

NAC transcription factors in autopolyploid *Saccharum spontaneum*: genome-wide identification, expression pattern and a 'Dry' orthologous gene

Hengbo Wang

Fujian Agriculture and Forestry University

Yongjun Wang

Fujian Agriculture and Forestry University

Naiyan Xiao

Fujian Agriculture and Forestry University

Xiuting Hua

Fujian Agriculture and Forestry University

Muqing Zhang

GuangXi University

Ray Ming

Fujian Agriculture and Forestry University

Jisen Zhang (✉ zjisen@fafu.edu.cn)

Fujian Agriculture and Forestry University <https://orcid.org/0000-0002-8435-6615>

Research article

Keywords: NAC transcription factor, Genome-wide, tissue-specific expression, Sugarcane

Posted Date: November 8th, 2019

DOI: <https://doi.org/10.21203/rs.2.17032/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The NAC acronym originates from these three genes (NAM, ATA1/2 and CUC2), and were proved to participate in plant development, sugar accumulation and stress tolerance. However, little information is available regarding these genes in sugarcane. The newly published sugarcane genome provided an opportunity to identify the SsNAC transcription factors.

Results: In this study, a total of 151 NAC genes including 327 alleles were identified in the autopolyploid *S. spontaneum* genome and were distributed unevenly on eight chromosomes with the majority located on Chr 5(54, 16.51%), Chr 1(51, 15.60%) and Chr 2(51, 15.60%). 71.6%(234) and 12.53%(41) of SsNAC genes were mainly derived from segmental duplication and tandem duplication. Phylogeny analysis of NAC TF proteins by comparing NAC between sugarcane and *Arabidopsis thaliana* suggested that the NAC family can be divided into 15 subgroups with one subgroup unclassified. RNA-seq data analysis revealed that 82 SsNAC were expressed in 15 segments of the developmental gradient of the leaf and 74 SsNAC were presented in 12 different developmental stages, respectively. Remarkably, SsNAC genes of the ATAF subgroup presented the highest expression levels among the subgroups, and seven of eight SsNAC genes from the ATAF subgroup excluding SsNAC30 were present at much higher expression levels in the developmental gradient of the leaf than those in different developmental tissues types, indicating that the ATAF subgroup may have significant roles and participate in leaf growth and development and photosynthesis. Importantly, the NAC1 subgroup SsNAC91 gene, being orthologous to Sobic.006G147400 (Dry gene), displayed significantly higher expression levels in premature and mature stems of the low sucrose and low water content species *S. spontaneum* as opposed to the orthologous gene in the high sugar and high water content species *S. officinarum* . This suggests that the SsNAC91 gene may also play important roles in cellulose biosynthesis and water transport in sugarcane as it does in sorghum.

Conclusions: This study provided the basis for the comprehensive genomic study of the SsNAC gene family and thus established a good foundation for the functional analyses of SsNAC genes which can be utilized to breed new varieties of sugarcane.

Background

During the process of evolution, plants encounter numerous stresses in nature, such as pathogen infection, drought, waterlogging and extreme temperatures, which can seriously affect growth, development and breeding. Plants have evolved numerous mechanisms for dealing with detrimental conditions [1]. Transcription factors (TFs) regulate genes expression and can silence genes to help enable the survival of many stress responses by binding to the cis-acting elements in the gene promoter regions that activate or repress gene expression. NACs are TFs and were originally described due to their shared DNA-binding domain (in petunia the NAM TF, and in *Arabidopsis* ATAF1, ATAF2 and CUC2 TFs). The NAC families have now been identified in many crops, including *Arabidopsis thaliana* [2], *Nicotiana tabacum* [3], *Glycine max* [4], *Oryza sativa* [5], *Zea mays* [6], *Vitis vinifera* [7], *Brassica rapa* [8] and *Populus trichocarpa* [9], *Sorghum bicolor* [10], *Medicago truncatula* [11], *Prunus mume* [12], *Sesamum indicum*

[13], *Pyrus* [14], *Fragaria vesca* [15], *Fagopyrum tataricum* [16]. Based on the conserved domain in the N-terminal region, the NAC family can be divided into five sub-domains (named A to E), while the transcription activation region (TAR) in the C-terminal region is highly diverse and may determine the different biological functions of the NAC members [2].

Many studies have shown that NAC transcription factors are involved in many biological functions in plants, including regulation of plant growth and development, abiotic stress response, photosynthesis, sugar accumulation and the induction of host resistance to pathogens. NAC transcription factors also promote the development of flower organs, the formation of lateral roots, the formation of secondary walls of cells, and leaf senescence [17]. In *Arabidopsis thaliana*, NACs were identified and classified with 14 subgroups in group I and 4 subgroups in group II according to homology with the rice NAC protein domain [2]. *Arabidopsis* ATAF1 expression is induced by tissue impairment, pathogenic microbial infection and salicylic acid or jasmonic acid to down-regulate expression [18]. It was proved to have a positive effect for regulating penetration resistance in basal defence [19]. The expression of *ATAF2* is induced by tissue damage, methyl jasmonate (MeJA) and salicylic acid (SA) [18]. In transgenic maize lines in which the *NAC7*(GRMZM2G114850) gene from maize was down-regulated by RNAi a delay in senescence and increase in both biomass and the accumulation of nitrogen were found in vegetative tissues [20]. In maize, *NAC7* was suggested to regulate genes involved in photosynthesis, chlorophyll degradation and protein turnover pathways based on studies using leaf protoplasts overexpressing *NAC7* and employing *nac7* RNAi leaves [20]. Recently, a NAC TF named the *Dry*, a single locus in sorghum was found to be the causal gene for control the pith stem traits and therefore relevant to sugar accumulation, which suggests that the *Dry* gene is an important first-layer master switch for secondary cell wall biosynthesis [21]. NAC secondary wall thickening promoting factor1 (NST1) and NST3, regulate the formation of secondary walls in woody tissues of *Arabidopsis thaliana* [22]. Homozygous double knockout NST1 and NST3 lines were prepared using T-DNA lines and it was found that, the inter fascicular fibers and secondary xylem of secondary wall growth were completely suppressed, whereas the formation of woody tissue cells was not affected [22]. In tomato, silencing of *SINAP2* or *SIORE1S02* (an orthologous gene of *AtORE1*) led to increased amounts of fruit with increased sugar content [23, 24]. 26 members of the NAC proteins family were identified using the SUCEST sugarcane EST database [25]. *SsNAC23* was strongly induced through chilling stress (at 4°C), indicating that this gene participated in the response to low temperature stress.

Sugarcane is significant for biofuel and sugar production and is considered a dual-purpose crop, in addition it is the crop that generates the most biomass in terms of worldwide production (1.9 billion tons of biomass produced in 2014 worldwide, FAOSTAT, 2014). Since the implementation of cross breeding, the focus on sugarcane research and breeding has been directed at selecting varieties based on three criteria: high biomass per unit area, high percentage of soluble sugar from bagasse, and high cell wall content for digestibility. Although *NAC* genes can affect secondary cell wall (SCW) biosynthesis, abiotic stress response, senescence, photosynthetic capability of leaves and sugar accumulation, it is still unclear which *SsNAC* genes are involved in secondary cell wall biosynthesis of biomass, abiotic stress responses and participate in photosynthesis metabolism. However, genome-wide identification and the

analysis of NAC protein expression has not been completed in sugarcane to date. In addition, the recent completion of the sugarcane genome sequence provides an excellent opportunity to identify and analyze the NAC TF gene family at the whole-genome level [26]. To date, only 26 members of the NAC protein have been identified using the SUCEST database[25]. In this study, a variety of bioinformatics approaches were used to identify and classify the *SsNAC* genes family in sugarcane according to similarities with *Arabidopsis thaliana* NAC TF proteins. A phylogenetic analysis was performed with 151 *SsNAC* TF genes, which analyzed gene structure, motif composition, chromosomal location and duplication events. The expression patterns of all *SsNAC* TF genes were comprehensively analyzed based on the RNA-seq transcriptome information. This study is helpful in understanding the functional characteristics of *SsNAC* and providing potential candidate genes to assist in the breeding high biomass sugarcane varieties.

Results

Identification of SsNAC genes in sugarcane

To identify *SsNAC* proteins accurately in *S. spontaneum*, an HMM search was performed using the NAC domain and 449 *SsNAC* proteins were identified genome-wide. Then, using these 137 NAC proteins of rice as query sequences [5], the 220 *SsNAC* homologous proteins were obtained by comparing the protein databases of *S. spontaneum*. The NAC protein sequences achieved by the two methods above (449 and 220 proteins) were analyzed, and repeats and the non-coding-domain - sequences of NAC were eliminated. The predicted protein sequences were further checked with NCBI-conserved domain and search Pfam database (<http://pfam.xfam.org/>) and were excluded for incomplete NAC conserved domains of proteins. Taken together, a total of 151 genes loci were predicted to encode putative NAC or NAC-like proteins and named as *SsNAC1-SsNAC151* based on order on chromosome information in the representative monoploid genome of *S. spontaneum* (Additional file 1). Detailed information on 151 *SsNAC* proteins was compiled by establishing the isoelectric point (PI), the length of amino acids (AA) and molecular weight (MW) (Additional file 1). The extremes in terms of protein sequence length are *SsNAC63* (145 aa) and *SsNAC65* (803 aa). The PI of the proteins ranged from 4.29 (*SsNAC29*) to 10.93 (*SsNAC89*) and the MWs ranged from 13.92 to 104.05 kDa.

S. spontaneum is an autopolyploid plant and gene allelic information is available. We further identified 327 alleles for the 151 *SsNAC* genes, and 91 of the 151 *SsNAC* genes have 2, 3, or 4 allelic genes. The exact composition is as follows: 39 (25.8%) with two alleles, 23 (15.2%) with three alleles and 29 (19.2%) genes with four alleles (Additional file 2).

Phylogenetic, structural, and motif analysis of SsNAC genes

To understand the evolutionary relationship and classification among *SsNAC* genes, an unrooted phylogenetic tree was constructed using the representative 151 predicted full-length *SsNAC* protein sequences excluding the alleles and 105 NAC proteins of *A. thaliana* (ANAC) [2] (Additional file 3). Based on their phylogenetic distribution with ANAC proteins, *SsNAC* proteins were divided into 15 subgroups (Fig. 1). Various subgroups corresponded to *A. thaliana* NAM, NAC1, OSNAC7, OSNAC8, TIP, ANAC011,

NAC2, ATAF, NAP/AtNAC3, SENU5, OSNAC022, TERN, ANAC063 and OsNAC003 subgroups [2], of which there were 16, 8, 15, 5, 0, 5, 6, 8, 9, 0, 18, 5, 0 and 17 *SsNAC* genes respectively. The subgroup OSNAC022 had the highest number of *SsNAC* proteins (18), but no *SsNAC* members from the TIP, SENU5 and ANAC063 subgroups were identified. In addition, *SsNAC* proteins with similar functions or composition tended to cluster together. Most subgroups were shared among NAC members from *A. thaliana* and *S. spontaneum*, but species-specific subgroups were also found. These results indicate that the functions and roles of *SsNAC* genes may differentiate compared with the ANAC proteins.

Subsequently, a phylogenetic tree was reconstructed using the full-length *SsNAC* protein sequences (151), and the *SsNAC* gene family was divided into fifteen subgroups, named I-XV (Fig. 2a). The largest subgroups I and XV had 16, 17 members respectively with the exception of the unclassified proteins (33). The smallest subgroups of both III and XIII had only 3 members each. It was not possible to classify eight members (*Sspon.002C0019710*, *Sspon.006D0007980*, *Sspon.005A0023730*, *Sspon.002A0014630*, *Sspon.002C0010760*, *Sspon.004A0012960*, *Sspon.002B0022580* and *Sspon.004A0012240*) into any subgroups is highlighted by the low bootstrap values (28–81). The structure of these genes were analyzed using the genomic DNA to ascertain the composition of introns and exons (Fig. 2b) and the same subgroups were found to have the same or similar gene structures. The number of exons ranged from 1–12. Although the majority of genes (126/151) have 1–4 exons, thirteen members have only one intron in the coding sequences in the *SsNAC* family. Subgroup II to subgroup VI and subgroup XV showed different gene structures with longer introns and more exon breaks than other subgroups.

To further determine the diversification of the *SsNAC* genes family, ten conserved motifs were identified by the MEME online software (Fig. 2c). In general, most members of *SsNAC* genes family were clustered together in the phylogenetic tree and shared common motif compositions, which indicated that these genes with similarly conserved motifs performed similar biological functions. Based on the domain composition of NAC family proteins in *A. thaliana*, the 133 of 151 *SsNAC* members contain a complete NAC DNA-binding domain (A-E) [2]. For example, motif 2 and 8 corresponded to domain A, motif 5 corresponded to domain B, motif 1 and 4 corresponded to domain C, motif 6 and 3 corresponded to domain D, motif 7 corresponded to domain E. Motifs 9 and 10 presented specific subgroups and substituted for some subdomain A-E. Subgroup XV possessed motif 9 and lacked motifs 8 and 5. Motif 10 was absent in most subgroups except for subgroup XIII, subgroup VIII and subgroup XII, which was located in the C-terminal of NAC TAR. Additionally, motif 9 and 10 in some specific subgroup were substituted respectively for motif 8 and 5, motif 4 and 6. These results suggested that functional divergence may occur in subgroup XIII and subgroup XV.

Chromosomal location and duplication events of SsNAC genes

To understand the genomic distribution of *SsNAC* genes on 32 chromosomes comprising 8 homologous group of 4 members each, the 151 representative *SsNAC* genes including 327 alleles in the autopolyploid *S. spontaneum*, were found to be unevenly distributed among 32 chromosomes (Additional file 2 and 4). The majority of *SsNAC* genes were located on Chr 5 (54, 16.51%), Chr 1 (51, 15.60%) and Chr 2 (51,

15.60%) (Additional file 4). But the chromosomes with the highest number of *SsNAC* genes were 18 on Chr 1D and Chr 3C, while only 3 and 2 *SsNAC* genes were distributed on Chr 7D and 8C respectively. The total number of chromosomes with 10 *SsNAC* gene loci or greater is 29 which includes a total of 256 genes (78.3%). A total of 13 chromosomes have less than 10 *SsNAC* gene loci and include 71 genes (21.7%). Chr 5 and 2 were rearranged for fissions of 4 ancestral chromosomes [26], explaining the high number of *SsNAC* genes observed on Chr 5 and 2.

Gene duplication events play an important role in the expansion of the gene family numbers. In order to identify the duplication events in *SsNAC* gene families, a synteny relationship was created with the alignment results of MCScanx software and the BLASTp method. According to this synteny map, there are a total of 279 pairs of *SsNAC* genes with a collinear relationship in the whole genome of *S. spontaneum*, including 133 pairs of non-alleles and 146 pairs of alleles (Fig. 3). Based on the principle of Holub [27], 234 *SsNAC* genes were derived from segmental duplication events (Additional file 5), and 41 *SsNAC* genes were assembled in 18 tandem duplication regions on Chr 1 (four clusters), Chr 2 (three clusters), Chr 3 (two clusters), Chr 5 (six clusters), Chr 6 (two clusters), Chr 8 (one clusters), whereas in Chr 4 and 7 there were no tandem duplication genes. Chr 5 had the highest gene numbers (14/41), with one of clusters (Chr 5D) containing four genes with tandem duplication. The fact that Chr 5 is a fusion of two ancestral chromosomes and is a hot spot for chromosome recombination explains why there are many tandem duplication events occurring on Chr 5 [26]. These results indicate that segregation duplication events may play an important role in expanding the number of *SsNAC* family members more than tandem duplication events.

Comparative analysis of NAC genes between S. spontaneum and its related species Sorghum bicolor

Sorghum and sugarcane are close relatives and shared a common ancestor about 8–9 million years ago [28]. For detailed analysis of the numbers and genetic structure of *NAC* genes between *Sorghum bicolor* and *S. spontaneum*, a phylogenetic tree was constructed using the protein sequences from the two species. We identified 110 *NAC* proteins of sorghum (*SbNAC*) using the 140 *NAC* proteins of rice to blast the sorghum protein database, which were downloaded from the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>). In comparison with the number of *SbNAC* genes, 41 additional *SsNAC* genes were identified, 17 of which are derived from the tandem duplication genes (Additional file 5 and 6). Other genes (24) mainly located on Chr. 1, 6 and 8, do not have an orthologous gene corresponding to sorghum, indicating that they were generated after the differentiation of sugarcane and sorghum. These results further suggested that tandem duplication and WGD or segmental duplication events are main factors that determined the expansion of the *NAC* gene family.

Furthermore, the exon/intron structural analysis was constructed between the 151 *SsNAC* and 110 *SbNAC* genes of sorghum using TBtools software. As shown in Additional file 6, the majority of *SbNAC* and *SsNAC* genes have one to six exons, such as 11 and 13 *NAC* genes with one exon, 12 and 37 with two exons, 64 and 65 with three exons, 9 and 11 with four exons, 7 and 6 with five exons and 4 and 8 with six exons respectively. These findings suggest that the number of gene exons with similar genetic

relationships between sugarcane and sorghum is very similar. The number of genes containing two exons in sorghum is significantly less than that of *S. spontaneum*. The results showed that exon fusion occurs on some branches of sugarcane (Additional file 6). Other *SsNAC* gene (11/151) exons ranged from 7–10, nine of which originated after sugarcane and sorghum differentiation. The results indicated that exon breakage occurred during the evolution of *SsNAC* genes family.

Expression patterns of SsNAC in the developmental gradient of the leaf

Sugarcane is a C₄, highly-photosynthetic crop that is bred and propagated to obtain sucrose. Leaves are the most important tissue organ as they produce sucrose by photosynthesis. To determine the function and expression patterns of the *SsNAC* genes in sugarcane according to maize leaf patterning [29], all *SsNAC* genes were analyzed based on the RNA-seq datasets of 15 developmental leaf stages in *S. spontaneum*, represented by a basal zone (1 cm, section 1 to 3), a transitional zone (5 cm, section 4 to 6), a maturing zone (10 cm, section 7–10) and mature zone (tip, 1 cm, section 11 to 15). 69 *SsNAC* genes were not detected in any of the analyzed leaf developmental stages (Transcripts Per Million, TPM <1), indicating that these genes may have special expression patterns and are not expressed in the analyzed leaf developmental stages. The other 82 *SsNAC* genes were expressed (TPM >1) in leaves at 15 different developmental stages, and these genes can generally be classified into four different patterns according to their expression pattern (low expression, low to high expression, high to low expression and double peak pattern). 27 *SsNAC* genes showed low expression patterns (1 < TPM < 8) and 39 *SsNAC* genes presented high expression patterns (TPM > 8) (Fig.4, Additional file 7), which all had expression levels that increased sharply from section 11 to 15 in the mature leaves. The results indicated that these genes may take part in photosynthesis. There were 10 *SsNAC* genes that had expression levels ranging from low to high in sections 3 to 5. The remaining 6 *SsNAC* genes had a double peak pattern. All highly expressed *SsNAC* genes displayed an increased expression level as the leaf matured, indicating that these genes may play an important role in leaf photosynthesis.

Only 39 genes of the 11 subgroups (NAM, NAC1, NAC2, ONAC8, ATAF, NAP, ONAC022, TERN, ONAC003, AtNAC3 and Unclassified) showed an increase in expression in the mature zones of the leaf, with the other *SsNAC* genes not being expressed in the same group, indicating that the *SsNAC* genes of different subgroups have functionally differentiated. Three *SsNAC* genes (*SsNAC45*, *SsNAC22* and *SsNAC14*) were highly expressed in mature zone of the leaf compared to other zones. These three genes have the average expression values of TPM over 610, suggesting that these three genes are dominant functional genes in the *SsNAC* TF gene family. Meanwhile, the top five genes with highest expression levels are members of the ATAF subgroup, with the ATAF subgroup gene functions being related to plant response biotic stress (wounding and fungi) [19]. The following genes belong to subgroup OSNAC7: *SsNAC08*, *SsNAC90*, *SsNAC142* and *SsNAC143*. Their transcript abundance gradually increased from base to transitional zone of the leaves and reduced quickly in the maturing and mature zone, suggesting that the some *SsNAC* genes of same subgroups showed similar function.

Expression pattern of SsNAC genes in different tissue types

Gene expression patterns in different tissue types can provide crucial clues to determine gene function. As an organ of sucrose transport and storage, stems play an important role in the growth and development of sugarcane, but only a few studies have analyzed the developmental stage of the stem. We investigated the gene expression levels of the *SsNAC* genes in different tissues (leaf and stem) and all genes were analyzed based on the RNA-seq datasets, which was constructed using 3 developmental stages (seedling, premature and mature) and 5 different tissues, including 2 leaf tissue types (leaf, leaf roll) and 3 stem types (seedling, premature and mature stem with different internode). A total of 77 *SsNAC* genes had no detectable expression in the examined tissues (TPM <1), suggesting these genes have functional redundancy or special spatial and temporal expression patterns not covered by our samples. Twenty-seven *SsNAC* genes displayed higher expression levels in pre-roll/leaf and mature roll/leaf than other tissues (premature-stem and mature-stem), suggesting an inducible expression pattern exists and also indicating the involvement of the leaf in morphogenesis and photosynthesis at various stages (Fig.5, Additional file 8). The expression levels of *SsNAC14*, *SsNAC22*, *SsNAC45*, *SsNAC31* and *SsNAC54* were much higher in pre-mature-leaf-roll, pre-mature-leaf, mature-leaf-roll and mature-leaf than in others tissues. These five genes presented the highest expression levels in different tissue (leaf-roll, leaf and stem) and the developmental leaf, suggesting they may play roles in the photosynthesis of sugarcane leaves. Interestingly, these five genes all belonged to ATAF subgroup which regulates the plant response to abiotic stress [30].

Three *SsNAC* genes (*SsNAC91*, *SsNAC58* and *SsNAC09*) had higher expression in the premature-stem (3/6/9) and mature-stem (3/6/9) than in other tissues. These results suggested that these genes may participate in transportation of nutrients or the morphogenesis of stem tissue. In terms of gene expression with different tissues, *SsNAC91* and *SsNAC58* presented similar expression patterns except for *SsNAC09* indicating a high expression in the seedling stage (leaf and stem), but they belonged to different subgroups (NAC1 and OSNAC003). Compared with *SsNAC91* and *SsNAC58*, *SsNAC09* was highly expressed in the early development stage of premature stem3 and mature stem3, indicating that this gene is involved in stem morphogenesis for nutrient transportation.

In addition, four *SsNAC* genes (*SsNAC27*, *SsNAC71*, *SsNAC129* and *SsNAC143*) belonged to a different subgroup (Unclassified, OSNAC003, OSNAC7), but showed high expression levels in the mature stem (3/6/9), indicating that the functions of these genes have a similar function to *SsNAC91* and that they may be involved in the biosynthesis of secondary walls, cellulose and hemicellulose. In general, analysis the *SsNAC* genes expression patterns in different developmental tissue reveals that the analyzed genes may play important roles in plant growth and stress responses.

Expression pattern analysis of SsNAC allelic genes

S. spontaneum is an autotetraploid species, and in general it should therefore have four alleles. The segregation of alleles in polyploidy crops that express at different levels is important in the breeding of specific varieties. To further analyze the role of the *SsNAC* allele in sugarcane growth and development, we defined whether the expression level of the alleles was more than two-fold different as a marker for

distinguishing the allelic gene differentially expressed as neutral or non-neutral. Twenty-nine *SsNAC* genes with four alleles were analyzed, and the expression levels determined using the previous method [26]. The number of non-neutral expression genes (17/29) was more than that of the neutral expression genes (12/29) (Additional file 9), which is highly similar to that found in a previous report [26]. However, the gene numbers of the two expression patterns varied based on the type of tissue being examined [26]. Comparison of allele expression revealed that there is significant functional redundancy among alleles in polyploidy species.

Gene expression analysis of the orthologous Dry (SsNAC91) in S. spontaneum and S. officinarum

Sugarcane stem is enriched in cellulosic, water and sugar, which are the main raw material for the sucrose and biofuels. However, few genes were significantly expressed to be found in stems and developmental tissues stage. *SsNAC91* has 83.22% similarity to the Dry protein of sorghum, which belongs to NAC1 subgroup and *SsNAC91* controls the pith/juicy stem if the dry gene function was acquired or lost in sorghum [21] (Fig.6B). *SsNAC91* was highly expressed in premature and mature stem tissue (6/9) but weakly expressed in premature and mature stem tissue (3) (Fig.6A, Additional file 10). The results showed that the gene was highly expressed in the middle and later stages of stem development and had a strong temporal and spatial specific expression pattern. The gene plays important roles in cell wall composition of the stem and is conserved in the majority cereals [21].

To further identify the expression level of *SsNAC91* in two founding *Saccharum* species, two sets of RNA-seq data from high sugar content *S. officinarum* and low sugar content *S. spontaneum* were used to analyze the gene expression level, including different tissues types and leaf gradient segments. According to the protein similarity, two loci were discovered on Chr 4 and 5, including *Sspon.004D0011690* (*SsNAC82*), *Sspon.005B0007561* (*SsNAC91*) from *S. spontaneum* and *Soffic.06A0022040*, *Soffic.06A0021960* from *S. officinarum*. Each locus has 2–3 alleles, but only one allele has a high-level expression, and other alleles have low expression or no expression is detectable. As shown Fig.6, the *Sspon.005B0007561* (*SsNAC91*) gene exhibited higher expression in the premature stem, mature stem, premature leaf roll and premature leaf than *Soffic.06A0022040*, indicating the genes had tissue specific expression and spatial, temporal expression patterns in stem. All alleles of the *SsNAC91* and *Soffic.06A0021960* presented very low expression levels in the developing leaves of the two species. There were significantly different expression patterns in the different tissues in the two species, the results are consistent with the phenomenon that *S. officinarum* has a high sugar content (juicy stem) whereas *S. spontaneum* has a highly fiber content (pith stem) as found in the sorghum [21].

Verification of gene expressional pattern based on RT-qPCR.

To verify the authenticity of the RNA-seq data, the four genes (*SsNAC45*, *SsNAC22*, *SsNAC14* and *SsNAC31*) that showed the highest expression levels in the mature zone of the leaf gradient and the representative four segments of developing leaf were analyzed by real-time quantitative PCR (RT-qPCR). The results of RT-qPCR were consistent with the RNA-seq data with the high correlation coefficient (0.94) (Additional file 11 and 12).

Discussion

The gene identification, structure and expansion of SsNAC families in S. spontaneum

To date, NAC proteins have been identified and verified in numerous plants, but there are few studies on the NAC family in sugarcane due to its polyploidy, complexity of genome. In this study, 151 NAC transcription factor genes were identified and named *SsNAC01-SsNAC151* in the representative monoploid genome. The number of *SsNAC* genes is very similar in different species: 151 in rice (466 Mb), 145 in sorghum (750 Mb), 148 in maize (2 300 Mb), 105 in *Arabidopsis thaliana* (125 Mb), 152 in tobacco (2.4 Gb), and 152 in soybean (1.025 Gb). The results showed that the NAC family genes are relatively conserved and stable after monocotyledonous and dicotyledonous differentiation, even taking into account ongoing diverse whole-genome duplication events (WGD) in different species. The *SsNAC* protein family was further divided into 15 different subgroups according to the classification of *ANAC* protein in *Arabidopsis thaliana*. These *SsNAC* genes were widely distributed in some subgroups except for TIP, SENU5 and ANAC063 subgroups, which did not include any *SsNAC* genes and the same results were reported regarding *SbNAC* in sorghum [10]. The three subgroups, TIP, SENU5 and ANAC063, were detected in rice, maize, *Arabidopsis*, Tartary buckwheat and prunus mume [2, 6, 12, 16]. Meanwhile, 33 *SsNAC* proteins were not divided into any subgroups according to *ANAC* protein homologs. These results suggest that these NAC genes may have been either lost in sugarcane or gained in, rice, maize, *Arabidopsis*, Tartary buckwheat and prunus mume after the divergence from the common ancestor, indicating the neo-functionalization of NAC families in sugarcane.

In addition to exploring the diversity of *SsNAC* genes, *SsNAC* proteins clustered in the same subgroup possessed similar motifs (Fig 2.). Most *SsNAC* proteins (15 subgroups) contain eight of ten motifs (motif 1 to 8) within the N-terminal region of the NAC domain. However, motif 10 in subgroup XIII in *Arabidopsis* has no *ANAC* proteins clustered together and motif 10 replaces motif 4 (part of subdomain C) and motif 6 (part of subdomain D), which may have generated a subgroup with a unique function [2, 12, 16]. In addition, motif 9 substitutes for motif 8 (part of subdomain A) and motif 5 (subdomain B) (Fig 2.), which belongs to the OSNAC003 subgroup containing ANAC010 and ANAC073 which are secondary cell wall associated NAC domain proteins [31]. The two genes are involved in the thickening of secondary cell walls in plant cells through the transcriptional activation of the expression of genes related to secondary wall synthesis [32, 33]. This phenomenon of the merging of conserved motifs has also been found in the NACs of other species [16]. The recombination of the motifs may contribute to the divergence of NAC families, thus providing the potential dynamics for the functional divergence of the NAC member in plants.

By analyzing chromosomal locations, we found that Chr1, Chr2, Chr3, Chr5 and Chr6 have greater numbers of *SsNAC* genes than others chromosomes (Table S2). All chromosomes were rearranged for fissions of some ancestral chromosomes except Chr1 [26]. For example, Chr2 was a recombination of three ancestral chromosomes (A7, A9 and A11L). Chr3 and Chr6 are the result of the fusion of segments of their ancestral chromosomes A3, A10 and A5, A12L respectively [26]. Six of the 18 clusters including

14 *SsNAC* genes in the whole genome were located on Chr 5, marking it as having a greater number of duplicated genes than others chromosomes. Similar to the previous studies in maize and sorghum, 21 gene clusters containing 46 NAC maize genes (*ZmNAC*) were detected on maize chromosomes and 19 *SbNAC* gene duplications were also identified in sorghum genome and were distributed in Chr1, 2, 3 and 6 [6, 10]. These chromosomes in sugarcane and sorghum had shared common ancestral chromosomes [26]. This finding indicates that tandem duplication plays a leading role in the expansion of the number of *SsNAC* genes, and it also provides evidence for chromosome rearrangement being one of the reasons for the expansion of gene number. Meanwhile, tandem duplication events happened before the ancestors of sorghum and sugarcane differentiated. The tandem duplication genes of sugarcane were much more numerous than in other diploid plant species and this was probably caused by the recent whole genome duplication in the *Saccharum* species.

Sugarcane, sorghum and maize belong to the *Saccharinae* group in the *Andropogoneae* tribe of grasses and are important crops for human food and sugar. Sorghum shares common ancestry with maize (12 to 15 million years ago) and sugarcane (5 to 9 million years ago) [28]. A total of 145 non-redundant *SbNAC* genes were found to be distributed across the 10 chromosomes [10]. The gene code alone is provided in the aforementioned study, and the sequence information and the *SbNAC* gene IDs are not related to any part of that study. Therefore, we re-identified the *SbNAC* genes based on the rice NAC sequence and found that the total number of identifications (110 *SbNAC*) did not match the results of the previous study (145 *SbNAC*) [10]. Compared with the number and structure of NAC protein in *SbNAC* and *SsNAC*, most members NAC genes cluster with each other in the region of the phylogenetic tree that contains similar genes structures in terms of their exons number and exons phase. The number of exons was very varied and ranged from 0–10, three exons was the average for the majority of NAC genes in sugarcane (66/151) and sorghum (64/110). This result was also reported for rice, sorghum [34, 35].

The gene expression pattern and potential function of SsNAC families in S. spontaneum

Different expression patterns of genes may lead to diverse functions during the evolution of species. The biological function of new proteins is preliminary identified by comparing the orthologous gene in *Arabidopsis* and rice [16]. In our study, we constructed two different RNA-seq datasets (developmental gradient of the leaf with 15 segments representing 4 different zones and different development tissues representing 12 tissues). In heat maps, 33 and 27 of *SsNAC* genes were highly expressed (TPM >10) (Fig 4 and Fig 5) in the developmental leaf and different developmental tissues respectively. A total of 18 genes (*SsNAC45*, *SsNAC22*, *SsNAC14*, *SsNAC31*, *SsNAC54*, *SsNAC47*, *SsNAC109*, *SsNAC63*, *SsNAC12*, *SsNAC21*, *SsNAC120*, *SsNAC83*, *SsNAC132*, *SsNAC58*, *SsNAC118*, *SsNAC122*, *SsNAC96*, *SsNAC85*) that belong to eight subgroups (ATAF, NAC2, NAM, NAP, OSNAC003, OSNAC022, OSNAC8 and Unclassified) were found to be highly expressed in the two different RNA-seq datasets, with the expression level in the leaf gradient being higher than in other tissues, indicating that these genes play an important role during the development and growth of leaves and tissues (stem and leaf-roll). Interestingly, most *SsNAC* (39/55) genes have a high peak expression at section 11 (mature zone) of the leaf (Fig 4), with this section mainly involving the differentiation and accumulation of chloroplasts in mesophyll cells and bundle sheath cells

[29]. Other *SsNAC* genes were expressed highly between section 3 (basal zone) and section 5 (translational zone) of the leaf gradient pattern. Basic cellular functions are active in this region, such as DNA synthesis, cell wall synthesis, and hormone signaling [29]. These results indicate that the *SsNAC* genes including the NAC2 subgroup may play an important role in leaf morphogenesis and chloroplast development and photosynthesis.

In this study, phylogenetic analyses suggested that *SsNAC78*, *SsNAC58*, *SsNAC118*, *SsNAC28* and *SsNAC40* were classified into the NAC2 subgroup, which was clustered with *ANAC050*, *ANAC052*, *ANAC053*, *ANAC082* and *ANAC078* which all had the trend of high expression levels at section 11 (mature zone) of the leaf. These results suggested that the NAC2 subgroup may play a role in photosynthesis in *Saccharum*. In *Arabidopsis*, *ANAC050* and *ANAC052* may play roles in transcriptional repression and regulate the flowering time by interacting with the histone demethylase JMJ14 [36]. *ANAC053* was identified as an NTL4 transcription factor which promotes ROS production by binding to the promoters of genes encoding ROS biosynthetic enzymes, while ABA induces leaf senescence under drought conditions [37]. Thus, it is possible that the NAC2 subgroup has other functions in the development of *Saccharum* in addition to its potential role in photosynthesis.

A total of 8 *SsNAC* genes belong to the ATAF subgroup from the phylogenetic tree, with 7 of these genes (*SsNAC45*, *SsNAC22*, *SsNAC14*, *SsNAC31*, *SsNAC54*, *SsNAC83* and *SsNAC132*) being highly expressed in the two RNA-seq datasets models with the exception being *SsNAC30*. Seven of the eight were expressed in the leaf gradient model much higher levels than in different tissue developmental stages. Expression of ATAF is induced by wounding, pathogenic fungi and hormones such as jasmonic acid (JA), salicylic acid (SA) and ethylene (ET) [18, 19, 38]. Findings showed that *SsNAC* genes of ATAF subgroup may have evolved new functions in sugarcane, such as participating in leaf growth and development, morphogenesis and photosynthesis. The leaf gradient material is the second leaf that has been grown for 11 days. The whole plant is relatively young and highly susceptible to surrounding abiotic stress, and some NAC genes may respond to stress. For example, *SsNAC83* displayed lower expression levels than the five genes in the subgroups, it was named *SsNAC23* in a previous study [25], and it is strongly induced at 4 °C but not at 12°C, demonstrating that *SsNAC83* plays important roles in the response to extreme low temperature stress [25].

Seven genes (*SsNAC12*, *SsNAC21*, *SsNAC57*, *SsNAC38*, *SsNAC85*, *SsNAC93* and *SsNAC100*) of sixteen *SsNAC* from the NAM subgroup have higher expression levels in the leaf than in the stem, are orthologous to *ANAC031* and regulate leaf development and the formation of the apical meristem [39, 40]. In addition, the NAM subgroup genes are involved in the development of vegetative organs and reproductive organs and also take part in cell division and leaf development [39–42]. Three (*SsNAC38*, *SsNAC57* and *SsNAC100*) of the four genes were highly expressed between section 2 (basal zone) and section 8 (maturing zone) in the leaf source-sink transition zone and also in vascular bundle cells, mesophyll cells and secondary cell walls which are forming and consequently differentiating with the expression of related genes continues to increase. Our results with the NAM subgroup genes were also confirmed the previous conclusion [39–42].

We also found that seven genes (*SsNAC91*, *SsNAC09*, *SsNAC58*, *SsNAC129*, *SsNAC071*, *SsNAC027* and *SsNAC143*) had higher expression levels in the premature (6/9) and mature stem (6/9) than in other tissues (Fig 5). These seven genes belong to 4 subgroups (NAC1, OSNAC022, OSNAC003 and unclassified), four of them genes (*SsNAC09*, *SsNAC58*, *SsNAC129* and *SsNAC143*) also presented a high expression level in the gradient of leaf pattern. *SsNAC143* was a member of the OSNAC7 subgroup and clustered with ANAC043 (NST1), ANAC012 (SND1). Secondary wall-associated NAC domain protein1 (SND1), is homologous to ANAC043(NST1) and ANAC066 (NST2) genes, which are key regulators of secondary wall synthesis in fibers of *Arabidopsis thaliana* stems [32, 43]. However, individual T-DNA gene knockouts of these two genes did not lead to a reduction in secondary cell wall thickness but the double T-DNA knockout did. SND1/NST1-RNAi plants lead to all three major secondary wall components (cellulose, xylan, and lignin) being significantly reduced suggesting that two genes had redundant functions in the regulation of secondary wall synthesis in fibers [43]. *SsNAC58*, *SsNAC129*, *SsNAC071* belonging to the OSNAC003 subgroup were similar to ANAC088(SND2) and ANAC010(SND3) involved in vascular development in *Arabidopsis thaliana* [32]. SND2 and SND3 are genes induced by NST1 and NST3 to further activate the synthesis of cell wall components and promote the formation of the secondary wall [32]. Together, these results directed us to the hypothesis that some *SsNAC* genes may participate in the vascular and secondary wall development in the leaf and stem of sugarcane.

SsNAC91 gene may play an important role in stem cellulose biosynthesis of Saccharum

Modern sugarcane cultivars are hybridized by *S. officinarum* and *S. spontaneum* hybrids, which have been repeatedly backcrossed with *S. officinarum* resulting in 80–90% of the genome coming from *S. officinarum* and 10–20% from *S. spontaneum* [44]. *S. spontaneum* has a higher cellulose content than *S. officinarum*, but *S. officinarum* has a higher sugar and water content than *S. spontaneum*. In this study, a total of 8 genes constituting the NAC1 subgroup were identified from the phylogenetic tree (Fig. 2), but the genes were not expressed except *SsNAC91*. The *SsNAC91* gene had a higher expression in the premature and mature stem of *S. spontaneum* than that of *Soffic.06A0022040* of *S. officinarum*, suggesting that this gene may also play important roles in cellulose biosynthesis and water transport. Expression levels of this gene were significantly different, which may at least in part account for the different levels of sucrose, cellulose and water content between *S. officinarum* and *S. spontaneum*.

Zhang et al.[21] also found that the *Dry* gene is involved in the synthesis of the secondary cell wall and water transport in sorghum. The lack of functional *Dry* protein affects the morphology, cell wall thickness and cell wall composition of vascular bundles and parenchyma cells in sweet sorghum stems. *Dry* may be a new type of secondary cell wall synthesis related genes. The role of this subgroup is to inhibit the growth of specific cells, promote the formation of tissue boundaries, and promote the production of apical (leaf) meristems [41, 42, 45]. It is hypothesized that when a functional *Dry* gene is not present-lost, the transportation of water in sorghum stalk is reduced. The sweet sorghum stalk accumulates a large amount of water, and water is the main transport carrier of photosynthetic products such as sugar, which in turn affects sugar transport. The homologous gene to the *Dry* gene in *Arabidopsis* is *ANAC074*, and it has been found that the expression of *ANAC074* increased during leaf senescence [46]. The homologous

gene of *Dry* gene in rice is Os4g43560 and it was found that under moderate drought conditions, the expression of Os4g43560 increased in the young panicle of drought-tolerant rice, but the expression did not change in the drought-tolerant rice [47]. The same results also showed that the senescence of the seeds in sorghum is higher than that of sweet sorghum[48]. The *NAC7* gene of maize was clustered into the NAC1 subgroup and expression was down-regulated by RNAi. This transgenic inbred line can prolong the photosynthetic period of maize, increase the production and biological yield, and finally achieve the goal of high yields of grain by maintaining the greening function of maize leaves and prolonging aging[20]. Therefore, determining the function of *Dry* genes in sugarcane can potentially provide excellent targets for improving the breeding efficiency of sugarcane to obtain increased sugar and biomass production.

Although many *SsNAC* genes are not expressed in these samples, they may have other expression patterns involved in growth and development. In this study, the analysis of expression of the *SsNAC* genes at different developmental stages and tissue types enabled expression and functional comparisons to be drawn with model plants, which will facilitate the understanding of the functional roles of individual genes and importantly provide information necessary to drive improvements in future sugarcane molecular breeding programs.

Conclusions

In our study, we performed a comprehensive analysis of *SsNAC* genes using phylogenetic, chromosomal location, duplication event data, and expression patterns analyses in different developmental tissues. Based on comparative genomics to identify *SsNAC* genes, 151 *SsNAC* genes were identified and distributed unevenly on eight chromosomes, a finding that will provide basic information for the functional characterization of *SsNAC* genes. Phylogenetic analysis suggested that the *SsNAC* protein family can be divided into 15 subgroups, with one subgroup unclassified. The majority of the 151 *SsNAC* genes, including 327 alleles, were located on Chr 5 (54, 16.51%), Chr 1 (51, 15.60%) and Chr 2 (51, 15.60%). Collinearity analysis indicated that 84.13% (275) of *SsNAC* genes were mainly derived from segmental duplication and tandem duplication. Furthermore, based on expression profiling of RNA-seq data analysis of all *SsNAC* genes with the leaf development and different tissues types, seven genes of the ATAF subgroup were highly expressed in the developmental stages of leaf and different tissue types, and with much higher expression levels in leaf development than in the different tissues types (leaf and stem). The result showed that the *SsNAC* genes of the ATAF subgroup may have participated in leaf growth and development, morphogenesis and photosynthesis. Meanwhile, the NAC1 subgroup *SsNAC91* gene, being orthologous to Sobic.006G147400 (*Dry* gene), displayed much higher expression in premature and mature stem of the low sugar and water content *S. spontaneum* than orthologous genes in the high sugar and water content *S. officinarum*. The results indicate that *SsNAC91* may also play important roles in cellulose biosynthesis and water transport in sugarcane and is an excellent species for further functional studies.

Materials And Methods

Plant material

The gene expression patterns of two *Saccharum* species, *S. spontaneum* SES208 (Ss, $2n = 8x = 64$, originated in USA) and *S. officinarum* LA Purple (So, $2n = 8x = 80$, originated in USA), were previously analyzed [49]. These two species were deposited national germplasm repository of sugarcane (Yunnan China) and planted in the sugarcane experiment field at Fujian Agricultural and Forestry University (Fuzhou China). To analyze the different developmental stages, tissue samples were collected from leaf roll, leaf, top internode, maturing internode and mature internode respectively. The above mentioned tissues representing the different developmental stages were collected from both LA purple and SES208 that were grown in the following conditions: light intensity of $350 \mu\text{mol}/\text{m}^2/\text{sec}$, 14:10L/D, 30°C L/ 22°C D and 60% relative humidity. Tissues were collected from 11-day-old second leaves of SES208 and 15-day-old second leaves of LA purple precisely 3 h into the light period. These leaves were cut into 15 1-cm segments and pooled from 4 plants per biological replicate with 3 biological replicates for each tissue type being prepared [49]. The different developmental tissue stages were prepared according to the methods previously described [50, 51].

Identification of the NAC gene family via bioinformatics approaches

Two methods were used to identify all putative NAC proteins in *S. spontaneum*. Firstly, the Hidden Markov Model (HMM) profile of the NAM domain PF02365 was downloaded from the Pfam database. The NAM domain was used to search the protein database of *S. spontaneum* with a value (e-value) cut-off at 1.0. Secondly, the NAC proteins of rice were downloaded as described previously [50, 51]. Then, the 140 NAC proteins of rice identified as query sequences were utilized to search the protein database of *S. spontaneum* and sorghum database. The genomic data of *S. spontaneum* was generated in our own laboratory. The sorghum protein database was obtained from phytozome web (<https://phytozome.jgi.doe.gov/pz/portal.html>). All non-redundant sequences identified by the two methods described above were consolidated and repeats and allele sequences were eliminated. Thirdly, these sequences were subjected to further analysis using the Pfam protein family database (<http://pfam.xfam.org/>) and the BLASTp program. *Arabidopsis thaliana* NAC (ANAC) protein sequences were downloaded from the TAIR 10 release (<http://www.arabidopsis.org>) [5]. In addition, the online ExPASy-ProtParam (<http://web.expasy.org/protparam/>) program was used to calculate the isoelectric point (PI), the length of amino acids (AA) and molecular weight (MW).

Phylogenetic, motif and gene structure analysis of SsNAC genes

An unrooted phylogenetic tree was constructed from *S. spontaneum* and *A. thaliana* NAC proteins using the MEGA7.0 software with Neighbor-Joining (NJ), Poisson distribution, pairwise deletion and a bootstrap value of 1000. The TBtools (default parameters) software was used for exploring the composition of the exons/introns of *SsNAC* genes [52]. MEME online software (<http://meme-suite.org/tools/meme>) was used

to analyze the motifs of SsNAC proteins with the parameters as follows: motif width set to 6–200 and number set to 10 [53].

Chromosomal locations, gene duplication and collinearity analysis of SsNAC genes

The physical location of SsNAC genes on the chromosomes was downloaded from http://www.life.illinois.edu/ming/downloads/Spontaneum_genome/. To analyze the duplication pattern for each SsNAC gene, the BLASTP program (E-value $<10^{-5}$) and Multiple Collinearity Scan toolkit (MCScanX) were used [54].

Expression analysis profile of SsNAC genes using the RNA-seq

According to the manufacturer's protocol (TruSeq® RNA, Illumina), the cDNA libraries were prepared using Illumina® TruSeq™ RNA Sample Preparation Kit (RS–122–2001(2), Illumina). The RNA-seq libraries were sequenced with 100 single reads on Illumina HiSeq 2500 at the Center for Genomics and Biotechnology at the Fujian Agriculture and Forestry University. Based on the TRINITY method combining Inchworm, Chrysalis, and Butterfly with independent modules, raw data was aligned to the reference genome [51].

Validation TPM values of SsNAC gene expression levels using qPCR

The five SsNAC genes (*SsNAC45*, *SsNAC22*, *SsNAC14*, *SsNAC31* and *SsNAC129*) were validated by qPCR. The following four tissues of *S. spontaneum* were chosen for samples: the first, sixth, tenth and fifth segment of 11-day old second leaves. Primers were designed using Integrated DNA Technologies online tools (<http://www.idtdna.com/Primerquest/Home/Index>). The eukaryotic elongation factor 1a (eEF–1a) served as the internal control [55]. The qPCR was repeated three times and the data analyzed was calculated using the $2^{-\Delta\Delta C_t}$ method. qPCR was performed using the method previously described [51].

List Of Abbreviations

NAC: NAM, ATA1/2 and CUC2

CDS: *Coding domain sequence*

eEF–1a: Eukaryotic elongation factor 1a

ROS: Reactive oxygen species

Declarations

We declared that two *Saccharum* species for the study are not necessary to apply any permissions.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The *S. spontaneum* genome project was deposited into Genbank with accession numbers: QVOL000000000.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Science and Technology Planting Project of Guangdong Province (2019B020238001); the National key research and development program [2018YFD1000104]; 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 [No. JA12082].

Authors' contributions

JZ and HW conceived the study and designed the experiments. HW, YW, NX, XH and JZ carried out the experiments. HW, YW, XH, MZ, RM and JZ analyzed the data. HW and JZ wrote the manuscript. All authors read and approved the final paper

Authors' information

Henbo Wang and Yongjun Wang contributed equally to this work.

References

1. Broun P: *Transcription factors as tools for metabolic engineering in plants. Current opinion in plant biology* 2004, 7(2):202–209.

- 2.Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Matsubara K, Osato N, Kawai J, Carninci P *et al*: *Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana*. *DNA research: an international journal for rapid publication of reports on genes and genomes* 2003, *10*(6):239–247.
- 3.Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, Laudeman TW, Timko MP: *Tobacco transcription factors: Novel insights into transcriptional regulation in the Solanaceae*. *Plant physiology* 2008, *147*(1):280–295.
- 4.Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K, Tran LSP: *Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress*. *DNA Research* 2011, *18*(4):263–276.
- 5.Fang Y, You J, Xie K, Xie W, Xiong L: *Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice*. *Molecular Genetics and Genomics* 2008, *280*(6):547–563.
- 6.Peng X, Zhao Y, Li X, Wu M, Chai W, Sheng L, Wang Y, Dong Q, Jiang H, Cheng B: *Genomewide identification, classification and analysis of NAC type gene family in maize*. *J Genet* 2015, *94*(3):377–390.
- 7.Wang N, Zheng Y, Xin H, Fang L, Li S: *Comprehensive analysis of NAC domain transcription factor gene family in Vitis vinifera*. *Plant cell reports* 2013, *32*(1):61–75.
- 8.Liu T, Song X, Duan W, Huang Z, Liu G, Li Y, Hou X: *Genome-Wide Analysis and Expression Patterns of NAC Transcription Factor Family Under Different Developmental Stages and Abiotic Stresses in Chinese Cabbage*. *Plant Molecular Biology Reporter* 2014, *32*(5):1041–1056.
- 9.Hu R, Qi G, Kong Y, Kong D, Gao Q, Zhou G: *Comprehensive Analysis of NAC Domain Transcription Factor Gene Family in Populus trichocarpa*. *BMC Plant Biology* 2010, *10*(1):145.
- 10.Kadier Y, Zu Y-y, Dai Q-m, Song G, Lin S-w, Sun Q-p, Pan J-b, Lu M: *Genome-wide identification, classification and expression analysis of NAC family of genes in sorghum [Sorghum bicolor (L.) Moench]*. *Plant Growth Regulation* 2017, *83*(2):301–312.
- 11.Ling L, Song L, Wang Y, Guo C: *Genome-wide analysis and expression patterns of the NAC transcription factor family in Medicago truncatula*. *Physiology and molecular biology of plants: an international journal of functional plant biology* 2017, *23*(2):343–356.
- 12.Zhuo X, Zheng T, Zhang Z, Zhang Y, Jiang L, Ahmad S, Sun L, Wang J, Cheng T, Zhang Q: *Genome-Wide Analysis of the NAC Transcription Factor Gene Family Reveals Differential Expression Patterns and Cold-Stress Responses in the Woody Plant Prunus mume*. *Genes* 2018, *9*(10).
- 13.Zhang Y, Li D, Wang Y, Zhou R, Wang L, Zhang Y, Yu J, Gong H, You J, Zhang X: *Genome-wide identification and comprehensive analysis of the NAC transcription factor family in Sesamum indicum*.

- 14.Ahmad M, Yan X, Li J, Yang Q, Jamil W, Teng Y, Bai S: *Genome wide identification and predicted functional analyses of NAC transcription factors in Asian pears*. 2018, 18(1):214.
- 15.Zhang H, Kang H, Su C, Qi Y, Liu X, Pu J: *Genome-wide identification and expression profile analysis of the NAC transcription factor family during abiotic and biotic stress in woodland strawberry*. *PloS one* 2018, 13(6):e0197892.
- 16.Liu M, Ma Z, Sun W, Huang L, Wu Q, Tang Z, Bu T, Li C, Chen H: *Genome-wide analysis of the NAC transcription factor family in Tartary buckwheat (*Fagopyrum tataricum*)*. *BMC genomics* 2019, 20(1):113.
- 17.Olsen AN, Ernst HA, Leggio LL, Skriver K: *NAC transcription factors: structurally distinct, functionally diverse*. *Trends in plant science* 2005, 10(2):79–87.
- 18.Wang XE, Basnayake B, Zhang HJ, Li GJ, Li W, Virk N, Mengiste T, Song FM: *The Arabidopsis ATAF1, a NAC Transcription Factor, Is a Negative Regulator of Defense Responses Against Necrotrophic Fungal and Bacterial Pathogens*. *Molecular Plant* 2009, 22(10):1227–1238.
- 19.Jensen MK, Rung JH, Gregersen PL, Gjetting T, Fuglsang AT, Hansen M, Joehnk N, Lyngkjaer MF, Collinge DB: *The HvNAC6 transcription factor: a positive regulator of penetration resistance in barley and Arabidopsis*. *Plant molecular biology* 2007, 65(1–2):137–150.
- 20.Jun Z, Kevin AF, John LVH, Rajeev G, Nick M, Jindong S, William BA, Yang W, Benjamin W, Hua M *et al*: *Identification and characterization of a novel stay-green QTL that increases yield in maize*. *Plant biotechnology journal* 2019.
- 21.Zhang L-M, Leng C-Y, Luo H, Wu X-Y, Liu Z-Q, Zhang Y-M, Zhang H, Xia Y, Shang L, Liu C-M: *Sweet Sorghum Originated through Selection of Dry, a Plant-Specific NAC Transcription Factor Gene*. *PLANT CELL* 2018, 30(10):2286–2307.
- 22.Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M: *NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis*. *Plant Cell* 2007, 19(1):270–280.
- 23.Lira BS, Gramegna G, Trench BA, Alves FRR, Silva EM, Silva GFF, Thirumalaikumar VP, Lupi ACD, Demarco D, Purgatto E *et al*: *Manipulation of a Senescence-Associated Gene Improves Fleshy Fruit Yield*. *Plant physiology* 2017, 175(1):77.
- 24.Ma X, Zhang Y, Turečková V, Xue G-P, Fernie AR, Mueller-Roeber B, Balazadeh S: *The NAC Transcription Factor SINAP2 Regulates Leaf Senescence and Fruit Yield in Tomato*. *Plant physiology* 2018, 177(3):1286.

25. Nogueira FTS, Schlögl PS, Camargo SR, Fernandez JH, De Rosa VE, Pompermayer P, Arruda P: *SsNAC23, a member of the NAC domain protein family, is associated with cold, herbivory and water stress in sugarcane. Plant Science* 2005, *169*(1):93–106.
26. Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J *et al*: *Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L. Nature genetics* 2018, *50*(11):1565–1573.
27. Holub EB: *The arms race is ancient history in Arabidopsis, the wildflower. Nature reviews Genetics* 2001, *2*(7):516–527.
28. Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann JC, Arruda P, D'Hont A: *Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. The Plant journal: for cell and molecular biology* 2007, *50*(4):574–585.
29. Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR *et al*: *The developmental dynamics of the maize leaf transcriptome. Nature genetics* 2010, *42*(12):1060–1067.
30. Lu P-L, Chen N-Z, An R, Su Z, Qi B-S, Ren F, Chen J, Wang X-C: *A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. Plant molecular biology* 2007, *63*(2):289–305.
31. Hussey SG, Mizrahi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA: *SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in Arabidopsis fibres and increases fibre cell area in Eucalyptus. BMC Plant Biology* 2011, *11*(1):173.
32. Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M: *The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. Plant Cell* 2005, *17*(11):2993–3006.
33. Zhong R, Demura T, Ye ZH: *SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. Plant Cell* 2006, *18*(11):3158–3170.
34. Sanjari S, Shirzadian-Khorramabad R, Shobbar Z-S, Shahbazi M: *Systematic analysis of NAC transcription factors' gene family and identification of post-flowering drought stress responsive members in sorghum. Plant cell reports* 2019, *38*(3):361–376.
35. Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S: *Genome-wide analysis of NAC transcription factor family in rice. Gene* 2010, *465*(1):30–44.
36. Ning Y-Q, Ma Z-Y, Huang H-W, Mo H, Zhao T-t, Li L, Cai T, Chen S, Ma L, He X-J: *Two novel NAC transcription factors regulate gene expression and flowering time by associating with the histone demethylase JMJ14. Nucleic acids research* 2015, *43*(3):1469–1484.

- 37.Lee S, Seo PJ, Lee HJ, Park CM: *A NAC transcription factor NTL4 promotes reactive oxygen species production during drought-induced leaf senescence in Arabidopsis. The Plant journal: for cell and molecular biology* 2012, 70(5):831–844.
- 38.Delessert C, Kazan K, Wilson IW, Van Der Straeten D, Manners J, Dennis ES, Dolferus R: *The transcription factor ATAF2 represses the expression of pathogenesis-related genes in Arabidopsis. Plant Journal* 2005, 43(5):745–757.
- 39.Nikovics K, Blein T, Peaucelle A, Ishida T, Morin H, Aida M, Laufs P: *The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell* 2006, 18(11):2929–2945.
- 40.Agata B, Magdalena Rs-S, Dorota B-W, Agnieszka P, Mitsuhiro A, Dorota K: *The CUP-SHAPED COTYLEDON2 and 3 genes have a post-meristematic effect on Arabidopsis thaliana phyllotaxis. Annals of botany* 2015, 115(5):807–820.
- 41.Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M: *Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. The Plant Cell* 1997, 9(6):841–857.
- 42.Souer E, vanHouwelingen A, Kloos D, Mol J, Koes R: *The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell* 1996, 85(2):159–170.
- 43.Zhong R, Richardson EA, Ye ZH: *Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. Planta* 2007, 225(6):1603–1611.
- 44.Hont AD, Grivet L, Feldmann P, Glaszmann JC, Rao S, Berding N: *Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. Molecular and General Genetics* 1996, 250(4):405–413.
- 45.Raman S, Greb T, Peaucelle A, Blein T, Laufs P, Theres K: *Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in Arabidopsis thaliana. Plant Journal* 2008, 55(1):65–76.
- 46.Podzimska-Sroka D, O'Shea C, Gregersen PL, Skriver K: *NAC Transcription Factors in Senescence: From Molecular Structure to Function in Crops. Plants (Basel)* 2015, 4(3).
- 47.Mohammed N, Akhter Most S, Kouji S, Ali M, Ramiah V, Rachid S, Arvind K, Hei L, Kotb A, Shoshi K: *Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64. Molecular genetics and genomics* 2012, 287(5):389–410.

- 48.Chuan-yuan. L, Huai-qing. H, Hai-chun. J: *Research Progress on the Stem Juiciness of Sweet Sorghum. Biotechnology Bulletin* 2019, 35(5):9–14.
- 49.Zhang J, Arro J, Chen Y, Ming R: *Haplotype analysis of sucrose synthase gene family in three Saccharum species. BMC genomics* 2013, 14:314.
- 50.Hu W, Hua X, Zhang Q, Wang J, Shen Q, Zhang X, Wang K, Yu Q, Lin YR, Ming R *et al*: *New insights into the evolution and functional divergence of the SWEET family in Saccharum based on comparative genomics.* 2018, 18(1):270.
- 51.Wang Y, Hua X, Xu J, Chen Z, Fan T, Zeng Z, Wang H, Hour AL, Yu Q, Ming R *et al*: *Comparative genomics revealed the gene evolution and functional divergence of magnesium transporter families in Saccharum.* 2019, 20(1):83.
- 52.Chen C, Xia R, Chen H, He Y: *TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. BioRxiv* 2018:289660.
- 53.Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: *MEME SUITE: tools for motif discovery and searching. Nucleic acids research* 2009, 37:W202-W208.
- 54.Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*: *MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research* 2012, 40(7):e49.
- 55.Ling H, Wu Q, Guo J, Xu L, Que Y: *Comprehensive selection of reference genes for gene expression normalization in sugarcane by real time quantitative rt-PCR. PloS one* 2014, 9(5):e97469.

Additional Files

Additional file 1 List of SsNAC gene information identified in *S. spontaneum* genome

Additional file 2 The 151 SsNAC genes with alleles in the autopolyploid *S. spontaneum* genome

Additional file 3 NAC protein information on the *Arabidopsis thaliana* genome

Additional file 4 The number of SsNAC genes on chromosomes with alleles

Additional file 5 Gene duplication types in the autopolyploid *S. spontaneum* genome

Additional file 6 Comparison of gene structure of the SsNAC and SbNAC genes between *S. spontaneum* and *Sorghum bicolor*

Additional file 7 RNA-seq data of SsNAC genes in developmental gradient of the leaf

Additional file 8 RNA-seq data of SsNAC genes under different development tissues

Additional file 9 Expression pattern analysis of allelic genes of *SsNAC*

Additional file 10 RT-qPCR verification of expression levels of *SsNAC91* in partial tissues of two *Saccharum* species

Note: Stems 3, 9, 15 and stems 3, 6, 9 were from *S. officinarum* LA-Purple and *S. spontaneum* SES 208, respectively.

Additional file 11 The primers for RT-qPCR verification of six *SsNAC* genes in *S. spontaneum*

Additional file 12 RT-qPCR verification of expression levels of *SsNAC45*, *SsNAC22*, *SsNAC14* and *SsNAC31* in the developing leaf of *S. spontaneum*

Figures

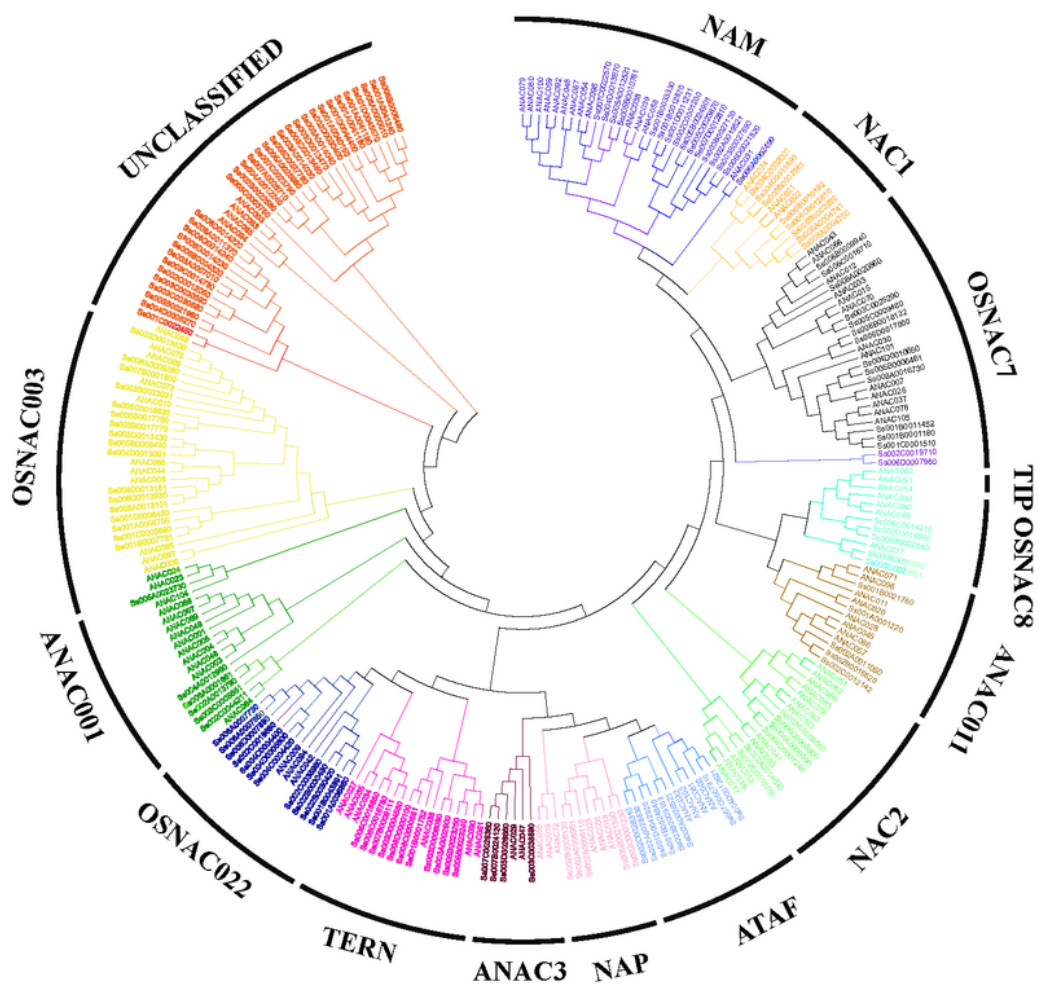


Figure 1

Phylogenetic analysis of SsNAC genes and ANAC genes. The tree divided the SsNAC proteins into 15 subgroups. The evolutionary history was determined using the Neighbor-Joining method. The evolutionary distances were computed using the Poisson correction method.

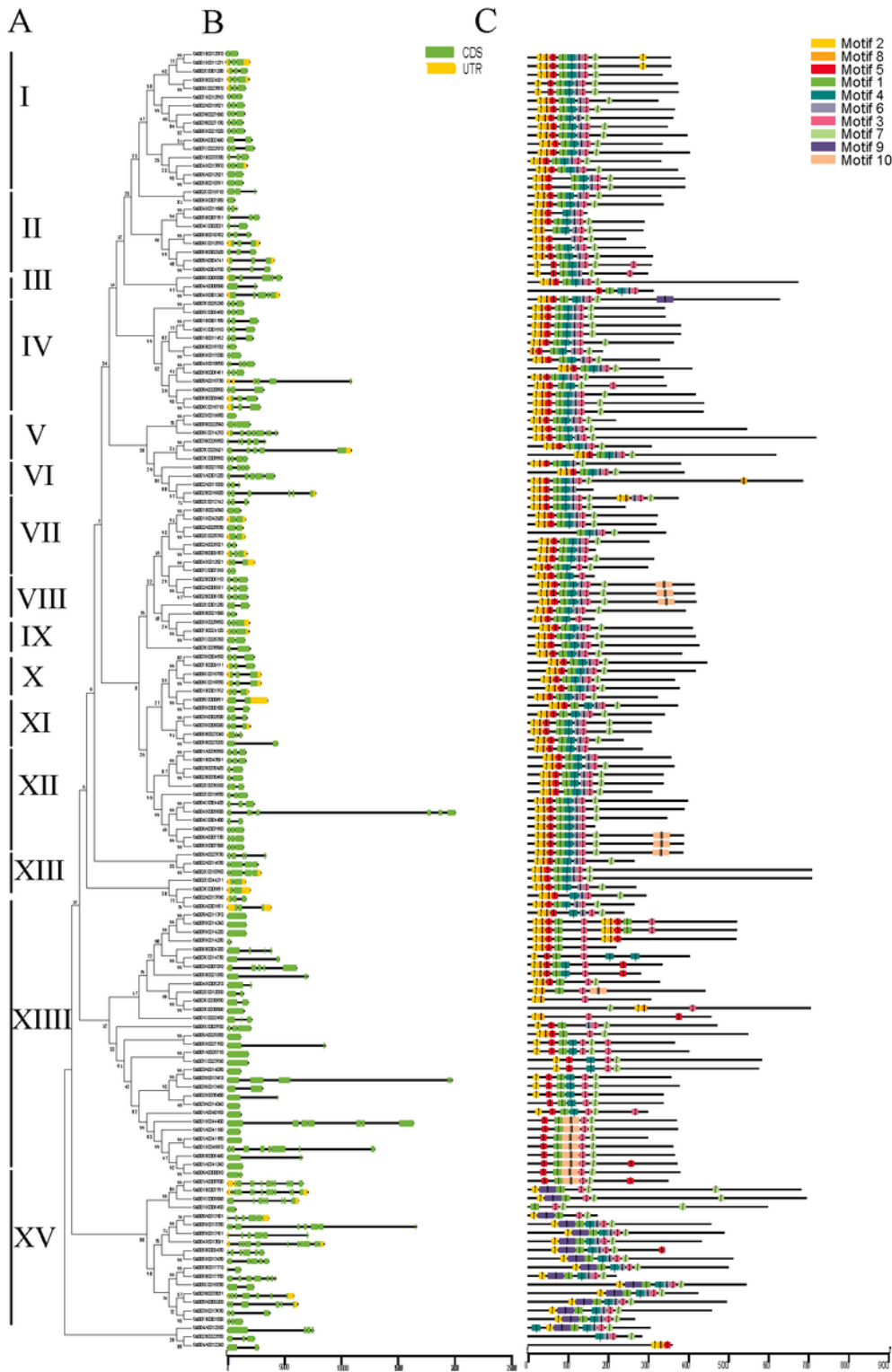


Figure 2

Phylogenetic relationship, gene structure, and motif pattern of SsNAC proteins A: The phylogenetic tree was built based on the full-length sequences of SsNAC proteins using MEGA 7.0 software. B: Exon-intron structure of SsNAC genes. C: The motif composition of SsNAC proteins were composed of 10 motifs.

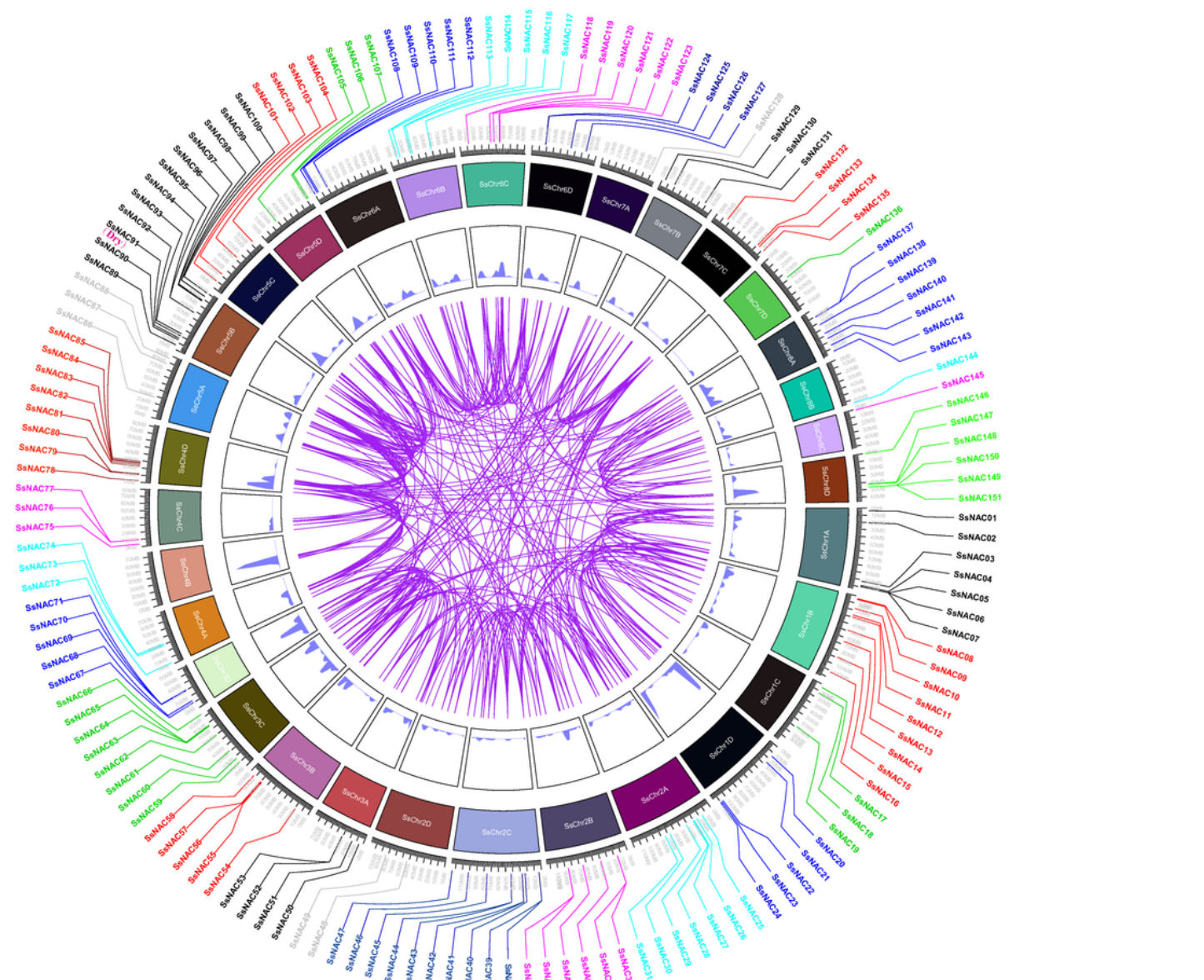


Figure 3

Schematic representations of the interchromosomal relationships of the SsNAC genes.

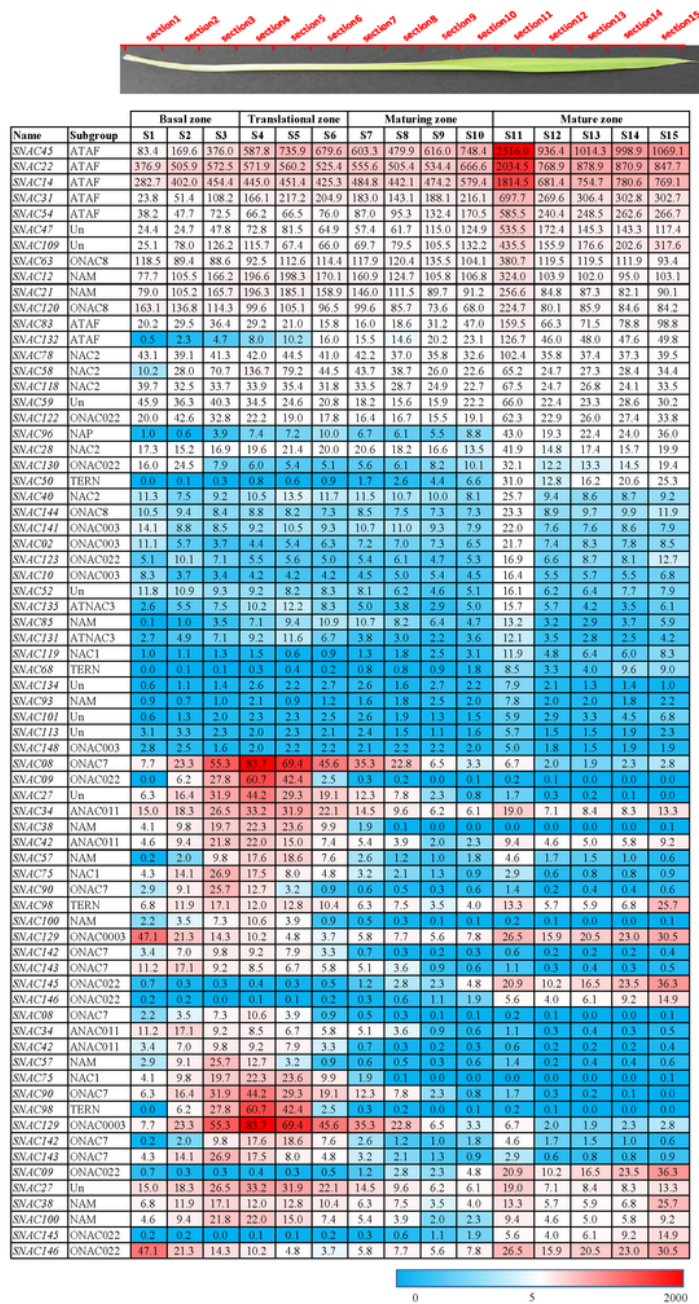


Figure 4

RNA-seq data of SsNAC genes in different leaf developmental stages

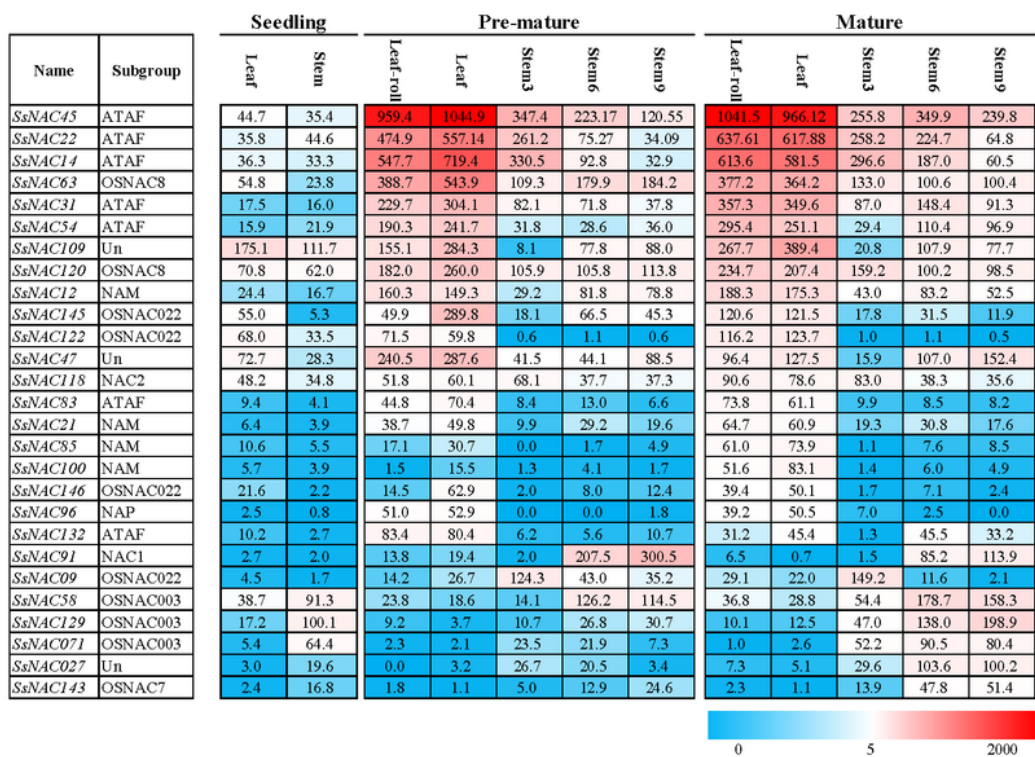


Figure 5

RNA-seq data of SsNAC genes in different tissues

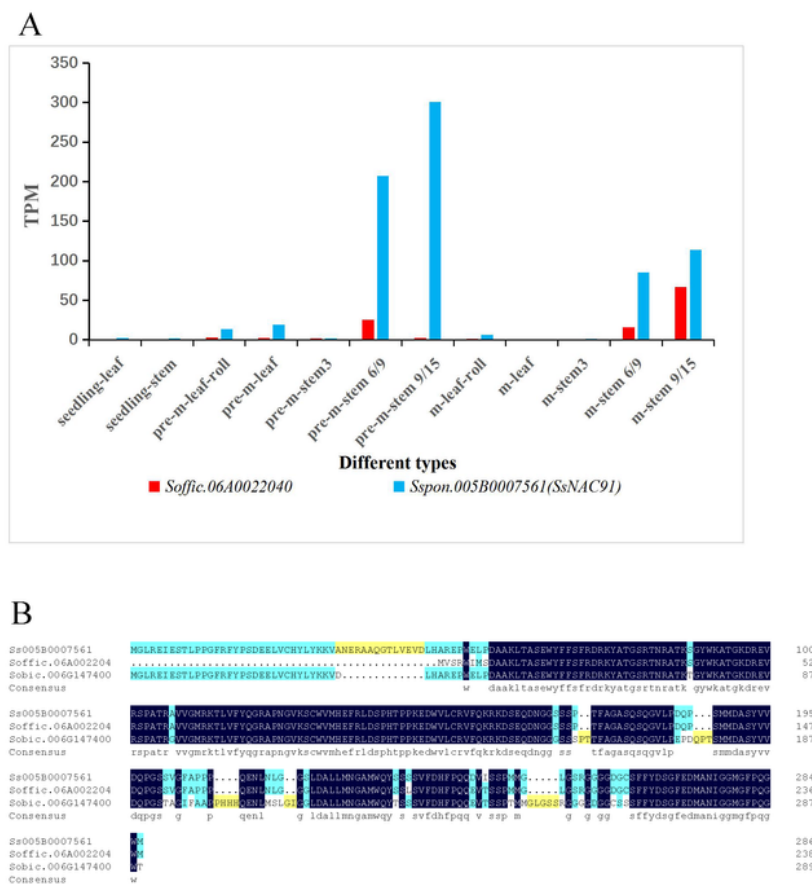


Fig. 6 The expressional pattern and protein sequence alignment of *Dry*

A. The *SsNAC91* expression in two founding *Saccharum* species; **B.** Protein sequence alignments of *Dry* from two *Saccharum* species (*S. officinarum* and *S. spontaneum*) and sorghum.

Note: Ss005B0007561 (*SsNAC91*) is from *S. spontaneum*, Soffic.06A0022040 is from *S. officinarum* and Sobic.006G147400.1 is from sorghum.

Figure 6

The expressional pattern and protein sequence alignment of *Dry*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile10.pdf](#)
- [Additionalfile8.xlsx](#)
- [Additionalfile9.xlsx](#)
- [Additionalfile7.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile1.xlsx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile12.pdf](#)
- [Additionalfile11.xlsx](#)
- [Additionalfile6.pdf](#)
- [Additionalfile2.xlsx](#)