationship of *ScTHI1* with photosynthetic tissues, as previously reported for *Alnus glutinosa* (*Ribeiro et al., 1996*), Arabidopsis (*Papini-Terzi et al., 2003*; *Ribeiro et al., 2005*), and the crops species cassava (*Mangel et al., 2017*), wheat (*Joshi et al., 2020*), and barley (*Joshi et al., 2020*). In addition, our results of CREs in the promoter region of *ScTHI1* reveals potential binding sites to TFs related to development process corroborating the expression data.

Our complementation assays revealed that the three full-length copies of *ScTHI1* could not complement the KBY5 strain. *Papini-Terzi et al. (2003)* also presented that this strain grows poorly after 4 days of cultivation when bearing the A184V construction. The ScTHI1-1 DelN and ScTHI1-2b DelN transformants were partially complemented but took longer (28 days) to develop, suggesting that the chloroplast transit peptide at the N-terminus of THI1 from sugarcane somehow interferes with the complementation efficiency in the KBY5 yeast strain. *ZmTHI1-1* and *ZmTHI1-2* can restore thiamine prototrophy in yeast (*Belanger et al., 1995*).

CONCLUSIONS

Taken together, the study of the sugarcane *THI1* supports the existence of multiple independent rounds of gene duplication events involving *THI1* orthologs. Each tribe presents its unique genomic *THI1* environment except maize which shares the same environment for the two gene copies. Expression of sugarcane *THI1* is redundant across tissues and developmental stages where the leaf presents the higher expression level and the root the least. This is consistent with the similarity observed at the core promoter of the paralog genes; however, subtle intensity changes demand dissecting the expression differences in more detail. Sugarcane gene copies are redundant at the transcription level, and two of the three copies are functionally redundant. Further studies are needed to explore the contribution of the levels of the thiazole ring in C4 photosynthetic plant tissues or potentially the relevance of the THI1 protein activity.

Accession numbers

Sequence data from this article can be found in NCBI. Accession numbers: KF184731.1; KF184959.1; KF184877.1; KF184851.1; KF184849.1; KF184930.1; KF184747.1; KF184858.1; KF184724.1; PRJEB26915

ACKNOWLEDGEMENTS

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Henrique Moura Dias conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Andreia Prata Vieira conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Erika Maria de Jesus analyzed the data, authored or reviewed drafts of the article, initial genomic studies, and approved the final draft.
- Nathalia de Setta analyzed the data, authored or reviewed drafts of the article, initial evolutionary analysis, and approved the final draft.
- Gesiele Barros analyzed the data, authored or reviewed drafts of the article, initial Bioinformatic pipeline, and approved the final draft.
- Marie-Anne Van Sluys conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available in the Supplemental Tables and at GenBank: KF184731.1, KF184959.1, KF184877.1, KF184851.1, KF184849.1, KF184930.1, KF184747.1, KF184858.1, KF184724.1, PRJEB26915.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.14973#supplemental-information.

REFERENCES

Ahn I-P, Kim S, Lee Y-H. 2005. Vitamin B1 functions as an activator of plant disease resistance. *Plant Physiology* **138(3)**:1505–1515 DOI 10.1104/pp.104.058693.

Anand L, Rodriguez Lopez CM. 2022. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinformatics* **23(1)**:1–9 DOI 10.1186/s12859-021-04556-z.

Armenteros A, Juan J, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, Nielsen H. 2019a. Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* **2(5)**:e201900429 DOI 10.26508/lsa.201900429.

Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Von Heijne G, Nielsen H. 2019b. SignalP 5. 0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37(4)**:410–423 DOI 10.1038/s41587-019-0036-z.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37(Web Server)**:202–208 DOI 10.1093/nar/gkp335.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16(1)**:37–48 DOI 10.1093/oxfordjournals.molbev.a026036.

Belanger FC, Leustek T, Chu B, Kriz AL. 1995. Evidence for the thiamine biosynthetic pathway in higher-plant plastids and its developmental regulation. *Plant Molecular Biology* **29(4)**:809–821 DOI 10.1007/BF00041170.

Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research* **49(D1)**:D344–D354 DOI 10.1093/nar/gkaa977.

Chabregas SM, Luche DD, Farias LP, Ribeiro AF, Van Sluys MA, Menck CFM, Silva-Filho MC. 2001. Dual targeting properties of the N-terminal signal sequence of Arabidopsis thaliana THI1 protein to mitochondria and chloroplasts. *Plant Molecular Biology* **46(6)**:639–650 DOI 10.1023/A:1011628510711.

Chabregas SM, Luche DD, Van Sluys M-A, Menck CFM, Silva-Filho MC. 2003. Differential usage of two in-frame translational start codons regulates subcellular localization of Arabidopsis thaliana THI1. *Journal of Cell Science* **116(2)**:285–291 DOI 10.1242/jcs.00228.

Chatterjee A, Abeydeera ND, Bale S, Pai PJ, Dorrestein PC, Russell DH, Ealick SE, Begley TP. 2011. Saccharomyces cerevisiae THI4p is a suicide thiamine thiazole synthase. *Nature* **478(7370)**:542–546 DOI 10.1038/nature10503.

de Setta N, Metcalfe CJ, Cruz GMQ, Ochoa EA, Van Sluys M-A. 2012. Noise or symphony: comparative evolutionary analysis of sugarcane transposable elements with other grasses. In: *Plant Transposable Elements*. Berlin, Heidelberg: Springer, 169–192.

de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GMQ, Del Bem LE, Vicentini R, Nogueira FTS, Campos RA, Nunes SL, Turrini PCG, Vieira AP, Ochoa Cruz EA, Corrêa TCS, Hotta CT, de Mello Varani A, Vautrin S, da Trindade AS, de Mendonça Vilela M, Lembke CG, Sato PM, de Andrade RF, Nishiyama MY, Cardoso-Silva CB, Scortecci KC, Garcia AAF, Carneiro MS, Kim C, Paterson AH, Bergès H, D'Hont A, de Souza AP, Souza GM, Vincentz M, Kitajima JP, Van Sluys MA. 2014. Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* **15(1)**:1–17 DOI 10.1186/1471-2164-15-540.

Dietz KJ, Zörb C, Geilfus CM. 2021. Drought and crop yield. *Plant Biology* **23(6)**:881–893 DOI 10.1111/plb.13304.

Dong W, Stockwell VO, Goyer A. 2015. Enhancement of thiamin content in Arabidopsis thaliana by metabolic engineering. *Plant and Cell Physiology* **56(12)**:2285–2296 DOI 10.1093/pcp/pcv148.

Dong W, Thomas N, Ronald PC, Goyer A. 2016. Overexpression of thiamin bipsynthesis genes in rice increases leaf and unpolished grain thiamin content but not resistance to Xanthomonas oryzae pv. oryzae. *Froniters in Plant Science* **7**:1–11 DOI 10.3389/fpls.2016.00616.

Duggleby RG. 2006. Domain relationships in thiamine diphosphate-dependent enzymes. *Accounts of Chemical Research* **39(8)**:550–557 DOI 10.1021/ar068022z.

Duggleby RG, McCourt JA, Guddat LW. 2008. Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiology and Biochemistry* **46(3)**:309–324 DOI 10.1016/j.plaphy.2007.12.004.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* **47(D1)**:D427–D432 DOI 10.1093/nar/gky995.

Eser BE, Zhang X, Chanani PK, Begley TP, Ealick SE. 2016. From suicide enzyme to catalyst: the iron-dependent sulfide transfer in Methanococcus jannaschii thiamin thiazole biosynthesis. *Journal of the American Chemical Society* **138(11)**:3639–3642 DOI 10.1021/jacs.6b00445.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8(3)**:186–194 DOI 10.1101/gr.8.3.186.

Fitzpatrick TB, Noordally Z. 2021. Of clocks and coenzymes in plants: intimately connected cycles guiding central metabolism? *New Phytologist* **230(2)**:416–432 DOI 10.1111/nph.17127.

Forlani G, Mantelli M, Nielsen E. 1999. Biochemical evidence for multiple acetoin-forming enzymes in cultured plant cells1Part 1 in the series "Acetoin Synthesis in Higher Plants".1. *Phytochemistry* **50(2)**:255–262 DOI 10.1016/S0031-9422(98)00550-0.

Formann S, Hahn A, Janke L, Stinner W, Sträuber H, Logroño W, Nikolausz M. 2020. Beyond sugar and ethanol production: value generation opportunities through sugarcane residues. *Frontiers in Energy Research* **8**:1–21 DOI 10.3389/fenrg.2020.579577.

Frank RAW, Leeper FJ, Luisi BF. 2007. Structure, mechanism and catalytic duality of thiamine-dependent enzymes. *Cellular and Molecular Life Sciences* **64(7–8)**:892–905 DOI 10.1007/s00018-007-6423-5.

Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G, Charron C, Hervouet C, Costet L, Yahiaoui N, Healey A, Sims D, Cherukuri Y, Sreedasyam A, Kilian A, Chan A, Van Sluys MA, Swaminathan K, Town C, Bergès H, Simmons B, Glaszmann JC, Van Der Vossen E, Henry R, Schmutz J, D'Hont A. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications* **9(1)**:1–10 DOI 10.1038/s41467-018-05051-5.

Godoi PHC, Galhardo RS, Luche DD, Van Sluys MA, Menck CFM, Oliva G. 2006. Structure of the thiazole biosynthetic enzyme THI1 from Arabidopsis thaliana. *Journal of Biological Chemistry* **281(41)**:30957–30966 DOI 10.1074/jbc.M604469200.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40(D1)**:1178–1186 DOI 10.1093/nar/gkr944.

Gordon D. 2003. Viewing and editing assembled sequences using Consed. *Current Protocols in Bioinformatics* **Chapter 11**:1–43 DOI 10.1002/0471250953.bi1102s02.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Research* **8(3)**:195–202 DOI 10.1101/gr.8.3.195.

Goyer A. 2010. Thiamine in plants: aspects of its metabolism and functions. *Phytochemistry* **71(14–15)**:1615–1624 DOI 10.1016/j.phytochem.2010.06.022.

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. 2007. Quantifying similarity between motifs. *Genome Biology* **8(1)**:1–9 DOI 10.1186/gb-2007-8-1-r1.

Hwang S, Cordova B, Chavarria N, Elbanna D, McHugh S, Rojas J, Pfeiffer F, Maupin-Furlow JA. 2014. Conserved active site cysteine residue of archaeal THI4 homolog is essential for thiamine biosynthesis in Haloferax volcanii. *BMC Microbiology* **14(1)**:1–13 DOI 10.1186/s12866-014-0260-0.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473(7345)**:97–100 DOI 10.1038/nature09916.

Joshi J, Beaudoin GAW, Patterson JA, García-García JD, Belisle CE, Chang LY, Li L, Duncan O, Millar AH, Hanson AD. 2020. Bioinformatic and experimental evidence for suicidal and catalytic plant THI4s. *Biochemical Journal* **477(11)**:2055–2069 DOI 10.1042/BCJ20200297.

Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annual Review of Biochemistry* **78(1)**:569–603 DOI 10.1146/annurev.biochem.78.072407.102340.

Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338(5)**:1027–1036 DOI 10.1016/j.jmb.2004.03.016.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30(4)**:772–780 DOI 10.1093/molbev/mst010.

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, Van Der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research* **46(D1)**:D260–D266 DOI 10.1093/nar/gkx1126.

Khan MT, Khan IA. 2019. *Sugarcane biofuels: status, potential, and prospects of the sweet crop to fuel the world.* Cham, Switzerland: Springer Nature.

Kong D, Zhu Y, Wu H, Cheng X, Liang H, Ling HQ. 2008. AtTHIC, a gene involved in thiamine biosynthesis in Arabidopsis thaliana. *Cell Research* **18(5)**:566–576 DOI 10.1038/cr.2008.35.

LeClere S, Rampey RA, Bartel B. 2004. IAR4, a gene required for auxin conjugate sensitivity in arabidopsis, encodes a pyruvate dehydrogenase E1α homolog. *Plant Physiology* **135(2)**:989–999 DOI 10.1104/pp.104.040519.

Lee S, Choi S, Jeon D, Kang Y, Kim C. 2020. Evolutionary impact of whole genome duplication in Poaceae family. *Journal of Crop Science and Biotechnology* **23(5)**:413–425 DOI 10.1007/s12892-020-00049-2.

Li CL, Wang M, Wu XM, Chen DH, Lv HJ, Shen JL, Qiao Z, Zhang W. 2016. THI1, a thiamine thiazole synthase, interacts with Ca2+-dependent protein kinase CPK33 and modulates the s-type anion channels and stomatal closure in arabidopsis. *Plant Physiology* **170(2)**:1090–1104 DOI 10.1104/pp.15.01649.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25(11)**:1451–1452 DOI 10.1093/bioinformatics/btp187.

Machado CR, Costa De Oliveira RL, Boiteux S, Praekelt UM, Meacock PA, Menck CFM. 1996. Thi1, a thiamine biosynthetic gene in Arabidopsis thaliana, complements bacterial defects in DNA repair. *Plant Molecular Biology* **31(3)**:585–593 DOI 10.1007/BF00042231.

Machado CR, Praekelt UM, De Oliveira RC, Barbosa ACC, Byrne KL, Meacock PA, Menck CFM. 1997. Dual role for the yeast THI4 gene in thiamine biosynthesis and DNA damage tolerance. *Journal of Molecular Biology* **273(1)**:114–121 DOI 10.1006/jmbi.1997.1302.

Mangel N, Fudge JB, Fitzpatrick TB, Gruissem W, Vanderschuren H. 2017. Vitamin B1 diversity and characterization of biosynthesis genes in cassava. *Journal of Experimental Botany* **68(13)**:3351–3363 DOI 10.1093/jxb/erx196.

Marsden SR, McMillan DGG, Hanefeld U. 2020. Assessing the thiamine diphosphate dependent pyruvate dehydrogenase E1 subunit for carboligation reactions with aliphatic ketoacids. *International Journal of Molecular Sciences* **21(22)**:1–15 DOI 10.3390/ijms21228641.

Medina-Silva R, Barros MP, Galhardo RS, Netto LES, Colepicolo P, Menck CFM. 2006. Heat stress promotes mitochondrial instability and oxidative responses in yeast deficient in thiazole biosynthesis. *Research in Microbiology* **157(3)**:275–281 DOI 10.1016/j.resmic.2005.07.004.

Ming R, Liu SC, Lin YR, Da Silva J, Wilson W, Braga D, Van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH. 1998. Detailed alignment of Saccharum and Sorghum chromosomes: comparative organization of closely related diploid and polyploid genomes. *Genetics* **150(4)**:1663–1682 DOI 10.1093/genetics/150.4.1663.

Papini-Terzi FS, Galhardo RDS, Farias LP, Menck CFM, Van Sluys MA. 2003. Point mutation is responsible for Arabidopsis tz-201 mutant phenotype affecting thiamin biosynthesis. *Plant and Cell Physiology* **44(8)**:856–860 DOI 10.1093/pcp/pcg104.

Partida VGS, Dias HM, Corcino DSM, Van Sluys MA. 2021. Sucrose-phosphate phosphatase from sugarcane reveals an ancestral tandem duplication. *BMC Plant Biology* **21(1)**:1–13 DOI 10.1186/s12870-020-02795-5.

Pourcel L, Moulin M, Fitzpatrick TB. 2013. Examining strategies to facilitate vitamin B1 biofortification of plants by genetic engineering. *Frontiers in Plant Science* **4**:1–8 DOI 10.3389/fpls.2013.00160.

Powell RV, Willett CR, Goertzen LR, Rashotte AM. 2019. Lineage specific conservation of cis-regulatory elements in Cytokinin Response Factors. *Scientific Reports* **9(1)**:1–11 DOI 10.1038/s41598-019-49741-6.

Praekelt UTAM, Byrne KL, Meacock PA. 1994. Gene of Saccharomyces cerevisiae. *Yeast* **10**:481–490 DOI 10.1002/(ISSN)1097-0061.

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology* **20**:1–23.

Rapala-Kozik M. 2011. *Vitamin B1 (thiamine): a cofactor for enzymes involved in the main metabolic pathways and an environmental stress protectant.* Amsterdam: Elsevier Ltd.

Rapala-Kozik M, Wolak N, Kujda M, Banas AK. 2012. The upregulation of thiamine (vitamin B1) biosynthesis in Arabidopsis thaliana seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response. *BMC Plant Biology* **12(1)**:1–14 DOI 10.1186/1471-2229-12-2.

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. G: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47(W1)**:W191–W198 DOI 10.1093/nar/gkz369.

Ribeiro DT, Farias LP, De Almeida JD, Kashiwabara PM, Ribeiro AFC, Silva-Filho MC, Menck CFM, Van Sluys MA. 2005. Functional characterization of the thi1 promoter region from Arabidopsis thaliana. *Journal of Experimental Botany* **56(417)**:1797–1804 DOI 10.1093/jxb/eri168.

Ribeiro A, Praekelt U, Akkermans ADL, Meacock PA, Van Kammen A, Bisseling T, Pawlowski K. 1996. Identification of agthi1, whose product is involved in biosynthesis of the thiamine precursor thiazole, in actinorhizal nodules of Alnus glutinosa. *Plant Journal* **10(2)**:361–368 DOI 10.1046/j.1365-313X.1996.10020361.x.

Rocha AG, Mehlmer N, Stael S, Mair A, Parvin N, Chigri F, Teige M, Vothknecht UC. 2014. Phosphorylation of Arabidopsis transketolase at Ser428 provides a potential paradigm for the metabolic control of chloroplast carbon metabolism. *Biochemical Journal* **458(2)**:313–322 DOI 10.1042/BJ20130631.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61(3)**:539–542 DOI 10.1093/sysbio/sys029.

Schena M, Picard D, Yamamoto KR. 1991. Vectors for constitutive and inducible gene expression in yeast. *Methods in Enzymology* **194**:389–398 DOI 10.1016/0076-6879(91)94029-C.

Schiestl RH, Gietz RD. 1989. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Current Genetics* **16(5–6)**:339–346 DOI 10.1007/BF00340712.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7(1)**:1–6 DOI 10.1038/msb.2011.75.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Teisher JK, Clark LG, Barberá P, Gillespie LJ, Zuloaga FO. 2017. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: an update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution* **55(4)**:259–290 DOI 10.1111/jse.12262.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Davis JI, Morrone O. 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution* **53(2)**:117–137 DOI 10.1111/jse.12150.

Souza GM, Van Sluys MA, Lembke CG, Lee H, Margarido GRA, Hotta CT, Gaiarsa JW, Diniz AL, Oliveira MDM, Ferreira SDS, Nishiyama MY, Ten-Caten F, Ragagnin GT, Andrade PDM, De Souza RF, Nicastro GG, Pandya R, Kim C, Guo H, Durham AM, Carneiro MS, Zhang J, Zhang X, Zhang Q, Ming R, Schatz MC, Davidson B, Paterson AH, Heckerman D. 2019. 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(12)**:1–18 DOI 10.1093/gigascience/giz129.

Strobbe S, Verstraete J, Stove C, Van Der Straeten D. 2021. Metabolic engineering of rice endosperm towards higher vitamin B1 accumulation. *Plant Biotechnology Journal* **19(6)**:1253–1267 DOI 10.1111/pbi.13545.

Svačina R, Sourdille P, Kopecký D, Bartoš J. 2020. Chromosome pairing in polyploid grasses. *Frontiers in Plant Science* **11**:1–17 DOI 10.3389/fpls.2020.01056.

Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proceedings of the National Academy of Sciences of the United States of America* **107(1)**:472–477 DOI 10.1073/pnas.0908007107.

Tunc-Ozdemir M, Miller G, Song L, Kim J, Sodek A, Koussevitzky S, Misra AN, Mittler R, Shintani D. 2009. Thiamin confers enhanced tolerance to oxidative stress in Arabidopsis. *Plant Physiology* **151(1)**:421–432 DOI 10.1104/pp.109.140046.

Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van De Peer Y, Coppens F, Vandepoele K. 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research* **46(D1)**:D1190–D1196 DOI 10.1093/nar/gkx1002.

Vettore AL, Da Silva FR, Kemper EL, Arruda P. 2001. The libraries that made SUCEST. *Genetics and Molecular Biology* **24(1–4)**:1–7 DOI 10.1590/S1415-47572001000100002.

Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MIT, Henrique-Silva F, Giglioti ÉA, Lemos MVF, Coutinho LL, Nobrega MP, Carrer H, França SC, Bacci M, Goldman MHS, Gomes SL, Nunes LR, Camargo LEA, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon MLPN, Ferro JA, Silveira HCS, Marini DC, Lemos EGM, Monteiro-Vitorello CB, Tambor JHM, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MMA, de Rosa VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SMZ, Nobrega FG, Menck CFM, Braga MDV, Telles GP, Cara FAA, Pedrosa G, Meidanis J, Arruda P. 2003. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Research* **13(12)**:2725–2735 DOI 10.1101/gr.1532103.

Vieira AP. 2018. Sugarcane thi1 homologs: a molecular and functional studies. Universidade de São Paulo, São Paulo/SP, Brazil. *Available at* https://teses.usp.br/teses/disponiveis/11/11151/tde-14082018-094015/pt-br.php.

von Heijne G, Gavel Y. 1988. Topogenic signals in integral membrane proteins. *European Journal of Biochemistry* **174(4)**:671–678 DOI 10.1111/j.1432-1033.1988.tb14150.x.

Whittaker A, Botha FC. 1997. Carbon partitioning during sucrose accumulation in sugarcane internodal tissue. *Plant Physiology* **115(4)**:1651–1659 DOI 10.1104/pp.115.4.1651.

Zhang X, Eser BE, Chanani PK, Begley TP, Ealick SE. 2016. From suicide enzyme to catalyst: the iron-dependent sulfide transfer in Methanococcus jannaschii thiamin thiazole biosynthesis. *Journal of the American Chemical Society* **138(11)**:3639–3642 DOI 10.1021/jacs.6b00445.

Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, Wai CM, Zheng C, Shi Y, Chen S, Xu X, Yue J, Nelson DR, Huang L, Li Z, Xu H, Zhou D, Wang Y, Hu W, Lin J, Deng Y, Pandey N, Mancini M, Zerpa D, Nguyen JK, Wang L, Yu L, Xin Y, Ge L, Arro J, Han JO, Chakrabarty S, Pushko M, Zhang W, Ma Y, Ma P, Lv M, Chen F, Zheng G, Xu J, Yang Z, Deng F, Chen X, Liao Z, Zhang X, Lin Z, Lin H, Yan H, Kuang Z, Zhong W, Liang P, Wang G, Yuan Y, Shi J, Hou J, Lin J, Jin J, Cao P, Shen Q, Jiang Q, Zhou P, Ma Y, Zhang X, Xu R, Liu J, Zhou Y, Jia H, Ma Q, Qi R, Zhang Z, Fang J, Fang H, Song J, Wang M, Dong G, Wang G, Chen Z, Ma T, Liu H, Dhungana SR, Huss SE, Yang X, Sharma A, Trujillo JH, Martinez MC, Hudson M, Riascos JJ, Schuler M, Chen LQ, Braun DM, Li L, Yu Q, Wang J, Wang K, Schatz MC, Heckerman D, Van Sluys MA, Souza GM, Moore PH, Sankoff D, VanBuren R, Paterson AH, Nagai C, Ming R. 2018. Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. *Nature Genetics* **50(11)**:1565–1573 DOI 10.1038/s41588-018-0237-2.

Moura Dias et al. (2023), *PeerJ*, DOI 10.7717/peerj.14973

27/27