Tropical Plant Biology (2019) 12:158–173
https://doi.org/10.1007/s12042-019-09225-3

Evolution and Expression Analysis of Starch Synthase Gene Families in *Saccharum spontaneum*

Panpan Ma[1] · Yuan Yuan[2] · Qiaochu Shen[1] · Qing Jiang[1] · Xiuting Hua[1] · Qing Zhang[1] · Muqing Zhang[3] · Ray Ming[1,4] · Jisen Zhang[1]

Received: 29 November 2018 / Accepted: 27 March 2019 / Published online: 14 May 2019

Abstract

Starch is one of two crucial products of photosynthetic carbon-assimilation and mainly functions as the unit of energy storage in most crops such as rice, maize and sorghum, whereas interestingly in sugarcane that unit of energy storage is sucrose. Mature sugarcane stalk tissue has a very large apoplastic volume and contains nearly 700 mM sucrose—which is among the highest recorded sucrose concentrations in plant tissue. We identified 9 genes of starch synthases (*SSs*) related to the starch synthesis pathway in the genome of *S. spontaneum*. Based on gene structure and phylogenetic analysis, *SSs* genes were clustered into five clades and were relatively conserved. In *S. spontaneum*, the *SS* is a very ancient gene family, in which, *SSIIIa* and *SSIIIb* originated from the ρ-whole genome duplications (WGDs), *SSIIb* and *SsIIc* originated from gene duplication after the split of monocots and dicots; *GBSSI* and *GBSSII* in Clade V and *SSIIa* in Clade II were retained from the ε-WGD, and the remaining two *SSs* (*SSI* and *SSIV*) were retained from the very ancient gene duplication event about 355–389 million year ago (Mya). In addition, we found all *SS* genes were under the influence of strong purification with a Ka/Ks ratio of less than 0.5 in *S. spontaneum*. In the 5 families, *SSIIIa, SSIIb* and *GBSSII* had relatively predominant expression levels in all the examined tissues from the two *Saccharum* species, indicating the three genes were the fundamental members in the non-storage tissues, leaf or stem, which is in agreement with previous studies. Interestingly, the expression levels of *SSs* in stems showed significantly higher values in *S. spontenum* than in *S. officinarum* at pre-mature and mature stages. These results were negatively correlated with the sucrose levels between the two *Saccharum* species. At the pre-mature and mature stages, the sucrose contents in stems from *S. officinarum* were much higher than in stems from *S. spontenum,* suggesting that *SSs* involved in the differential of carbohydrate metabolism between the two *Saccharum* species. Besides, the expression of *SSs* displayed a clearly consistent trend

Communicated by: Philippe Lashermes

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12042-019-09225-3) contains supplementary material, which is available to authorized users.

✉ Jisen Zhang
zjisen@fafu.edu.cn; zjisen@126.com

Panpan Ma
shuijing3916@qq.com

Yuan Yuan
yuanyuanfjnu@163.com

Qiaochu Shen
shenqiaochu1994@163.com

Qing Jiang
jq_0923@163.com

Xiuting Hua
xiutinghua@126.com

Qing Zhang
zhangqing970@126.com

Ray Ming
rming@life.uiuc.edu

[1] Center for Genomics and Biotechnology, Haixia Institute of Science and Technology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, College of Crop Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[2] College of Life Sciences, Fujian Normal University, Fuzhou 350108, China

[3] Guangxi Key Lab for Sugarcane Biology, Guangxi University, Nanning 530005, Guangxi, China

[4] Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

in line with normal distribution under the diurnal rhythms of *S. spontaneum*. Moreover, the expression pattern of *SSIIIa, SSIIb* and *GBSSII* displayed a clearly consistent trend in both *Saccharum* species and in maize, rice, which was in accordance with photosynthetic intensity across leaf gradients. This result suggested the functional constraints for the *SSs* gene family in *Gramineae*. Our results are valuable for further functional analysis of *SSs* genes and provided the foundation for carbohydrate metabolism in sugarcane.

Keywords Gene expression · Gene evolution · *Saccharum spontaneum* · Starch synthetases

Introduction

Starch is the most widespread and abundant storage carbohydrate together with sucrose in higher plants. There are two types of starch molecular (glucose polymers): amylopectin and amylose. Serving as an energy source, starch is synthesized transitorily in the plastids or chloroplasts in leaves, or in the specialized plastids in starch-storing tissues (Feike et al. 2016; Ingkasuwan et al. 2012; Preiss 2004; Smith and Stitt 2007). The transitory starch is accumulated in the day and degraded at night to supply leaf respiration and sucrose synthesis, which is vital for normal plant growth. If not used quickly, storage starch will maintain a longer-term storage, which is remobilized to support phases of growth, seed germination, seedling establishment, or contribute to stress resistance (Lloyd and Kossmann 2015; Pfister and Zeeman 2016; Skeffington et al. 2014; Streb and Zeeman 2012; Sweetlove et al. 1996).

Sugarcane (*Saccharum spp.*) belonging to the grass family Poaceae, a significant food and bioenergy crop of many countries, contributes to 80% of the sugar and 40% of ethanol production worldwide (Moore et al. 2013; Zhang et al. 2014). As a C4 crop, sugarcane is the most photosynthetic efficient field crop with an ideal energy yield system. Modern sugarcane cultivars are highly polyploid derived from an interspecific cross between *Saccharum officinarum,* having high sucrose and low fiber content, and *Saccharum spontaneum,* with low sucrose and high fiber content (D'Hont 2005; Grivet et al. 2003; Moore et al. 2013; Piperidis et al. 2010). Mature sugarcane stalk tissue with a very large apoplastic volume contains nearly 700 mM sucrose, which is among the highest recorded sucrose concentrations in plant tissue (Sulpice et al. 2009). This high concentration of sucrose, rather than starch (another major flux from photosynthetic carbon), in the cytosol and apoplast poses an interesting question. It is unknown in sugarcane whether starch synthesis and degradation, due to its unusually high sucrose content in stems differing from other Poaceae, acts on the regulation of sucrose accumulation and even carbon metabolism.

Starch synthetases (SSs), containing two kinds of enzyme families, the soluble SSs and the granule-bound starch synthetases (GBSSs), are the key players in starch biosynthesis. The major activity of starch synthetases is to elongate the non-reducing ends of glucose chains in starch synthesis, ADP-D-glucose + $(1,4\text{-}\alpha\text{-D-glucosyl})_{(n)} \rightarrow$ ADP + $(1,4\text{-}\alpha\text{-D-glucosyl})_{(n+1)}$ (Dian et al. 2005; Leterrier et al. 2008; Streb and Zeeman 2012; Zeeman et al. 2010). It is well accepted that SSs specifically synthesize amylopectin and GBSS specifically catalyze amylose biosynthesis. SS that transfers glucosyl residues from ADP-glucose to non-reducing ends of glucan chains by forming a α-1, 4-glucosidic linkage. In addition to the action of SS, GBSS can contribute to side chains of amylopectin to form long chains of amylopectin (Zeeman et al. 2010). Some researchers found SSII and SSIII have partly overlapping steps in the determination of amylopectin structure (Zhang et al. 2008). The SS is localized in plastid stroma, with at least four distinct classes of SSs in rice (SS1, SS2, SS3 and SS4), and each class is encoded by one or more genes (Hirose and Terao 2004). Two isoforms of GBSS (GBSSI, GBSSII) have been identified in several plants. GBSSI is largely located in plant storage organs, such as seed, roots or tubers (Kumar Behera and Sahoo 2009; Stitt and Zeeman 2012; Zhang et al. 2014), whereas GBSSII is in non-storage tissues, such as leaves and stems (Vrinten and Nakamura 2000).

Recently, the whole genome sequences of wild sugarcane species, *S. spontaneum*, have been published by our group following the genome information of several monocots such as rice, maize, and sorghum (Matsumoto et al. 2005; Paterson et al. 2009; Schnable et al. 2009). Together with integrated physical and genetic maps, these genome sequence databases provide unique perspectives for genome-wide investigations of gene families in sugarcane. Identifying and characterizing genes related to starch synthesis, such as *SSs*, would also provide the opportunity to pursue fundamental evolutionary and expression differences between monocots and eudicots (Cheng et al. 2012), and further pathway studies underlying correlation on carbohydrate metabolism. In this study, the identification and characterization of sugarcane *SSs* genes, with the polygenetic evolution and expression analysis, have aimed to perform a comprehensive study to clarify the impact of *SSs* genes in the regulatory mechanisms governing starch and sucrose accumulation in the leaves of sugarcane.

To identify the genetic variations in starch synthetases, the structure and evolutionary pressure of the *SSs* genes were investigated. Expression of the *SSs* genes was then analyzed to gain insight into the molecular mechanisms underlying starch synthesis. Finally, in order to understand the function

of starch synthetases in sugarcane, the differential expression analysis of *SSs* genes between sugarcane and close species was investigated.

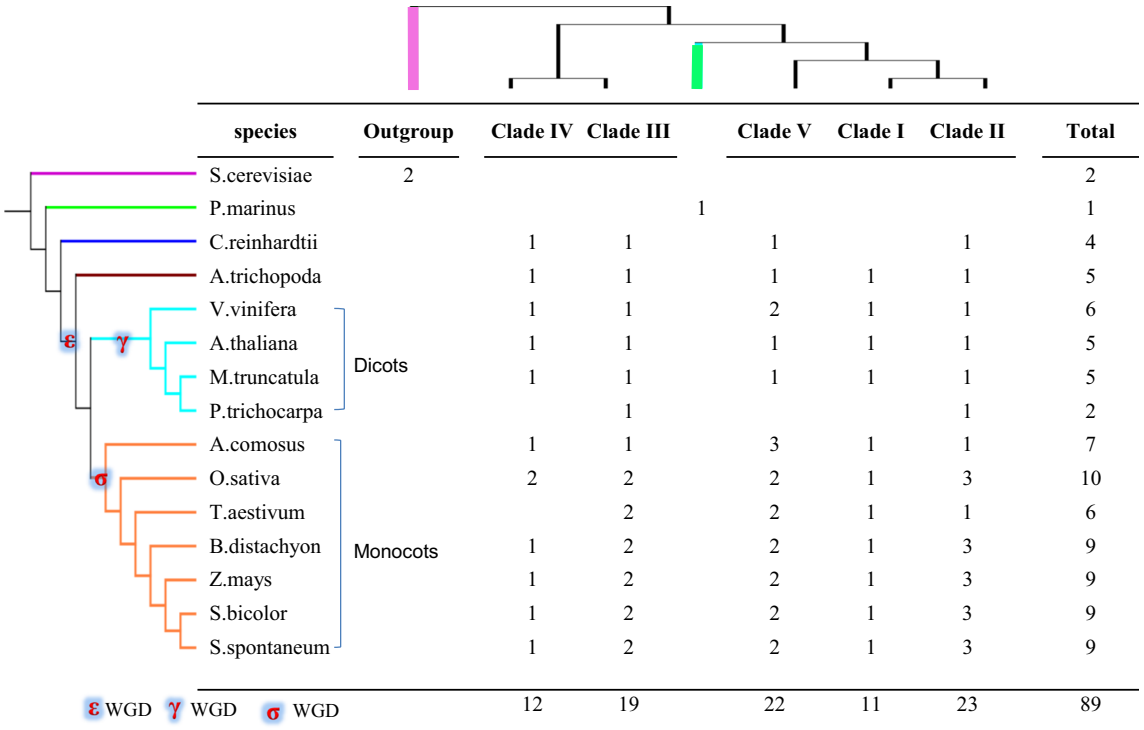
Results

Identification of *SSs* Genes in Sugarcane and Other Species

The sorghum sequences were used as query sequences by BLASTp to search the orthologs from the genome of AP85–441 (the haplotype of SES-208, *Saccharum spontaneum*) (Zhang et al. 2018). Genes of starch synthetases in *Saccharum* were then designated as *Saccharum SSs, SpSSs* in *S. spontaneum, SoSSs* in *S. officinarum* and *SrSSs* in *S. robustum*, for consistency with the sorghum SSs genes' name (*SbSSs*) (Nougué et al. 2014). The numbers of *SSs* genes found in all studied species were indicated combing with the phylogenetic relationships (Fig. 1). These genes containing the Glyco_transf_5 domain (GT5, PF08323) and Glycos_transf_1 domain (GT1, PF00534) were selected from the proteome datasets. Moreover, 89 *SSs* genes were confirmed with high homology, of which there were 9 in *S. spontaneum*, an average about 8~9 in 6 other monocots, an average about 5 in 4 dicots, 5 in *Amborella trichopoda*, 4 in *Chlamydomonas reinhardtii*, 1 in *Prochlorococcus marinus* and 2 in *Saccharomy cescerevisiae* (Fig. 1).

The 9 *SSs* genes of sugarcane shared an alignment similarity of above 92% to corresponding sorghum sequences. The range of alignment was from 92.55% to 97.14% with an average value of 94.55% (Table 1). In *SSs* genes, the numbers of subfamily members in different gene families differed from 1 to 3. Compared with the *SSs* gene families from sorghum, *SSs* genes in sugarcane showed a similar molecular weight ranging from 66.26–190.29 kDa and the isoelectric point from 4.95–6.78. The analyses of the protein sequences of *SpSS* genes harbored catalytic conserved GT5 and GT1 domains, except for *SpSSIIIa* which has a GT4 domains instead of GT1 and *SpSSIIIb* which has an additional GT4 domain. More details about other features, including protein domains and isoelectric point (pI) information, are presented in Table 1.

Total 24 allelic sequences for the nine *SpSSs* were identified in *S. spontaneum* genome (Table S1). The gene alleles of *SpSSs* generally presented high conservation (Fig. S2) for exon/intron numbers and exon sizes with slight variations. Six of the gene alleles, *SpSSIIa-h3*/*SpSSIIa-h1*, *SpSSIIa-h2*, *SpSSI-h1, SpSSIIIa-h1* and *SpGBSSII-h3* were observed to have size variation for their exon/intron. Because the high conservation among the alleles in the autopolyploid *S. spontaneum* genome, the gene alleles would not be used for further comparative genomics analysis.



species	Outgroup	Clade IV	Clade III	Clade V	Clade I	Clade II	Total
S.cerevisiae	2						2
P.marinus			1				1
C.reinhardtii		1	1	1		1	4
A.trichopoda		1	1	1	1	1	5
V.vinifera		1	1	2	1	1	6
A.thaliana		1	1	1	1	1	5
M.truncatula		1	1	1	1	1	5
P.trichocarpa			1			1	2
A.comosus		1	1	3	1	1	7
O.sativa		2	2	2	1	3	10
T.aestivum			2	2	1	1	6
B.distachyon		1	2	2	1	3	9
Z.mays		1	2	2	1	3	9
S.bicolor		1	2	2	1	3	9
S.spontaneum		1	2	2	1	3	9
		12	19	22	11	23	89

Fig. 1 Phylogenetic relationships of starch synthetases (*SSs*) for 4 dicots, 7 monocots, *A. trichopoda* and 3 outgroup species (*Odrade* et al., 2014). The number of SSs genes found in the genome of each species is indicated. The branches of dicots, monocots, *A. trichopoda* and outgroup species are marked in sky blue, orange, brown, blue, green and purple, respectively. The clusters of all *SSs* genes are presented on the top. WGD: Whole-genome duplication

Table 1 Sequence similarity and physicochemical properties of starch synthetases genes between sorghum and sugarcane. The identity value is the similarity of protein sequences in sugarcane aligned to corresponding orthologs in sorghum

	Sorghum						*S. spontaneum*					Identity
	Gene name	Gene ID	Size (aa)	Domains	pI	Mw (kDa)	Gene ID	Size (aa)	Domains	pI	Mw (kDa)	
Clade I	*SbSSI*	*Sobic.010G047700.1*	648	GT5,GT1, GT1–4	5.66	71.08	*Sspon.008A0018461*	660	GT5,GT1,GT1–4	5.86	70.54	96.45
Clade II	*SbSSIIa*	*Sobic.001G239500.1*	685	GT5,GT1, GT1–4	5.34	77.27	*Spon.001C0022820*	688	GT5,GT1, GT1–4	5.43	85.73	93.82
	SbSSIIb	*Sobic.004G238600.1*	705	GT5,GT1, GT1–4	6.04	76.25	*Sspon.004D0005621*	703	GT5,GT1,GT1–4	5.88	75.81	94.46
	SbSSIIc	*Sobic.010G093400.1*	771	GT5,GT1, GT1–4	6.13	83.75	*Sspon.008A0014201*	780	GT5,GT1,GT1–4	5.99	84.86	92.56
Clade III	*SbSSIIIa*	*Sobic.006G221000.1*	1192	CBM53 (3x), GT5,GT4	5.29	134.24	*Sspon.005A0003941*	1186	CBM53 (3x), GT5,GT4	5.25	134.19	94.8
	SbSSIIIb	*Sobic.007G068200.1*	1685	CBM53 (3x), GT5,GT4, GT1	5.01	190.02	*Sspon.006D0012320*	1684	CBM53 (3x), GT5,GT4,GT1	4.95	190.29	92.55
Clade IV	*SbSSIV*	*Sobic.009G209300.2*	910	GT5,GT1,GT4	6.05	103.11	*Sspon.007B0004030*	910	GT5,GT1,GT4	6.12	103.00	97.14
Clade V	*SbGBSSI*	*Sobic.010G022600.1*	609	GT5,GT1, GT1–4	6.37	66.07	*Sspon.008B0020751*	632	GT5,GT1, GT1–4	6.78	66.26	96.72
	SbGBSSII	*Sobic.002G116000.1*	608	GT5,GT1, GT1–4	6.23	66.51	*Sspon.002B0022811*	635	GT5,GT1, GT1–4	6.30	66.70	96.05

pI: isoelectric point Mw: molecular weight

Phylogenetic Analysis of *SSs* Genes Between *S. spontaneum* and Other Plants

To better characterize phylogenetic inference of *SSs* among all representative species, containing 6 monocots (*Ananas comosus, Brachypodium distachyon, Oryza sativa, Sorghum bicolor, Zea mays* and *Triticum aestivum*), 4 dicots (*Medicago truncatula, Populus trichocarpa*, *Vitis vinifera* and *Arabidopsis thaliana*), *A. trichopoda* and 3 outgroup species (*C. reinhardtii, P. marinus* and *S. cerevisiae*), we built phylogenetic trees with 89 *SSs* protein sequences.

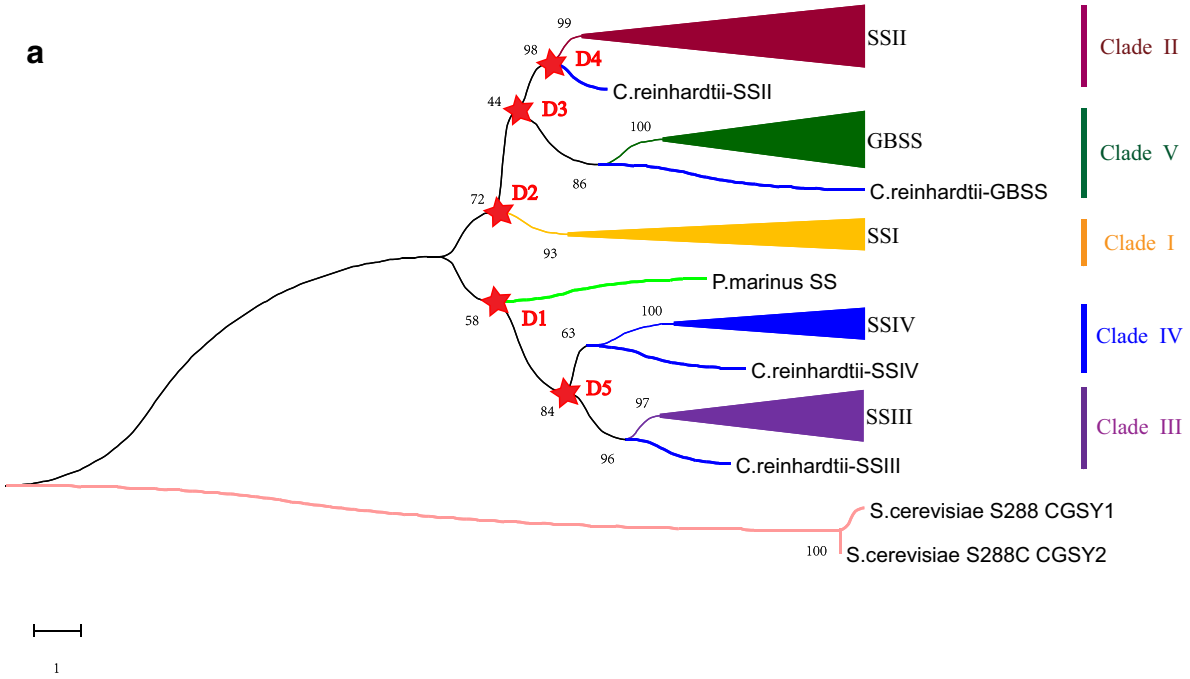
The phylogeny of the *SSs* indicated that prior to the angiosperm radiation five duplication events occurred (Fig. 2a) leading to the specialization of distinct functions. Among all five clades, it was duplications of D2 and D3 that led to three paralogous clades, clade II, clade I and clade V; a similarly duplication of D5 led to clades III and IV. All enzyme families differentiated to have special functions before the development of seed plants. In the grass family branch, *SSII* were further grouped into three subfamilies of *SSIIa, SSIIb and SSIIc, SSIII* into *SSIIIa and SSIIIb, GBSS* into *GBSSI* and *GBSSII* (Fig. 2b).

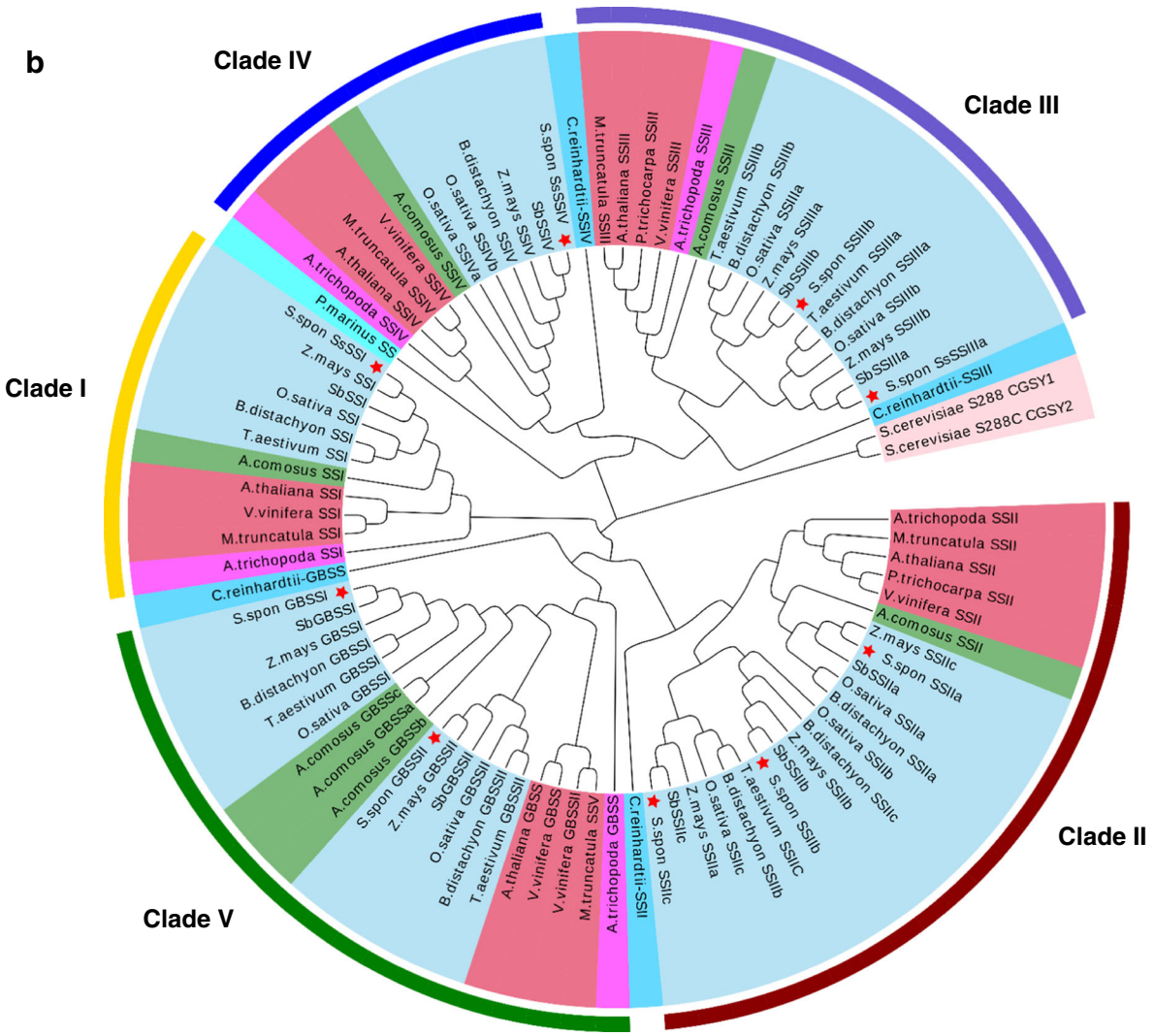
The Ka/Ks (nonsynonymous to synonymous substitution) ratios were used to investigate the possible direction and extent of selection. The Ks ratios were calculated by analyzing orthologous gene pairs of *SSs* in different clades among three species, maize, sorghum and *S. spontaneum* (Table 2), respectively. The gene duplication divergence time estimation based on Ks suggested that the *SSs* gene subfamilies were split about 355–389 Mya (Table 2), this being in the *Carboniferous* period, indicating that the *SSs* gene subfamilies are ancient gene families.

Comparative Analysis of *SSs* Gene Structure in *O. sativa, Z. mays, S. bicolor and S. spontaneum*

To further explore the evolutionary mechanisms, we compared the gene structure of *SpSSs* with their orthologs among three close species of monocot: *O. sativa, Z. mays* and *S. bicolor*. Based on their exon/intron characteristics, we classified the *SSs* into five clades and nine groups (Fig. 3), which was consistent to the classification above. The genes presented high conservation for exon/intron numbers and exon sizes

Fig. 2 **a** A schematic diagram showing the relationship of the starch synthetases (*SSs*) five clades based on the phylogenetic tree constructed by the ML method. The *SSs* clustered into five clades (I-V) marked on the right side. The five red stars represent five duplication events (D1-D5). **b** Phylogenetic tree of starch synthases in different plants exhibiting compartmental specialization by the ML method. The branches'label background of dicots, monocots, *A. trichopoda*, *A. comosus* and outgroup species (*C.reinhardtii, P.marinus, S.cerevisiae*) are marked in sky blue, crimsonorange, magentabrown, forest greenblue, deep sky bluebottle green, aquagreen and pinkpurple, respectively


a
SSII
C.reinhardtii-SSII
GBSS
C.reinhardtii-GBSS
SSI
P.marinus SS
SSIV
C.reinhardtii-SSIV
SSIII
C.reinhardtii-SSIII
S.cerevisiae S288 CGSY1
S.cerevisiae S288C CGSY2
D1
D2
D3
D4
D5
99
98
44
100
86
72
93
58
63
100
84
97
96
100
Clade II
Clade V
Clade I
Clade IV
Clade III
1


b
Clade IV
Clade III
Clade I
Clade II
Clade V
C.reinhardtii-SSIV
M.truncatula SSIV
A.thaliana SSIV
P.trichocarpa SSIV
V.vinifera SSIV
A.trichopoda SSIV
A.comosus SSIV
T.aestivum SSIV
B.distachyon SSIV
O.sativa SSIVb
Z.mays SSIV
SbSSIV
S.spon SsSSIV
P.marinus SS
S.spon SsSSI
Z.mays SSI
SbSSI
O.sativa SSI
B.distachyon SSI
T.aestivum SSI
A.comosus SSI
A.thaliana SSI
V.vinifera SSI
M.truncatula SSI
A.trichopoda SSI
C.reinhardtii-GBSS
S.spon GBSSI
SbGBSSI
Z.mays GBSSI
B.distachyon GBSSI
T.aestivum GBSSI
O.sativa GBSSI
A.comosus GBSSc
A.comosus GBSSa
A.comosus GBSSb
S.spon GBSSII
Z.mays GBSSII
SbGBSSII
O.sativa GBSSII
B.distachyon GBSSII
T.aestivum GBSSII
A.thaliana GBSS
V.vinifera GBSS
V.vinifera GBSSII
M.truncatula SSV
A.trichopoda GBSS
C.reinhardtii-SSII
S.spon SSIIc
SbSSIIc
Z.mays SSIIa
O.sativa SSIIc
B.distachyon SSIIc
T.aestivum SSIIC
S.spon SSIIb
SbSSIIb
Z.mays SSIIb
O.sativa SSIIb
B.distachyon SSIIa
O.sativa SSIIa
SbSSIIa
S.spon SSIIa
Z.mays SSIIc
A.comosus SSII
V.vinifera SSII
P.trichocarpa SSII
A.thaliana SSII
M.truncatula SSII
A.trichopoda SSII
A.trichopoda SSIII
M.truncatula SSIII
A.thaliana SSIII
P.trichocarpa SSIII
V.vinifera SSIII
A.comosus SSIII
T.aestivum SSIIIb
B.distachyon SSIIIb
O.sativa SSIIIb
Z.mays SSIIIb
SbSSIIIb
S.spon SSIIIb
T.aestivum SSIIIa
B.distachyon SSIIIa
O.sativa SSIIIa
Z.mays SSIIIa
SbSSIIIa
S.spon SsSSIIIa
C.reinhardtii-SSIII
S.cerevisiae S288 CGSY1
S.cerevisiae S288C CGSY2

in 4 species, except for *SSIIa*, whose orthologs showed major differences between inter-species in the presence and size of exons and introns.

In poaceae, the gene structure of *SpSSI, SpSSIIb SpSSIIc, SpGBSSI, SpGBSSII, SpSSIV* and *SpSSIIIa* were highly conserved. The structure of *SSIIa, SSIV* and *GBSSII* in *S. spontaneum* was closer to homologues in maize than in sorghum while other genes were closer to those in sorghum than in maize (Fig. 3). In *SSIIa*,there were some variation between 4 species, a loss of 1–2 exons in sorghum and the addition of second exons in rice. In *SSIIc*, *S. spontaneum*, sorghum and rice had much larger second exon, suggesting this exon may be a fusion which was split into the second and third exons in maize. Regarding intron size, *SSIIa* of *S. spontaneum* harbored larger sixth introns than their orthologs in sorghum, maize and rice, which is also the case in *SSIIIa*. Therefore, starch synthetases genes were highly conserved gene families.

To deduce the evolutionary force after the split of sorghum and *Saccharum*, the Ka/Ks ratios of all sugarcane genes were calculated by analyzing orthologous gene pairs of *SSs* between sorghum and *S. spontaneum* (Fig. S1), respectively. All genes of the starch synthesases had a Ka/Ks ratio of less than 0.5, with these genes identified as being significantly different from 0.5 ($p \leq 0.05$), indicating the evolution of *SSs* genes were under the influence of a strong purifying (stabilizing) selection.

The Strategy of Expression Analysis of *SSs* in *Saccharum*

The Different Tissues of Three Developmental Stages To investigate the *SSs* genes functional diversity between the three *Saccharum* species, we performed comparative transcriptome profiling based on the RNA-seq datasets at three developmental phases (seedling, pre-mature stage and mature stage) and in five different tissues including two leaves (mature leaf and leaf roll) and 3 stalks (mature, maturing and immature) (Fig. 4). In the 5 families analyzed, *SSIIc, SSIIIb* and *GBSSI* genes were expressed at very low or undetectable levels (FPKM<10) compared with other isoforms, indicating *SSIIc, SSIIIb* and *GBSSI* displayed low transcript levels in non-storage tissues. In contrast, *SSIIIa, GBSSII* had relatively high expression levels in all the examined tissues from the three species, indicating the three genes were the fundamental members in the non-storage tissues, leaf or stem of *Saccharum*. Meanwhile, we found the expression of *SSIIIa* and *GBSSII* had significant tissue differences between leaves (much higher) and stems (lower) in three developmental stages among three *Saccharum* species, particularly the expression of *GBSSII* in seedling and pre-mature stages. The expressions of *SpSSIIIa* and *SpGBSSII* were verified by qRT-PCR, the results showed that correlation coefficients between FPKM and qRT-PCR were $R^2 = 0.96$ and $R^2 = 0.70$ for *SsSSIIIa* and *SsGBSSII,* respectively (Fig. S3).

In general, the expression of genes had higher transcript levels in *S. spontenum* than in *S. officinarum* in all the examined tissues. Interestingly, at pre-mature stages both in leaf and stem tissue the expression levels of *GBSSII* showed dramatically higher values in *S. spontenum* than in *S. officinarum* and *S. robustum*. But at the mature stage GB*SSII* in different tissues showed lower expression in *S. officinarum* than in *S. spontenum* and *S. robustum*.

The Segments of Developmental Leaf To further explore the *SSs* genes functional divergence in the source tissues, we exploited the continuous developmental gradient of the leaf to describe the transcriptome of *SSs* in the three *Saccharum* species (Fig. 5). With reference to previous methods in our group, the *Saccharum* leaf was divided into four zones similar to the rice and maize leaf (Li et al. 2010a):a basal zone (base, sink tissue), a transitional zone (going through the sink-source transition), a maturing zone and a mature zone (active C_4 photosynthetic zones, fully differentiated), which was finally cut into 15 clips in one centimeter pieces. Similarly, with the expression level at different developmental stages, the genes of *SSIIIa, SSIIb* and *GBSSII* presented relatively high levels in all the examined genes among three *Saccharum* species, especially the *GBSSII* which had 2–3 times the expression levels of *SSIIIa* and *SSIIb*. In contrast, *SSIIc*, *SSIIIb* and *GBSSI* showed little and even no expression (Fig. 5), suggesting a very small role in vegetative tissue of *Saccharum*. Of the three highly expressed genes, the peak value of expression in 15 segments was in 9–12 clips differing from genes in line with normal distribution. Expression increased gradually from the tip to the peak point and increased gradually from peak point to base of the leaf in both *Saccharum* species. Importantly *S. robustum* expression of *SSIIIa* was highest followed by *S. sponteneum,* expression of *GBSSII* and then *S. officinarum* expression of *SSIIb.* Remarkably, *GBSS* and *SSIIb* had the highest expression levels in the 9 and 10 clips of maturing zones, whereas other genes in mature zones were among three *Saccharum* species.

The Diel Cycles To explore the expression patterns of the *SSs* during the diel cycles, we designed an experimental program for RNA-seq sampling from both *S. officianrum* and *S. spontanenum* at 2-h intervals over the first 24-h period, followed by 4-h intervals over the next 24-h. Consistently, the genes of *SSIIIa, SSIIb* (replaced by *SSIIa* in *S. officianrum*) and *GBSSII* in *S. spontanenum* presented relatively high levels over the two 24-h periods (Fig. 6a), whereas, *SSIIc, SSIIIb, SSIV* and *GBSSI* showed undetectable levels in both *Saccharum* species, further supporting the functional location of starch synthetase genes. Of the relative highly expressed genes, *SSIIIa, GBSSII, SSIIb* and *SSIIa* (Fig. 6b),

Table 2 Divergence time of starch synthetases gene families in five Clades in among 3 monocot species, *Z. mays, S. bicolor* and *S. spontaneum*

Clade	Median Ks	Gene pairs used	Divergence time (Mya)
Clade I/Clade II	4.60 ± 0.71	27	376.8
Clade I/Clade III	4.57 ± 0.71	18	374.8
Clade I/Clade IV	4.56 ± 0.03	9	373.8
Clade I/Clade V	4.51 ± 0.89	18	369.8
Clade II/Clade III	4.65 ± 0.25	54	381.3
Clade II/Clade IV	4.64 ± 0.22	27	380.7
Clade II/Clade V	4.49 ± 0.26	54	368.0
Clade III/Clade IV	4.74 ± 0.04	18	388.4
Clade III/Clade V	4.49 ± 0.1	34	367.8
Clade IV/Clade V	4.42 ± 0.03	18	362.1

the transcriptome expression profiles displayed normal distribution in both *Saccharum* species. Remarkably, the peak points of the expression profile SSs genes differed between two *Saccharum* species, *S. spontanenum* was at 10 o'clock and *S. officianrum* was at 12 o'clock. Meanwhile the expression level of *GBSSII* in *S. spontanenum* was much higher than in *S. officianrum*. All results were verified by datasets sampling at 4-h intervals over an additional 24-h.

Discussion

In this study, we first identified 9 genes from 5 starch synthase subfamilies in *S. spontaneum*. Comparative genomics suggested that the *SpSSs* genes underwent divergence or functional specialization before the emergence of the seed plant which evolved conservatively with strong purifying (stabilizing) selection. The gene expressional analysis showed that *GBSSII, SSIIIa* and *SSIIb* were dominantly expressed in *Saccharum,* presenting high expression in the source leaf with a positive correction to the photosynthetic intensity, and response to the diel cycles. Similar expression patterns were observed in the continuous developmental gradient of the leaf for the *SSs* in maize and rice, indicating the functional constraint for the *SSs* families in *Gramineae*. These results present ideas and detailed information for further understand the special characteristics in *Saccharum* with regard to low starch and high sucrose.

Evolutionary Conservation and Divergence of *SSs* Genes in *Saccharum*

In this study, the starch synthase families were phylogenetically clustered into five clades (I-V) (Fig. 2), which were consistent with previous research (Nougué et al. 2014; Pan et al. 2011). The occurrence of whole-genome duplications (WGDs) had been considered a driving force for the evolution of organisms. It has been well characterized that the WGDs (ε) in 150 to 270 Mya were the most recent evolutionary events and generated successful lineages of land plants, the recent common ancestor of extant angiosperms (Edger and Pires 2009; Jiao et al. 2014; Jiao et al. 2011). Therefore, we took *A. trichopoda* into account, a typical angiosperm, as the reference of phylogenetic analysis. In our study, the divergence of all five clades occurred before the emerging of *A. trichopoda*, indicating that the genes of *SSs* all experienced duplication, which is consistent with previous studies (Batra et al. 2017; Cheng et al. 2012; Dian et al. 2005). Moreover, the divergence times of the five clades were estimated in this study (Table 2), which was accompanied with the emergence of the *Gymnospermae* about 355–389 Mya, indicating that the last common ancestors (LCAs) of five clades were split at the similar period. The expansions of subfamilies of *SSs* were very ancient events, which may have resulted in the functional divergence for these subfamilies.

For each subfamily in the clades, there is a clear separation between the monocots and eudicot lineage (Fig. 2b), that is to say additional duplications occurred subsequent to the angiosperm radiation. Therefore we could speculated that the *SSs* genes must have undergone divergence or functional specialization before monocots and eudicots split (165Mya), which is in keeping with *Cheng*'s study (Cheng et al. 2012). In this study, we observed that Clade III, Clade V and Clade II from the grass family branch contained more gene numbers in monocots than in dicots. Clade I and Clade IV contained one subclade or two subclades for dicot and monocot plant species, respectively (Fig. 2b). As for the reason, we assumed that the dicot plant underwent gene loss after the first gene duplication event and before the divergence of dicots and monocots. In addition, Clade II had a recent gene duplication after the splits of monocots and dicots, generating the subclade containing *SpSSIIb* and *SpSSIIc*. In Clade III, gene duplications were observed to have occurred after the splits of the *Gramineae* and pineapple (Fig. 2b), indicating the LCA of *SSIIIa* and *SSIIIb* originated from the pan-grass ρ-WGD as pineapple had no ρ-WGD (Ming et al. 2015b). It is interesting that the *SS*

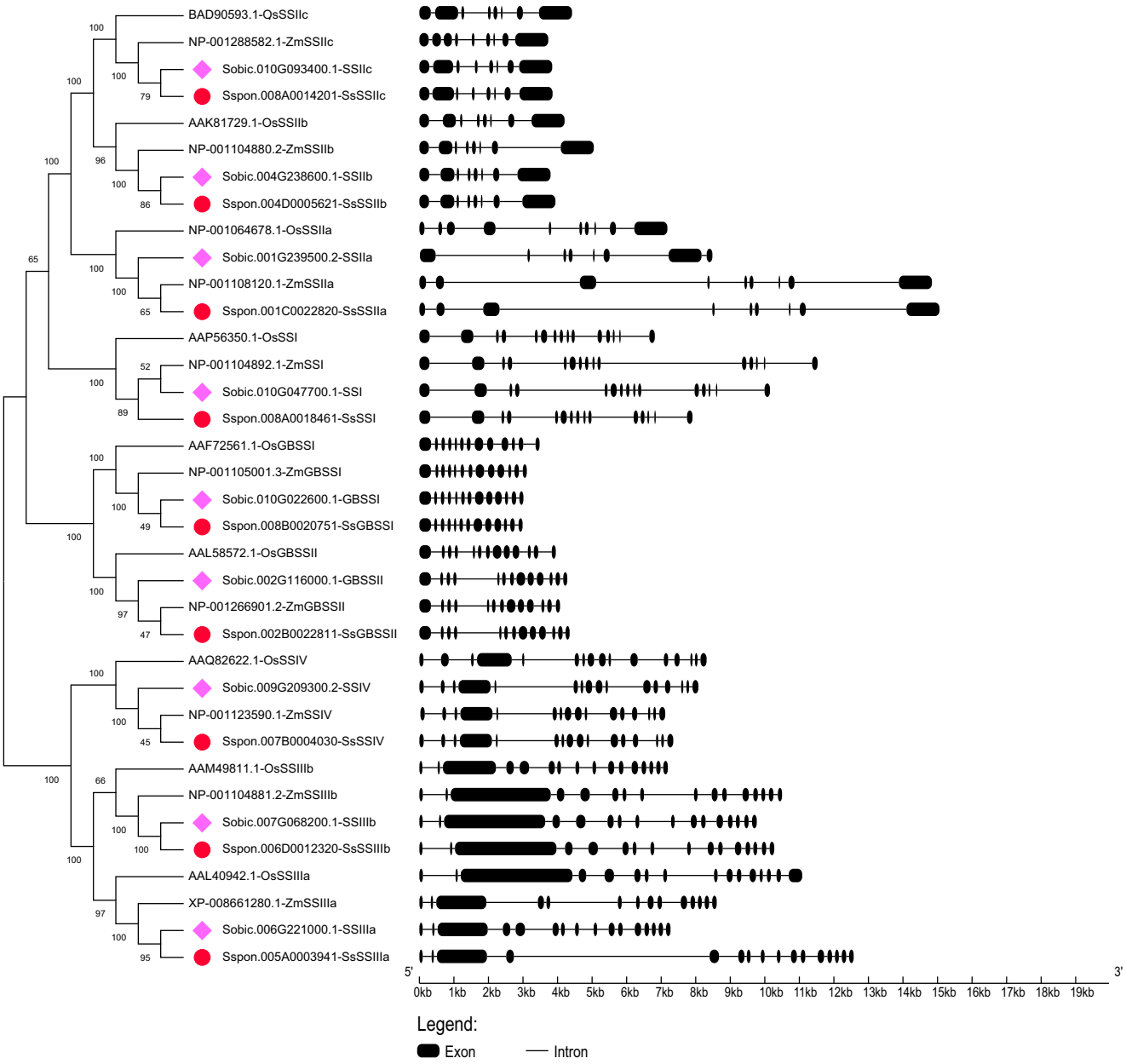


Fig. 3 The comparison of gene structure of starch synthetases in 4 monocot species, *O.sativa, Z. mays, S. bicolor* and *S. spontaneum*. Gene structure starts from translation start sites to stop sites. Diagram is drawn to scale. Exons are represented by boxes and introns as lines

gene families likely had no recent duplication events after the ρ-WGD in the examined monocot species, suggesting the functional constraint for the *SSs* families. Based on this deduction, the SSs gene families are thought to have no functional redundancy in monocot species including *S. spontaneum,* which is consistent with Ka/Ks analysis indicating that the SSs genes are under strong purifying (stabilizing) selection.

The phylogeny and the divergence estimation provided the evidence for the evolution of *SpSS*. The *SS* is a very ancient gene family, in which, *SpSSIIIa* and *SpSSIIIb* originated from the ρ-WGD, *SpSSIIb* and *SpSSIIc* originated from gene duplication after the split of monocot and dicot. *SpGBSSI* and *SpGBSSII* in clade V and *SpSSIIa* in Clade II were retained from the ε-WGD, and the two genes, *SpSSI* and *SpSSIV,* were retain from a very ancient gene event about 355–389 Mya.

Based on the analysis of gene structure, the characteristics of exon/introns in *S. spontaneum* of *SSI, SSIIb, SSIIc, GBSSII, GBSSI, SSIV, SSIIIa, SSIIIb* genes were conserved in five monocots. These results further verified that starch biosynthesis emerged from the metabolism of cytosolic glycogen and a pattern of neo-functionalization occurred in the ancestral paralog of monocots (Ball et al. 2011; Deschamps et al. 2008; Zeeman et al. 2010).

	leaf	stem	leaf-roll	m-leaf	stem3	stem6/9/8	stem9/15/13	leaf-roll	m-leaf	stem3/3	stem6/9/8	stem9/15/13
SpSSI	8.3	7.0	7.9	11.7	6.6	4.6	9.4	60.6	59.6	8.4	12.8	17.3
SoSSI	10.4	3.2	5.9	9.5	1.2	4.8	5.3	20.9	13.8	2.7	7.2	12.4
SrSSI	4.7	3.3	7.5	5.8	3.3	2.1	5.8	59.9	78.4	15.6	23.4	23.9
SpSSIIa	33.3	11.7	11.3	9.5	7.8	3.4	3.6	18.8	17.6	5.9	4.9	3.9
SoSSIIa	40.2	24.1	20.3	8.4	5.5	5.4	5.8	30.3	17.4	10.4	13.7	10.4
SrSSIIa	25.7	13.0	0.9	0.2	0.8	1.8	1.3	26.6	15.8	6.2	7.6	8.0
SpSSIIb	57.4	21.8	37.3	34.8	14.8	11.8	13.1	14.6	15.5	11.8	11.5	8.0
SoSSIIb	36.3	13.5	29.4	25.4	2.8	2.4	3.2	11.2	6.8	3.4	4.1	2.9
SrSSIIb	33.0	15.2	3.7	1.7	2.5	2.5	3.0	23.8	11.1	11.4	5.8	4.4
SpSSIIc	0.5	1.8	0.8	1.6	2.7	3.9	9.3	0.6	0.3	1.7	3.0	6.7
SoSSIIc	0.5	1.2	0.3	0.1	0.6	0.4	0.8	0.2	0.1	0.1	4.0	7.9
SrSSIIc	0.4	1.1	0.3	0.9	1.3	5.5	8.3	2.2	2.7	2.9	9.1	8.9
SpSSIIIa	51.1	33.7	52.2	39.2	19.0	38.5	38.9	28.5	24.3	15.6	14.0	18.7
SoSSIIIa	64.8	39.9	40.5	17.4	8.5	12.4	15.7	38.8	24.8	12.7	16.7	16.3
SrSSIIIa	51.2	29.9	34.8	19.2	0.5	1.1	4.5	48.2	40.0	15.9	25.7	25.8
SpSSIIIb	0.0	0.4	0.2	0.0	1.1	0.1	0.3	0.2	0.4	0.7	0.2	0.5
SoSSIIIb	0.4	0.8	0.5	0.2	1.3	1.7	1.2	0.8	0.6	2.3	1.5	2.8
SrSSIIIb	0.2	0.8	0.1	0.0	0.1	0.4	0.6	0.6	0.5	1.5	1.8	1.4
SpSSIV	8.6	2.8	3.8	3.6	0.5	1.7	1.9	5.2	6.0	1.2	2.2	1.9
SoSSIV	4.9	4.5	4.3	2.5	0.8	1.7	2.6	4.5	3.5	1.1	1.4	2.1
SrSSIV	4.8	2.5	1.2	2.1	0.2	0.6	2.8	5.0	5.4	0.8	2.6	2.3
SpGBSSI	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SoGBSSI	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SrGBSSI	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SpGBSSII	332.3	77.5	507.6	343.5	177.4	378.0	480.9	140.9	108.5	89.5	114.8	159.1
SoGBSSII	343.9	47.5	230.1	189.9	5.9	3.3	5.9	75.0	31.1	17.6	51.6	16.9
SrGBSSII	100.2	36.7	79.6	45.6	4.9	4.9	9.6	160.6	130.9	49.3	80.5	74.5

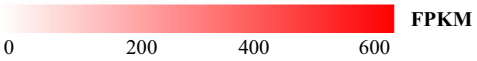


Fig. 4 The expression pattern of starch synthetases genes based on FPKM in different tissues of three developmental stages in three *Saccharum* species

Gene Expression During the Diel Cycles of *SSs* Genes in *Saccharum*

The analysis of transcriptome expression profiles could provide the most direct data to enable a thorough inquiry into gene function to be undertaken. Several authors found that the *SSII* and *SSIII* genes had a dominant function in the synthesis of amylopectin and also agreed with the role of *SSI* enzyme in the synthesis of small chains, while the *SSIV* and *GBSS* have an effect on the number of starch granules (Valdez et al. 2008). In our study, the expression levels of *Saccharum SSs* were extracted from three sugarcane transcriptome databases. The *GBSSII, SSIIIa, SSIIb* genes displayed relatively higher gene expression levels among all of the *SS* genes in three sugarcane species, suggesting the *GBSS* is the key isoform involved in the synthesis of starch granules and the role of *SSIIIa* and *SSIIb* is in the synthesis of soluble starch.

Based on the transcriptome database in different tissues at three growth stages, there were significant differences in gene expression between source (leaf) and sink tissues (stem). *SSIIIa* and *GBSSII* had relatively high expression levels, while, *SSIIc, SSIIIb* and *GBSSI* displayed low transcript levels in non-storage tissues, which is in agreement with previous studies (Stitt and Zeeman 2012; Zhang et al. 2014). The variations in expression in tissues could be explained by the general trend that plant leaves accumulate more carbohydrates especially temporary starch stores and then transport it to stem sink organs for plant growth (Moore et al. 2013). Especially at seedling stage, the expression of *GBSSII* in the leaf was about 3–8 times that in the stem; precisely 4 times in *S. spontenum,* 3 times in *S. robustum*, and unexpectedly 8 times in *S. officinarum.* We further measured the starch content of leaf and stem in sorghum and three *Saccharum* species. The content of starch in the leaf was higher than that in the stem, whereas the expression of three major genes (*SpGBSSII, SpSSIIIa, SpSSIIb*) in leaf was significantly higher than in stem (Fig. S4). Furthermore, the difference of starch content between leaf and stem is obvious just in *S. officinarum,* and starch content in *S. officinarum* is apparently higher than in *S. spontenum* and *S. robustum.* We suspected that in seedlings the synthesis of starch is mainly used in plant growth and the high expression of *SSs* lead to the accumulation of starch in *S. officinarum.*

Leaf	basal zone			transitional zone			maturing zone				mature zone				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gene															
SpSSI	18.01	21.98	13.42	5.53	3.30	2.59	3.32	3.62	10.68	7.50	10.68	14.18	16.85	17.67	17.66
SoSSI	5.13	9.30	6.91	3.62	1.81	1.81	2.29	4.39	6.28	6.92	7.29	8.41	7.75	11.91	12.60
SrSSI	14.66	17.46	11.21	6.05	5.87	4.91	3.62	2.91	3.54	5.27	5.29	6.98	7.85	8.42	10.28
SpSSIIa	16.54	12.88	10.83	11.62	14.28	19.35	24.67	27.98	41.29	41.08	41.29	41.81	39.54	32.83	26.38
SoSSIIa	24.07	25.59	25.90	18.94	20.57	27.32	34.22	36.43	34.94	28.33	25.60	25.15	25.71	25.06	29.89
SrSSIIa	24.51	21.37	17.93	11.68	12.04	12.61	14.99	17.96	20.05	20.42	21.37	20.70	22.36	21.35	22.41
SpSSIIb	17.42	16.03	15.09	16.82	22.55	30.54	37.24	40.69	69.62	67.48	**69.62**	69.13	62.25	56.93	46.37
SoSSIIb	16.94	19.42	19.51	18.98	30.39	44.51	70.03	82.64	**90.46**	87.49	81.49	80.67	66.76	60.07	42.23
SrSSIIb	19.21	15.92	14.95	15.40	18.96	22.80	29.64	40.18	45.39	46.68	55.86	54.93	**57.72**	48.48	43.08
SpSSIIc	0.83	7.89	7.72	8.55	10.95	14.83	18.38	19.91	34.38	32.72	34.38	33.65	29.90	28.00	22.94
SoSSIIc	0.05	0.05	0.05	0.57	0.50	0.63	0.37	0.29	0.17	0.13	0.04	0.01	0.03	0.07	0.02
SrSSIIc	2.21	1.70	1.31	1.06	0.99	0.92	0.78	0.70	0.38	0.32	0.39	0.29	0.37	0.17	0.19
SpSSIIIa	23.22	20.78	15.84	17.81	24.37	31.72	45.01	53.46	80.20	72.08	80.20	**86.95**	81.45	70.57	57.83
SoSSIIIa	22.69	23.19	22.47	26.32	35.98	45.31	55.10	61.63	76.32	71.91	71.54	**78.94**	54.28	65.08	53.87
SrSSIIIa	30.96	29.12	22.75	26.92	32.48	40.85	59.84	74.88	83.93	91.46	92.82	**96.54**	91.89	76.27	75.13
SpSSIIIb	1.13	0.78	0.27	0.15	0.14	0.12	0.09	0.06	0.06	0.08	0.06	0.05	0.06	0.08	0.12
SoSSIIIb	0.19	0.07	0.05	0.05	0.03	0.03	0.04	0.03	0.02	0.02	0.01	0.01	0.01	0.02	0.02
SrSSIIIb	0.47	0.42	0.19	0.12	0.05	0.07	0.07	0.07	0.02	0.07	0.05	0.07	0.06	0.01	0.03
SpSSIV	1.10	1.56	2.33	3.09	4.76	5.49	6.56	6.70	8.90	8.12	8.90	9.10	9.03	8.11	8.02
SoSSIV	0.76	1.63	2.70	3.20	3.33	3.21	3.13	3.21	3.67	3.33	4.32	4.68	4.14	4.57	4.05
SrSSIV	1.03	1.85	2.48	2.61	2.84	2.70	3.59	2.82	2.62	3.18	2.75	2.92	2.99	2.93	2.96
SpGBSSI	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.07	0.16
SoGBSSI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SrGBSSI	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.02
SpGBSSII	56.41	40.16	32.43	34.23	56.58	87.11	126.56	175.49	287.89	274.20	287.89	**290.52**	266.76	195.82	149.03
SoGBSSII	6.82	2.91	3.95	5.78	18.26	55.66	126.30	187.02	239.88	**268.73**	216.26	187.53	148.61	150.15	148.69
SrGBSSII	36.61	26.15	21.48	21.76	29.75	49.37	84.07	120.34	**161.60**	159.59	165.78	158.06	150.09	138.37	116.01

FPKM
0 150 300

Fig. 5 The expression pattern of starch synthetases genes based on FPKM at developmental leaf in three *Saccharum* species

Compared with *S. officinarum* and *S. robustum*, in *S. spontenum GBSSII* displayed very high expression levels in the pre-mature stages which is the period of sugar accumulation. The results for sugar accumulation have a reverse relationship with sucrose levels in the three species. The sucrose content in stems from *S. officinarum* were much higher than stems from *S. robustum,* and there was little, even almost no sucrose content in the stems from *S. spontenum* (Bonnett 2013). Therefore, the variations have a clear trend of gene expression indicating that there is a negative correlation between expression levels of the starch synthetase and sucrose content.

Based on the expression of *SSs* as a continuous developmental gradient in the leaf, the transcriptome expression profiles displayed a clearly consistent trend among three *Saccharum* species in accordance with photosynthetic intensity, which is similar to the expression pattern of *SSs* in maize and rice leaf (Fig. 7) from the published study (Wang et al. 2014), indicating the functional constraints for the *SSs* gene family in Gramineae.

Based on the expression of *SSs* during the diel cycles in *Saccharum*, the transcriptome expression profiles of four highly expressed genes (Fig. 5b) displayed a clearly consistent trend in line with a normal distribution, indicating that diurnal rhythms were involved in regulating *SSs* expression. As the previous study found (Bläsing and Stitt 2005; Muller et al. 2014; Seaton et al. 2014), the circadian clock participates in the regulation of carbon metabolism and the dynamic adjustment of starch turnover in response to stress in several plant species through unknown mechanisms in a changing environmental conditions.

Materials and Methods

Plant Materials

Three wild species of *Saccharum*, SES208 (*S. spontaneum*, Ss, 2n = 8× =64), LA-Purple (*S. officinarum*, So, 2n = 8× = 80), and Molokai6081 (*S. robustum*, Sr, 2n = 8× = 80) were used in the study (Ming et al. 2015a). Plants were grown in the field on the campus of Fujian Agricultural and Forestry University (Fuzhou, China, 119°16′48″E, 26°4′48”N) in the February of 2015. Tissue samples from leaf roll, leaf, top internode (i.e., internode number 3), maturing internode (i.e., internode number 9 for LA Purple and Molokai6081, internode number 6 for SES208) and mature internode (i.e., internode number 15 for LA Purple and

a

Gene\ time(h)	First day												Second day						
	6:00	8:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	0:00	2:00	4:00	6:00	10:00	14:00	18:00	22:00	2:00	6:00
SpSSI	6.27	15.00	19.56	22.03	**39.59**	37.71	30.39	19.11	9.08	6.76	7.48	7.11	4.43	21.68	**37.03**	27.02	10.69	4.76	6.22
SoSSI	16.60	17.83	28.16	53.53	**61.87**	44.91	32.41	31.91	22.71	21.79	24.65	23.10	34.14	47.13	**78.33**	46.10	55.31	44.34	33.15
SpSSIIa	26.61	28.24	28.98	11.79	4.38	1.08	0.53	0.50	0.75	1.33	4.25	23.92	31.01	21.05	5.15	1.04	1.02	9.75	33.31
SoSSIIa	31.81	57.31	**82.66**	23.18	8.84	4.98	3.66	3.35	3.05	2.95	3.25	3.35	4.51	**60.37**	30.80	5.81	4.95	5.40	29.42
SpSSIIb	52.45	**61.66**	58.69	20.76	6.77	0.30	0.12	0.09	0.19	0.78	4.88	40.73	37.51	**52.10**	7.79	0.19	0.23	8.03	39.46
SoSSIIb	8.69	17.84	20.17	25.26	12.26	5.62	3.59	2.48	1.51	1.80	2.11	2.59	2.78	14.98	24.35	5.28	4.25	3.60	6.75
SpSSIIc	0.64	0.94	1.03	1.73	1.67	1.50	1.10	1.19	1.90	1.38	1.00	0.67	0.66	0.93	1.33	1.31	1.31	0.73	0.78
SoSSIIc	0.35	0.64	0.51	0.40	0.20	0.28	0.35	0.56	0.35	0.25	0.66	0.39	0.82	0.56	0.22	0.57	0.52	0.38	0.43
SpSSIIIa	28.34	47.15	**88.35**	53.32	28.43	10.49	5.73	3.70	3.63	5.36	12.21	26.69	25.00	**69.72**	37.47	7.64	4.23	13.54	28.66
SoSSIIIa	47.16	52.87	57.80	61.59	58.15	**69.13**	48.57	46.29	41.16	43.51	48.58	44.67	54.30	59.43	**85.82**	47.38	59.84	63.51	54.78
SpSSIIIb	0.00	0.01	0.01	0.03	0.01	0.02	0.03	0.01	0.01	0.02	0.04	0.01	0.02	0.00	0.03	0.01	0.01	0.01	0.00
SoSSIIIb	1.73	1.46	1.95	1.30	1.13	1.25	1.23	0.95	1.07	1.01	0.83	0.79	1.39	0.98	0.78	0.93	1.13	0.71	1.33
SpSSIV	8.90	6.43	6.92	8.41	6.11	6.59	7.26	9.91	9.68	11.63	13.31	8.79	10.96	6.70	7.09	7.34	9.99	11.82	7.81
SoSSIV	11.53	9.17	8.26	6.12	5.74	7.61	7.23	7.22	9.71	8.82	8.47	7.75	7.68	7.62	5.90	6.50	7.83	7.83	7.18
SpGBSSI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
SoGBSSI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SpGBSSII	107.90	411.95	**586.35**	229.28	113.70	14.39	8.60	2.16	1.25	2.51	19.72	155.51	69.33	**650.00**	81.37	1.63	0.79	20.14	73.56
SoGBSSII	31.23	74.23	106.85	**312.56**	306.65	167.14	117.04	99.59	46.42	59.31	47.89	49.67	79.42	154.15	**327.22**	110.90	103.26	110.36	51.78

FPKM

0 350 700

b

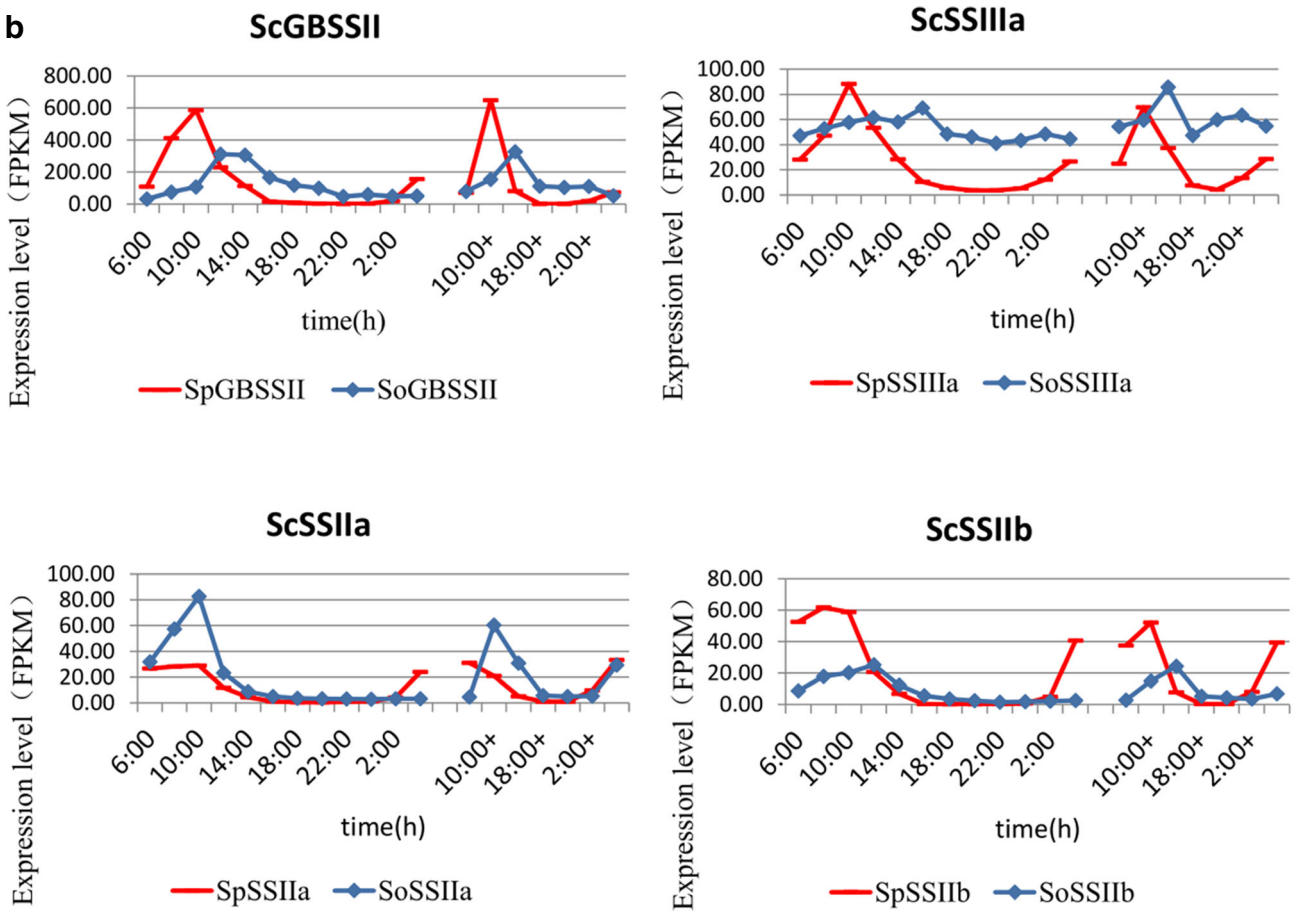


Fig. 6 **a** The expression pattern of starch synthetases genes based on FPKM during the diel cycles in *S. spontaneum* and *S. officinarum*. **b** The expression pattern of four starch synthetases genes (*GBSSII, SSIIIa, SSIIa, SSIIb*) with high expression based on FPKM during the diel cycles

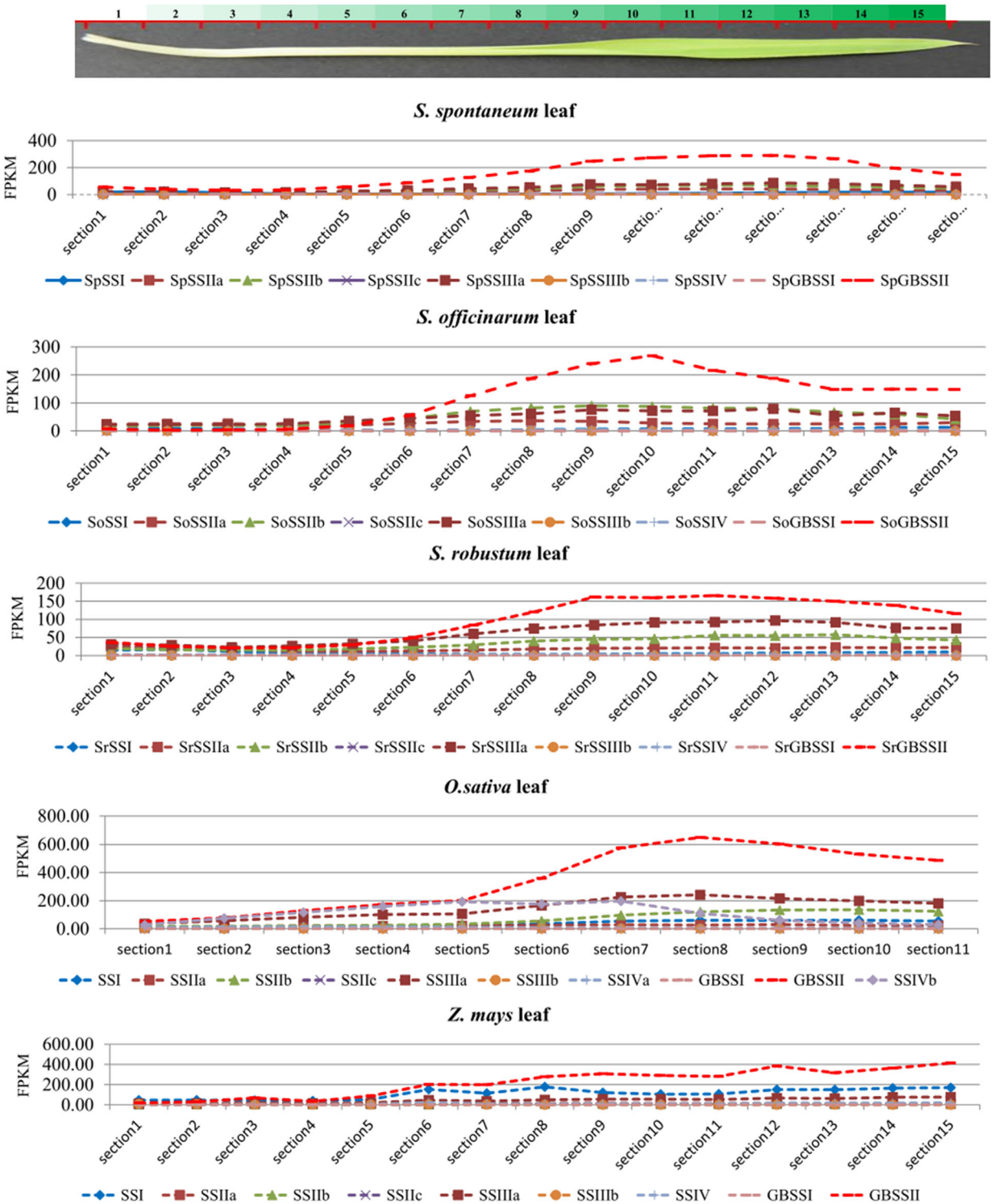


Fig. 7 The comparison of expression patterns of starch synthetases genes in developing leaves of maize, rice, *S. spontaneum, S. officinarum* and *S. robustum*

Molokai6081, internode number 9 for SES208) were collected from premature 7-month-old sugarcane plants and 11-month-old mature sugarcane plants from different branches of the same individuals (as replicates). Internodes were numbered from the top, as previously described (Tao et al. 2017), and the corresponding internode number for the different *Saccharum* species, due to the variation in number of stems, was also established according to the previously described approach (Wang et al. 2013). Stem and leaf tissues from the seedlings of the three sugarcane species were

collected from 35-day-old plants. The developing leaves of SES-208 were 11-day-old second leaves, leaves of LA-purple were 15-day-old second leaves and Molokai6081 were 12-day-old second leaves. Leaves for diurnal sampling were collected from mature stage *S. spontaneum* and *S. officinarum* every 2 h between 6 a.m. and 4 a.m. of the next day in March, 2017. Following on with the second cycle for obtaining samples, the leaves were collected every 4 h between 6 a.m. the next day and 4 a.m. on the third day. Harvested tissue was immediately frozen in liquid nitrogen and stored at −80 °C prior to RNA isolation.

Identification of Genes Encoding *SSs* in Monocots and Dicots

Protein sequences of *SSs* genes in *Arabidopsis thaliana, Oryza sativa, Sorghum bicolor, Zea mays, Triticum aestivum, Populus trichocarpa* and *Chlamydomonas reinhardtii, Prochlorococcus marinus, Saccharomy cescerevisiae* (three outgroup) from previous studies (Campbell et al. 2016; Nougué et al. 2014; Schwarte et al. 2015; Yan et al. 2015), were downloaded from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) database. The sequences of *Arabidopsis thaliana* and *Sorghum bicolor SSs* genes were then used to search *SSs* from the Phytozome database (version10.0; https://phytozome.jgi.doe.gov/pz/portal.html) for *Brachypodium distachyon, Ananas comosus, Medicago truncatula, Amborella trichopoda* and *Vitis vinifera*.

To confirm these homologous proteins, the hidden Markov model was used to perform searches on the Pfam database (http://pfam.xfam.org/search) based HMMER program (Finn et al. 2014). We obtained sequences of *Chlamydomonas reinhardtii* and *Synechoccus sp. CC9311* as outgroups from the Biocyc database (https://biocyc.org/gene-search.shtml). These sequences containing the catalytic domains (Gly5 and Gly1) with matches achieving similarity scores of 80.0, probability scores >50.0 and e-value <10–4, were collected.

Identification and Sequence Analysis of SSs Genes in Sugarcane

The orthologs of sorghum in sugarcane were initially identified based on protein sequences from BLASTp in the Linux system (https://blast.ncbi.nlm.nih.gov/Blast.cgi) followed by manually check. The phylogenetic trees were constructed with Muscle alignment using default parameters in MEGA-X program. Then the genes in sugarcane, which shared high sequence similarity and were phylogenetically distributed together, were selected. Protein sequences were further verified by searching the Conserved Domain Database in NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the PFAM database (http://pfam.xfam.org/search). The *SSs* genes of sugarcane possessing two conserved catalytic domains (Gly5 and Gly1) were considered to be genuine *SSs* genes.

Phylogenetic Analysis of *SSs* Genes in All Studied Species

Based on the alignment of protein sequences, the phylogenetic tree of *SSs* gene families was constructed using the maximum likelihood (ML) method. The construction of the ML tree was carried out using MEGA-X with the 100 nonparametric boostrap replicates, γ-distribution, and WAG amino acid substitution model (Guindon et al. 2003). The reliability of internal branches of the tree was valued by the bootstrapping of 1000 replicates. Subsequently, non-synonymous (Ka) and synonymous (Ks) substitution ratios of the pairs of orthologous genes from sorghum and *S. spontaneum* were calculated in the Linux system using the methods in website (https://github.com/tangerzhang/FAFU-cgb/tree/master/easy_KaKs). *Fisher*'s exact test for small samples was applied to justify the validity of Ka and Ks calculated by the method (Wang et al. 2009). The Ks values of pairs of different clades among maize, sorghum and *S. spontaneum* were calculated to determine the divergence time (T) by T = Ks/ (2*6.1*10^-9) *10^-6 Mya (Hu et al. 2018).

Diversity of Sequence, Gene Structure and Physicochemical Properties

Multiple sequence alignments were performed using MUSCLE with default parameters for *SSs* proteins. The exon-intron structures for the *SSs* genes were graphed using the online tool Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/)(Hu et al. 2015). ExPASy (http://web.expasy.org/compute_pi/) was used to predict the isoelectric point and relative molecular mass of SSs proteins.

Expression Profiles Analysis of *SSs* Genes in *Saccharum*

For the investigation of different developmental stages, three *Saccharum* species, LA-Purple (*S. officinarum*), SES-208 (*S. spontaneum)* and Molokai6081 (*S. robustum*) were used. The experiment was performed as previously described (Zhang et al. 2016). For the leaf gradient experiment *Saccharum* plants were grown with lamps at light intensity of 350 μmol/m^2/s, 14:10 l/D, 30 °C L/22 °C D and 60% relative humidity. Tissue was collected after planting 3 h into the L period as detailed by Li et al.' (Li et al. 2010b). The second leaves were cut into 15 segments per 1 cm from 11-day-old SES-208, from 15-day-old LA-Purple and from 13-day-old Molokai6081. The data of leaf transcriptomes in maize and rice were collected from the GEO database in

NCBI (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54274).

For the diel time course experiment, the mature leaf sample of *S. spontaneum* was collected from one plant as one replicate, and three biological replicates were performed for RNA-seq library construction. The collection of *S. officinarum* leaves was also consistent with the above approach. The previously described approach was detailed by Ming et al. (Ming et al. 2015a).

The methods for cDNA preparation were performed according to the manufacturer's protocol (TruSeq RNA, Illumina). The normalization of sequencing data were performed in accordance with our previous work (Zhang et al. 2016). Raw data was aligned to reference gene models using the method combining HISAT and cufflinks, and then the FPKM value was calculated. The production of RNA-seq data of three *Saccharum* species all used the genome of AP85–441 (published in November, 2018 by our group) for reference.

Experimental Validation of *SSs* Genes by PCR and Gene Expression Levels by qRT-PCR

Extraction of sugarcane genomic DNA was performed by CTAB method. Gene-specific primer pairs were designed using the Primer 5 software. The reaction system of PCR for Premix Taq. PCR reaction program was 95 °C for 3 min, 95 °C for 30s, 57 °C for 30s, 72 °C for 2 min, 34 cycles, 72 °C 5 min, 4 °C save. The expression level of two SSs genes (*SsSSIIIa* and *SsGBSSII*) in four tissues (internodes 9, 15, leaves and leaf roll in LA-Purple, and internodes 6, 9, leaves and leaf roll in SES-208) of two *Saccharum* species were validated by qRT-PCR. The expression level of each gene relative to housekeeping genes were calculated using the $2^{-\Delta\Delta Ct}$ with three replicates per sample (Pfaffl 2001). Then we made the correlated analyses between qRT-PCR values with FPKM values. In this experiment *GAPDH* and *eEF* were used as the reference genes. The primer sequences for *SpSS* were GGTCGGGTGAAGCTATGTTTA (*SSIIIa*-F), CAAGGTTCGAAGATGGAAGGA (*SSIIIa*-R) in *S. spontaneum,* and in *S. officinarum* were GGGTATGC TTCAAGACCTACTG (*SSIIIa*-F), CTCACCATAGAGTG ACCCATTG (*SSIII*a-R), AAGCCTGTTGTAGGGAGAAA G (*GBSSII*-F) and CCAGAAGTGAGTTCCTTGACATA (*GBSSII*-R). The primer sequences for the gene *GAPDH* were CACGGCCACTGGAAGCA (F), TCCTCAGGGTTCCT GATGCC (R), and for the gene *eEF* were TTTCACAC TTGGAGTGAAGCAGAT (F), GACTTCCTTCACAA TCTCATCATAA (R).

Acknowledgements This project was supported by grants from the 863 program (2013AA100604), NSFC (31201260, 31760413 and 31660420), Science and Technology Major Project of Guangxi (AA17202025) and Fujian Provincial Department of Education (No. JA12082). We are grateful for editing the language of Irene Lavagi.

Authors' Contributions Jisen Zhang designed the experiments. Jisen Zhang and Panpan Ma conceived the study. Panpan Ma, Yuan Yuan, Qiaochu Shen, Qing Jiang, Xiuting Hua, Qing Zhang, Muqing Zhang, Ray Ming performed the experiments and analyzed the data. Panpan Ma and Jisen Zhang wrote the manuscript. All authors read and approved the final article.

Compliance with Ethical Standards

Conflict of Interests The authors declare no competing financial interests.

References

Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J Exp Bot 62. https://doi.org/10.1093/jxb/erq411

Batra R et al (2017) Comparative Analysis of AGPase Genes and Encoded Proteins in Eight Monocots and Three Dicots with Emphasis on Wheat. Front Plant Sci 8:19. https://doi.org/10.3389/fpls.2017.00019

Bläsing OE, Stitt M (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. Plant Cell 17:3257

Bonnett GD (2013) Developmental Stages (Phenology). In: Sugarcane: Physiology, Biochemistry, and Functional Biology. John Wiley & Sons Ltd, pp 35–53. doi:https://doi.org/10.1002/9781118771280.ch3

Campbell BC et al (2016) Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. Plant Biotechnol J 14:2240–2253. https://doi.org/10.1111/pbi.12578

Cheng J et al (2012) Diversification of genes encoding granule-bound starch synthase in monocots and dicots is marked by multiple genome-wide duplication events. PLoS One 7:e30088. https://doi.org/10.1371/journal.pone.0030088

D'Hont A (2005) Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenetic and Genome Research 109:27–33

Deschamps P et al (2008) Metabolic symbiosis and the birth of the plant kingdom. Mol Biol Evol 25. https://doi.org/10.1093/molbev/msm280

Dian W, Jiang H, Wu P (2005) Evolution and expression analysis of starch synthase III and IV in rice. J Exp Bot 56:623–632. https://doi.org/10.1093/jxb/eri065

Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Research An International Journal on the Molecular Supramolecular & Evolutionary Aspects of Chromosome Biology 17:699

Feike D et al (2016) The Starch Granule-Associated Protein EARLY STARVATION1 Is Required for the Control of Starch Degradation in Arabidopsis thaliana Leaves. Plant Cell 28:1472–1489. https://doi.org/10.1105/tpc.16.00011

Finn RD et al (2014) Pfam: the protein families database. Nucleic Acids Res 42:D222

Grivet L, Glaszmann J-C, D'Hont A (2003) Molecular evidence for sugarcane evolution and domestication. Paper presented at the

Molecular Biology Workshop, Montpellier, France, 2003-04-07 / 2003-04-11

Guindon S, Gascuel O, Rannala B (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol 52:696–704. https://doi.org/10.1080/10635150390235520

Hirose T, Terao T (2004) A comprehensive expression analysis of the starch synthase gene family in rice (Oryza sativa L.). Planta 220:9–16. https://doi.org/10.1007/s00425-004-1314-6

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31:1296

Hu W et al (2018) New insights into the evolution and functional divergence of the SWEET family in Saccharum based on comparative genomics. BMC Plant Biol 18:270. https://doi.org/10.1186/s12870-018-1495-y

Ingkasuwan P et al (2012) Inferring transcriptional gene regulation network of starch metabolism in Arabidopsis thaliana leaves using graphical Gaussian model. BMC Syst Biol 6:100–100. https://doi.org/10.1186/1752-0509-6-100

Jiao Y, Li J, Tang H, Paterson AH (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell 26:2792–2802

Jiao Y et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97

Kumar Behera K, Sahoo S (2009) Rapid In vitro Micro propagation of Sugarcane (*Saccharum officinarum* L. cv- Nayana) Through Callus Culture vol 7

Leterrier M, Holappa LD, Broglie KE, Beckles DM (2008) Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: functional and evolutionary implications. BMC Plant Biol 8:98. https://doi.org/10.1186/1471-2229-8-98

Li P et al (2010a) The developmental dynamics of the maize leaf transcriptome. Nat Genet 42:1060

Li P et al (2010b) The developmental dynamics of the maize leaf transcriptome. Nat Genet 42:1060–1067. https://doi.org/10.1038/ng.703

Lloyd JR, Kossmann J (2015) Transitory and storage starch metabolism: two sides of the same coin? Curr Opin Biotechnol 32:143–148. https://doi.org/10.1016/j.copbio.2014.11.026

Matsumoto T et al (2005) The map-based sequence of the rice genome. Nature 436:793–800

Ming R et al (2015a) The pineapple genome and the evolution of CAM photosynthesis. Nat Genet 47:1435

Ming R et al (2015b) The pineapple genome and the evolution of CAM photosynthesis. Nat Genet 47:1435–1442. https://doi.org/10.1038/ng.3435

Moore PH, Paterson AH, Tew T (2013) Sugarcane: The Crop, the Plant, and Domestication. In: Sugarcane: Physiology, Biochemistry, and Functional Biology. John Wiley & Sons Ltd, pp 1–17. doi:https://doi.org/10.1002/9781118771280.ch1

Muller LM, von Korff M, Davis SJ (2014) Connections between circadian clocks and carbon metabolism reveal species-specific effects on growth control. J Exp Bot 65:2915–2923. https://doi.org/10.1093/jxb/eru117

Nougué O, Corbi J, Ball SG, Manicacci D, Tenaillon MI (2014) Molecular evolution accompanying functional divergence of duplicated genes along the plant starch biosynthesis pathway. BMC Evol Biol 14:103–103. https://doi.org/10.1186/1471-2148-14-103

Pan X, Yan H, Li M, Wu G, Jiang H (2011) Evolution of the Genes Encoding Starch Synthase in Sorghum and Common Wheat. Molecular Plant Breeding. https://doi.org/10.5376/mpb.2011.02.0009

Paterson AH et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551 doi:https://doi.org/10.1038/nature07723. https://www.nature.com/articles/nature07723#supplementary-information

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45

Pfister B, Zeeman SC (2016) Formation of starch in plant cells. Cellular and Molecular Life Sciences: CMLS 73:2781–2807. https://doi.org/10.1007/s00018-016-2250-x

Piperidis G, Piperidis N, D'Hont A (2010) Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. Mol Gen Genomics 284:65–73. https://doi.org/10.1007/s00438-010-0546-3

Preiss J (2004) Plant starch synthesis:3–56 doi:10.1533/9781855739093.1.3

Schnable PS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schwarte S, Wegner F, Havenstein K, Groth D, Steup M, Tiedemann R (2015) Sequence variation, differential expression, and divergent evolution in starch-related genes among accessions of Arabidopsis thaliana. Plant Mol Biol 87:489–519. https://doi.org/10.1007/s11103-015-0293-2

Seaton DD, Ebenhoh O, Millar AJ, Pokhilko A (2014) Regulatory principles and experimental approaches to the circadian control of starch turnover. J R Soc Interface 11:20130979. https://doi.org/10.1098/rsif.2013.0979

Skeffington AW, Graf A, Duxbury Z, Gruissem W, Smith AM (2014) Glucan, Water Dikinase Exerts Little Control over Starch Degradation in Arabidopsis Leaves at Night. Plant Physiol 165:866–879. https://doi.org/10.1104/pp.114.237016

Smith AM, Stitt M (2007) Coordination of carbon supply and plant growth. Plant Cell Environ 30:1126–1149. https://doi.org/10.1111/j.1365-3040.2007.01708.x

Stitt M, Zeeman SC (2012) Starch turnover: pathways, regulation and role in growth. Curr Opin Plant Biol 15:282–292. https://doi.org/10.1016/j.pbi.2012.03.016

Streb S, Zeeman SC (2012) Starch metabolism in Arabidopsis. The Arabidopsis book 10:e0160. https://doi.org/10.1199/tab.0160

Sulpice R et al (2009) Starch as a major integrator in the regulation of plant growth. Proc Natl Acad Sci U S A 106:10348–10353. https://doi.org/10.1073/pnas.0903478106

Sweetlove LJ, Burrell MM, ap Rees T (1996) Starch metabolism in tubers of transgenic potato (Solanum tuberosum) with increased ADPglucose pyrophosphorylase. Biochem J 320:493–498

Tao Y et al (2017) Whole-Genome Analysis of Candidate genes Associated with Seed Size and Weight in Sorghum bicolor Reveals Signatures of Artificial Selection and Insights into Parallel Domestication in Cereal Crops. Front Plant Sci 8:1237. https://doi.org/10.3389/fpls.2017.01237

Valdez HA, Busi MV, Wayllace NZ, Parisi G, Ugalde RA, Gomez-casati DF (2008) Role of the N-Terminal starch-binding domains in the kinetic properties of Starch Synthase III from Arabidopsis thaliana. Biochemistry 47. https://doi.org/10.1021/bi702418h

Vrinten PL, Nakamura T (2000) Wheat Granule-Bound Starch Synthase I and II Are Encoded by Separate Genes That Are Expressed in Different Tissues. Plant Physiol 122:255–264. https://doi.org/10.1104/pp.122.1.255

Wang DP, Wan HL, Song Z, Yu J (2009) γ -MYN: a new algorithm for estimating Ka and Ks with consideration of variable substitution rates. Biol Direct 4:1–18

Wang J, Nayak S, Koch K, Ming R (2013) Carbon partitioning in sugarcane (Saccharum species). Front Plant Sci 4:201. https://doi.org/10.3389/fpls.2013.00201

Wang L et al (2014) Comparative analyses of C(4) and C(3) photosynthesis in developing leaves of maize and rice. Nat Biotechnol 32:1158–1165. https://doi.org/10.1038/nbt.3019

Yan L, Wei S, Wu Y, Hu R, Li H, Yang W, Xie Q (2015) High-Efficiency Genome Editing in Arabidopsis Using YAO Promoter-Driven CRISPR/Cas9 System. Mol Plant 8:1820–1823. https://doi.org/10.1016/j.molp.2015.10.004

Zeeman SC, Kossmann J, Smith AM (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. Annu Rev Plant Biol 61:209–234. https://doi.org/10.1146/annurev-arplant-042809-112301

Zhang J et al (2018) Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. Nat Genet 50:1565–1573. https://doi.org/10.1038/s41588-018-0237-2

Zhang J, Zhou M, Walsh J, Zhu L, Chen Y, Ming R (2014) Sugarcane Genetics and Genomics. John Wiley & Sons Ltd, Hoboken

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. BMC Genomics 17:88. https://doi.org/10.1186/s12864-016-2419-6

Zhang X, Szydlowski N, Delvalle D, D'Hulst C, James MG, Myers AM (2008) Overlapping functions of the starch synthases SSII and SSIII in amylopectin biosynthesis in Arabidopsis. BMC Plant Biol 8:96. https://doi.org/10.1186/1471-2229-8-96