PCP
PLANT & CELL PHYSIOLOGY

Regular paper

Genome-Wide Identification and Expression Profile Analysis of *WRKY* Family Genes in the Autopolyploid *Saccharum spontaneum*

Zhen Li[1], Xiuting Hua[1], Weiming Zhong[1], Yuan Yuan[1,2], Yongjun Wang[1], Zhengchao Wang[3], Ray Ming[1,2,4] and Jisen Zhang[1,2,*]

[1]Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Center for Genomics and Biotechnology, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2]Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning, Guangxi 530004, China
[3]Provincial Key Laboratory for Developmental Biology and Neurosciences, College of Life Sciences, Fujian Normal University, Fuzhou 350007, China
[4]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
*Corresponding author: E-mail, zjisen@fafu.edu.cn; Fax, 86-591-88202395.
(Received May 7, 2019; Accepted December 8, 2019)

WRKY is one of the largest transcription factor families in plants and plays important roles in the regulation of developmental and physiological processes. To date, the *WRKY* gene family has not been identified in *Saccharum* species because of its complex polyploid genome. In this study, a total of 294 sequences for 154 *SsWRKY* genes were identified in the polyploid *Saccharum spontaneum* genome and then named on the basis of their chromosome locations, including 13 (8.4%) genes with four alleles, 29 (18.8%) genes with three alleles and 41 (26.6%) genes with two alleles. Among them, 73.8% and 16.0% of the *SsWRKY* genes originated from segmental duplications and tandem duplications, respectively. The *WRKY* members exhibited conserved gene structures and amino acid sequences among the allelic haplotypes, which were accompanied by variations in intron sizes. Phylogenetic and collinearity analyses revealed that 27 *SsWRKYs* originated after the split of sorghum and *Saccharum*, resulting in a significantly higher number of *WRKYs* in sugarcane than in the proximal diploid species sorghum. The analysis of RNA-seq data revealed that *SsWRKYs'* expression profiles in 46 different samples including different developmental stages revealed distinct temporal and spatial patterns with 52 genes expressed in all tissues, four genes not expressed in any tissues and 21 *SsWRKY* genes likely to be involved in photosynthesis. The comprehensive analysis of *SsWRKYs'* expression will provide an important and valuable foundation for further investigation of the regulatory mechanisms of *WRKYs* in physiological roles in sugarcane *S. spontaneum*.

Keywords: Expression profiles • Genome wide • Phylogenetic analysis • *Saccharum spontaneum* • *WRKY* family.

Introduction

WRKY is one of the largest transcriptional regulator families in plants and named after the WRKY domain, approximately 60 amino acid residues that is highly conserved among family members (Eulgem et al. 2000). The two most defining structural characteristics of WRKY domain are WRKYGQK heptapeptide sequence and a zinc finger motif C_2H_2 (C-X_{4-5}-X_{22-23}-H-X_1-H) or C_2HC (C-X_7-C-X_{23}-H-X_1-C) (Rushton et al. 1995, Eulgem et al. 2000). The WRKYGQK heptapeptide sequence has some variants including WKKYGQK, WRKYGKK (Wang et al. 2014), WRKYGEK, WRKYGRK, WKRYGQK, WSKYEQK (Wu et al. 2005) and WKKYGEK (GiacomelLi et al. 2010). Based on the number of WRKY domains and the features of their zinc finger motif, WRKYs are usually divided into three groups (I, II and III). WRKY proteins with two domains belong to group I, while the proteins with one WRKY domain belong to group II with a C_2H_2 zinc finger, or to group III with a C_2HC zinc finger. Group I is subsequently divided into two subgroups, subgroup Ia containing C_2H_2 zinc finger and subgroup Ib containing C_2HC zinc finger, while group II is divided into IIa, IIb, IIc, IId and IIe based on their phylogenetic relations (Eulgem et al. 2000, Xie et al. 2005). Therefore, the WRKY family in higher plants is divided into groups I, IIa + IIb, IIc, IId + IIe and III (Zhang and Wang 2005, Rushton et al. 2008), while WRKYs with an incomplete WRKY domain are assigned to group IV (Xie et al. 2005).

The first WRKY cDNA, *SPF1*, was cloned from sweet potato (*Ipomoea batatas*) in 1994 (Ishiguro and Nakamura 1994). With the availability of whole genome information, *WRKYs* have been identified genome wide from various plant species. These include the *WRKY* gene family of *Arabidopsis thaliana* (Eulgem et al. 2000), cucumber (*Cucumis sativus*) (Ling et al. 2011), tomato (*Solanum lycopersicum*) (Huang et al. 2012), grape (*Vitis vinifera* L.) (Guo et al. 2014), rice (*Oryza sativa*) (Ross et al. 2007), *Brachypodium distachyon* (Tripathi et al. 2012), maize (*Zea mays*) (Wei et al. 2012), wheat (*Triticum aestivum* L.) (Ning et al. 2017) and pineapple (*Ananas comosus*) (Xie et al. 2018).

Previous studies have demonstrated that WRKY transcription factors play important roles in regulating the various development and physiological processes (Wu et al. 2016). For instance, *AtWRKY71* accelerates flowering and controls shoot branching in Arabidopsis (Guo et al. 2015, Yu et al. 2016).

Plant Cell Physiol. 61(3): 616–630 (2020) doi:10.1093/pcp/pcz227, Advance Access publication on 12 December 2019, available online at https://academic.oup.com/pcp

AtWRKY18 and *AtWRKY53* have a positive effect on leaf senescence. However, overexpression of *AtWRKY18* and *AtWRKY53* can result in delayed senescence (Potschin et al. 2014). Furthermore, *WRKYs* have been found to be involved in regulating various biotic and abiotic stress defenses, specifically *NaWRKY3* and *NaWRKYY6* coordinate responses to herbivory in tobacco (*Nicotiana attenuata*) (Skibbe et al. 2008) and overexpression of *OsWRKY76* in rice plants significantly increases the susceptibility to *Magnaporthe oryzae* but improves tolerance to cold stress (Yokotani et al. 2013). In addition, *AtWRKY46*, *AtWRKY54* and *AtWRKY70* are involved in brassinosteroid-regulated plant growth and negative regulation of drought tolerance (Chen et al. 2017a).

Sugarcane (*Saccharum* spp.) is a classic C_4 crop with the highest photosynthetic rates among crops and is responsible for about 80% of sugar and 40% of ethanol production worldwide (Lam E et al. 2009, Zhang et al. 2013). Based on the conventional taxonomy, *Saccharum* includes the following three founding species: *Saccharum spontaneum*, *Saccharum robustum* and *Saccharum officinarum* (Irvine 1999). Modern sugarcane cultivar was derived from a cross between *S. spontaneum* and *S. officinarum*. Both *S. spontaneum* and *S. officinarum* are considered to be the founding species for sugarcane studies. However, *S. spontaneum* AP85-441 ($1n = 4x = 32$), generated from the anther of octoploid SES208 ($2n = 8x = 64$) (Moore et al. 1989, Zhang et al. 2018b), is the only autopolyploid with allele-defined genome data availability in *Saccharum*. Due to the potential collapsing of homologous sequences that occurred during the assembly of the tetraploid *S. spontaneum* genome, it is particularly challenging to accurately identify all members of certain gene families based on sequence search alone, especially for the large gene families, such as *WRKY*.

According to previous studies (Li et al. 2010, Phukan et al. 2016), *WRKYs* play a pivotal role in the regulation of sugar metabolism and photosynthetic processes in plants. However, no related research on *WRKYs* was available in sugarcane due to its complicated genome. This study marks the identification of the *WRKY* family in an autopolyploid plant, with 154 *WRKYs* being identified through a combination of comparative genomics and whole genome sequence resources of *S. spontaneum*. *SsWRKY1* to *SsWRKY154* were named based on the loci of these genes. The evolutionary relationship and collinearity of *SsWRKY* genes were investigated. Subsequently, the gene structure of *SsWRKYs* and *SbWRKYs* (*Sorghum bicolor WRKYs*) were elucidated and the variations within allelic genes of *SsWRKYs* were explored. In addition, this study comprehensively determined the expression profiles of *SsWRKY* genes in various tissues. Taken together, our study will provide a foundation for further investigations into the evolutionary processes and physiological functions of *SsWRKY* genes in *S. spontaneum*.

Results

Identification of *WRKY* in *S. spontaneum* genome

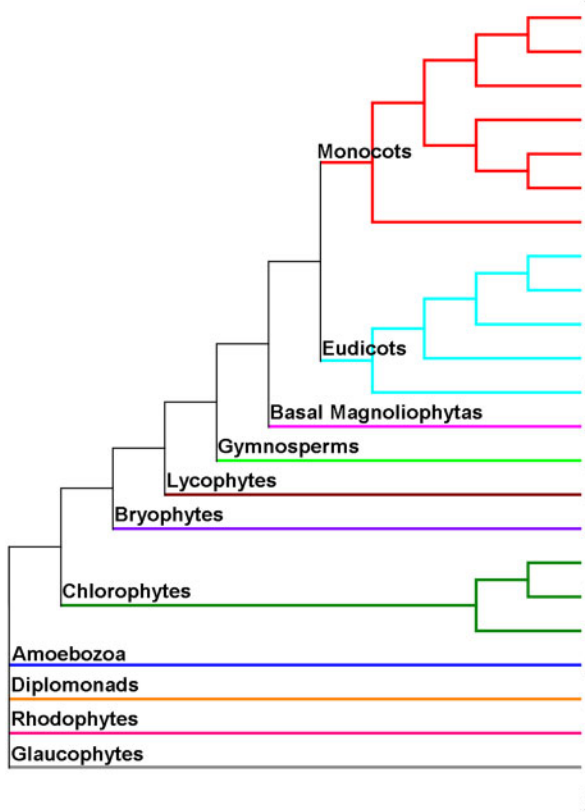
WRKYs are found extensively in plants and even in some non-plants, such as fungi and slime molds. Here, 23 species representing the 11 lineages were selected to analyze the sugarcane *WRKY* family in a comprehensive and comparative way. After searching with the HMMER software and deleting the members lacking the WRKYGQK heptapeptide, 1,360 *WRKYs* were obtained (**Fig. 1**). *Saccharum spontaneum* contained the second largest number of *WRKYs* among the examined species. In addition, subgroups IIc and III were larger than the remaining subgroups in *S. spontaneum*, which was similar to the other monocot plant species (**Fig. 1**).

A total of 314 gene sequences were originally obtained using both PF03106 [a Hidden Markov Model (HMM) profile of WRKY DNA-binding domain (DBD)] and our own *S. spontaneum* genome database with BLASTP. Among them, 11 did not have a WRKY domain-coding sequence, and were therefore excluded from further analysis. Forty-seven gene sequences were re-annotated manually with the assistance of FGENESH (**http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind**), nine sequences were identified without WRKYGQK sequences and were deleted. Furthermore, 83 *WRKYs* have 2, 3 or 4 allelic genes, including 13 (8.4%) genes with four alleles, 29 (18.8%) genes with three alleles and 41 (26.6%) genes with two alleles (**Supplementary Fig. S1**). Based on their chromosomal locations, we named these 294 *SsWRKY* gene alleles as *SsWRKY1* to *SsWRKY154*, and additional -1 to -4 were added to the gene name for their alleles (**Supplementary Table S1**). Among them, *Sspon.003B0016500* and *Sspon.008A0006220* were separated into two genes after re-annotation and their sequences contained the WRKY functional domain, meaning that they were named as *SsWRKY63-2a* and *SsWRKY63-2b* and *SsWRKY143-1a* and *SsWRKY143-1b*, respectively (**Supplementary Table S1**).

Detailed information about the SsWRKYs is listed in **Supplementary Table S1**, including the deduced protein length, the molecular weight (MW), the theoretical isoelectric point (p*I*), the aliphatic index (AI), the grand average of hydropathicity (GRAVY) and the instability index (II). The length of putative SsWRKYs ranged from 92 to 949 amino acids, and the MW of the proteins ranged from 10,290.53 to 214,275.80 Da, whereas the p*I* ranged from 4.71 to 10.37, which is similar to grapevine (Wang et al. 2014), implying that different SsWRKYs might operate in various microenvironments. The values of GRAVY were all negative, suggesting that SsWRKYs were all hydrophilic, similarly to wheat (Ning et al. 2017).

Classification and phylogenetic analysis of *SsWRKYs*

To further categorize and investigate the evolutionary relationship of *SsWRKY* genes, we constructed an unrooted phylogenetic tree with 172 WRKY domains in sugarcane and 109 in rice using the neighbor-joining (NJ) and maximum likelihood (ML) methods, respectively (**Fig. 2** and **Supplementary Fig. S2**). Based on the classification of *OsWRKYs* and the primary structure features of SsWRKY proteins, all 154 *SsWRKY* genes were classified into three major groups. The 17 *SsWRKYs* with two WRKY domains were assigned to group I, which consists of two subgroups. Subgroup Ia *SsWRKYs* had C_2H_2 zinc finger motifs of C-X_4-C-X_{22-23}-H-X_1-H, and subgroup Ib *SsWRKYs* contained C_2HC zinc fingers (C-X_7-C-X_{23}-H-X_1-C). Interestingly,



Species	I	II a	II b	II c	II d	II e	III	IV	Total
B. distachyon	17	3	6	22	6	8	22	2	86
T. aestivum	31	11	7	45	17	10	45	0	166
O. sativa	13	4	8	19	6	11	31	4	96
Z. mays	27	7	11	29	14	17	31	0	136
S. bicolor	11	5	8	23	7	11	31	0	96
S. spontaneum	17(34)	6(20)	12(24)	40(66)	11(22)	12(25)	51(97)	5(6)	154(294)
A. comosus	14	2	7	11	7	5	7	2	55
P. trichocarpa	23	5	9	40	13	4	10	0	104
C. sativus	8	4	4	20	8	7	6	0	57
A. thaliana	14	3	8	18	7	9	13	0	72
V. vinifera	12	3	8	16	6	6	6	2	59
S. lycopersicum	15	5	8	18	6	16	11	1	80
A. trichopoda	6	2	4	8	3	4	5	1	33
P. sitchensis	1	0	1	4	3	0	0	0	9
S. moellendorffii	8	0	6	7	7	2	4	0	34
P. patens	13	0	10	56	16	0	19	2	116
C. reinhardtii	1	0	0	0	0	0	0	0	1
V. carteri	1	0	2	0	0	0	0	0	3
C. variabilis	1	0	0	0	0	0	0	0	1
D. discoideum	1	0	0	0	0	0	0	0	1
G. lamblia	1	0	0	0	0	0	0	0	1
C. merolae	0	0	0	0	0	0	0	0	0
C. paradoxa	0	0	0	0	0	0	0	0	0
summary	235	60	119	376	137	122	292	19	1360

Fig. 1 The evolutionary relationship of 23 species among 11 lineages. The numbers in parenthesis detail the number of alleles of *SsWRKYs* in *S. spontaneum*. The phylogenetic tree was constructed using MEGA 7.0 based on the evolutionary relationship of 23 species obtained from NCBI (**https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi**).

SsWRKY153 had three domains, which were absent in other plants, suggesting that *SsWRKY153* may be a *WRKY* gene specific to sugarcane, this gene will be further investigated. Group II was comprised of 81 *WRKYs* with a single WRKY domain and a zinc finger motif of C-X_{4-5}-C-X_{23}-H-X_1-H. The present study further divided group II into five subgroups based on the presence of specific sequences in their zinc finger motifs, subgroup IIa with a CX_5CPVKKK(L/V)Q motif, subgroup IIb with a CX_5CPVRKQVQ motif, subgroup IIc with a CX_4C motif, subgroup IId with a CX_5CPARKHVE motif and subgroup IIe with a CX_5CPARK(Q/M)V(E/D) motif, each containing 6, 12, 40, 11 and 12 proteins, respectively. Fifty-one *SsWRKYs* with a single WRKY domain were assigned to group III due to their C_2HC zinc finger motif (C-X_7-C-X_{23-31}-H-X_1-C). Finally, five *SsWRKYs* had only partial WRKY domains and were classified as group IV.

We found that the C-terminal domains of subgroup Ia *SsWRKYs* clustered into clade IaC and the N-terminal domains of subgroup Ia were clustered into another clade IaN (**Supplementary Fig. S3**). Subgroups IIa and IIb clustered into one clade, with subgroups IId and IIe clustering into another, and this phenomenon provided the evidence for the previous proposal to merge subgroups IIa and IIb into a single subgroup, as well as to merge subgroups IId and IIe into a single subgroup. Subgroup IIc clustered into three clades, IIc1, IIc2 and IIc3, with most *SsWRKYs* of this subgroup clustered into the IIc1 clade; four subgroup IIc *SsWRKYs* clustered into IIc3 clade with the N-terminal domain of *SsWRKY99*. In addition, 10 subgroup IIc SsWRKY domains (*SsWRKY26*, *-98*, *-100*, *-106*, *-107*, *-110*, *-111*, *-112*, *-148* and *-150*) clustered within the IaN clade and *SsWRKY29* clustered within the IaC clade, indicating that group I *WRKY* genes may represent the ancestral form of other *WRKYs*. Therefore, *SsWRKY26*, *-98*, *-100*, *-105*, *-106*, *-107*, *-110*, *-111*, *-112*, *-148* and *-150* may have evolved from the genes of subgroup Ia through the loss of the C-terminal domain, while *SsWRKY29* may have evolved from genes of subgroup Ia through the loss of the N-terminal domain. Both WRKY domains of subgroup Ib (**Supplementary Fig. S3**, green diamond) are imbedded within the cluster of group III (**Supplementary Fig. S3**, red-filled triangle), supporting the conclusion that subgroup Ib evolved from the duplication of single group III WRKY domain in Gramineae (Brand et al. 2013).

Genomic location and duplication events among *SsWRKY* genes

The genome chromosome location information of *SsWRKYs* showed 286 of the 294 *SsWRKY* gene alleles distributed throughout the eight chromosomes of *S. spontaneum*, with the remaining eight gene alleles located in the unanchored scaffolds. Many *SsWRKY* genes were located on chromosome 3 (31, 20.13%), chromosome 7 (27, 17.53%) and chromosome 2

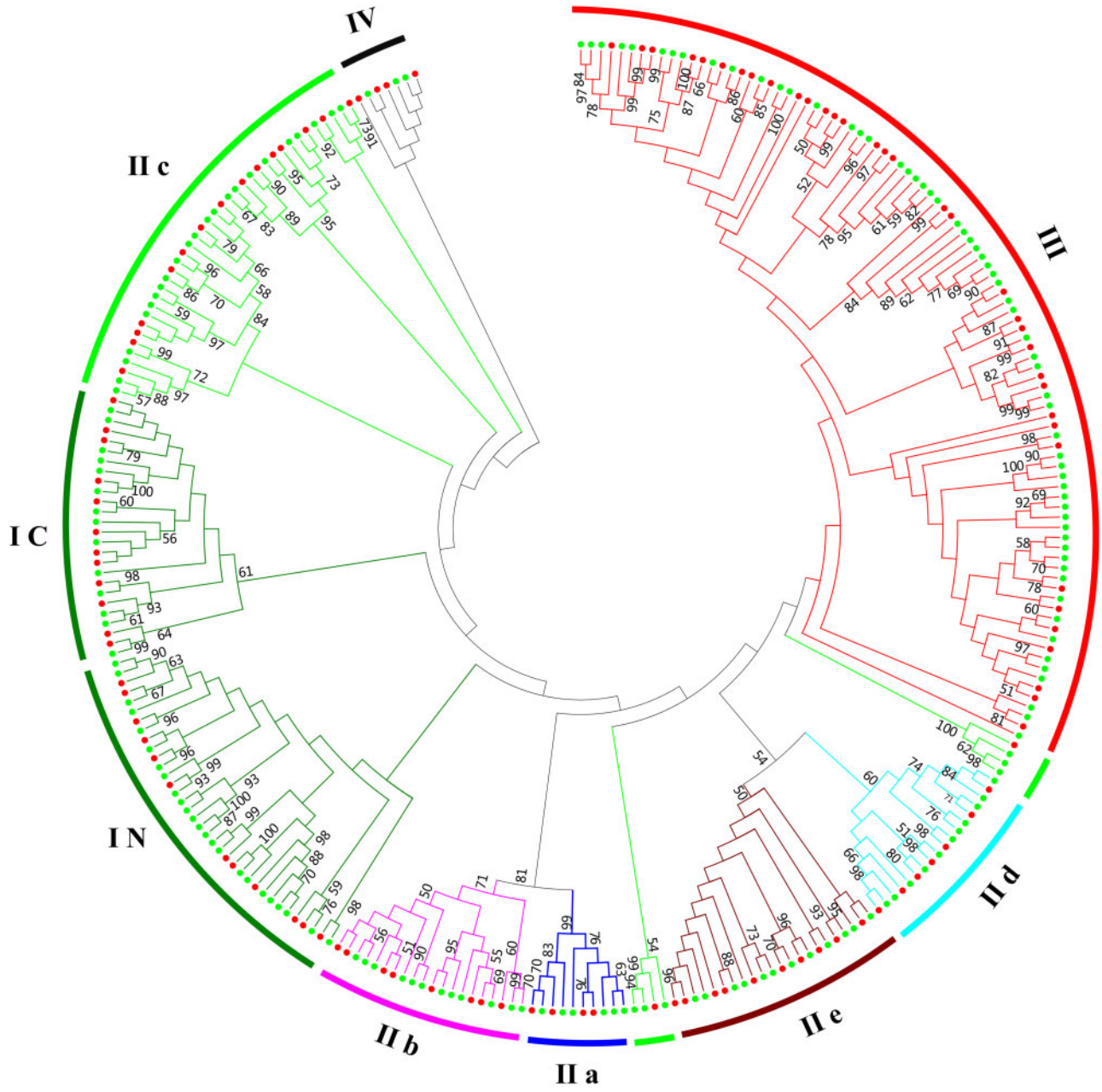


Fig. 2 Phylogenetic tree of WRKY domains from AP85-441 (*S. spontaneum*) and rice (*O. sativa*). The unrooted NJ tree was constructed based on the WRKY domains from AP85-441 and rice using MEGA7.0 with 1,000 boosted replicates. Branches with <50% bootstrap support were collapsed. The different colored arcs indicate different groups (or subgroups) of the WRKY domain. 'N' or 'C' indicates the N-terminal or C-terminal WRKY domains of group I, respectively. The green and red solid circles represent the WRKY domain from AP85-441 and rice, respectively.

(24, 15.58%), while chromosomes 4 and 8 are the locations for only 8 (5.19%) and 9 (5.84%) genes, respectively. In addition, *SsWRKYs* of subgroup IIc and group III were present in all chromosomes, while the members of group I were not found on chromosomes 4 and 8, the *WRKY* genes of subgroup IIb were not present in chromosomes 1 and 5, subgroup IId *WRKY* genes were not found on chromosomes 7 and 8 and subgroup IIe *WRKY* genes were not found on chromosomes 2 and 6. Importantly, the members of subgroup IIa were only found on chromosomes 2, 4 and 8 and the genes of group IV were only found on chromosomes 1, 2, 3 and 7 (**Table 1**).

Gene duplication events of the *WRKY* genes have been identified in some plants, such as wheat and grape (Guo et al. 2014, Ning et al. 2017). In this study, we identified 549 *WRKY* collinear pairs in *S. spontaneum* by using BLASTP and MCScanx software, with 442 pairs of nonalleles and 107 pairs of alleles (**Fig. 3** and **Supplementary Table S2**). According to the methods of Holub (2001), a chromosomal region within 200 kb containing two or more genes is defined as a tandem duplication event. Considering the difference of genome size between *S. spontaneum* and Arabidopsis, we examined the extended genomic regions of 500 kb. Eighty *SsWRKY* gene alleles were clustered

Table 1 The number of *SsWRKY* genes on chromosomes in different groups

Chromosome		Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8
I	Genes	2	3	2	0	2	5	2	0
	Alleles	3	6	3	0	4	2	6	0
IIa	Genes	0	2	0	1	0	0	0	3
	Alleles	0	7	0	4	0	0	0	7
IIb	Genes	0	1	5	1	0	1	3	1
	Alleles	0	0	13	3	0	0	2	0
IIc	Genes	4	3	12	2	1	9	6	2
	Alleles	8	2	15	3	2	4	11	0
IId	Genes	4	1	0	2	2	1	0	0
	Alleles	7	3	0	1	3	0	0	0
IIe	Genes	2	0	4	2	1	0	1	2
	Alleles	2	0	10	3	0	0	2	3
III	Genes	4	12	7	0	11	2	14	1
	Alleles	4	19	12	0	16	3	21	2
IV	Genes	1	2	1	0	0	0	1	0
	Alleles	0	0	2	0	0	0	0	0
Summary	Genes	17	24	31	8	17	18	27	9
	Alleles	24	37	55	14	25	9	42	12

into 34 tandem duplication event regions on chromosome 1B (one cluster), 1D (two clusters), 2A (one cluster), 2B (two clusters), 2C (one cluster), 2D (two clusters), 3A (five clusters), 3B (two clusters), 3C (three clusters), 3D (one cluster), 4C (one cluster), 5A (one cluster), 5B (one cluster), 5C (two clusters), 5D (one cluster), 6A (one cluster), 7A (two clusters), 7B (one cluster), 7C (one cluster), 8A (one cluster) and 8D (one cluster) (**Supplementary Table S3**). Chromosome 3 had 11 clusters, indicating a hot spot of *WRKY* gene duplication. In addition to the tandem duplication events, 217 gene alleles were found to derive from segmental duplication events (**Supplementary Table S4**). These results suggested that segmental duplication events may play a more critical role in some *SsWRKY* genes possibly being generated by gene duplication events.

We then estimated the divergence time among tandem-duplicated *SsWRKY* genes and between *SsWRKY–SbWRKY* based on the pairwise Ks (**Supplementary Table S5**). The divergence time of 17 *SsWRKY* tandem duplication pairs ranged from 0.205 to 99.020 mya, illustrating that these *SsWRKYs* were generated by recent gene duplication events in *S. spontaneum*, while the other 12 tandem duplication pairs were ancient, due to the divergence time ranging from 141.393 to 310.639 mya. For example, *SsWRKY67* and *SsWRKY54* formed a tandem duplication pair as well as their orthologous *SbWRKY24* and *SbWRKY25* in sorghum and *OsWRKY10* and *OsWRKY97* in rice. However, some *SsWRKY* tandem duplication pairs had only one orthologous *SbWRKY*, such as *SsWRKY70* and *SsWRKY71-2* (**Supplementary Fig. S4**). This phenomenon may also be responsible for the fact that the number of *WRKY* genes in *S. spontaneum* is more than that of sorghum. In addition, the divergence time between *SsWRKY4-3*, *SsWRKY5-1*, *SsWRKY5-3*, *SsWRKY20*, *SsWRKY21-1*, *SsWRKY21-2*, *SsWRKY22-1*, *SsWRKY31-3*, *SsWRKY45-1*, *SsWRKY46*, *SsWRKY47-1*, *SsWRKY48*, *SsWRKY49-1*, *SsWRKY59-1*, *SsWRKY59-2*, *SsWRKY63-2a*, *SsWRKY63-2b* and *SsWRKY96* and their orthologous *SbWRKYs* ranges from 4.497 to 7.479 mya (**Supplementary Table S5**), which are shorter than that of *S. spontaneum* and sorghum (7.779 mya) (Zhang et al. 2018a), suggesting that these 14 genes originated from tandem duplication after the divergence of *S. spontaneum* and sorghum.

Multiple sequence alignment of SsWRKY proteins

The defining feature of WRKY transcription factors is the WRKY domain, which interacts with the W-box (C/T)TGAC(T/C), thereby activating a large number of defense-related genes (Eulgem et al. 2000). A multiple sequence alignment of the core WRKY domain with approximately 60 amino acids of all 294 SsWRKY proteins is provided in **Supplementary Fig. S5**. A total of 329 WRKY domains were found in SsWRKYs, with most domains containing the highly conserved WRKY motif. However, 20 domains possessed one to three mismatched amino acids within the WRKY motif, such as WMKY, WKKY, WRKL, WRRT, WEKF, WTTY, WKKN, WKAC, WEAC, WKNL and WTNL. Similarly, more than 80% of SsWRKYs contained the heptamer WRKYGQK, while WRKYGEK and WRKYGKK were the most common variants, found in 19 and 14 proteins, respectively. Other variants, such as WRRTCRR (SsWRKY42-1) and WMKYGQK (SsWRKY102 and SsWRKY108 and SsWRKY153), were presented in group I. WRKYGHK (SsWRKY143-4) was found in subgroup IIa. In subgroup IIc, SsWRKY5-1/-5-3/-4-2/-4-3/-73 contained a WKKYGQK sequence, SsWRKY111/-112 contained a WKACAQN sequence and WEACAQN, WTNLGLQ and WEACAQN were found in SsWRKY100, SsWRKY107 and SsWRKY110, respectively. In group III, WEKFGEK was found in SsWRKY34 and WTTYSQK was found in SsWRKY32; strikingly, the alleles of SsWRKY56 contained a single WRKL.

The zinc finger motif (C-X_{4-5}-X_{22-23}-H-X_1-H or C-X_7-C-X_{23-31}-H-X_1-C) at the C-terminus of the WRKY domain is a further important structural characteristic of WRKYs (Eulgem et al. 2000). However, some variants present in this motif (**Supplementary Fig. S5**), C-X_5-C-X_{29}-HNH, C-X_4-C-X_{24}-HQH and C-X_4-C-X_{25}-HSH, were found in group II. Notably, group III possessed motifs of C-X_5-C-X_{27}-HIC, C-X_5-C-X_{28}-HTC, C-X_8-C-X_{24}-HTC, C-X_6-C-X_{37}-HTC and C-X_7-C-X_{23}-HNH, which were not observed in other plants. Some SsWRKYs contained an incomplete zinc finger motif, which could be due to mistakes in genome annotations, or pseudogenes without biological functions.

Allelic gene analysis of *SsWRKYs*

The results of the gene structure analysis indicated that the number and size of exons/introns among *SsWRKY* alleles were highly conserved, albeit with some slight differences. In group I, the second intron of *SsWRKY27-4* was larger than the other three alleles of *SsWRKY27*. The first intron of *SsWRKY2-1* was larger than *SsWRKY2-2*, and the last exon was split. The number

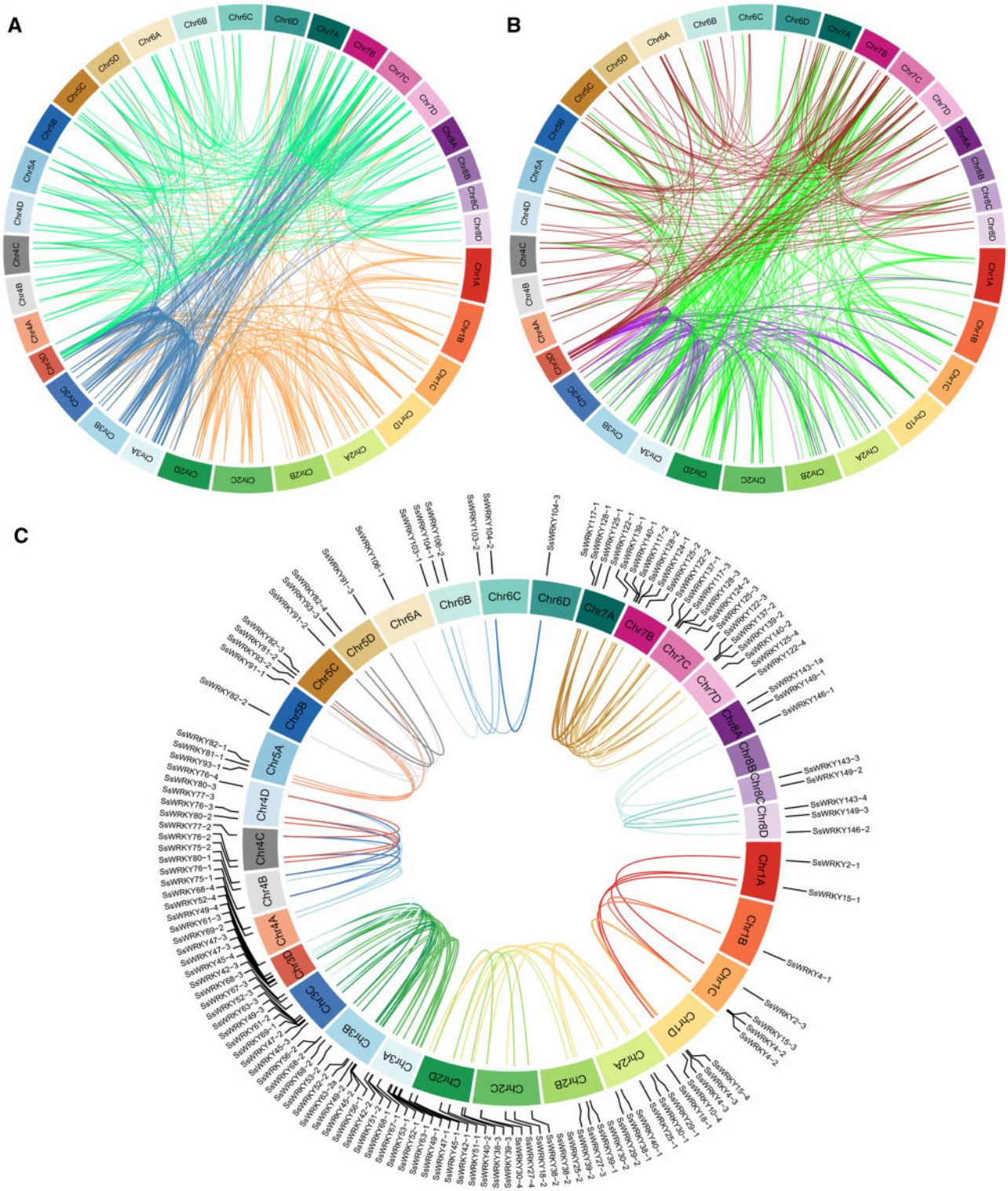


Fig. 3 Schematic representations highlighting the interchromosomal relationships of the *SsWRKY* genes. (A) Collinear analysis of all 294 *SsWRKY* gene alleles. (B) Collinear analysis of nonalleles. (C) Collinear analysis of alleles.

of exons in *SsWRKY2-3* was significantly less than the other two alleles, which may be due to the incomplete WRKY domain. In group II, *SsWRKY68-1* lost an intron compared with *SsWRKY68-2* and *SsWRKY68-3* had an intron inserted. The last introns of *SsWRKY80-1* and *SsWRKY81-1* were larger than in their other alleles, while the second intron of *SsWRKY10-4* was smaller than the other alleles. In group III, the first exon of *SsWRKY91-1* was split into two exons, as was the last exon of *SsWRKY91-2*. In addition, the last exons of *SsWRKY1-3* and *SsWRKY39-2* were split compared with their alleles. These results suggest that exon

insertion and splitting may have occurred in *SsWRKY* alleles (**Supplementary Fig. S6**).

To identify the evolutionary forces acting on the 87 *SsWRKY* genes with alleles, the ratio of the non-synonymous substitution rate to the synonymous substitution rate (Ka/Ks) was calculated (**Fig. 4** and **Supplementary Fig. S7**). With the exception of *SsWRKY2*, *SsWRKY52* and *SsWRKY131*, the Ka/Ks ratios of *SsWRKY* alleles were <1, suggesting that these *SsWRKY* genes were under purifying selection.

Comparative analysis of *WRKYs* between *S. spontaneum* and *S. bicolor*

Sorghum bicolor is the closest related diploid to *S. spontaneum*; the comparison of gene structures between these two species will provide clues to the evolutionary gene events caused by polyploidization. To further analyze the differences in *WRKYs* between *S. spontaneum* and *S. bicolor*, the exon/intron structural analysis for the 154 *SsWRKYs* and 96 *SbWRKYs* was constructed. As shown in **Fig. 5** and **Supplementary Fig. S8**, 148 of the 154 *SsWRKY* genes possessed two to six exons (29 with two exons, 74 with three exons, 19 with four exons, 15 with five exons and 11 with six exons), with *SsWRKY104* and *SsWRKY109* having only one exon, *SsWRKY42* and *SsWRKY130* having seven exons and both *SsWRKY70* and *SsWRKY142* having eight exons. These findings indicated that both gains and losses of exons occurred during the evolution of the *WRKY* family. The intron phases within the same groups were conserved in the DBD. Intron phases were divided into three groups based on the position of the intron within a codon (Xie et al. 2005): a phase 0 intron interrupts and falls between

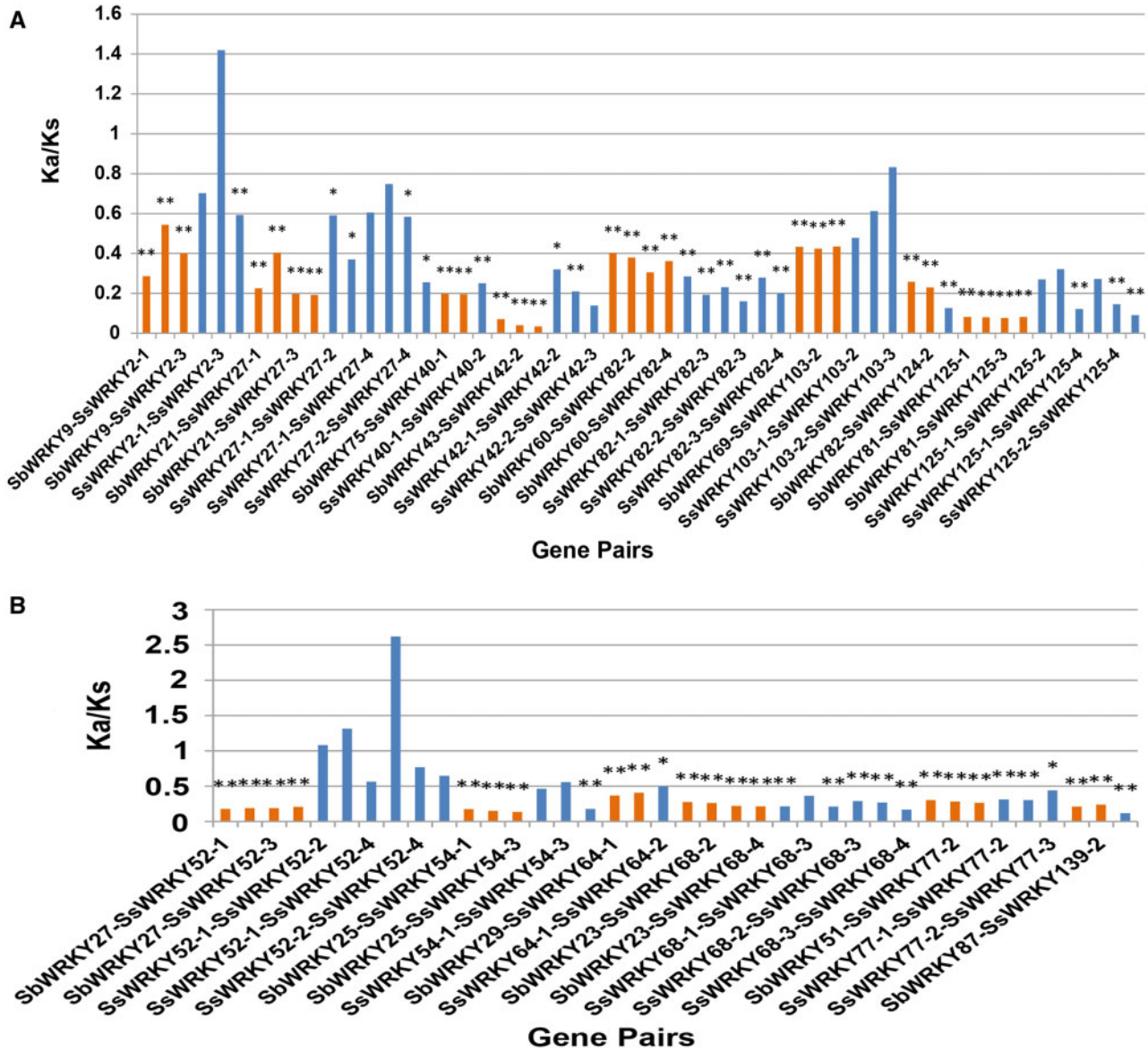


Fig. 4 The Ka/Ks of *SsWRKY* alleles and *SsWRKY–SbWRKY*. (A) The Ka/Ks of *WRKYs* in group I. (B) The Ka/Ks of *WRKYs* in subgroup IIb. The blue indicates the Ka/Ks of *SsWRKY* alleles, and the orange indicates the Ka/Ks of gene orthologs between sorghum and *S. spontaneum*. The *P*-value of <0.05 is indicated by *. The *P*-value of <0.01 is indicated by **.

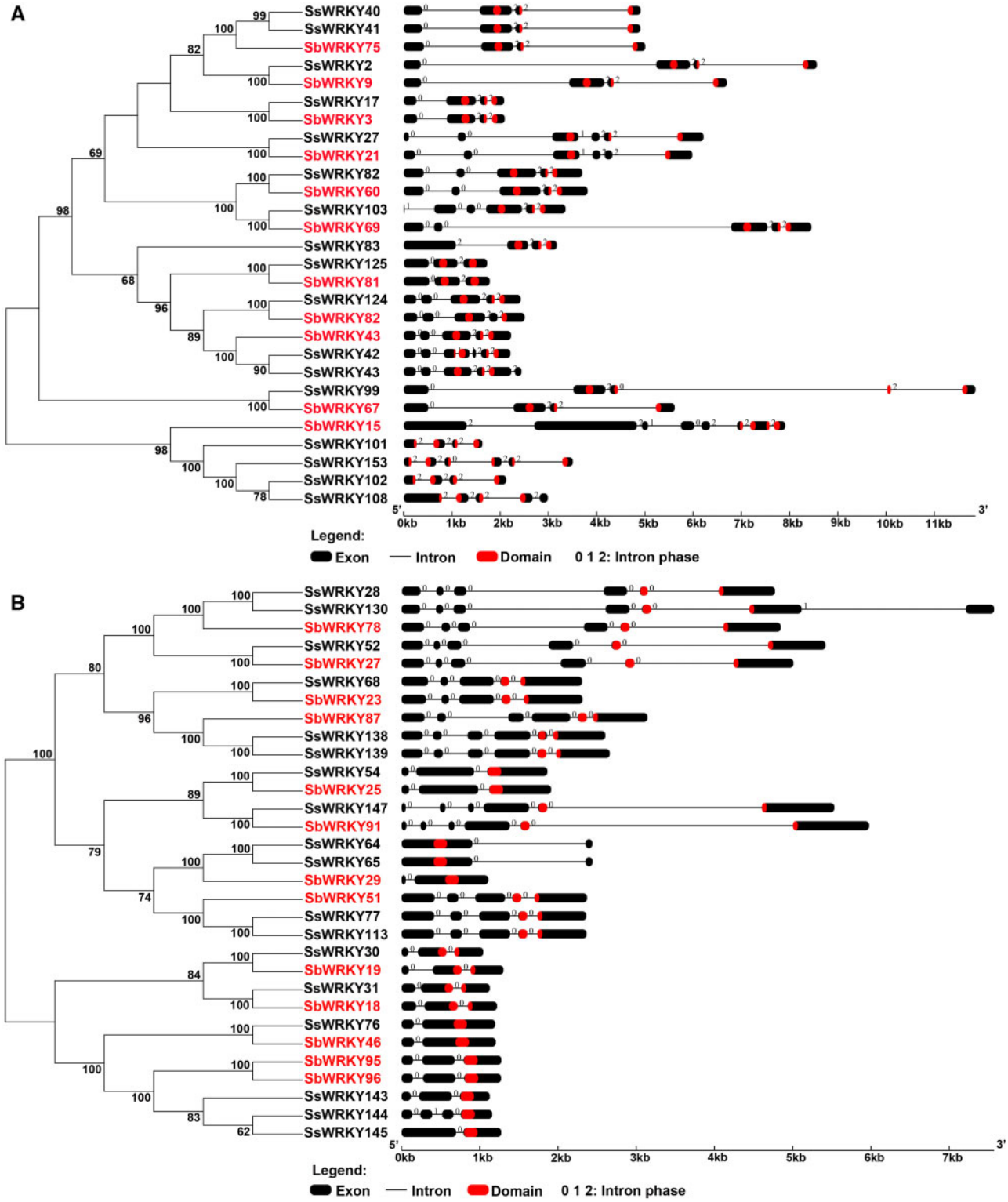


Fig. 5 Comparison of gene structure of the *WRKYs* between AP85-441 (*S. spontaneum*) and *S. bicolor*. (A) Comparison of the *WRKY* gene structures in group I. (B) Comparison of *WRKY* gene structures in subgroups IIa and IIb. Exons and introns are represented by black boxes and lines, respectively. The WRKY domains are highlighted by red boxes, and the number indicates the phases of corresponding introns: phase 0 introns interrupt two codons; phase 1 introns lie after the first base in a codon and phase 2 introns interrupt a codon between the second and third nucleotides.

two codons, a phase 1 intron is inserted after the first base in a codon and a phase 2 intron interrupts a codon between the second and third nucleotides. Almost all the sequences encoding C-terminal WRKY domains of the subgroup Ia *SsWRKY* genes contained a phase 2 intron, which is an R-type intron (Wu et al. 2005), whereas no N-terminal WRKY domains had phase 2 introns (**Fig. 5A**). On the contrary, sequences encoding both C-terminal and N-terminal WRKY domains of subgroup Ib genes contained R-type introns. The R-type intron was also present in the WRKY domain of subgroups IIc, IId and IIe and group III of *SsWRKYs*, suggesting the origin of the R-type intron preceded the majority of gene duplication in this family. In addition, the phase 0 intron (V-type intron; Wu et al. 2005) was observed in subgroups IIa and IIb (**Fig. 5B**), which provided additional evidence to support the designation of groups.

Consistent with *SsWRKYs*, *SbWRKYs* (detailed information of *SbWRKYs* provided in **Supplementary Table S6**) also exhibited similar exon/intron structures in the identical subgroups. However, some variations in exon numbers and intron sizes were observed between *SbWRKYs* and their orthologous *SsWRKYs* (**Fig. 5** and **Supplementary Fig. S8**). *SsWRKY2* contained a larger first intron than the orthologous *SbWRKY9* in group I (**Fig. 5A**). In subgroup IIa + b, the second introns of *SsWRKY138/139* were smaller than those of the orthologous *SbWRKY87* (**Fig. 5B**). *SsWRKY116* contained a larger first exon than *SbWRKY89* in subgroup IIc. In subgroup IId + e, intron expansion occurred in *SsWRKY3* and *SsWRKY74* compared with their orthologous *SbWRKY* genes. In addition, *SsWRKY25* and *SsWRKY142* possessed four exons, while their orthologous genes had only three exons in group III (**Supplementary Fig. S8**). These results demonstrate that the gaining and splitting of exons and introns expansion occurred in the *SsWRKY* family.

In comparison to sorghum, an additional 58 WRKYs were identified in *S. spontaneum*. Of them, 16 genes, including *SsWRKY26*, *SsWRKY32*, *SsWRKY83*, *SsWRKY95*, *SsWRKY110* and *SsWRKY115*, were observed to lack the orthologous *SbWRKY* genes (**Fig. 5** and **Supplementary Fig. S8**), suggesting that these *WRKYs* may have been generated after the divergence of *S. spontaneum* and sorghum. Furthermore, 23 additional *SsWRKYs* were derived from tandem duplication events (**Supplementary Table S3**).

Expression profiles of *SsWRKY* genes in *S. spontaneum*

To assess the temporal expression patterns of the sugarcane *WRKY* genes, our present study analyzed the transcriptome data of 12 different tissues and developmental stages of *S. spontaneum* (**Supplementary Fig. S9** and **Table S7**). The transcripts of 15 *SsWRKY* genes were not detected in any of these 12 samples, indicating that these genes were pseudogenes or had special expression patterns not examined in our libraries. Fifty-nine *SsWRKY* genes were expressed in all 12 samples tested (FPKM > 0), and 36 genes showed constitutive expression (FPKM > 2 in all samples). Expression profiles of *SsWRKYs* showed that *SsWRKY76* and *SsWRKY93* were highly expressed in all tissues at different developmental stages of *S. spontaneum*, suggesting that these two genes are key functional genes (**Fig. 6** and **Supplementary Table S7**). The expression levels of *SsWRKY67*, *SsWRKY68*, *SsWRKY71*, *SsWRKY78* and *SsWRKY79* increased with the maturity of the leaves, indicating that they may play an important role during the growth and development of leaves. *SsWRKY39* and *SsWRKY104* exhibited the highest expression levels in the leaves of maturing *S. spontaneum* (**Fig. 6** and **Supplementary Table S7**), while a decrease from the top to bottom of the stem was found during sugar accumulation, revealing that these two genes may play an important role in the regulation of sugar metabolism in the stem and the development of their leaves in *S. spontaneum*.

For clarifying the functional differentiation of *WRKY* genes in the source tissues (photosynthesis tissues) of *S. spontaneum*, our study exploited the continuous developmental gradient of leaf blades. According to the study of maize leaves by Li et al. (2010), the leaves were divided into four zones, including a basal zone (base, 1 cm above the leaf two ligule, sink tissue), a transitional zone (5 cm, 1 cm below the leaf one ligule, undergoing the sink-source transition), a maturing zone (10 cm, 4 cm above the leaf one ligule) and a mature zone (tip, 1 cm below the leaf two tip, fully differentiated and active C4 photosynthetic zones). The sink region is considered as the tissue from the basal zone to the sixth segment, and the source region is the tissue region from the seventh segment to the tip. Six *SsWRKYs* displayed undetectable levels indicating that these genes may play a very limited role during the development of leaves in *S. spontaneum* (**Supplementary Fig. S10** and **Table S7**). Within the gene family, *SsWRKY125* had the highest levels of expression in the leaf blade and specifically in the mature zone (fully differentiated and active C4 photosynthetic zones). The transcript abundances of *SsWRKY68/76* in the different segments gradually increased from the base to tip of the leaves while gradually reduced in *SsWRKY56*. The expression levels of *SsWRKY71* peaked in the transitional zone and were lowest in the basal zone and maturing zone. In contrast, *SsWRKY140* exhibited a peak expression in the basal region and the distal region of the leaf, while displaying lowest expression in the transitional region. Moreover, five *SsWRKY* genes (*SsWRKY21/22/39/93/109*) in the basal region of the leaves showed the highest transcript abundances (**Fig. 7** and **Supplementary Table S7**). These results were further confirmed by quantitative real-time PCR (qRT-PCR) experiments for a subset of these genes in three leaf segments of *S. spontaneum* (**Supplementary Fig. S11**).

To investigate the diel expression patterns of the *WRKYs*, we also collected RNA-seq samples at 2 h intervals over a 24-h period and 4-h intervals over an additional 24 h from *S. spontaneum*. Transcripts of four *WRKY* genes (*SsWRKY91*, *SsWRKY92*, *SsWRKY95* and *SsWRKY115*) were not detected in all samples, suggesting that these genes may play minor roles in the diel changes in physiological processes in *S. spontaneum* (**Supplementary Fig. S12** and **Table S7**). *SsWRKY93* and *SsWRKY45* showed high expression levels over the two 24-h cycles, with *SsWRKY45* having a peak expression at the start of the night period (18:00) and *SsWRKY93* significantly expressing at 8:00 (**Fig. 8** and **Supplementary Table S7**). Moreover, *SsWRKY40* was observed to have a diel peak expression at 18:00. *SsWRKY125*, *-76*, *-140* and *-129* had higher expression in the light

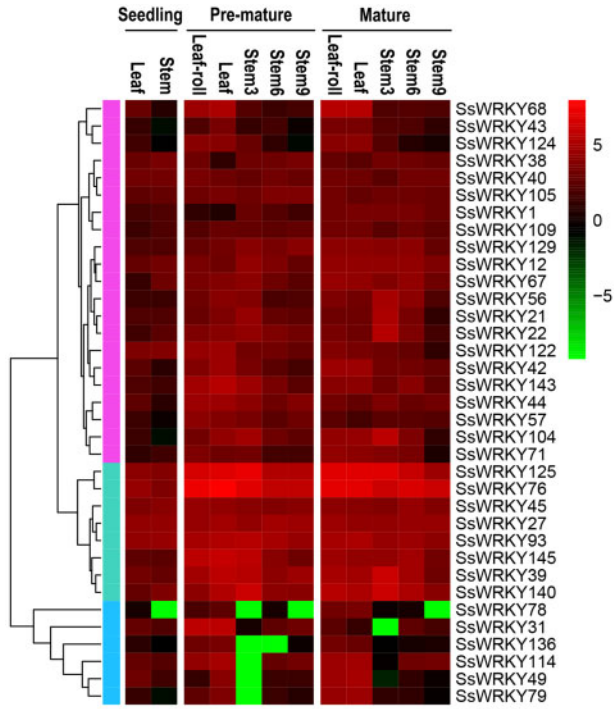


Fig. 6 Expression pattern of *WRKY* genes in different tissues and developmental stages in *S. spontaneum*. Seedling, premature and mature indicate 35-day-old, 9-month-old and 12-month-old *S. spontaneum*, respectively.

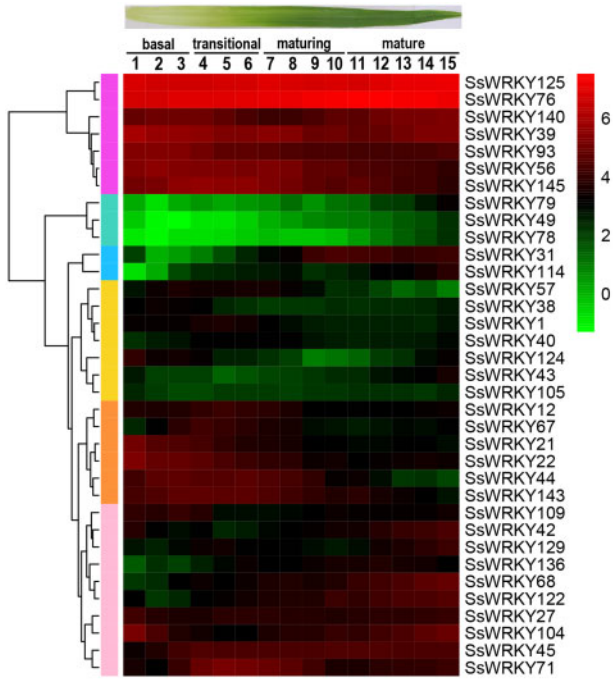


Fig. 7 Expression pattern of *WRKY* genes across leaf gradients of *S. spontaneum*.

period compared to that in dark, while *SsWRKY38* and *SsWRKY105* were expressed at peak levels during the night. Additionally, five genes, *SsWRKY44*, *SsWRKY56*, *SsWRKY71*, *SsWRKY143* and *SsWRKY145*, displayed high expression levels at 8:00, 10:00, 12:00 and 16:00 and were barely expressed during other periods (**Fig. 8** and **Supplementary Table S7**). Taken together, these findings point to the potential importance of *SsWRKYs* in diurnal rhythms. It is notable that 21 *SsWRKYs* showed specific expression levels in leaf gradient segments and in the leaves during the diurnal cycles, suggesting that these genes were likely involved in the photosynthesis process of *S. spontaneum*. We further reviewed the close orthologous genes of these 21 *SsWRKYs* in Arabidopsis using BLASTP (*E*-value $< 1e^{-20}$) and phylogenetic analysis of WRKY proteins in *S. spontaneum* and Arabidopsis, and 13 orthologous *AtWRKYs* were previously reported in functional studies (**Supplementary Table S8** and **Fig. S13**). Among them, *AtWRKY40*, which is the orthologous of *SsWRKY76* and *SsWRKY145*, inhibits expression of the light-harvesting chlorophyll *a*/*b*-binding proteins (Liu et al. 2013), the orthologous *AtWRKY46* of *SsWRKY21* and *SsWRKY22* regulates light-dependent stomatal opening in guard cells (Ding et al. 2014) and the other orthologous genes regulate the biotic stress and abiotic stress, or function in the other processes of development in Arabidopsis.

For 13 *SsWRKY* genes with four alleles in *S. spontaneum*, we analyzed the allelic expression of these *WRKYs* using the method of Zhang et al. (2018b). Of these 13 genes, seven *SsWRKY* genes displayed non-neutral expression named allelic differentially expressed genes, while the remaining six genes displayed neutral expression. Our analysis showed that *SsWRKY* alleles may play different roles in the growth and development of *S. spontaneum*.

Discussion

This present study represents the first identification and characterization of *WRKYs* from whole genome sequences of sugarcane. Previous extensive studies have been performed in the *WRKY* gene family of many other plant species, including Arabidopsis (Eulgem et al. 2000), rice (Ross et al. 2007), grape (Guo et al. 2014), soybean (Yang et al. 2017), maize (Wei et al. 2012) and *B. distachyon* (Wen et al. 2014).

WRKY transcription factors regulate many physiological processes, and various abiotic and biotic stress responses (Rushton et al. 2010, Phukan et al. 2016). In this study, 154 *WRKY* members were identified from whole genome sequences of the *S. spontaneum* and were classified into group I (17 genes), group II (81 genes), group III (51 genes) and group IV (five genes) based on the phylogenetic relationship with rice and the primary structure of the WRKY domain. Although groups I, II and III *WRKYs* have been well characterized in many plants, group IV *WRKYs* are rarely reported in the literature. The *SsWRKYs* of group IV contained an incomplete domain, and the WRKYGQK motif was only identified after re-annotation using FGENESH (**http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind**), indicating that these genes may have lost their function as *WRKYs*.

This study identified 1,360 *WRKY* genes in *S. spontaneum*, and another 22 representative species were used to further explore the evolution of the *WRKY* family. Interestingly,

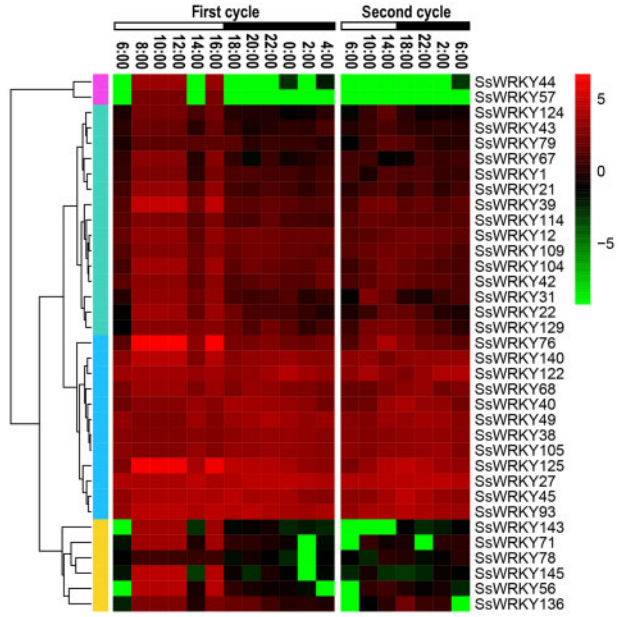


Fig. 8 Expression pattern of *WRKY* genes at different time periods in leaves of *S. spontaneum*.

compared with Arabidopsis (genome size 125 Mb), cucumber (genomic size 367 Mb), rice (genome size 480 Mb) and sorghum (genomic size 760 Mb), *S. spontaneum* (monoploid genome size 843 Mb) had larger numbers of *WRKY* family members (Ling et al. 2011, Zhang et al. 2016). In addition, the number of *WRKYs* in many higher plants was more than that of some lower plants, suggesting that *WRKY* genes might play a significant role in the evolution of simpler unicellular to more complex multicellular forms. Thus, we speculated that the number of *WRKY* genes expanded as plants evolved possibly due to genome duplication, which is in agreement with a previous study (Ning et al. 2017). Interestingly, the number of *WRKYs* in *S. spontaneum* was similar to that in wheat (C3) and maize (C4) and higher than that in sorghum (C4) and rice (C3), indicating that the number of *WRKYs* may not be related to C3 or C4 metabolism. Understandably, the gene expansion of *SsWRKY* mainly occurred after the divergence of the two C4 close relative plant species, sorghum and *S. spontaneum*.

As a proximal species of sugarcane, the number of *WRKYs* in *S. bicolor* (96) was less than that of *S. spontaneum* (154). Previous studies showed that segmental and tandem duplication might be important in the expansion of the *WRKY* gene family in species, such as white pear (Huang et al. 2015), peanut (Song et al. 2016b), *Brassica napus* (He et al. 2016) and pineapple (Xie et al. 2018). The results of collinearity analysis showed that 217 (73.8%) of the *SsWRKY* genes originated from segmental duplication and 47 (16.0%) originated from tandem duplication, revealing the high segmental and low tandem duplications presented in *SsWRKY* genes, which is consistent with those in wheat (Ning et al. 2017), grapevine (Guo et al. 2014) and soybean (Song et al. 2016a). Therefore, both tandem and segmental duplications contributed to the expansion of *SsWRKY* family, but the segmental duplication may play a more critical role. The duplications of individual genes, chromosomal segments or entire genomes could provide a primary source of material for the origin of evolutionary novelties, including new gene functions and expression patterns (Lynch and Conery 2000). Thus, the gene expansion of *S. spontaneum* may contribute to the neofunctionality of *WRKYs*.

Multiple sequence alignments of SsWRKYs revealed many variations in WRKY domains, including WRKYGEK, WRKYGKK, WRKYGHK, WEKFGEK, WRRTCRR, WEACAQN, WKNLGMQ, WTNLGLQ, WKACAQN, WTTYSQK, WKKNQIH and WRKL. The WRKYGQK motif showed binding site preferences with the W-box, while the *WRKY* genes without the WRKYGQK motif may recognize binding sequences other than W-box as previous reports suggest (Ciolkowski et al. 2008). In the case of NtWRKY12, which contains a WRKYGKK motif, it appeared to bind a SURE-like element (TTTCCAC), as opposed to a W-box element (van Verk et al. 2008). In soybean, GmWRKY6 and GmWRKY21 contained a WRKYGKK motif, which could not bind normally to the W-box (Zhou et al. 2008). The mutation in the WRKYGQK motif influenced the normal interaction of *WRKY* genes with the downstream target genes, suggesting that these amino acids may be important for binding site recognition. Notably, the variants of WRKYGQK motif were mainly observed in subgroup IIc and group III of sugarcane, indicating that the *WRKYs* of subgroup IIc and group III may have a variety of biological functions.

The diversification of exon-intron structures plays an important role in the evolution of many gene families, and the rearrangement and fusion of different chromosome fragments result in the gain and loss of exons-introns (Xu et al. 2012, Guo et al. 2013). Our current study showed that *SsWRKYs* belonging to the same group share similar exon-intron structures, with the exception of the group III genes: *SsWRKY97-1* possessed 13 exons and *SsWRKY97-2* possessed only five exons. In addition, *SsWRKY* genes in subgroup IIc contained 2–6 exons, while approximately 75% (30/40) had 2–3 exons. Sorghum is one of the closest diploid relatives of the *Saccharum* genus. Comparative analysis of the orthologous between *SsWRKYs* and *SbWRKYs* made it possible to investigate the specific evolutionary events after the polyploidization of *Saccharum* (Zhang et al. 2016). Among the *SsWRKY* genes, the intron of *SsWRKY169* was larger than its orthologous *SbWRKY58* and *SsWRKY116* contained a larger first exon than its orthologous gene *SbWRKYY89*. Moreover, the number of exons/introns in *SsWRKY121* and *SsWRKY151* were greater than that of their orthologous genes. These results suggested that *SsWRKY* family was undergoing gene expansion following polyploidization in *S. spontaneum*.

Gene expressional patterns are highly correlated with their functions in plants (Niehrs and Pollet 1999). In our study, the expression profiles revealed the different expression patterns for each *SsWRKY* in different tissues, providing a valuable resource for future functional research. Sun et al. (2003) found that SUSIBA2 (HvWRKY46) was involved in sugar signaling in barley, while in our study, *SsWRKY21*, *SsWRKY22*, *SsWRKY39*, *SsWRKY56*, *SsWRKY104*, *SsWRKY125* and *SsWRKY140* presented a peak expression level in stem 3 during the maturity, and the transcripts in stems gradually decreased from the distal to near-terrestrial end, indicating that these genes may participate in

sugar metabolism. In monocots, leaf development and photosynthetic differentiation proceeded from base to tip in a highly regular and continuous manner (Sharman 1942, Leech et al. 1973, Evert et al. 1996). According to the research of Li et al. (2010), the mesophyll and bundle sheath cells in the basal region of the leaf were undifferentiated and their basic cellular functions were still activated in this region, e.g. DNA synthesis, cell wall synthesis and hormone signaling. The current study found that the expression levels of *SsWRKY21*, *SsWRKY22*, *SsWRKY39*, *SsWRKY56*, *SsWRKY93* and *SsWRKY124* were higher in the basal region of the leaves, indicating the involvement of these genes in basic cellular activities. In the region of transition from sink to source tissues, the transcript abundance of genes associated with the photosynthetic machinery was increased (Li et al. 2010) and the expression levels of *SsWRKY71* and *SsWRKY145* were dominant in the region of transition from sink to source tissues, suggesting that these two genes were associated with the photosynthetic machinery. The distal region of the leaf was fully differentiated and had the highest levels of photosynthesis, with *SsWRKY68*, *SsWRKY122* and *SsWRKY125* exhibiting peak expression in the distal region of the leaves, further indicating that these three *WRKY* genes may be highly correlated with the photosynthetic reactions in leaves. The 24-h 'biological clock' in plants known as circadian rhythm gives plants an innate ability to measure time, allowing them to anticipate daily changes in the environment and to coordinate the developmental and metabolic processes induced by the environmental factors, such as light and temperature (Green et al. 2002, Michael et al. 2003, Dodd et al. 2005, Ni et al. 2009). It is well documented that the expression levels of several plant genes are affected by diurnal control (Giuliano et al. 1988, Becker et al. 1992, Xu et al. 2017). In *S. spontaneum*, the expression of *SsWRKY38*, *SsWRKY40*, *SsWRKY45*, *SsWRKY56*, *SsWRKY67*, *SsWRKY71*, *SsWRKY145* and another 22 *SsWRKY* genes fluctuated in the mature leaves during the diurnal cycle, suggesting that the transcriptional levels of these *SsWRKY* genes may be subject to diurnal control. Interestingly, *SsWRKY12*, *SsWRKY27*, *SsWRKY40*, *SsWRKY39*, *SsWRKY44*, *SsWRKY45*, *SsWRKY56*, *SsWRKY67*, *SsWRKY68*, *SsWRKY71*, *SsWRKY93*, *SsWRKY109*, *SsWRKY114*, *SsWRKY124*, *SsWRKY125*, *SsWRKY129* and *SsWRKY140* correlate with photosynthesis in sugarcane rather than their orthologous *WRKYs* in Arabidopsis, revealing that *WRKY* genes function differently in monocots and dicots. In conclusion, the expression profiles of *SsWRKYs* in this study provided rich resources to further investigate the functions of *SsWRKYs* in *S. spontaneum*.

Conclusions

In this present study, 154 *SsWRKY* genes were initially identified from the whole *S. spontaneum* genome and were categorized into three main groups. These *SsWRKY* genes were unevenly distributed among 32 chromosomes of *S. spontaneum*. Phylogenetic and collinearity analyses provided some valuable clues to the evolutionary characteristics of *SsWRKY* genes. The collinearity analysis indicated gene duplication contributed to the expansion of the *WRKY* gene family in *S. spontaneum*. Expression analysis suggested that *SsWRKY* genes may have played a key role in the process of development, sugar metabolism and photosynthesis in *S. spontaneum*. Overall, this systematic analysis provided valuable information to understanding the biological functions of the *SsWRKY* genes in sugarcane.

Materials and Methods

Plant materials

Sugarcane seedling stem and leaf tissues were obtained from 35-day-old *S. spontaneum* (SES-208, $2n = 8x = 64$). Tissue samples from leaf roll, leaf, top internode (internode number 3), maturing internode (internode number 6) and mature internode (internode number 9) were collected from premature 9-month-old and mature 12-month-old *S. spontaneum* as previously described (Zhang et al. 2016, Chen et al. 2017b). The 11-day-old *S. spontaneum* seedlings were grown with lamps at a light intensity of 350 μmol/m^2/s, 14:10 L/D, 30°C L/22°C D and 60% relative humidity and used for harvesting the second leaves, which were cut into fifteen 1-cm segments according to the method of Li et al. (2010). Samples were pooled from an average of four plants per biological replicate, and three biological replicates in total were collected on different dates.

Finally, as described by Ming et al. (2015), the middle 4-cm part of the first leaves of *S. spontaneum* was collected from a field at Fujian Agriculture and Forestry University (Fuzhou, Fujian, 119°16′48″E, 26°4′48″N) during the sugar accumulation stage every 2 h (6:00, 8:00, 10:00, 12:00, 14:00, 16:00, 18:00, 20:00, 22:00, 24:00, 2:00, 4:00) and every 4 h (6:00, 10:00, 14:00, 18:00, 20:00, 2:00, 6:00) from March 2 to 3, 2017.

These tissues were wrapped in tin foil paper and frozen using liquid nitrogen and stored −80°C for RNA extraction.

Sequence retrieval

WRKY sequences of rice, *A. thaliana*, maize, wheat, *B. distachyon*, grape, tomato, poplar (*Populus trichocarpa*) and cucumber were obtained from rice genome annotation project (**http://rice.plantbiology.msu.edu/**) (Kawahara et al. 2013), TAIR (The Arabidopsis Information Resource) (**http://www.arabidopsis.org/**) and Plant Transcription Factor Database (**http://planttfdb.cbi.pku.edu.cn/**) (Jin et al. 2017). The genomic data of *S. bicolor* and other species were obtained from Phytozome (**https://phytozome.jgi.doe.gov/pz/portal.html**) and NCBI (**https://www.ncbi.nlm.nih.gov/**). The genomic data of *S. spontaneum* was generated in our own laboratory (accession number in GenBank: QVOL00000000) (Zhang et al. 2018b).

Identification of WRKY in *S. spontaneum*

A HMM profile of the WRKY DBD (PF03106) was obtained from the Pfam protein family database (**http://pfam.xfam.org/**) (Finn et al. 2016) and used to identify putative WRKYs from the *S. bicolor* genome sequence through the BLASTP program and default parameters (Ling et al. 2011). Next, the selected sorghum WRKYs were used as query sequences to carry out BLASTP searches against the predicted sequences of *S. spontaneum*. The sequences with *E*-value $<1e^{-10}$ were selected for further analysis, and the HMM profile of WRKY domains was applied as the query to survey all potential proteins. Finally, the physical and chemical properties including number of amino acids (NA), MW, theoretical pI, GRAVY, AI and II of the putative SsWRKYs were calculated using the online ExPASy-ProtParam tool (**http://web.expasy.org/protparam/**). Manual annotation was performed for the genes that were incorrectly predicted.

Multiple sequence alignment, phylogenetic analysis and classification of sugarcane WRKYs

A total of 294 predicted SsWRKY proteins were included in multiple sequence alignments using DNAMAN. The sequences of domains of SsWRKYs and OsWRKYs were aligned using MUSCLE in MEGA 7.0 with default parameters (Kumar et al. 2016). The phylogenetic tree (**Fig. 2**) based on the alignments was

constructed using MEGA 7.0 with the NJ method with the bootstrap test replicated 1,000 times, the Poisson model and pairwise deletion. To verify the classification of the phylogenetic tree based on NJ, an additional phylogenetic tree (**Supplementary Fig. S2**) was constructed using the ML method with 1,000 bootstrap replicates, the Jones-Taylor-Thornton (JTT) model, gamma distribution and partial deletion (Kumar et al. 2016). Based on the multiple sequence alignment and the previously reported classification of *OsWRKYs*, the *SsWRKYs* were assigned to different groups and subgroups (Ross et al. 2007, Rushton et al. 2010, Xu et al. 2016).

Exon-intron structure of *SsWRKY* genes

The exon-intron structures of sugarcane *WRKY* genes were determined based on their coding sequence alignments and their respective genomic sequences, while diagrams were obtained from the online program Gene Structure Display Server (**http://gsds.cbi.pku.edu.cn/**) (Hu et al. 2015).

Calculation of Ka/Ks

The non-synonymous (Ka) and synonymous (Ks) substitution ratios were calculated by the easy_Kaks calculation program (Zhang 2018). Meanwhile, Fisher's exact test for small samples was applied to verify the Ka and Ks calculated by this method (Wang et al. 2009). The divergence time (T) was calculated by $T = \mathrm{Ks}/(2 \times 6.1 \times 10^{-9}) \times 10^{-6}$ Mya (Lynch and Conery 2000).

Chromosomal locations and collinearity analysis for all *SsWRKYs*

The physical location of *SsWRKYs* on the chromosomes was obtained from the database of *S. spontaneum* genome. To analyze the duplication pattern for each *SsWRKY* gene, the BLASTP program (*E*-value $<1e^{-5}$) and Multiple Collinearity Scan toolkit (MCScanX) were used (Wang et al. 2012).

Analysis of the expression profiles of WRKYs in *S. spontaneum* based on RNA-seq

The cDNA library preparation was performed according to the manufacturer's protocol (TruSeq® RNA; Illumina). The RNA-seq libraries were pooled and sequenced with 100-nt paired-end primers using an Illumina HiSeq2500 platform at the Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University. Raw data were aligned to reference gene models using TRINITY (Griffith et al. 2018). RNA-seq quantitative analysis was completed through Trinity Transcript Quantification, and the fragments per kilobase of exon model per million mapped reads (FPKM) value of the gene was calculated using the RESM (RNA-Seq by Expectation-Maximization) method.

Experimental validation of *SsWRKY* gene expression levels by qRT-PCR

The expression levels of four *SsWRKY* genes (*SsWRKY56*, *SsWRKY76*, *SsWRKY93* and *SsWRKY140*) in three tissues (the sixth, 10th and 15th segments of 11-day-old second leaves) of *S. spontaneum* were validated by qRT-PCR. Integrated DNA Technologies (**http://www.idtdna.com/Primerquest/Home/Index**) were used to design gene-specific primer pairs (**Supplementary Table S9**). Total RNA of tissues was extracted using a Plant RNA kit (R6827-01; Omega Bio-Tek). First-strand cDNA was synthesized with the TransScript All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (One-Step gDNA Removal) (TransGen Biotech). The real-time qPCR was performed with SYBR green (TaKaRa Biotechnology) on a Multicolor Real-Time PCR Detection System (Bio-Rad). Reaction parameters for thermal cycling were 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s, finally a melting curve (65–95°C, at the increments of 0.5°C) performed to confirm the PCR specificity. The glyceraldehyde-3-phosphate dehydrogenase gene and eukaryotic elongation factor 1a were selected as internal standards for normalization, and tree replicates were used (Ling et al. 2014). The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001).

Supplementary Data

Supplementary data are available at PCP online.

Funding

The National Key Research and Development Program [2018YFD1000104]; the National High-tech R&D Program [2013AA100604]; the National Natural Science Foundation of China [31201260, 31760413 and 31660420]; the Science and Technology Major Project of Guangxi [AA17202025]; and the Fujian Provincial Department of Education [JA12082].

Disclosures

The authors have no conflicts of interest to declare.

References

Becker, T.W., Caboche, M., Carrayol, E. and Hirel, B. (1992) Nucleotide sequence of a tobacco cDNA encoding plastidic glutamine synthetase and light inducibility, organ specificity and diurnal rhythmicity in the expression of the corresponding genes of tobacco and tomato. *Plant Mol. Biol.* 19: 367–379.

Brand, L.H., Fischer, N.M., Harter, K., Kohlbacher, O. and Wanke, D. (2013) Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and in vitro binding assays. *Nucleic Acids Res.* 41: 9764–9778.

Chen, J., Nolan, T.M., Ye, H., Zhang, M., Tong, H., Xin, P., et al. (2017a) Arabidopsis WRKY46, WRKY54, and WRKY70 transcription factors are involved in brassinosteroid-regulated plant growth and drought responses. *Plant Cell* 29: 1425–1439.

Chen, Y., Zhang, Q., Hu, W., Zhang, X., Wang, L., Hua, X., et al. (2017b) Evolution and expression of the fructokinase gene family in *Saccharum*. *BMC Genomics* 18: 197.

Ciolkowski, I., Wanke, D., Birkenbihl, R.P. and Somssich, I.E. (2008) Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol. Biol.* 68: 81–92.

Ding, Z.J., Yan, J.Y., Xu, X.Y., Yu, D.Q., Li, G.X., Zhang, S.Q., et al. (2014) Transcription factor WRKY46 regulates osmotic stress responses and stomatal movement independently in Arabidopsis. *Plant J.* 79: 13–27.

Dodd, A.N., Salathia, N., Hall, A., Kevei, E., Toth, R., Nagy, F., et al. (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science (New York, NY)* 309: 630–633.

Eulgem, T., Rushton, P.J., Robatzek, S. and Somssich, I.E. (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5: 199–206.

Evert, R.F., Russin, W.A. and Bosabalidis, A.M. (1996) Anatomical and ultrastructural changes associated with sink-to-source transition in developing maize leaves. *Int. J. Plant Sci.* 157: 247–261.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44: D279–D285.

Giacomelli, J.I., Ribichich, K.F., Dezar, C.A. and Chan, R.L. (2010) Expression analyses indicate the involvement of sunflower WRKY transcription factors in stress responses, and phylogenetic reconstructions reveal the existence of a novel clade in the Asteraceae. *Plant Sci.* 178: 398–410.

Giuliano, G., Hoffman, N.E., Ko, K., Scolnik, P.A. and Cashmore, A.R. (1988) A light-entrained circadian clock controls transcription of several plant genes. *EMBO J.* 7: 3635–3642.

Green, R.M., Tingay, S., Wang, Z.Y. and Tobin, E.M. (2002) Circadian rhythms confer a higher level of fitness to Arabidopsis plants. *Plant Physiol.* 129: 576–584.

Griffith, M., Griffith, O., Haas, B., Walker, J., Wagner, A., Ainscough, B., et al. (2018) TRINITY. **https://github.com/trinityrnaseq/trinityrnaseq/wiki** (October 1, 2018, date last accessed).

Guo, C., Guo, R., Xu, X., Gao, M., Li, X., Song, J., et al. (2014) Evolution and expression analysis of the grape (*Vitis vinifera* L.) WRKY gene family. *J. Exp. Bot.* 65: 1513–1528.

Guo, R., Xu, X., Carole, B., Li, X., Gao, M., Zheng, Y., et al. (2013) Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape. *BMC Genomics* 14: 554.

Guo, D., Zhang, J., Wang, X., Han, X., Wei, B., Wang, J., et al. (2015) The WRKY transcription factor WRKY71/EXB1 controls shoot branching by transcriptionally regulating RAX genes in Arabidopsis. *Plant Cell* 27: 3112–3127.

He, Y., Mao, S., Gao, Y., Zhu, L., Wu, D., Cui, Y., et al. (2016) Genome-wide identification and expression analysis of WRKY transcription factors under multiple stresses in *Brassica napus*. *PLoS One* 11: e0157558.

Holub, E.B. (2001) The arms race is ancient history in Arabidopsis, the wildflower. *Nat. Rev. Genet.* 2: 516–527.

Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J. and Gao, G. (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics (Oxford, England)* 31: 1296–1297.

Huang, S., Gao, Y., Liu, J., Peng, X., Niu, X., Fei, Z., et al. (2012) Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. *Mol. Genet. Genomics* 287: 495–513.

Huang, X., Li, K., Xu, X., Yao, Z., Jin, C. and Zhang, S. (2015) Genome-wide analysis of WRKY transcription factors in white pear (*Pyrus bretschneideri*) reveals evolution and patterns under drought stress. *BMC Genomics* 16: 1104.

Irvine, J.E. (1999) *Saccharum* species as horticultural classes. *Theor. Appl. Genet.* 98: 186–194.

Ishiguro, S. and Nakamura, K. (1994) Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 59 upstream regions of genes coding for sporamin and b-amylase from sweet potato. *Mol. Gen. Genet.* 244: 563–571.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., et al. (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45: D1040–D1045.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (New York, NY)* 6: 4.

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33: 1870–1874.

Lam, E., Shine, J., DA Silva, J., Lawton, M., Bonos, S., Calvino, M., et al. (2009) Improving sugarcane for biofuel engineering for an even better feedstock. *GCB Bioenergy* 1: 251–255.

Leech, R.M., Rumsby, M.G. and Thomson, W.W. (1973) Plastid differentiation, acyl lipid, and fatty acid changes in developing green maize leaves. *Plant Physiol.* 52: 240–245.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., et al. (2010) The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* 42: 1060–1067.

Ling, J., Jiang, W., Zhang, Y., Yu, H., Mao, Z., Gu, X., et al. (2011) Genome-wide analysis of WRKY gene family in *Cucumis sativus*. *BMC Genomics* 12: 471.

Ling, H., Wu, Q., Guo, J., Xu, L. and Que, Y. (2014) Comprehensive selection of reference genes for gene expression normalization in sugarcane by real time quantitative RT-PCR. *PLoS One* 9: e97469.

Liu, R., Xu, Y.-H., Jiang, S.-C., Lu, K., Lu, Y.-F., Feng, X.-J., et al. (2013) Light-harvesting chlorophyll a b-binding proteins, positively involved in abscisic acid signalling, require a transcription repressor, WRKY40, to balance their function. *J. Exp. Bot.* 64: 5443–5456.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25: 402–408.

Lynch, M. and Conery, J. (2000) The evolutionary fate and consequences of duplicate genes. *Science (New York, NY)* 290: 1151–1155.

Michael, T.P., Salome, P.A., Yu, H.J., Spencer, T.R., Sharp, E.L., McPeek, M.A., et al. (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science (New York, NY)* 302: 1049–1053.

Ming, R., VanBuren, R., Wai, C.M., Tang, H., Schatz, M.C., Bowers, J.E., et al. (2015) The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47: 1435–1442.

Moore, P., Nagai, C. and Fitch, M. (1989) Production and evaluation of sugarcane haploids. *Proc. Int. Soc. Sugar Cane Technol.* 20: 599–607.

Ni, Z., Kim, E.D., Ha, M., Lackey, E., Liu, J., Zhang, Y., et al. (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457: 327–331.

Niehrs, C. and Pollet, N. (1999) Synexpression groups in eukaryotes. *Nature* 402: 483–487.

Ning, P., Liu, C., Kang, J. and Lv, J. (2017) Genome-wide analysis of WRKY transcription factors in wheat (*Triticum aestivum* L.) and differential expression under water deficit condition. *PeerJ* 5: e3232.

Phukan, U.J., Jeena, G.S. and Shukla, R.K. (2016) WRKY transcription factors: molecular regulation and stress responses in plants. *Front. Plant Sci.* 7: 760.

Potschin, M., Schlienger, S., Bieker, S. and Zentgraf, U. (2014) Senescence networking WRKY18 is an upstream regulator, a downstream target gene, and a protein interaction partner of WRKY53. *J. Plant Growth Regul.* 33: 106–118.

Ross, C.A., Liu, Y. and Shen, Q.J. (2007) The WRKY gene family in rice (*Oryza sativa*). *J. Integr. Plant Biol.* 49: 827–842.

Rushton, P.J., Bokowiec, M.T., Han, S., Zhang, H., Brannock, J.F., Chen, X., et al. (2008) Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. *Plant Physiol.* 147: 280–295.

Rushton, P.J., Macdonald, H., Huttly, A.K., Lazarus, C.M. and Hooley, R. (1995) Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of a-Amy2 genes. *Plant Mol. Biol.* 29: 691–702.

Rushton, P.J., Somssich, I.E., Ringler, P. and Shen, Q.J. (2010) WRKY transcription factors. *Trends Plant Sci.* 15: 247–258.

Sharman, B.C. (1942) Developmental anatomy of the shoot of *Zea mays* L. *Ann. Bot.* 6: 245–284.

Skibbe, M., Qu, N., Galis, I. and Baldwin, I.T. (2008) Induced plant defenses in the natural environment: *Nicotiana attenuata* WRKY3 and WRKY6 coordinate responses to herbivory. *Plant Cell* 20: 1984–2000.

Song, H., Wang, P., Hou, L., Zhao, S., Zhao, C., Xia, H., et al. (2016a) Global analysis of WRKY genes and their response to dehydration and salt stress in soybean. *Front. Plant Sci.* 7: 9.

Song, H., Wang, P., Lin, J.Y., Zhao, C., Bi, Y. and Wang, X. (2016b) Genome-wide identification and characterization of WRKY gene family in peanut. *Front. Plant Sci.* 7: 534.

Sun, C., Palmqvist, S., Olsson, H., Borén, M., Ahlandsberg, S. and Jansson, C. (2003) A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. *Plant Cell* 15: 2076–2092.

Tripathi, P., Rabara, R.C., Langum, T.J., Boken, A.K., Rushton, D.L., Boomsma, D.D., et al. (2012) The WRKY transcription factor family in *Brachypodium distachyon*. *BMC Genomics* 13: 270.

van Verk, M.C., Pappaioannou, D., Neeleman, L., Bol, J.F. and Linthorst, H.J. (2008) A Novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. *Plant Physiol.* 146: 1983–1995.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., et al. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40: e49.

Wang, M., Vannozzi, A., Wang, G., Liang, Y.H., Tornielli, G.B., Zenoni, S., et al. (2014) Genome and transcriptome analysis of the grapevine (*Vitis vinifera* L.) WRKY gene family. *Hortic. Res.* 1: 14016.

Wang, D., Wan, H., Zhang, S. and Yu, J. (2009) Gamma-MYN a new algorithm for estimating Ka and Ks with consideration of variable substitution rates. *Biol. Direct* 4: 20–18.

Wei, K.F., Chen, J., Chen, Y.F., Wu, L.J. and Xie, D.X. (2012) Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Res.* 19: 153–164.

Wen, F., Zhu, H., Li, P., Jiang, M., Mao, W., Ong, C., et al. (2014) Genome-wide evolutionary characterization and expression analyses of WRKY family genes in *Brachypodium distachyon*. *DNA Res*. 21: 327–339.

Wu, K.L., Guo, Z.J., Wang, H.H. and Li, J. (2005) The WRKY family of transcription factors in rice and Arabidopsis and their origins. *DNA Res*. 12: 9–26.

Wu, Z.J., Li, X.H., Liu, Z.W., Li, H., Wang, Y.X. and Zhuang, J. (2016) Transcriptome-wide identification of *Camellia sinensis* WRKY transcription factors in response to temperature stress. *Mol. Genet. Genomics* 291: 255–269.

Xie, T., Chen, C., Li, C., Liu, J., Liu, C. and He, Y. (2018) Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. *BMC Genomics* 19: 490.

Xie, Z., Zhang, Z.L., Zou, X., Huang, J., Ruas, P., Thompson, D., et al. (2005) Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol.* 137: 176–189.

Xu, G., Guo, C., Shan, H. and Kong, H. (2012) Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. USA* 109: 1187–1192.

Xu, Z., Ma, J., Qu, C., Hu, Y., Hao, B., Sun, Y., et al. (2017) Identification and expression analyses of the alanine aminotransferase (AlaAT) gene family in poplar seedlings. *Sci. Rep.* 7: 45933.

Xu, H., Watanabe, K.A., Zhang, L. and Shen, Q.J. (2016) WRKY transcription factor genes in wild rice *Oryza nivara*. *DNA Res.* 23: 311–323.

Yang, Y., Zhou, Y., Chi, Y., Fan, B. and Chen, Z. (2017) Characterization of soybean WRKY gene family and identification of soybean WRKY genes that promote resistance to soybean cyst nematode. *Sci. Rep.* 7: 17804.

Yokotani, N., Sato, Y., Tanabe, S., Chujo, T., Shimizu, T., Okada, K., et al. (2013) WRKY76 is a rice transcriptional repressor playing opposite roles in blast disease resistance and cold stress tolerance. *J. Exp. Bot*. 64: 5085–5097.

Yu, Y., Liu, Z., Wang, L., Kim, S.G., Seo, P.J., Qiao, M., et al. (2016) WRKY71 accelerates flowering via the direct activation of FLOWERING LOCUS T and LEAFY in *Arabidopsis thaliana*. *Plant J.* 85: 96–106.

Zhang, X. (2018) Easy_KaKs. **https://github.com/tangerzhang/FAFU-cgb/tree/master/easy_KaKs** (October 1, 2018, date last accessed).

Zhang, J., Arro, J., Chen, Y. and Ming, R. (2013) Haplotype analysis of sucrose synthase gene. *BMC Genomics* 14: 314.

Zhang, Q., Hu, W., Zhu, F., Wang, L., Yu, Q., Ming, R., et al. (2016) Structure, phylogeny, allelic haplotypes and expression of sucrose transporter gene families in *Saccharum*. *BMC Genomics* 17: 88.

Zhang, Y. and Wang, L. (2005) The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol. Biol*. 5:1.

Zhang, J., Zhang, Q., Li, L., Tang, H., Zhang, Q., Chen, Y., et al. (2018a) Recent polyploidization events in three *Saccharum* founding species. *Plant Biotechnol. J.* 17:264–274.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018b) Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50: 1565–1573.

Zhou, Q.Y., Tian, A.G., Zou, H.F., Xie, Z.M., Lei, G., Huang, J., et al. (2008) Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnol. J.* 6: 486–503.