1 | INTRODUCTION

Considerable success has been achieved in bioengineering accumulation of oil in the leaves of crops (Andrianov et al., 2010; Vanhercke et al., 2014). Sugarcane is the most productive crop per unit of arable land annually, with a total production of about 1.95 billion tons from an area of 26.8 million ha, resulting in an average fresh cane yield of 72.8 tonnes/ha worldwide (FAOSTAT, 2019; http://www.fao.org/faostat/en/#data/QC). Fiber Sorghums have also proved highly productive, yielding up to 40 tonnes/ha dry matter (Gill et al., 2014). Considerable success has been achieved in bioengineering large accumulations of oils (triacylglycerides) in the leaves of both sorghum and sugarcane (Vanhercke et al., 2019; Zale et al., 2016). However, 80%–90% of plant biomass in these crops is stem, not leaf, at harvest. If similar high-levels of oil accumulation could be achieved in the stems of these plants then this would open the way to economically...
viable production of biodiesel and jet fuel at scale (Huang et al., 2016; Kumar et al., 2018). Converting these to highly productive oil crops therefore requires targeting the expression of transgene combinations that have achieved oil accumulation in leaves to the stem. This requires identifying gene promoters that are specific to the maturing stem, which can then be fused to the transgenes for this enhanced oil synthesis. The promoters currently used in genetically engineered sugarcane are mostly derived from highly expressed, constitutive genes, namely maize Ubi1 (Hansom et al., 1999), CaMV35S (Potenza et al., 2004), maize PepC (Harrison et al., 2011), maize cab-m5 (Petrasovits et al., 2012), rice RUBQ2 (Liu et al., 2003; Petrasovits et al., 2012), and Cestrum YLCV (Kinkema, Geijskes, deLucca, et al., 2014). However, constitutive expression of transgenes throughout a plant for metabolic engineering often has pleiotropic effects, such as stunted growth of TAG accumulating sorghum (Vanhercke et al., 2019), arrested growth and chlorosis in sugarcane lines with the highest production of polyhydroxybutyrate (Petrasovits et al., 2012) and reduction in biomass, lignin, Brix, and juice content in lines hyperaccumulating triacylglycerol (TAG; Parajuli et al., 2020). Therefore, metabolic engineering of sugarcane to convert carbohydrates into high-value products should be restricted to stems, particularly the pith parenchyma cells, where stored photoassimilates accumulate to tremendous levels (Komor, 2000) are available for efficient conversion.

Several putative stem promoters have been studied in sugarcane. The promoter of dirigent-like protein c22-a, which is specific to the mature culm (stem), was isolated, but no GUS reporter gene activity was detected in transient or stable transgenic lines (Goshu Abraha, 2005). Later, promoters of DIRIGENT (SHDIR16) and O-METHYLTRANSFERASE (SHOMT) were shown to drive the GUS reporter gene specifically in the stem vasculature tissues (no expression in parenchyma cells) in sugarcane and rice (Damaj et al., 2010). The promoter of the sugarcane Loading Stem Gene (ScLSG) drives luciferase reporter gene expression preferentially in the stem, with notable non-specific expression in other tissues, such as root (Moyle & Birch, 2013a, 2013b). The full promoter (5.7 kb, pAI57) of the sugarcane ScR1MYB1, which is expressed in the mature stem, is able to drive stronger luciferase reporter gene expression in older stems (Mudge et al., 2013) while a short promoter version (1.1 kb) is non-functional (Mudge et al., 2009). Similar to ScLSG, ScR1MYB1 transcripts were abundant in roots (Mudge et al., 2009). Mudge et al. (2013) compared a suite of stem-preferential promoters, including pLSG1, pAI57, and pH07 (promoter of a CBL-interacting protein kinase, Casu et al., 2004) with the constitutive promoter pUBi to drive expression of isomaltulose (IM) synthase, pAI57 outperformed the other stem promoter candidates and seems to show similar improvements of IM production as pUBi in a field trial (Mudge et al., 2013). Despite the fact that promising progress has been made in characterizing stem-specific promoters, there is a need to identify more stem-specific promoters. A larger suite of stem-specific promoters would allow more options for bioengineering and increase the possibility of stacking together a few genes with stem-specific expression patterns. To improve the production of biofuel carbon, an ideal stem-specific promoter should drive high and specific expression in parenchyma cells of the stem, especially during the mature stage when photoassimilates are highly accumulated in the pith.

Sugarcane has been extensively bred to achieve high sugar content. As a trade-off, fiber content has been decreased, and the plant is thus less resilient to unfavorable environmental conditions (Jackson, 2005). Commercial sugarcane hybrids carry ~80% of the S. officinarum genome (the sugary ancestor) and 10%–23% of the Saccharum spontaneum genome (the resilient ancestor; D’Hont et al., 1996; Moore & Botha, 2013). However, energy cane selections, which have a higher proportion of the S. spontaneum genome, are selected for several traits, such as higher fiber, higher biomass, higher production on marginal land, greater resistant to harsh environment and improved ratooning (Matsuoka, 2017; Matsuoka et al., 2014). Energy cane is bred to make better use of marginal land in order to avoid competition for land with other crops, although it contains lower levels of sugar relative to sugarcane (Matsuoka, 2017; Matsuoka et al., 2014).

Modern sugarcane cultivars are polyploid hybrids with chromosome numbers ranging from 2n = 100 to 130, while the monoploid chromosome set was estimated at 10 (D’Hont et al., 1996; Moore & Botha, 2013). Thus, 10–13 homologous alleles are expected to be present for most loci. Recently, an allele-defined genome of autopolyploid sugarcane using a haploid version of S. spontaneum became available (Zhang et al., 2018), offering an excellent opportunity to speed up the molecular studies of sugarcane. In this study, we analyzed an RNA-seq atlas of S. spontaneum tissues (Zhang et al., 2016) to identify genes that are specifically expressed in stems across various developmental stages. From the RNA-seq atlas, a putative tonoplast sugar transporter (TST) homolog was identified as a potential stem-specific gene at various developmental stages. An extensive BLAST search was carried out to identify all the TST genes and alleles in the S. spontaneum genome. In total, five TST genes and 21 alleles were identified. Real-time qPCR analysis confirmed the TST gene expression levels across various tissues at two developmental stages (immature and mature) in both sugarcane and energy cane. TST1 and TST2b were strongly and almost exclusively expressed in the stem of both sugarcane and energy cane. Dominant alleles for both candidate genes were identified. RNA in situ hybridization showed abundant TST1 and TST2b RNA transcripts in the pith parenchyma cells of the
mature stem. The promoters of TST1 (the pTST2b-1P allele) and TST2b (including pTST2b-1A and pTST2b-1C alleles) were subsequently cloned. The activities of both pTST2b-1A and pTST2b-1C were confirmed to drive GUS reporter gene expression exclusively in the stem using the model plant Arabidopsis thaliana (Meinke et al., 1998).

2  |  MATERIALS AND METHODS

2.1  |  Plant materials and growth conditions

Sugarcane cultivar CP88-1762 and energy cane cultivar UFCP 82-1655 were maintained in the Plant Biology Greenhouse at the University of Illinois at Urbana-Champaign. Plants were grown under controlled temperature (28°C/22°C, day/night) with a 14-h light (approximately 300–950 µmol m⁻² s⁻¹)/10-h dark regimen provided by a mix of natural photoperiod and light conditions. Arabidopsis Col-0 plants were grown under controlled temperature (22°C) with a 16-h light (130–150 µmol m⁻² s⁻¹)/8-h dark photoperiod. The hydroponic culture of Arabidopsis was carried out as previously described (Zeng et al., 2018) with minor adjustments. T₃ homozygous transgenic Arabidopsis seeds were germinated in a small growth tube filled with selection medium (hygromycin at 25 µg/ml) in the hydroponic system. Plants were grown in the hydroponic system under controlled temperature (22°C) with a 16-h light (100–110 µmol m⁻² s⁻¹)/8-h dark photoperiod.

2.2  |  Tau index (τ) analysis of S. spontaneum RNA-seq

Tissue specificity analysis using published S. spontaneum (SES208) RNA-seq data (Zhang et al., 2016) was performed using the Tau (τ) index (Kryuchkova-Mostacci & Robinson-Rechavi, 2017; Yanai et al., 2005) implemented in R with a script modified from a recent study (Hetti-Arachchilage et al., 2020). Expression values were grouped into four categories: seedling stem, seedling leaf, mature stem, and mature leaf. The τ index was calculated from those four values. For the purposes of this analysis, pre-mature and mature tissues were both considered mature when compared with the seedling tissue, and zeros in the expression values were converted to 1 × 10⁻¹⁰ in order to avoid dividing by zero errors. The highly stem-specific gene list was created by filtering for both a Tau index of 0.98 or higher and a maximum expression value falling into the mature stem category. The heatmap was constructed using the pheatmap package from Bioconductor in R based on log2-normalized FPKM values (Perry, 2016). Both columns and rows were organized using hierarchical clustering based on distance mapping.

2.3  |  Identification and phylogenetic analysis of TST family members

Protein sequences were retrieved from BLAST search in S. spontaneum AP85-441 genome (http://www.life.illinois.edu/ming/downloads/Spontaneum_genome/) using Sspon.08G0024630-1C protein sequence as bait. Arabidopsis, sugar beet and sorghum TST protein sequences were retrieved from Phytozome 12 (https://phytozome.jgi.doe.gov/pz/portal.html) according to previous studies (Bihmidine et al., 2016; Jung et al., 2015; Wormit et al., 2006). Sequences were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/; Madeira et al., 2019). The phylogenetic tree was constructed using the Maximum Likelihood method based on the Jones model (Jones et al., 1992) with the bootstrap consensus inferred from 1000 replicates (Felsenstein, 1985). The tree with the highest log likelihood (−9614.0509) was shown. The percentage of trees in which the associated taxa clustered together was shown next to the branches. The initial tree was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A Gamma distribution was used to model evolutionary rate differences among sites (four categories; +G, parameter = 0.7947). All positions with less than 90% site coverage were eliminated. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).

2.4  |  Sample collections from various tissues at two developmental stages

Samples from various tissues, including different internodes, young leaf (leaf roll), mature leaf (the third fully expanded leaf from top), and root were collected on four separate sampling days. Samples were taken when the cane reached a maximum of five internodes (for immature stage samples) or 16 internodes (for mature stage samples) according to their own developmental growth status (internodes were numbered top down). On each sampling day (around noon), approximately 10 g of samples from at least three individual plants grown in separate pots were collected and flash frozen in liquid nitrogen. Samples (except for those prepared for RNA in situ hybridization) were grounded into powders before storing at −80°C.

2.5  |  RNA isolation and RT-qPCR

RNA was isolated using Trizol (Invitrogen) as instructed by the manufacturer. We confirmed RNA integrity by electrophoresis in a 1.5% agarose gel. First-strand cDNA was synthesized from 1 µg of RNA using oligo(dT) and MuLV reverse transcriptase (NEB). Real-time qPCR was
performed using PowerUp™ SYBR™ master mix (Applied Biosystems) according to the manufacturer’s instructions on a CFX96 Real-Time PCR Detection System (Bio-Rad) using gene-specific primers (P1-P16, Table S1). The expression values were normalized to GAPDH expression values in each repeat using 2^{-\Delta\Delta CT} method (Livak & Schmittgen, 2001). The analyses were based on three biological replicates. Semi-quantitative RT-PCR was carried out using DreamTaq DNA Polymerase (ThermoFisher) and an equal amount of internode cDNA (pooled from three individual plants) in reactions subjected to 30 cycles of PCR using alleles-specific primers (P17-P40).

2.6 RNA in situ hybridization

The stem samples were embedded in paraffin as previously described (Casu et al., 2003; Long & Barton, 1998) with modifications. Energy cane stems from internode 5 and internode 16 (rind removed) were hand-cut into approximately 2-mm-thick disks before being further cut into small wedges. Samples were immediately fixed in a solution of 2% formaldehyde and 0.5% glutaraldehyde in 50 mM Na-PIPES buffer pH 7.2 for 2 h at room temperature in a vacuum desiccator. Samples were removed and placed into fresh fixation solution and shaken gently overnight at 4°C. Samples were softened using 4% ethylenediamine in 50 mM Na-PIPES buffer pH 7.2 for 1 week at 4°C before dehydrating in a series of ethanol and ethanol/Histoclear, followed by embedding in paraffin. The following RNA hybridization process was carried out according to an Abcam protocol (https://www.abcam.com/protocols/ish-in-situ-hybridization-protocol). RNA probes were synthesized in vitro (MAXIscript™ SP6/T7 Transcription Kit, ThermoFisher) from plasmid pGEM-T that was carrying either 214 or 683 bp of cDNA (P41-P48) specific to TST1 and TST2b, respectively. Probes were labeled with fluorescein-12-UTP and were subsequently hydrolyzed to approximately 100 nucleotides. Hybridized probes were detected with Anti-Fluorescein-AP (Sigma-Aldrich) and the substrates 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium.

2.7 Cloning procedures

The promoter regions (5000 bp upstream from ATG) of the alleles of TST1-1P, TST2b-1A, and TST2b-1C were synthesized (GenScript) with internal DraIII, BbsI, and BsaI restriction sites modified without affecting putative cis-regulatory elements in promoters (MatInspector). Constructs were assembled using the Golden Gate cloning protocol before subcloning into a binary vector (pL2V-1; Engler et al., 2014; Weber et al., 2011). Construct diagrams can be found in Figure 6a. The codon-optimized GUS sequence for sugarcane expression (scoGUS; Kinkema, Geijsskes, Shand, et al., 2014) and PvUbiII terminator (Mann et al., 2011) were kindly provided by Dr. Fredy Altpeter. For stable transformation of Arabidopsis, the binary vectors were transformed into Agrobacterium tumefaciens GV3101 by electroporation and then transformed into the Col-0 wild-type using the floral dip method (Clough & Bent, 1998). T0-transformed lines were selected based on hygromycin resistance, and gene insertion was verified by PCR analysis and GUS staining. Out of 24 T0 plants generated for each construct, 14 pTST2b-1A:GUS lines, 8 pTST2b-1C:GUS lines, and 16 p35S:GUS lines were GUS positive, while all pTST1-1P:GUS lines are negative as empty vector control lines are. Three GUS positive lines and three GUS negative lines carrying a single transgene insertion were propagated to T1 homozygous seeds. T1 lines were grown hydroponically to test GUS staining at the whole plant level. Native promoters (~5 kb upstream of ATG) of TST2b-1A from sugarcane CP88-1762 and energy cane UFCP 82-1655 were amplified using the same pair of primers (P49 and P50). The native promoters were cloned into pGEM-T and their sequences were confirmed by Sanger sequencing using three independent colonies.

2.8 GUS histochemical analysis

GUS staining was performed as previously described (Chen et al., 2012). Three-week-old homozygous T1 transgenic plants (before bolting) and 5-week-old whole plants (fully grown) were collected for histochemical GUS staining. The plants were stained for 48 h. A flatbed scanner was used for imaging results. Three independent lines of each transformation were scanned, and one of the representative images was shown.

3 RESULTS

3.1 Tau analysis to identify genes specifically expressed in the stem

Tau (τ) index analysis of tissue-specific gene expression has been proposed to be the most robust algorithm for RNA-seq data (Hetti-Arachchilage et al., 2020; Kryuchkova-Mostacci & Robinson-Rechavi, 2017). Here we used the \( \tau \) index (\( \tau > 0.98 \)) to identify stem-specific genes and genes highly expressed in the mature stem of sugarcane. The top 150 stem-specifically/preferentially expressed genes were graphically presented according to their expression across all tissues at various developmental stages (seedling, premature, and mature stages) in S. spontaneum using a heatmap (Figure 1). For the top 25 candidate genes, the FPKM values from the original RNA-seq data as well as the predicted functions of their homologs in Arabidopsis or rice were tabulated (Table S2).
The gene Sspon.08G0024630-1C (highlighted in red in Figure 1), a putative TST homolog, showed not only a highly stem-specific expression pattern, but also a progressive increase in expression over the course of internode development at both the pre-mature and mature stages (Figure 1; Table S2). This expression pattern is consistent with both the increased sugar accumulation seen in the internodes towards the bottom of the stem (Zhang et al., 2020) and the predominant storage of sugars in the tonoplast of pith parenchyma cells (Komor, 2000). These consistencies place this putative tonoplast sugar transporter at the top of our candidate gene list. Two other genes, Sspon.08G0000580-3D and Sspon.05G0023050-2P (a homolog responsive to desiccation 22 and a putative wound-induced protein, respectively), also met our criteria (namely, an older internode-specific pattern at both premature and mature stages) and deserve further investigation in future studies.

Among the top 25 candidate genes (Table S2), the remaining genes were specifically expressed in older internodes only at the premature stage and peaked in the intermediate internode (internode 6) at the mature stage. However, caution should be taken when interpreting results from the RNA-seq dataset, as no replicates of data from premature and mature stages were found (Zhang et al., 2016).

3.2 | Phylogenetic analysis to identify TST genes and alleles in S. spontaneum

Since the TST gene family contains multiple members in other plant species (Bihmidine et al., 2016; Jung et al., 2015; Wormit et al., 2006), we were confident that TST should have more than one member in S. spontaneum, which was confirmed by a recent
study (Zhang et al., 2020). After extensive BLAST searches in the *S. spontaneum* genome using the Ssp0.08G0024630-1C protein sequence as bait, 21 putative TST homologs in total were identified. These TST protein sequences were aligned and analyzed for their phylogenetic relationship, along with the published sorghum, Arabidopsis and sugar beet TSTs (Bihmide et al., 2016; Jung et al., 2015; Wormit et al., 2006; Data S1). As shown in Figure 2, TST3 members were grouped together from all tested species, while TST1/TST2 from monocots (sugarcane and sorghum) were grouped separately from TST1/TST2 in dicots (Arabidopsis and sugar beet), indicating a potential functional divergence for TST1/2 between the monocots and dicots. Sugarcane TST genes were phylogenetically named after the sorghum TSTs. As a result, five TST genes (TST1, TST2a, TST2b, TST3a, TST3b), each including three to five alleles, were delineated. There were high degrees of homozygosity among the alleles of each TST gene (Figure S1). Notably, two TST2b homologous genes were not annotated, but were manually curated from chromosome 3A and tig00020005 of the *S. spontaneum* genome.

3.3 | RT-qPCR analysis of TST genes in sugarcane and energy cane at immature and mature stages

As the genetic backgrounds vary among *S. spontaneum* (a wild species) and the cultivated sugarcane and energy cane (which inherited disproportionate amounts of genomes from *S. officinarum* and *S. spontaneum*), the expression patterns in sugarcane and energy cane may be different from those seen in the *S. spontaneum* RNA-seq data. TST gene expression was examined using RT-qPCR at the immature and mature stages in both sugarcane and energy cane to identify the most robust and tissue-specific TST promoter. We also included two other putative stem-specific genes, *LSG* and *R1MYB1* (the Z1 allele, its promoter was known as pA157; Mudge et al., 2013), for comparisons. As each TST gene has multiple alleles in *S. spontaneum*, which has been used to generate modern sugarcane cultivars (Meng et al., 2020), we would reasonably predict that multiple TST alleles are present in the sugarcane and energy cane cultivars. To strategically narrow down which TST genes are of most interest, we first tested TST gene expression using primers specifically targeting conserved regions of each TST gene, not individual TST alleles. Samples were taken from young leaves, mature leaves and roots as well as internodes, with internodes numbered from the top of the plant at each stage (immature and mature), such that older internodes are higher numbered internodes at each stage.

For sugarcane at the immature stage (Figure 3a), *TST1* and *TST2b* were highly expressed in older stems (IN5), with *TST2b* even similar to the housekeeping gene GAPDH in expression (relative expression close to 1), while they were undetectable in other tissues, including mature leaf, young leaf and root. By contrast, *TST2a* was not stem-specific, *TST3a* was extremely low in the stem, as illustrated by the value relative to GAPDH (0.0016 in IN5; about 500-fold lower than 2b), and *TST3b* was not detected in any tested tissues in the immature stage. The *LSG* gene was preferentially expressed in the stem, but also expressed fairly highly in other tissues. The *R1MYB1* gene was expressed in older internodes (Figure 3a, IN5), but also in roots. However, the transcript level of *R1MYB1* was not strong compared with the housekeeping gene (note the differences in scale). Considering the tissue specificity and the expression strength relative to GAPDH, *TST2b* appeared to be the best candidate, followed by *TST1*, among the tested genes at the immature stage of sugarcane. The mean Ct values of the housekeeping gene GAPDH in each sample were shown in Figure S2. The expression level of the internal control was stable between the two developmental stages and also in the various tissues, except for roots, which were about three cycles higher than the other tissues.

For sugarcane at the mature stage (Figure 3b), the expression patterns of the tested genes were largely the same as those observed at the immature stage. The transcript level of *TST2b* was highly stem-specific and was increased substantially in the stems at this stage, compared to at the immature stage. For example, *TST2b* at IN3 increased from 0.74 to 3.20, suggesting that *TST2b* may be significantly involved in the process of sugar accumulation in older internodes as
sugarcane plants mature. The expression patterns of the rest of the tested genes were similar to those at the immature stage, except for TST3b, which was now detected at an extremely low level (similar to TST3a) in the mature stage.

In energy cane, the gene expression patterns across the different tissues at both the immature and mature stages were overall similar to those of sugarcane (Figure 3c,d). Although TST2b was still highly stem-specific and strongly expressed in older internodes (IN5 in Figure 3c and IN16 in Figure 3d), its expression decreased from the immature stage to the mature stage (decreased from 2.78 to 1.00 at IN3), which differed from what we observed in sugarcane (Figure 3a,b). Energy cane was bred to have less sugar content and it also has a higher proportion of the S. spontaneum genome than sugarcane, which may contribute to the differential expression of TST2b across developmental stages between the two species.
3.4 | \textit{TST1} and \textit{TST2b} RNA transcripts are abundant in pith parenchyma cells

RNA-seq and RT-PCR analysis showed that \textit{TST1} and \textit{TST2b} were specifically expressed in the stem and strongly expressed in mature internodes with high sugar accumulation. To further examine the spatial expression patterns of \textit{TST1} and \textit{TST2b} at the cellular level, RNA in situ hybridization was conducted using an antisense RNA probe against either \textit{TST1} or \textit{TST2b} in stem tissues. The presence of RNA transcripts was visualized as dark blue stain on cross-sections of internodes. \textit{TST1} transcripts were detected in high sugar-accumulating pith parenchyma cells and vascular bundles (including xylem parenchyma cells, phloem parenchyma cells, sclerenchymatous fibers, and weakly in sieve elements/phloem companion cells) in both internode 5 and internode 16 (Figure 4a). \textit{TST1} transcripts were more abundant in internode 16 than in internode 5. \textit{TST2b} transcripts were detected only in pith parenchyma cells of internode 5 (Figure 4b) but were visualized in both pith parenchyma cells and vascular bundles in internode 16. Similar to \textit{TST1}, \textit{TST2b} transcripts were also more abundant in internode 16 than in internode 5, which is consistent with our RT-qPCR results (Figure 3). In control experiments using sense RNA probes, only weak background color was developed (Figure 4a,b).

3.5 | Allele analysis using RT-PCR and RT-qPCR

As polyploid sugarcane hybrids carry multiple alleles for each \textit{TST} gene. In \textit{S. spontaneum} genome, five alleles of \textit{TST1} (\textit{TST1-1A}, \textit{TST1-2B}, \textit{TST1-3C}, \textit{TST1-4D} and \textit{TST1-1P}) and four alleles of \textit{TST2} (\textit{TST2b-1A}, \textit{TST2b-1C}, \textit{TST2b-3A} and \textit{TST2b-tig}) were identified. Allele-specific primers were used to identify which allele might be preferentially expressed within \textit{TST1} and \textit{TST2b}. Considering the high level of homozygosity of alleles for a given \textit{TST} gene (Figure S1), we were able to re-design only one forward or reverse primer that specifically targeted one allele. As shown in Figure 5a, one or more alleles of \textit{TST2b} and \textit{TST1} of different size were amplified when compared with only the one amplicon for each \textit{TST} gene when the conserved pairs of primers were used. No differences in the expression patterns of the \textit{TST} alleles were observed between sugarcane and energy cane (Figure 5a). For the \textit{TST2b} gene, the alleles 3A and tig did not produce RT-PCR amplicons (3A and tig), while 1A and 1C were amplified from the internodes of both sugarcane and energy cane (Figure 5a, left). \textit{TST2b-1A} was highly expressed at an expression level similar to the mixed level of \textit{TST2b}. Thus, \textit{TST2b-1A} is likely to be the main contributor to the \textit{TST2b} levels detected by q-PCR and RNA in situ hybridization. For \textit{TST1}, only the pair of primers targeting three alleles, namely 1P, 2B, and 4D, were able to produce amplicon levels similar to that the conserved \textit{TST1} amplicon (Figure 5a, right), while primers targeting any single allele failed to amplify any products. This lack of amplification when targeting a single allele could be due to low levels of expression or insufficient similarity between the genome sequences we used to design primers targeting the untranslated regions and the cultivars providing our samples. We were not able to design gene-specific PCR primers to differentiate 1P, 2B, and 4D, as they have very high similarity in their coding sequences.

The sequences of the promoters of \textit{TST1-2B} and \textit{TST1-4D} were drastically different from the other three alleles, \textit{TST1-1P}, \textit{TST1-1A} and \textit{TST1-3C} (Figure S3). It has been observed that gene retention and fractionation occur in polyploid plants, resulting in duplicated genes residing in areas...
of differential gene density and being expressed at different levels (Cheng et al., 2018). The differences in the promoter sequences of the TST1 alleles may be due to past or ongoing fractionation (deletion) of cis-acting sites. Without being able to successfully detect individual allele expression of TST1-1P, TST1-2B, and TST1-4D, we riskily decided to start with TST1-1P at first, as it is one of the three alleles with similar promoters.

To confirm whether the expression patterns of the putative dominant alleles were consistent with the overall gene expression patterns described in Figure 3, RT-qPCR analysis was performed using allele-specific primers for samples from different tissues at immature and mature stages in sugarcane (Figure 5b) and energy cane (Figure 5c). The gene expression patterns of the putative dominant alleles, TST2b-1A3A and TST1-1P2B4D, consistently showed high levels of stem-specific expression in older internodes (IN5 at the immature stage and IN16 at the mature stage).

3.6 | TST2b promoters direct GUS gene expression toward the stem in Arabidopsis

The combined results of the RT-qPCR analysis, RNA in situ hybridization and allele characterization, the alleles of most interest were TST1-1P, TST2b-1A, and TST2b-1C, as they were stem-specific, strongly expressed in old internodes, and abundantly present in pith parenchyma cells. As promoters exert the predominant control of specific gene expression (Cooper, 2000), we hypothesized that the promoters of these genes conferred these specific features. To evaluate the specificity and robustness of these promoters, we cloned the corresponding promoter regions (5 kb upstream of ATG) from the S. spontaneum genome and added each to constructs carrying the reporter gene β-glucuronidase (GUS) for Arabidopsis transformation (Figure 6a). GUS driven by the constitutive 35S promoter served as a positive control, while the empty vector served as a negative control. Plants were grown in a hydroponic system in order to stain the whole plant, including roots. Three-week-old and 5-week-old plants were subjected to histochemical GUS staining. In the younger plants carrying a TST promoter construct, no GUS signals were detected, but GUS signals were constitutively present in p35S:GUS plant (Figure 6b, top). In older plants, strong GUS signals were visualized in stems, and weak signals were detected in primary veins of rosette leaves in pTST2b-1A:GUS and pTST2b-1C:GUS transgenic lines (Figure 6b, bottom), but no signals were detected in other tissues, including siliques (Figure 6b, insets), roots, rosette leaf mesophyll cells, and cauline leaves. The GUS signals were mainly found in the cortical cells in the vicinity of interfascicular fibers and the phloem in stem cross-sections of both pTST2b-1A:GUS and pTST2b-1C:GUS transgenic lines (Figure S4). GUS signals were observed throughout the plant in p35S:GUS lines with weak staining in stems. No GUS staining was observed in the TST1-1P:GUS lines or in empty vector (EV) control lines.
The availability of the *S. spontaneum* genome sequence provides substantial advantages for research. However, the promoters from it may not be favored for biotechnology, as *S. spontaneum* is a wild species. We assembled the ~5-kb genome sequence of *S. spontaneum* with at least 98.1% identity to the *pTST2b-1A* from sugarcane and 97.8% to the *pTST2b-1A* from energy cane, with the few mismatches (SNP, insertions and deletions) highlighted in red. The reference genome was sequenced from a haploid carrying four sets of monoploid chromosomes (Zhang et al., 2018), and modern sugarcane cultivars are polyploid hybrids with 10–13 alleles could be found for most loci (D’Hont et al., 1996; Moore & Botha, 2013). In this haploid dataset, we found 21 alleles in the *TST* gene family. Thus, the number of alleles in commercial sugarcane or energy cane cultivars is expected to be higher. Due to the high polyploidy and complex heterozygosity of the sugarcane genome, accrued defects in promoters as well as homology-dependent gene silencing may inhibit the expression of many alleles (Moore & Botha, 2013). Of the five *TST1* alleles and the four *TST2b* alleles, *TST1-1P*, *TST2b-1A*, and *TST2b-1C* were the preferentially expressed alleles. Expressed sequence tag (EST) studies indicate that multiple alleles are actively expressed at a locus (Grivet et al., 2003), and dominant alleles have been reported. For example, eight alleles of the R1MYB1 transcriptional factor were analyzed in sugarcane cultivar Q117, with one allele (Z1) estimated to contribute 40% of the RNA transcripts expressed in mature internodes, whereas four other alleles contribute <1% of the expressed transcripts (Mudge et al., 2009). The promoter of the Z1 allele (pAA57) can drive luciferase reporter gene expression in mature internodes (Mudge et al., 2013). In the same study, pAA57 and pUbi were used to drive a sucrose isomerase gene, which converts sucrose to high-value isomaltulose, in sugarcane. However, no significant differences were observed in isomaltulose content or total sugar content between pAA57 and pUbi transgenic sugarcane. This may be related to the strong, non-specific expression of Z1 detected in root (Figure 3). We also included another putative stem-specific gene, *LSG* (loading stem gene), in our RT-qPCR analysis. It is indeed highly expressed in internodes, but it is also highly expressed in root as well as young leaves (Figure 3). There were nine *LSG* alleles identified in sugarcane (Moyle & Birch, 2013a), and four out the six cloned *LSG* promoters were able to drive luciferase reporter gene expression in mature internodes and in root (Moyle & Birch, 2013b). In our study, both *TST2b-1A* and *TST2b-1C* promoters could drive GUS reporter gene specifically in the Arabidopsis stem, with no GUS signals observed in siliques, root, cauline leaves, or rosette leaves, excluding some major veins (weakly stained). The *TST1-1P* promoter, however, failed to activate GUS reporter gene activity in transgenic Arabidopsis. This exogenous promoter may not work well in the model plant (a dicot plant). However, this is unlikely as the two other promoters from sugarcane functioned well in Arabidopsis. The amplification cycles of TST1 genes were similar to those of the reference gene during qPCR in canes, while the GUS signal in Arabidopsis stems was weak compared to the positive control in Arabidopsis. These observations suggest that spatial localization within stem tissue may vary between Arabidopsis and canes. The in situ RNA signal was stronger in the vascular tissue than in the pith cells in canes, while the GUS signal was not strong in the vascular tissues in Arabidopsis, and thus the GUS signal in the pith may be too weak to be visualized. Additionally, the weak signal in Arabidopsis may be due to the use of the scoGUS gene, which may not be effectively expressed in Arabidopsis since it was optimized for expression in sugarcane and can result in 32- to 50-fold higher GUS expression than the standard GUS (Kinkema, Geijskes, deLucca, et al., 2014; Kinkema, Geijskes, Shand, et al., 2014). Another possible explanation is that the pith cells in the Arabidopsis stem
are not a sugar-accumulating cell type compared to those in canes and the architectures of Arabidopsis and canes are different. Nonetheless, it is worth validating these promoters in sugarcane and energy cane.

The TST family genes are involved in vacuolar sugar accumulation in different species, such as Arabidopsis (Schulz et al., 2011; Wormit et al., 2006), sugar beet (Jung et al., 2015), tomato and apple (Zhu et al., 2021). However, none of the previously identified TST genes are stem-specifically expressed. Sorghum that uses the stem as a major sugar-storage organ is a close relative of sugarcane. However, sorghum TST genes are highly expressed in both leave and stem. (Bihmidine et al., 2016). Nevertheless, the promoters identified here are expected to be useful in sorghum, another important biofuel grass (Mullet et al., 2014; Rooney et al., 2007). Although the stem of Arabidopsis does not accumulate high levels of sugar, as do sugarcane and sorghum, the TST2b-1A and TST2b-1C promoters still conveyed stem-specific expression. Therefore, these promoters could also be used in many other species for stem-specific expression of other target genes unrelated to sugar accumulation.

ACKNOWLEDGEMENTS
We thank Dr. Moonsub Lee for maintaining sugarcane CP88 and thank Deborah Black for general help and advice during plant growth in the greenhouse. We thank Dr. Freddy Altpeter for providing us ScoGUS and PvUbiII terminator level 0 plasmids, and the energy cane plants. We thank Qing Zhang for providing us the updated expression table for the S. spontaneum RNA-seq data. We also thank Dr. Stephen Long for editing this manuscript. The information, data, and work presented herein was funded in part by the Biological and Environmental Research (BER) program, U.S. Department of Energy, under Award Number DE-SC0018254. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS
Y. L. conducted hydroponic culture of transgenic Arabidopsis. C. M. W. performed the extensive BLAST searches in the S. spontaneum genome and manually curated the CDS and promoter sequences. G. B. conducted Tau analysis and made the corresponding heatmap. J. W. designed, conducted the rest of the experiments, analyzed data and wrote the manuscript. L.-Q. C. conceived the project, designed experiments, interpreted data, and wrote the manuscript.

DATA AVAILABILITY STATEMENT
The data that support the findings of the present study are available on request from the corresponding author.

REFERENCES
Andrianov, V., Borisjuk, N., Pogrebnyak, N., Brinker, A., Dixon, J., Spitsin, S., Flynn, J., Matysyczuk, P., Andryszyak, K., Laurrell, M., Golovkin, & M. Koprówski, H. (2010). Tobacco as a production platform for biofuel: Overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. Plant Biotechnology Journal, 8(3), 277–287. https://doi.org/10.1111/j.1467-7652.2009.00458.x
Bihmidine, S., Julius, B., Dweikat, I., & Baum, D. (2016). Tonoplast Sugar Transporters (StTSTs) putatively control sucrose accumulation in sweet sorghum stems. Plant Signaling & Behavior, 11(1). https://doi.org/10.1080/15592324.2015.1117721
Casu, R., Dimmock, C., Chapman, S., Grof, C., McIntyre, C., Bonnett, G., & Manners, J. (2004). Identification of differentially expressed transcripts from maturing stem of sugarcane by silico analysis of stem expressed sequence tags and gene expression profiling. Plant Molecular Biology, 54(4), 503–517. https://doi.org/10.1023/B:PLAN.0000038255.96128.41
Casu, R., Grof, C., Rae, A., McIntyre, C., Dimmock, C., & Manners, J. (2003). Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Molecular Biology, 52(2), 371–386. https://doi.org/10.1023/A:1023957214644
Chen, L., Qu, X., Hou, B., Sosso, D., Osorio, S., Fernie, A., & Frommer, W. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science, 335(6065), 207–211. https://doi.org/10.1126/science.1213351
Cheng, F., Wu, J., Cai, X., Liang, J., Freeing, M., & Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. Nature Plants, 4(5), 258–268. https://doi.org/10.1038/s41477-018-0136-7
Clough, S., & Bent, A. (1998). Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant Journal, 16(6), 735–743. https://doi.org/10.1046/j.1365-313x.1998.00343.x
Cooper, G. (Ed.). (2000). Regulation of transcription in eukaryotes. The cell: A molecular approach (2nd ed., pp. 268–286). Sinauer Associates, Inc.
Damaj, M. B., Kumpatla, S. P., Emani, C., Beremand, P. D., Reddy, A. S., Rathore, K. S., Buenrostro-Nava, M. T., Curtis, I. S., Thomas, T. L., & Mirkow, T. E. (2010). Sugarcane DIRIGENT and O-METHYLTRANSFERASE promoters confer stem-regulated gene expression in diverse monocots. Planta, 231(6), 1439–1458. https://doi.org/10.1007/s00425-010-1138-5
D’Hont, A., Grivet, L., Feldmann, P., Rao, S., Berding, N., & Glaszmann, J. (1996). Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp) by molecular cytogenetics. Molecular & General Genetics, 250(4), 405–413. https://doi.org/10.1007/BF02174028
Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014). A golden gate modular cloning toolbox for plants. ACS Synthetic Biology, 3(11), 839–843. https://doi.org/10.1021/sb4001504
Mudge, S., Osabe, K., Casu, R., Bonnett, G., Manners, J., & Birch, R. (2009). Efficient silencing of reporter transgenes coupled to known functional promoters in sugarcane, a highly polyplody crop species. *Planta*, 229(3), 549–558. https://doi.org/10.1007/s00425-008-0852-8

Mullet, J., Morishige, D., McCormick, R., Truong, S., Hilley, J., McKinley, B., Anderson, R., Olson, S. N., & Rooney, W. (2014). Energy sorghum – A genetic model for the design of C4 grass bioenergy crops. *Journal of Experimental Botany*, 65(13), 3479–3489. https://doi.org/10.1093/jxb/eru229

Parajuli, S., Kannan, B., Karan, R., Sanahuja, G., Liu, H., Garcia-Ruiz, E., Kumar, D., Singh, V., Zhao, H., Long, S., Shanklin, J., & Altpeter, F. (2020). Towards oilcane: Engineering hyperaccumulation of triacylglycerol into sugarcane stems. *Global Change Biology Bioenergy*, 12(7), 476–490. https://doi.org/10.1111/gcbb.12684

Perry, M. (2016). Flexible heatmaps for functional genomics and sequence features. *R Package Version*, 1(11).

Petrasovits, L. A., Zhao, L., McQuarler, R. B., Snell, K. D., Somleva, M. N., Patterson, N. A., Nielsen, L. K., & Brumbley, S. M. (2012). Enhanced polyhydroxybutyrate production in transgenic sugarcane. *Plant Biotechnology Journal*, 10(5), 569–578. https://doi.org/10.1111/j.1747-6591.2012.00686.x

Potenza, C., Aleman, L., & Sengupta-Gopalan, C. (2004). Targeting transgene expression in research, agricultural, and environmental applications: Promoters used in plant transformation. *Vitro Cellular & Developmental Biology-Plant*, 40(1), 1–22. https://doi.org/10.1079/1VP2003477

Rooney, W., Blumenthal, J., Bean, B., & Mullet, J. (2007). Designing sorghum as a dedicated bioenergy feedstock. *Biofuels, Bioproducts and Biorefining*, 1(2), 147–157. https://doi.org/10.1002/bbb.15

Schulz, A., Beyhl, D., Martin, I., Wormit, A., Neuhaus, E., Boschet, G., Büttner, M., Schneider, S., Sauer, N., & Hedrich, R. (2011). Proton-driven sucrose symport and antiport are provided by the vacuolar transporters SUC4 and TMT1/2. *Plant Journal*, 68(1), 129–136. https://doi.org/10.1111/j.1365-313X.2011.04672.x

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Vanhercke, T., Belide, S., Taylor, M., El Tahchy, A., Okada, S., Rolland, V., & Petrie, J. (2019). Up-regulation of lipid biosynthesis increases the oil content in leaves of *Sorghum bicolor*. *Plant Biotechnology Journal*, 17(1), 220–232. https://doi.org/10.1111/pbi.12959

Vanhercke, T., El Tahchy, A., Liu, Q., Zhou, X.-R., Shrestha, P., Divi, U. K., Rai, J.-P., Mansour, M. P., Nichols, P. D., James, C. N., Horn, P. J., Chapman, K. D., Beaudoin, F., Ruiz-López, N., Larkin, P. J., Feyter, R. C., Singh, S. P., & Petrie, J. R. (2014). Metabolic engineering of biomass for high energy density: Oilseed-like triacylglycerol yields from plant leaves. *Plant Biotechnology Journal*, 12(2), 231–239. https://doi.org/10.1111/pbi.12131

Weber, E., Engler, C., Gruetznner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, 6(2), https://doi.org/10.1371/journal.pone.0016765

Wormit, A., Trentmann, O., Feifer, I., Lohr, C., Tjaden, J., Meyer, S., & Neuhaus, H. (2006). Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport. *The Plant Cell*, 18(12), 3476–3490. https://doi.org/10.1105/tpc.106.047290

Yanai, I., Benjamini, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Opir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., & Shmueli, O. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, 21(5), 650–659. https://doi.org/10.1093/bioinformatics/bti042

Zale, J., Jung, J. H., Kim, J. Y., Pathak, B., Karan, R., Liu, H., Chen, X., Wu, H., Candreva, J., Zhai, Z., Shanklin, J., & Altpeter, F. (2016). Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. *Plant Biotechnology Journal*, 14(2), 661–669. https://doi.org/10.1111/pbi.12411

Zeng, H., Xia, C., Zhang, C., & Chen, L.-Q. (2018). A simplified hydroponic culture of Arabidopsis. *Bio-Protocol*, 8(24), e3121. https://doi.org/10.21769/BioProtoc.3121

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., & Ming, R. (2018). Allele-defined genome of the autoployploid sugarcane *Saccharum spontaneum* L. *Nature Genetics*, 50(12), 1754. https://doi.org/10.1038/s41588-018-0293-7

Zhang, Q., Hu, W., Zhu, F., Wang, L., Yu, Q., Ming, R., & Zhang, J. (2016). Structure, phylogeny, allelic haplotypes and expression of sucrose transporter gene families in Saccharum. *BMCG Genomics*, 17, https://doi.org/10.1186/s12864-016-2419-6

Zhang, Q., Hua, X., Liu, H., Yuan, Y., Shi, Y., Wang, Z., & Zhang, J. (2020). Evolutionary expansion and functional divergence of sugar transporters in Saccharum (*S. spontaneum* and *S. officinarum*). *Plant Journal*, 105, 884–906. https://doi.org/10.1111/tpj.15076

Zhu, L., Li, B., Wu, L., Li, H., Wang, Z., Wei, X., Ma, B., Zhang, Y., Ma, F., Ruan, Y.-L., & Li, M. (2021). MdERDL6-mediated glucose efflux to the cytosol promotes sugar accumulation in the vacuole through up-regulating TSTs in apple and tomato. *Proceedings of the National Academy of Sciences of the United States of America*, 118(1), e2022788118. https://doi.org/10.1073/pnas.2022788118

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Wang, J., Li, Y., Wai, C. M., Beuchat, G., & Chen, L.-Q. (2021). Identification and analysis of stem-specific promoters from sugarcane and energy cane for oil accumulation in their stems. *GCB Bioenergy*, 00, 1–13. https://doi.org/10.1002/gcbb.12872