BMC Plant Biology

RESEARCH

Open Access

Comprehensive analysis of the NAC transcription factor gene family in *Sophora tonkinensis* Gagnep

Ximei Liang[1,2,3], Shuangshuang Qin[1,2], Guili Wei[1,2], Xiaoyun Guo[1,2], Shugen Wei[1,2,3,4*], Fan Wei[1,2*] and Ying Liang[1,2*]

Abstract

Background *Sophora tonkinensis* Gagnep. has long been utilized in the treatment of anti-inflammatory and pain-relieving, with its principal active compounds being alkaloids and flavonoids. NAC transcription factors, a large family of plant-specific regulators, play pivotal roles in growth, development, stress responses, and secondary metabolism. However, comprehensive genome-wide characterization of *S. tonkinensis* NAC gene family (*StNAC*) remains unexplored.

Results This study identified 85 NAC proteins from the *S. tonkinensis* genome database. Phylogenetic analysis revealed that StNAC proteins were categorized into 15 subgroups based on their homology with *Arabidopsis thaliana* NAC proteins. Gene structure analysis demonstrated a variation in intron numbers ranging from 1 to 7, with a majority of *StNAC* genes containing 2–3 introns. Chromosomal distribution analysis indicated an uneven spread of *StNAC* genes across 9 chromosomes, with the highest number of *StNAC* genes on Chr3. Detection of 4 tandem duplicates and 32 segmental duplicates revealed that segmental duplication primarily drive *StNAC* genes amplification. Prediction of cis-regulatory elements suggested the involvement of *StNAC* genes in growth, stress responses, and hormone regulation. Gene expression analysis showed substantial variability expression of *StNAC* genes across different tissues. Notably, eight *StNAC* genes were identified as significantly associated alkaloid and flavonoid levels. qRT-PCR validation indicated that five genes were highly expressed in tissues, corroborating transcriptome data.

Conclusion These findings offer valuable insights for further functional characterization of NAC genes and their potential roles in alkaloid and flavonoid biosynthesis in *S. tonkinensis*.

Keywords *Sophora tonkinensis* Gagnep, NAC transcription factor, Gene family, Alkaloid, Flavonoid

*Correspondence:
Shugen Wei
weishugen2@163.com
Fan Wei
wfmanuscript@163.com
Ying Liang
198yingzi@163.com
[1]Guangxi Key Laboratory of Medicinal Resources Protection and Genetic Improvement, National Center for Traditional Chinese Medicine (TCM) Inheritance and Innovation, Guangxi Botanical Garden of Medicinal Plants, No. 189 Changgang Road, Xingning District, Nanning 530023, People's Republic of China
[2]National Engineering Research Center for Southwest Endangered Medicinal Materials Resources Development, Guangxi Botanical Garden of Medicinal Plants, Nanning, China
[3]College of pharmacy, Guangxi University of Chinese Medicine, Nanning, China
[4]Guangxi Key Laboratory of High Quality Formation and Application of Genuine Medicinal Materials / Guangxi Traditional Chinese Medicine Breeding Technology Innovation Center, Guangxi Botanical Garden of Medicinal Plants, No. 189 Changgang Road, Xingning District, Nanning 530023, People's Republic of China

Background

The NAC gene family, one of the largest plant-specific transcription factor families, derives its name from the founding members *NAM* (no apical meristem), *ATAF1/2* (*Arabidopsis thaliana* transcription activation factor 1/2), and *CUC2* (cup-shaped cotyledon 2), which govern developmental processes and stress responses [1, 2]. Structurally, NAC proteins feature a conserved N-terminal DNA-binding domain (subdivided into subdomains A–E) and a variable C-terminal transcriptional regulatory region (TRR) [3, 4]. While subdomains A, C, and D are evolutionarily conserved, the flexibility of subdomains B and E enables functional diversification, including roles in secondary metabolism regulation [3, 5].

NAC transcription factors are pivotal regulators of specialized metabolite biosynthesis, particularly alkaloids and flavonoids. In *Camellia sinensis*, *CsNAC086* directly activates *chalcone isomerase* (*CHI*) and *flavonol synthase* (*FLS*), driving flavonoid accumulation [6]. Similarly, *OpNAC1* in *Ophiorrhiza pumila* orchestrates camptothecin biosynthesis by binding to the *TDC* (tryptophan decarboxylase) promoter [7]. Stress-responsive NACs further modulate metabolite synthesis indirectly. For instance, drought-induced *TrNAC002* in white clover upregulates *CHS* and *CHI*, amplifying flavonoid production as a protective mechanism [8]. Mechanical damage, such as tobacco topping, activates jasmonic acid-indole-3-acetic acid (JA-IAA) crosstalk to elevate *NtNAC-R1* expression, enhancing nicotine biosynthesis while promoting lateral root growth [9]. These findings underscore NACs as molecular hubs integrating developmental cues and environmental stress with specialized metabolism.

In *Sophora tonkinensis* Gagnep. (*S. tonkinensis*), a medicinal plant prized for quinolizidine alkaloids (matrine, oxymatrine) and flavonoid glycosides (flavonoids, dihydroflavones, flavonols, genistin, genistein, trifolirhizin), the biosynthetic pathways remain partially characterized [5, 10–13]. The pathway of quinolizidine alkaloid synthesis involves decarboxylation of lysine as a precursor to cadaverine, followed by oxidative deamination to produce 5-aminovaleraldehyde, which is then spontaneously cyclized to Δ^1-Piperideine. This process is obtained by a reaction catalyzed by lysine decarboxylase and copper aminocarboxylase [14]. While the number of genes involved in the flavonoid anabolic pathway is large, and the regulatory mechanism is complex. It is now widely accepted that the flavonoid synthesis pathway consists of the phenylpropane pathway from the substrate phenylalanine through the catalytic and hydroxylation processes of various enzymes, like phenylalaninammonialyase (PAL), cinnamate 4-h ydroxylase (C4H), 4-coumarate: coenzyme A ligase (4CL), CHS, chalcone reductase (CHR), CHI, and others [15–17]. Notably, research has shown that NAC genes are involved in the biosynthetic pathways of alkaloids and flavonoids in medicinal plants. E.g.,weighted gene co-expression network analysis in *Litsea coreana* identified NAC transcription factors as hub genes co-expressed with *PAL*, *C4H*, and *CHI* [18], suggesting conserved regulatory roles in flavonoid biosynthesis. There is also *CsNAC086* in tea tree, OpNAC1 found in *Ophiorrhiza pumila*, *TrNAC002* from white clover, and the NAC transcription factor of Amomum tsao-ko fruit [6–8, 19]. Despite this, NAC genes in *S. tonkinensis* remain uncharacterized, limiting mechanistic insights into their metabolite regulation.

Here, the NAC gene family in *S. tonkinensis* (named: StNAC) were characterized using the whole genome of this plant (unpublished). Members of the *StNAC* gene family were identified and analyzed using bioinformatics methods. Furthermore, we analyzed the expression patterns of these genes in different plant tissues and their correlation with metabolites. These results bridge the gap between NAC functional diversity and specialized metabolism in *S. tonkinensis*, providing a foundation for targeted metabolic engineering.

Materials and methods

Characterization of the StNAC genes

Firstly, a Hidden Markov Model (HMM) was constructed using HMMER 3.0 software with 105 known *A. thaliana* NAC protein family sequences [5]. This model was then employed to search all coding protein sequences of *S. tonkinensis* to identify potential NAC gene family sequences. Reference NAC gene family sequences were obtained by sequence alignment of *S. tonkinensis* proteins, with an e-value threshold set to 1e-5, and potential NAC sequences were identified using blastp (version: ncbi-blast-v2.10.1+). The identified potential sequences were compiled as candidate NAC family protein sequences. These candidate sequences were subsequently annotated with structural domains using PfamScan (version: v1.6) and Pfam A (version: v33.1) databases [20, 21]. Sequences containing the structural domains PF02365 and PF01849 were then classified as final NAC gene family sequences.

Sequence and phylogenetic analysis

All coding sequence (CDS) of *StNAC* genes were analyzed for gene length, amino acid length, relative molecular weight, isoelectric point, hydrophilicity, stability, and other physicochemical properties using ExPASy (https://www.expasy.org). Multiple sequence alignment was perfo rmed using MAFFT (version v7) with default parameters to optimize residue homology [22]. The resulting alignment was subsequently utilized to construct a maximum likelihood (ML) phylogenetic tree in MEGA10 software. Evolutionary model selection was based on the Jones-Taylor-Thornton (JTT) substitution matrix, with partial

deletion handling of gaps and ambiguous sites (site coverage cutoff: 50%). Topological robustness was assessed through 1,000 bootstrap replicates. Final tree annotation and visualization were executed using the iTOL v6 platform (https://itol.embl.de/), incorporating hierarchical clustering of clades.

Transmembrane structural domain analysis and subcellular localization

Properties such as hydrophobicity, charge bias, helix length, and topological constraints of the transmembrane domains in the NAC gene family were analyzed using DeepTMHMM (version 1.0.8), a deep learning-based program for predicting transmembrane helices. Subcellular localization of *S. tonkinensis* NAC gene family members was determined using the tool available at https://wolfpsort.hgc.jp/.

Gene structure and conserved motif analysis

Gene structure analysis of 85 *StNAC* genes were conducted using General Feature Format (GFF) files and visualized with Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/). Conserved motifs were identified using MEME software (version v5.0.5, http://meme-suite.org/), with motif widths set between 6 and 50 and a maximum of 15 motifs sought.

Identificaiton of cis-elements in the promoter region of genes

The 2000 bp upstream region of the gene was considered as the promoter regulatory sequence, and transcription factors binding sites within this promoter region were predicted using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), which provides labeli ng and visualization of binding site locations on the gene promoter map.

Chromosome localization and covariance analysis

A physical map of the chromosomes was created using MG2C (http://mg2c.iask.in/mg2c_v2.1/) based on the chromosomal locations of *S. tonkinensis* NAC family genes. Covariance analysis was conducted using MCScanX with the following parameters: MATCH_SCORE = 50, MATCH_SIZE = 5, GAP_PENALTY = -1, OVERLAP_WINDOW = 5, E_VALUE = 1e-05, and MAX GAPS = 25. This analysis aimed to detect segmental and tandem duplications in the data. The non-synonymous substitution rate (Ka) and synonymous substitution rate (Ks) were evaluated using KaKs_Calculator (version 2.0), with the Ka/Ks ratio representing the rate of nonsynonymous to synonymous substitutions between protein-coding genes.

Expression of the *StNAC* genes in different tissues

Total RNA was extracted from the *S. tonkinensis* tissue using TRIzol® Reagent according the manufacturer's instructions. The *S. tonkinensis* RNA-seq transcriptome librariy was prepared following Illumina® Stranded mRNA Prep, Ligation (San Diego, CA) using 1 μg of total RNA. After quantified by Qubit 4.0, the sequencing library was performed on NovaSeq X Plus platform (PE150) using NovaSeq Reagent Kit. Expression patterns of *StNAC* genes in root, stem, leaf, and seed were investigated by the Majorbio Cloud Platform (https://cloud.majorbio.com/page/tools/) with retrieving tra nscriptomic data from previous studies (unpublished), which used these gene IDs as queries. Transcript abundance was reported as fragments per kilobase of exon model per million mapped reads (FPKM). To cluster genes with similar expression profiles, hierarchical clustering was performed on log2 (FPKM + 1) RNA-seq data using Cluster version 3.0, and the results were visualized with Java TreeView.

Correlation analysis of the *StNAC* genes

To further investigate the association of *StNAC* genes with alkaloid and flavonoid constituents, the contents of matrine, oxymatrine, genistin, genistein, and trifolirhizin in different tissues were obtained from our team. The contents of the five metabolites were detected by high performance liquid chromatography (HPLC). Their extraction and quantification methods were based on the Pharmacopoeia of the People 's Republic of China (2020 edition) [23].The correlation between *StNAC* genes and the levels of these compounds was analyzed using Spearman's rank correlation in R version 3.6.2, with $p < 0.05$ was considered statistically significant.

Expression analysis of *StNAC* genes by qRT-PCR

Tissues from roots, stems, leaves, and seeds of *S. tonkinensis* cultivated at the Guangxi Botanical Garden of Medicinal Plants were collected, with each tissue type sampled in triplicate. Samples were immediately wrapped in foil and snap-frozen in liquid nitrogen, then stored at -80 °C for RNA extraction. Total RNA was extracted using the FastPure Universal Plant Total RNA Isolation Kit (Vazyme, China). Reverse transcription was performed with HiScript® III RT SuperMix for qPCR (Vazyme, China), and real-time quantitative PCR (RT-qPCR) was conducted using ChamQTM Universal SYBR® qPCR Premix (Vazyme, China). Three biological replicates were prepared for each sample. Amplification specificity was confirmed by melting curve analysis, and relative expression levels were calculated using the $2^{-\Delta\Delta CT}$ method. Duncan's multiple range test (SPSS version 17.0) was employed to assess significant differences between samples. Primer design for specific *StNAC* genes

primers were carried out using Primer version 5.0 (Table S8).

Statistical analysis

Three biological replicates were included in the experiments. Statistical significance was assessed using Student's t-test in R software, with a significance level set at $p < 0.05$.

Results

Characterization of the NAC transcription factors in *S. tonkinensis*

A genome-wide search for the NAC gene family in *S. tonkinensis* was conducted using HMMSearch and BLASTP methods. After removing redundant sequences and confirming the presence of conserved NAC or NAM structural domains, a total of 85 putative NAC genes were identified and designated as *StNAC1* to *StNAC85* (Fig. 1). All identified proteins contained the conserved NAM domain (PF02365), which supports the accuracy of these predictions. The length of encoded proteins ranged from 147 amino acids (StNAC15) to 662 amino acids (StNAC7), with an average length of 355 amino acids. The molecular weight (MW) of the StNAC proteins varied from 17,379.62 Da (StNAC15) to 76,021.85 Da (StNAC8), with a mean of 40,257.67 Da. The predicted isoelectric point (pI) of the StNAC proteins values ranged from 4.7 (StNAC27) to 9.43 (StNAC63), averaging 6.83, indicating a mix of acidic and basic residues. The aliphatic index ranged from 44.98 to 79.51, reflecting a high proportion of aliphatic chains and suggesting that these proteins are fat-soluble. These variations in protein characteristics imply diverse functional roles. Subcellular localization predictions indicated that most StNAC proteins were located in the nucleus, with others found in the cytoplasm (10), chloroplasts (7), peroxisomes (4), and a minimal numbers in the plasma membrane and mitochondria (1), suggesting specialized functions in various

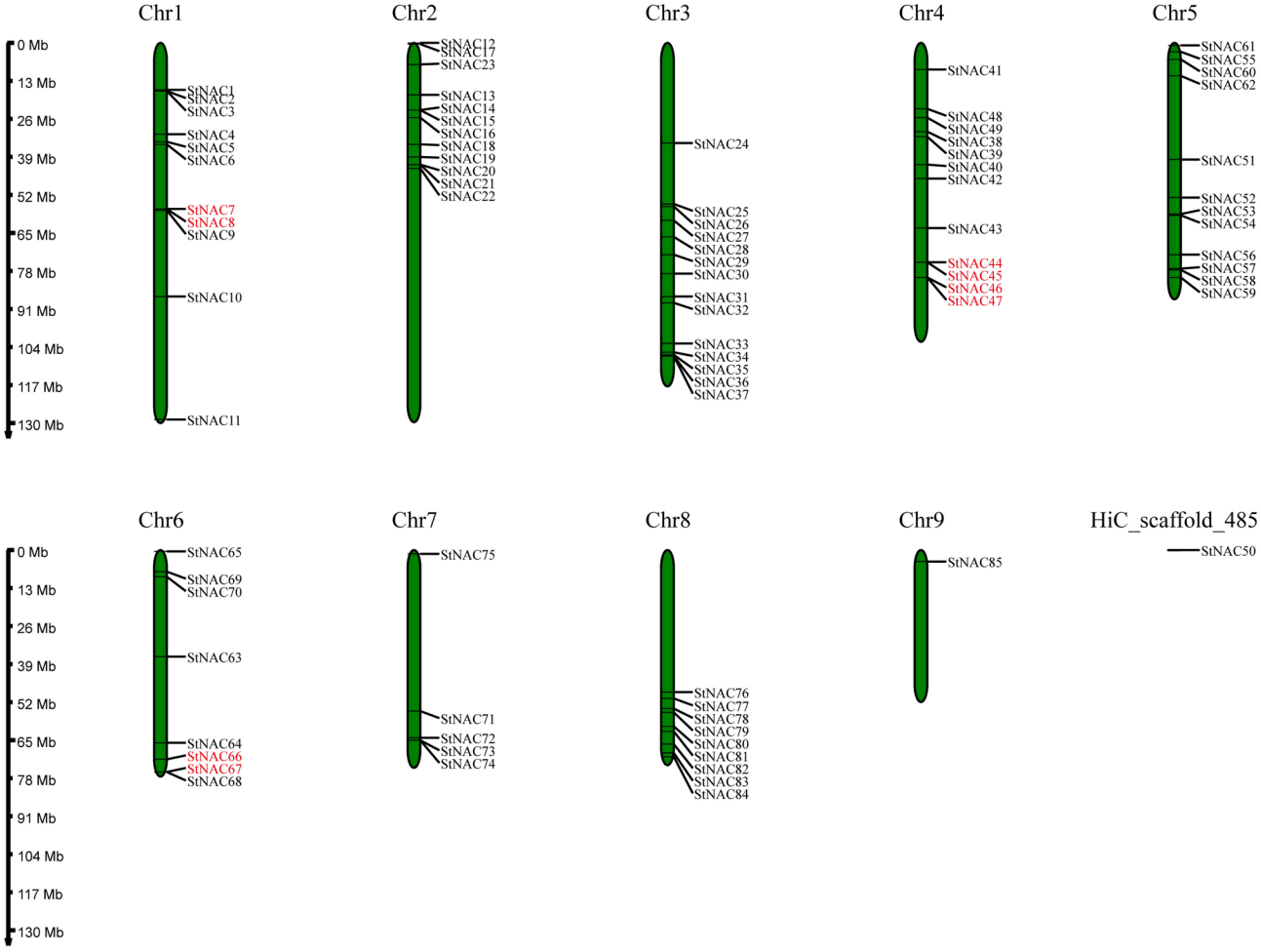

Fig. 1 Chromosomal location of the NAC family genes in *S. tonkinensis*. The vertical lines represent the chromosomes of *S. tonkinensis*. Chromosome numbering was located at the top of the chromosome. The scale on the left stands for the length of the chromosome. Tandem duplicated genes are highlighted in red

subcellular compartments. However, it should be noted that these predictions are computational in nature and have not been experimentally validated (Table S1).

The 85 *StNAC* genes exhibited non-random chromosomal distribution across nine chromosomes in *S. tonkinensis* (Fig. 1 and Table S1). Notably, *StNAC50* was mapped to overlapping clusters within unassembled genomic scaffolds. Chr3 displayed the highest gene density, containing 14 *StNAC* loci with relatively uniform spacing in its mid-distal region, while there is only one gene on Chr9. Chr2, 4, and 5 each accommodated 12 genes, and the remaining four chromosomes contained 5–9 genes with dispersed genomic arrangements. Furthermore, 8 genes formed a cluster of 4 tandem duplicates, highlighting tandem duplication as a significant factor in the expansion of the NAC gene family in *S. tonkinensis* (Table S1).

Multiple sequence comparison and phylogenetic analysis of *StNAC* genes

To investigate the sequence characteristics of conserved regions in StNAC proteins, multiple sequence alignments of all NAC proteins were performed. Most StNAC proteins' N-termini feature five conserved domains, designated motifs A through E (Fig. S1). However, some StNAC members lack one or more of these domains. For instance, proteins in subgroups C1/2/3/4 (except StNAC26/64) contain only motifs D and E. Several StNAC proteins, such as StNAC31/79/11/17/58, exhibit incomplete motif A, while StNAC26 and StNAC64 have incomplete motif D. These variations in motif presence suggest functional diversity among the *StNAC* gene family members in *S. tonkinensis*.

To explore the evolutionary relationships among *S. tonkinensis* NAC transcription factors, a phylogenetic tree was established using the NAC protein sequences from *S. tonkinensis* and *A. thaliana*. All members from both species were categorized into 15 subgroups that were designated as C1-15, respectively (Fig. 2). Homologous pairs, showing close phylogenetic relationships, were identified and clustered together. This classification into subgroups, based on evolutionary relationships to *A. thaliana*, offered insights into the potential functions of these *StNAC* genes [5]. The results indicated that *StNAC* proteins were distributed across all 15 subgroups, suggesting that *S. tonkinensis* did not experience significant NAC gene loss. The number of *StNAC* genes in each subgroup ranged from 2 to 10. The C11 subgroup had the largest number of members, while the C2/5 subgroups had only 2 *StNAC* genes each. The clustering of multiple NAC proteins from the same species likely results from segmental duplication events in their genomes [24].

Structural analysis of the *StNAC* gene family

The conserved regions within the 85 *StNAC* genes were delineated through multiple sequence comparisons (Fig. S1 and Fig. 3). The analysis revealed a high degree of conservation at the N-terminal, while the C-terminal exhibited significant variability, consistent with previous findings. Five conserved domains were identified, all located at the N-terminus, whereas no conserved domains have been found in the vicinity of the C-terminus. Examination of the conserved structural domains in the 85 *StNAC* genes indicated that subdomains A, C, and D were conserved, whereas subdomains B and E were divergent. Detailed scrutiny showed that subdomains B and E are not conserved in C12 sequences. The conservation of subdomains A, C, and D suggests their essential role in the function of NAC gene family, while the variability in subdomains B and E may reflect unique features of specific NAC genes.

Motif 2 + 8, motif 4, motif 1, motif 5 + 3, and motif 6 correspond to subdomains A, B, C, D, and E of the conserved domain in NAC transcription factor, respectively (Fig. 3B and Fig. S2). Analysis indicated that most conserved motifs were concentrated in the N-terminal region of the NAC domain, underscoring their significance in *StNAC* genes function. Frequent motifs, including motifs 2/3/5/6/8, appeared across nearly all subgroups from C1 to C14, which highlights their importance. Conversely, motifs 7/9/10/11/13/15, found only in subgroups C3/9/10, suggested newly emerged structural domains with potentially specialized functions. Overall, motifs within the same subgroups generally shared similar compositions and arrangements, which may correlate with the specific functions of these subgroups.

To elucidate the evolutionary dynamics of NAC genes in *S. tonkinensis*, DNA sequences were compared, and the exon-intron organization within the open reading frames of *StNAC* genes was examined (Fig. 3C). The number of introns in *StNAC* genes ranged from 1 to 7, with 57 *StNAC* genes had 2 introns, which represented the majority of the identified genes (67.06%). Conversely, *StNAC71*, featuring 8 introns, constituted the smallest proportion (1.18%) (Table S3). Members within the same evolutionary branch generally exhibited a conserved number of introns. For instance, all *StNAC* genes in subgroups C9/10/11/12/14/15 each contained 2 introns. In contrast, subgroup C2 ranged from 1 to 4 introns, subgroup C5 varied from 5 to 6 introns, and other subgroups displayed a broader range from 1 to 7 introns (Fig. 3 and Table S3). Overall, *StNAC* genes within the same branch shared similar exon-intron structures and motif compositions. This structural and compositional variability among branches suggested that the evolutionary divergence and specialization of *StNAC* genes likely occurred early in the evolution of *S. tonkinensis*.

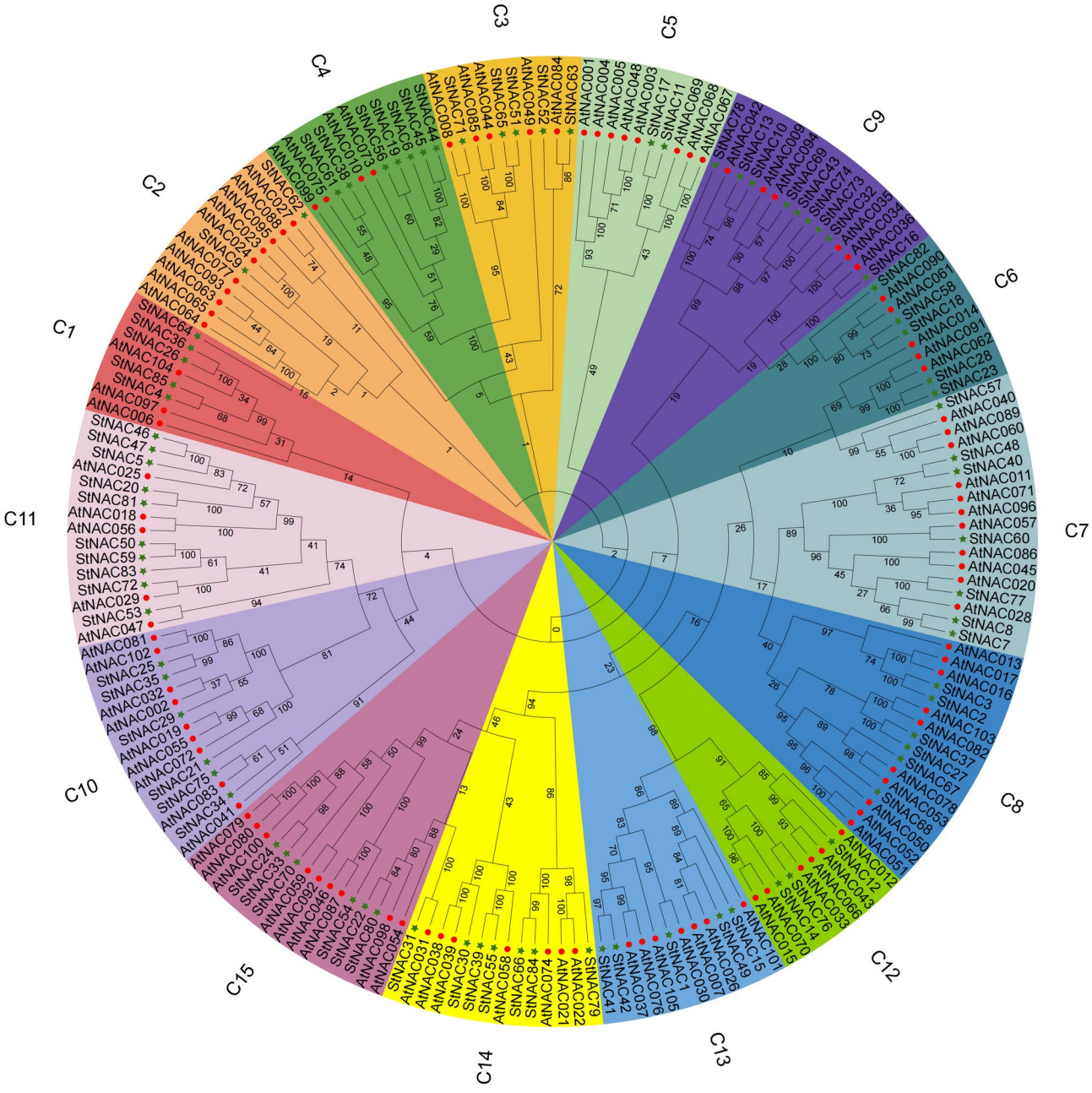

Fig. 2 Phylogenetic relationship among NAC superfamily genes of *S. tonkinensis* and *A. thaliana*. Construction of a rootless phylogenetic tree with maximum likelihood (ML) method These proteins were divided into 15 subgroups and each of them was assigned a different color. The numerical values on the phylogenetic tree branches represent bootstrap support percentages, with higher values indicating stronger statistical confidence in the corresponding clade topology. The Red circles indicated NAC genes in *A. thaliana*, and green pentagrams showed NAC genes in *S. tonkinensis*

Cis-elements analysis of *StNAC* genes

A 2000 bp genomic sequence upstream of the transcription start sites (TSS) of *StNAC* genes was analyzed to predict cis-acting elements. This analysis identified 3,010 cis-elements, averaging 35 elements per *StNAC* gene. These elements were predominantly associated with phytohormone responsiveness (1,006), stress response (740), light responsiveness (926), and plant growth and development (102) (Fig. 4 and Table S4). Furthermore, integrated analysis of cis-regulatory elements and expression data revealed that the promoter regions of *StNAC58/61*—genes significantly negatively correlated with genistin accumulation—harbor the MBSI cis-acting element, a MYB-binding site implicated in flavonoid biosynthetic regulation. This suggested that *StNAC58/61* may act as negative regulators of genistin biosynthesis. Additionally, green tissue-specific promoters were found to predominantly contain multiple light-responsive elements (LREs),

Fig. 3 Analysis of motifs and gene structure of *StNAC*. (**A**) Phylogenetic tree of *StNAC* gene family members. (**B**) Conserved motifs of *StNAC* genes. each motif was depicted by a colored box, the position of each motif can be estimated using the scale at the bottom, and different subclades of the rootless tree were marked with different colors. (**C**) Predicted gene structure of *StNAC* genes. CDSs, untranslated regions (UTRs), and introns were shown using blue boxes, green bars, and grey lines

including G-box, GT1 motif (or GATA motif), AT-rich elements, and Z-box, which are critical for transcriptional regulation. In this study, *StNAC4* (highly expressed in stems) contained 6 G-box motifs, while *StNAC61* harbored 6 GT1 motifs. Leaf-preferential *StNAC* genes, particularly *StNAC17*, exhibited abundant LREs, with five G-box and three GT1 motifs. Notably, *StNAC29*/*24*/*33*, which showed seed-specific expression, contained characteristic cis-regulatory motifs such as the GCN4 motif or RY motif, both associated with seed-specific promoter activity.

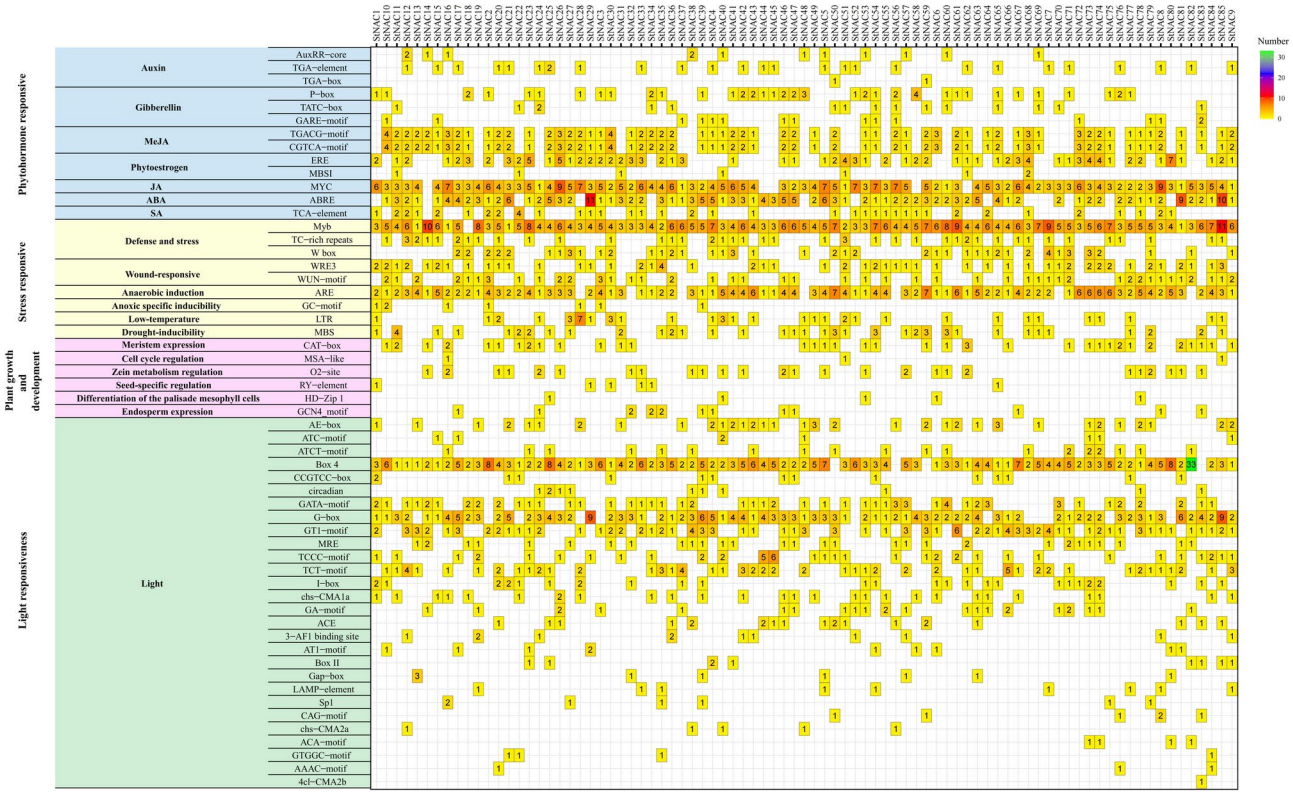

Fig. 4 Cis-elements analysis of NAC genes in *S. tonkinensis*. The color of the box from yellow to green signifies the number of cis-acting elements, and the different colored modules on the left side denote the different classes of cis-acting elements

Prediction of amino acid transmembrane structure and hydrophobicity analysis of *StNAC* genes

Prediction of the transmembrane structures for the 85 *StNAC* genes revealed that only *StNAC3* possessed a transmembrane domain. This domain, which spanned residues 583–599 in the amino acid sequence, was situated exclusively within the C-terminal region and comprised 17 amino acids. The remaining *StNAC* genes lacked transmembrane structures (Fig. S3). The hydrophilicity index ranged from – 1.127 to -0.337, with a lower value indicating greater hydrophilicity. Notably, *StNAC81* exhibited the highest hydrophilicity among the *StNAC* genes (Table S1).

Secondary and tertiary structure analysis

The secondary structure of StNAC proteins was analyzed using the SOPMA online program, revealing that the *StNAC* gene family predominantly consisted of α-helices, extended strands, *β*-turns, and random coils. Random coils were most prevalent, ranging from 43.35% (StNAC64) to 72.05% (StNAC48). The proportion of α-helices varied between 11.40% (StNAC56) and 29.95% (StNAC8). Extended strands were present at 8.18% (StNAC69) to 28.32% (StNAC80), while *β*-turns ranged from 1.32% (StNAC16) to 9.68% (StNAC19). Among the 85 StNAC members, 18 exhibited a structure ratio of random coil > extended strand ≥ α-helix > *β*-turn, including StNAC6/10/11/15/19. The remaining 67 members had a structure ratio of random coil > α-helix > extended strand > *β*-turn, indicating that side-chain interactions significantly influenced StNAC proteins (Fig. S4 and Table S5).

The tertiary structure of the StNAC proteins was predicted using Swiss-model, and the family was categorized into six groups (A-F). Group E contained 45 members, group D had 14, and group A included 8 members, while groups B/C/F each had 6 members. Overall, *StNAC* gene family members predominantly displayed random coils, with other structural elements distributed throughout, aligned with secondary structure predictions (Fig. S5).

Analysis of gene duplication in *StNAC* genes

Gene duplication plays a key role in the evolutionary processes of organisms, providing a foundation for physiological and morphological variations in plants [25] The study demonstrated that both segmental duplication/whole genome duplication (SD/WGD) and tandem duplication (TD) play crucial roles in gene expansion and functional differentiation across multiple gene families [26]. To explore tandem duplications within the NAC gene family of *S. tonkinensis*, sequence comparisons were employed to identify such duplications (Fig. 5). Analysis revealed

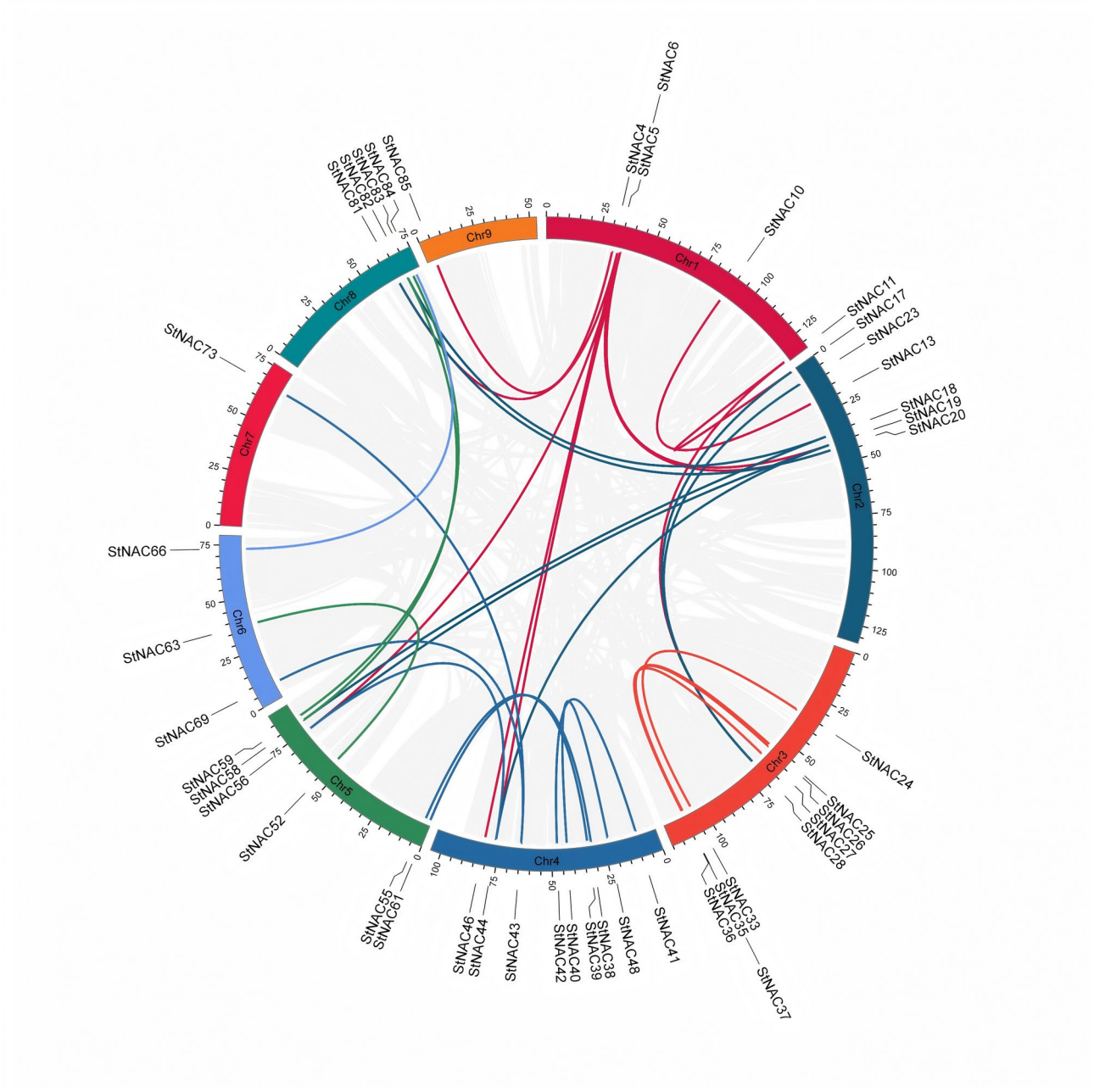

Fig. 5 Collinearity analysis of the NAC gene family in *S. tonkinensis*. Gray lines indicate all homozygous blocks in the genome. Segmental repeats of NAC gene pairs on different stainings were indicated and labeled with different color colors

that 8 *StNAC* genes (9.41%, forming 4 gene pairs) underwent TD events, while 45 genes (52.94%, comprising 32 gene pairs) experienced SD events, indicating SD as the primary driver for *StNAC* family expansion. Subsequent investigation of expression patterns in various *S. tonkinensis* tissues showed that among 32 segmentally duplicated gene pairs, four pairs (*StNAC23*/*28*, *StNAC24*/*33*, *StNAC11*/*28*, and *StNAC56*/*6*) exhibited similar expression profiles in roots, stems, leaves, and seeds, suggesting post-duplication functional redundancy. Notably, two tandemly duplicated pairs (*StNAC7*/*8* and *StNAC46*/*47*) displayed divergent expression patterns, with *StNAC46* showing peak expression in leaves versus *StNAC47*'s preferential expression in roots, potentially indicative of neofunctionalization. The remaining tandem pairs maintained comparable expression profiles, possibly implying subfunctionalization. These observations collectively suggest that successive duplication events in *StNAC* genes may facilitate functional specialization through redundancy, neofunctionalization, and subfunctionalization.

Furthermore, all duplicated gene pairs exhibited Ka/Ks ratios < 1, demonstrating strong purifying selection pressure acting on *StNAC* genes (Table S6).

Co-linearity analysis of NAC genes between genomes

To elucidate the evolutionary relationships of NAC gene families across different species, a comparative isobaric map was constructed for *S. tonkinensis* and *A. thaliana* (Fig. 6 and Table S7). This analysis revealed 79 homologous NAC gene pairs between the two species. Additionally, 20 *StNAC* genes were identified as being associated with at least two homologous gene pairs, indicating their potential significance in the evolution of the NAC gene family.

Expression of *StNAC* genes in different tissues

Expression profiles of *StNAC* genes across four different tissues were analyzed using publicly available RNA-seq data (Fig. 7). Results indicated that 78 *StNAC* genes were highly expressed in at least one tissue. The highest expression levels were observed in stems (49 genes), followed by roots, leaves, and seeds, suggesting their potential role as key regulators in root development. It was also noted that *StNAC29* and *StNAC33* exhibited significantly high expression in seeds, which correlated with the presence of a single RY element in their promoter regions. *StNAC35/30/65/25* showed the highest expression in roots, highlighting their involvement in root development. Additionally, six genes (*StNAC8/50/59/18/41/3*) were prominently expressed in stems and leaves, while seven genes (*StNAC5/34/27/72/51/68/67/28*) were notably expressed in both roots and stems. These tissue-specific *StNAC* genes present promising candidates for further functional studies to elucidate their regulatory roles.

Correlation analysis of the *StNAC* genes

To better characterize the upstream regulatory NAC genes involved in alkaloid and flavonoid biosynthesis, correlation analysis was performed using RNA-Seq data and metabolite content. Significant correlations were observed between NAC gene expression and the contents of matrine, oxymatrine, genistin, genistein, and trifolirhizin (Fig. 8). Specifically, *StNAC77* and *StNAC81*, which exhibited strong associations with matrine content, were highly expressed. Three genes significantly correlated with oxymatrine were specifically highly expressed in the seeds of *S. tonkinensis*, including *StNAC24*, *StNAC33*, and *StNAC83*. Additionally, the significant expression of *StNAC32*, *StNAC46*, and *StNAC79* was strongly correlated with genistein concentration.

Expression analysis of the *StNAC* genes by qRT-PCR

The expression of *StNAC* genes involved in alkaloid and flavonoid biosynthesis was analyzed by examining their expression in four different tissues (Fig. 9). Most genes linked to oxymatrine content, including *StNAC83*, *StNAC24*, and *StNAC33*, exhibited high expression specifically in the seeds of *S. tonkinensis*, and aligned with the RNA-seq data. Equally, the gene *StNAC32*, which showed a strong correlation with genistein concentration, were notably expressed in stem and leave. Conversely, genes linked to matrine content, including *StNAC77* and *StNAC81*, exhibited high expression specifically in the stem and seed. Yet, the trend of *StNAC81* expression levels in root and leave was not consistent with the RNA-seq results. *StNAC79* and *StNAC46* were remarkably expressed in stem and seed, respectively, which also did not agree with the RNA-seq results. Potential reasons for these discrepancies may include sample preparation variability, sequencing depth limitations, and experimental deviations.

Discussion

The NAC gene family represents one of the most extensive groups of plant transcription factors, playing vital roles in various processes including plant nutrition, secondary metabolite synthesis, responses to biotic and abiotic stresses, and hormone signaling pathways [24, 27–29]. This research aims to characterize and thoroughly analyze the NAC gene in *S. tonkinensis*. The results of the analyzed physicochemical properties demonstrate that *StNAC* genes varied in amino acid sequence length, molecular weight, and isoelectric point. That reflected the flexibility and functional diversity of *StNAC* genes. The subcellular localization of proteins provides essential insights into their functional mechanisms. In this study, computational tools were employed to predict

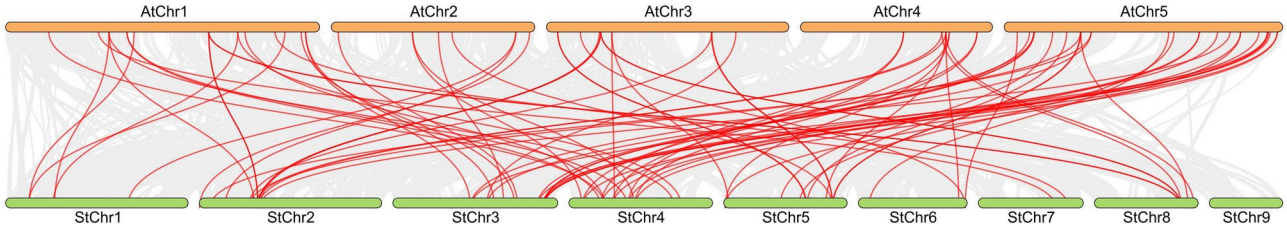

Fig. 6 Collinearity analysis of the NAC gene family between *S. tonkinensis* (StChr1-StChr9) and *A. thaliana* (AtChr1-AtChr5). Gray lines indicate co-lined blocks within the *S. tonkinensis* and *A. thaliana* genomes, and red lines highlight co-lined NAC gene pairs between the *S. tonkinensis* and *A. thaliana*

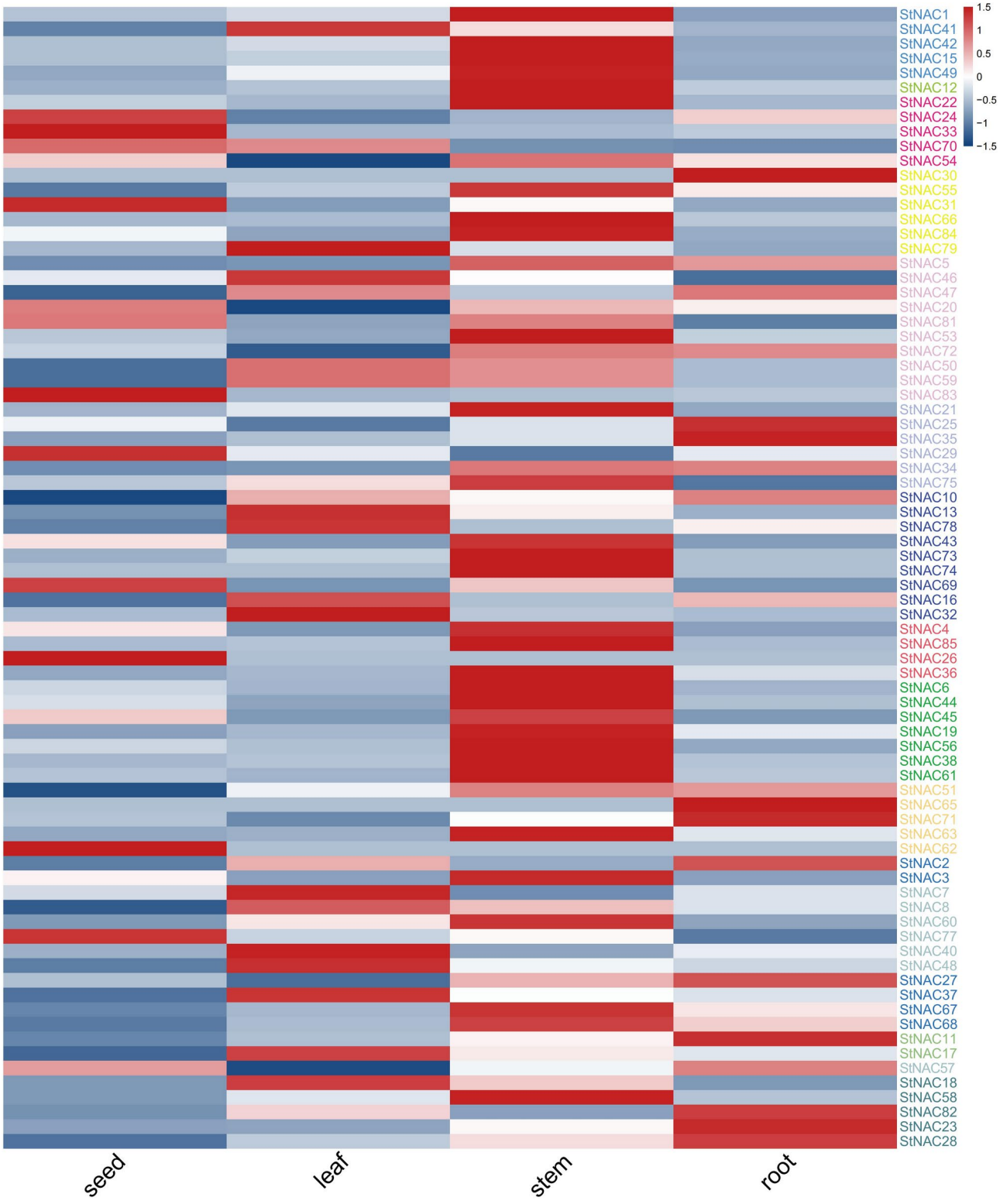

Fig. 7 Expression levels of *StNAC* genes in roots, stems, leaves, and seeds using RNA-seq. Red to blue color represents gene expression from high to low. Values of 1.5 and − 1.5 represents relative expression values

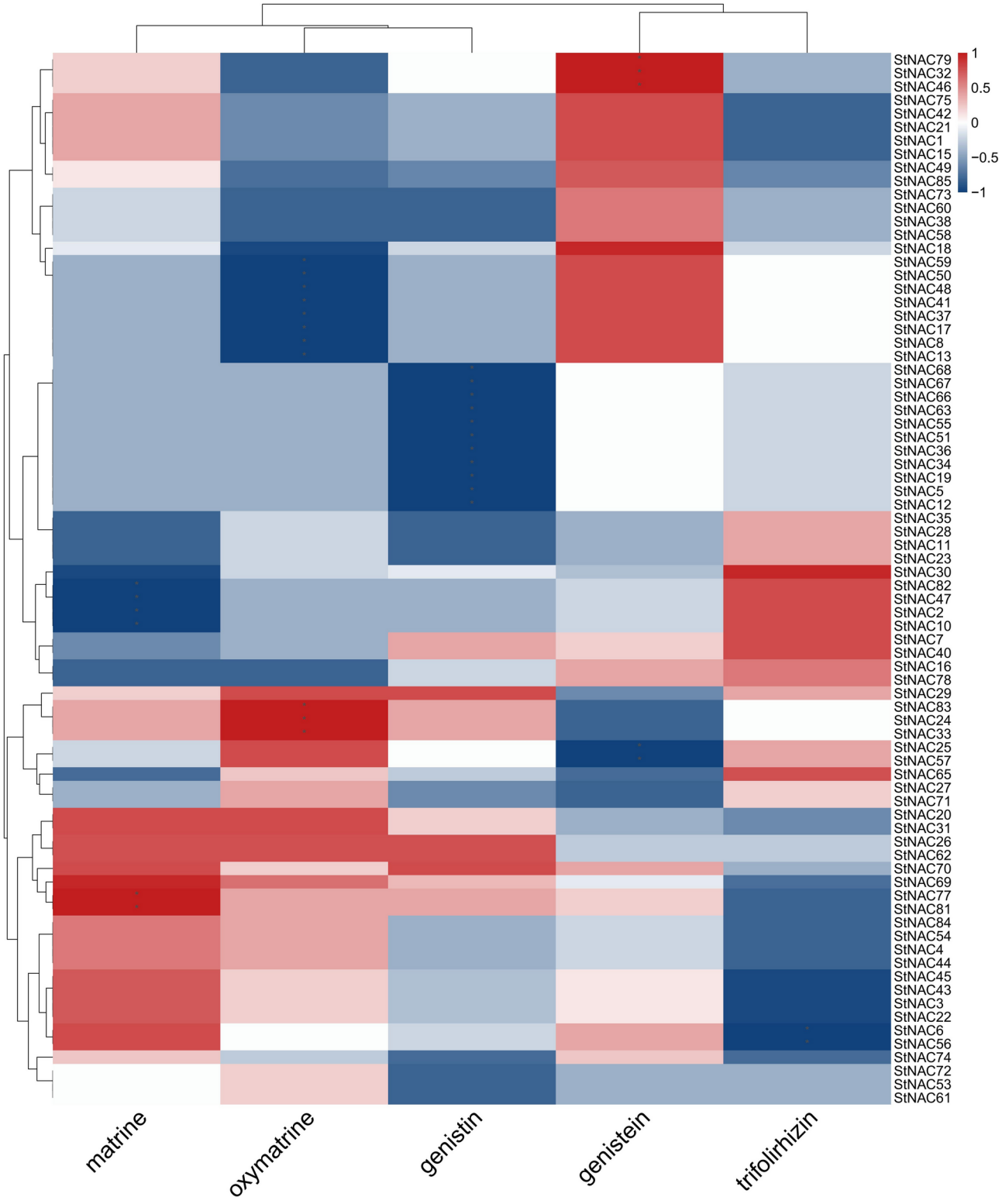

Fig. 8 Correlation analysis between *StNAC* genes and principal constituents content was analyzed by Spearman rank correlation analysis in R (v3.6.2), and $p < 0.05$ (*) was considered statistically significant

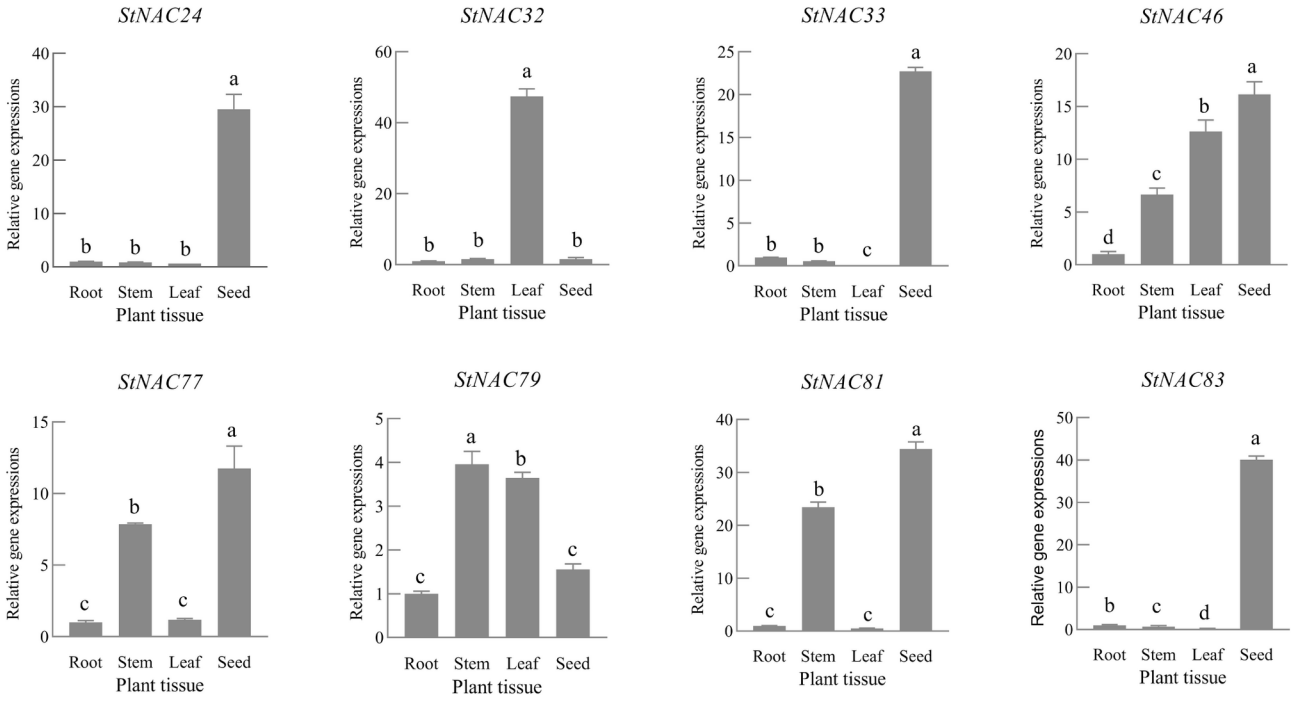

Fig. 9 Expression of eight *StNAC* genes in various tissues of *S. tonkinensis*. The mean ± SD of three independent repetitions is indicated by the error bars. Different lowercase letters (a, b, c, and d) denote significant differences ($p < 0.05$)

the localization of StNAC proteins. Notably, these predictions are inherently limited by their inability to account for dynamic cellular processes, protein interactions, or post-translational modifications that may influence localization [30–32]. To conclusively determine StNAC protein localization, experimental validation through methods such as GFP fusion assays or immunohistochemical microscopy remains indispensable, as demonstrated in analogous studies of plant transcription factors [33]. Such approaches will strengthen the foundation for investigating *StNAC* functions and interaction networks. The number of NAC transcription factor genes varies significantly across species. 85 *StNAC* genes were identified in the *S. tonkinensis* genome, a number comparable to *Ammopiptanthus mongolicus* (86), *Dendrobium nobile* (85), and *Salvia miltiorrhiza* (84) [24, 34, 35]. It is speculated that they may share some degree of similarity in the regulation of gene expression. Meanwhile, phylogenetic analysis of *S. tonkinensis* and *A. thaliana* categorized 85 *StNAC* genes into 15 subgroups (C1-C15) following the *A. thaliana* NAC protein classification framework. This subgroup organization aligns with the evolutionary patterns observed in *A. mongolicus* (15 subgroups) [24], but diverges from classifications reported in *S. miltiorrhiza* (9 subgroups) [35]. These interspecies discrepancies in subgroup numbers likely reflect a conservation and differentiation characteristics of NAC genes among different species. A total of 79 homologous gene pairs were identified between *S. tonkinensis* and *A. thaliana*, indicating a strong resemblance and close evolutionary relationship between these gene pairs. Despite this similarity, *StNAC* genes displayed considerable variability in physicochemical properties and uneven chromosome distribution, consistent with observations in *A. mongolicus*, *D. nobile* and *S. miltiorrhiza* [24, 34, 35].

The NAC domain governs DNA-binding activity at the N-terminus and plays pivotal roles in plant-pathogen interactions, stress adaptation, and critical physiological processes [36]. Structural studies reveal that the highly conserved subdomains C and D harbor nuclear localization signals and mediate promoter-specific DNA binding, forming the core functional module for transcriptional regulation [5, 37]. This implies that conserved motifs within these subdomains (e.g., motif 1, motif 5, and motif 3) may facilitate recognition of shared cis-regulatory elements. In contrast, subdomains B and E exhibit lower sequence conservation, contributing to functional diversification of NAC transcription factors [38]. Subdomain E, in particular, has been implicated in developmental stage modulation and tissue specificity [5]. Notably, motif 6 within subdomain E of StNAC46 may play essential roles in growth regulation. Subdomain A facilitates protein dimerization, with both homo- and heterodimerization documented across species [4, 39, 40]. For instance, heterodimer formation between ONAC127 and ONAC129 in rice is critical for seed development, as evidenced by nuclear colocalization via bimolecular fluorescence complementation [41]. Similarly, the A subdomain

of TsNAC1 from *Thellungiella salsuginea* interacts with the zinc-finger domain of TsHD1 to enhance thermotolerance, demonstrating evolutionary conservation of dimerization mechanisms [42]. These observations suggest that motifs 2 and 8, associated with subdomain A, mediate functional dimerization among StNAC proteins or with interacting partners. The C-terminal transcriptional regulatory domain (TRD) displays remarkable sequence divergence across species, driving a trans-activation region [38, 43]. Most *StNAC* genes possesed two introns and three exons, with members within the same phylogenetic branch exhibiting similar exon numbers. This structural similarity aligns with NAC genes in Oat and *Eucommia ulmoides* [44, 45], suggesting that the genetic composition of the NAC gene family in S. tonkinensis mirrors that of previously studied species.

Gene duplication acts as a pivotal mechanism in plant evolution, generating genetic diversity that serves as raw material for functional innovation [26]. This process can occur through multiple mechanisms, including WGD, TD, and SD, each contributing distinctively to gene family expansion [26]. In *S. tonkinensis*, the prevalence of SD over TD as the primary driver of *StNAC* family expansion (45 genes vs. 8 genes; Table S6) aligns with observations in other plant gene families, where segmental duplication events often provide raw material for evolutionary novelty [24, 44, 46]. The retention of duplicated *StNAC* genes under strong purifying selection (Ka/Ks < 1) suggests selective constraints to preserve ancestral functions, while their divergent expression profiles (e.g., *StNAC46*/*47* in leaves vs. roots) exemplify evolutionary trajectories to escape redundancy. Following duplication, genes may undergo distinct evolutionary fates, one copy retains the original function, while the other accumulates mutations leading to neofunctionalization or partitions ancestral roles through subfunctionalization [47]. For instance, in *Populus euphratica*, duplicated laccase genes (LACs) exhibit functional redundancy in post-translational modifications but diverge in stress-responsive roles under salinity [48]. Similarly, in Populus, for tandem duplication clusters containing three genes (*PNAC020*/*022*, *PNAC033*/*035*, *PNAC141*/*143*), the expression patterns of two membersare nearly identical, while the third member exhibits significant diversification, which may indicate neofunctionalization [49]. In *S. tonkinensis*, the coexistence of redundant (*StNAC23*/*28*) and divergent (*StNAC58*/*18*) gene pairs within the same family aligns with the "birth-and-death" model of gene family evolution [50]. Genes after tandem duplication usually behave in a genealogically specific manner compared to segmentation and whole genome duplication [51]. This can be confirmed in the phylogenetic results of the tandem duplication genes in this study, where *StNAC7/8*, *StNAC44/45*, *StNAC46/47*, and *StNAC67/68* are located in the same set of sub-branches, respectively. The results were the same for genes with segmental duplications.

Genes within the same subgroup often exhibit similar biological activities and functions [24, 44]. The 85 *StNAC* genes were classified into 15 subgroups, each characterized by analogous gene structures and motif arrangements. The functional roles of NAC genes in *S. tonkinensis* can be inferred from their *A. thaliana* homologs, providing a basis for further functional studies. Members within the same evolutionary branch tend to have highly specific protein sequences and similar functions. NAC and MYB transcription factors constitute a transcriptional regulatory network for secondary wall biosynthesis in ducts and fibers, in which *ANAC043*, *ANAC066*, and *ANAC012* act as master switches in this regulatory network as these factors co-regulate secondary cell wall biosynthesis in stem fibers [52–54]. Consequently, *StNAC12*/*76*/*14* from the C12 subgroup are likely involved in secondary wall synthesis in *S. tonkinensis* roots. The NAC triad *AtNAC017*, *AtNAC082* (C8), and *AtNAC090* (C6) negatively regulate leaf senescence by modulating salicylic acid levels and reactive oxygen species [55]. Similarly, *StNAC58*/*18*/*2*/*3*/*27*/*37* from the C6 and C8 subgroups may possess analogous functions. Furthermore, *ANAC019* (*AT1G52890*), *ANAC055* (*AT3G15500*), and *ANAC072* (*AT4G27410*) from the C10 subgroup are induced by drought, high salinity, and abscisic acid (ABA) [56]. Thus, it is hypothesized that *StNAC25*/*35*/*29*/*21*/*75*/*34* within the same subgroup are likely drought- and salt-responsive genes, playing a pivotal role in the survival of *S. tonkinensis* under adverse environmental conditions.

Cis-elements in gene promoters are crucial for regulating gene expression by serving as binding sites for transcription factors [57]. Analysis of 2000 bp promoter regions identified 3,010 cis-elements (average 35 per gene), predominantly linked to light responsiveness (30.76%), phytohormone signaling (33.43%), stress adaptation (24.58%), and developmental processes (3.39%) (Fig. 4). Light-responsive elements (G-box, GT1, Z-box) were relatively abundant, suggesting light quality/quantity may dynamically regulate *StNAC* expression. For instance, *StNAC4* (stem-specific) and *StNAC17* (leaf-preferential) harbored six G-box and three GT1 motifs, respectively, aligning with their tissue-specific expression [58]. Phytohormone-related elements included ABA, GA, MeJA, and auxin response motifs, consistent with NAC-mediated hormonal regulation reported in *Scutellaria baicalensis* Georgi, *Haloxylon ammodendron*, and cotton [59–61]. Notably, *StNAC58*/*61*, negatively correlated with genistin accumulation, contained MBSI motifs-MYB-binding sites implicated in flavonoid repression-supporting their role as regulators of

secondary metabolism [62]. Tissue-specific promoters further revealed functional specialization: *StNAC29* (seed-expressed) carried RY motifs associated with seed development, while *StNAC50*/*59* (stem/leaf-expressed) featured G-box/GT1 elements driving green tissue expression [63]. Additionally, several cis-acting elements related to abiotic stress responses [64], such as LTR, MBS, GATA-motif, TGACG-motif, CGTCA-motif, and ARE, were identified in the *StNAC* promoters. These findings suggest that *StNAC* genes likely play a role in the biotic stress resistance mechanisms of *S. tonkinensis*. These findings collectively highlight how *StNAC* cis-element diversity underpins their roles in environmental adaptation and organ-specific regulation.

The material basis of the pharmacological efficacy of *S. tonkinensis* is mainly alkaloids and flavonoids. NAC transcription factors play important regulatory roles in the biosynthesis of plant alkaloids and flavonoids [7, 65–67]. In order to investigate the effect of *StNAC* gene on alkaloid and flavonoid biosynthesis, a correlation analysis expression heatmap was constructed, including matrine, oxymatrine, genistin, genistein, and trifolirhizin. We found significant correlations between 8 *StNAC* genes and these metabolites. Specific compounds correspond to related these genes with some similarity in gene structure and promoters, such as *StNAC24/33/83*, with the same conserved motifs and number of introns. Promoter analysis showed that they all contain cis-acting elements in response to JA and ABA, signifying that the expression of these *StNAC* genes may be induced by both ABA and JA. To determine whether the eight *StNAC* genes identified in our study similarly promote alkaloid and flavonoid biosynthesis or contribute to tissue specificity in *S. tonkinensis*, we verified their expression levels. Interestingly, the flavonoid-related genes *StNAC32/46/81*,all of them were highly expressed in the aboveground parts of *S. tonkinensis*, namely, *StNAC32* was efficiently expressed in leaves, *StNAC46* was significantly expressed in tissues other than roots, and *StNAC79* was expressed at significantly higher levels in stems and leaves than in roots and seeds. Comparatively, the tissue expression specificity of genes related to alkaloid synthesis was more restricted, with the oxymatrine-related genes *StNAC24/33/83*, all of which were strongly expressed in the seeds, and the matrine-related genes *StNAC77/81*, which were prominently expressed in the stems and leaves. Besides, the results for the three genes *StNAC46*, *StNAC79*, and *StNAC81* were contradicted between RNA-seq and qPCR. The discrepancies in *StNAC46*, *StNAC79*, and *StNAC81* expression levels between RNA-seq and qPCR likely stem from methodological divergence and biological variability. RNA-seq captures genome-wide transcript abundance, while qPCR quantifies localized expression of specific isoforms, leading to inherent sensitivity differences [68, 69]. Technical factors such as RNA extraction protocols (e.g., column-based vs. TRIzol), RNA integrity (RIN), and degradation effects may bias RNA-seq toward truncated transcripts but minimally impact qPCR targeting short amplicons [70, 71]. Biological variability—including tissue sampling precision (bulk vs. cell-specific), temporal fluctuations (e.g., diurnal rhythms), or delayed post-harvest processing—could further amplify discrepancies, particularly for transiently expressed genes [72, 73]. Normalization approaches also contribute: RNA-seq relies on global methods (e.g., FPKM) to adjust for sequencing depth, whereas qPCR depends on reference genes (e.g., *Actin*) whose instability under experimental conditions may skew results [68, 69, 74–77]. To enhance consistency, appropriate standardized methods can be used, technical replications can be performed, and experimental setups can be carefully considered.

Generally speaking, specific compounds correspond to related these genes with some similarity in gene structure and promoters. Overexpression of *ANAC078* in *A. thaliana* significantly upregulated the transcript levels of flavonoid biosynthesis genes, such as *AtPAP1*, *AtCHS*, *AtCHI*, and *AtF3H* [78]. *StNAC32*/*46*/*79* identified in this study exhibited sequence identities of 47.4%, 41.6%, and 44.4% with *ANAC078*, respectively, based on full-length coding sequence alignment. Notably, these identities are higher than those previously reported for *MdNAC42* [79]. This suggests that *StNAC32/46/79* may be novel members of the NAC gene involved in the regulation of flavonoid synthesis. However, research on the regulation of NAC genes in alkaloid biosynthesis is very limited. Fortunately, functional predictions of *StNAC* genes can be strengthened through targeted experimental validation. To identify interacting proteins, yeast two-hybrid screening or co-immunoprecipitation (Co-IP) assays could be applied, as demonstrated in studies of *Haloxylon ammodendron* HaNAC1 and OsNAC016 in rice, which revealed interactions with stress-responsive kinases and transcription factors [80, 81]. Overexpression or CRISPR-Cas9-mediated knockdown of *StNAC* genes in *S. tonkinensis* hairy root cultures or heterologous systems like *Nicotiana benthamiana* could directly assess their roles in genistin biosynthesis. For example, transient overexpression of *AdNAC20* (linked to early-bolting) followed by comparative transcriptomics would clarify its regulatory impact on lignin and coumarin pathways [82]. Phenotypic consequences of *StNAC* manipulation could be evaluated under abiotic stress (e.g., drought, salinity) using established protocols for NAC gene functional analysis in model plants [83, 84]. Additionally, promoter-GUS fusions or tissue-specific RNA-seq in transgenic lines may uncover developmental roles, as shown for *VND7* in xylem differentiation [85]. In the future, we will

implement these approaches to address *S. tonkinensis*-specific biosynthetic networks.

Conclusions

In this study, 85 *StNAC* genes were systematically identified within the *S. tonkinensis* genome, displaying uneven distribution across nine chromosomes. Phylogenetic analysis classified these *StNAC* genes into 15 distinct subgroups. Comprehensive analyses of gene structure, motif organization, chromosomal localization, duplication events, evolutionary relationships, cis-acting elements, and tissue-specific expression patterns were performed. Notably, eight *StNAC* genes exhibited significant expression profiles in secondary metabolic pathways linked to alkaloids (matrine and oxymatrine) and flavonoids, implying their potential regulatory roles in the biosynthesis of these bioactive compounds in *S. tonkinensis*. The consistently high expression of five *StNAC* genes (*StNAC24*, *StNAC32*, *StNAC33*, *StNAC77*, and *StNAC83*) in relevant tissues may support their involvement in modulating these pathways. Future research should focus on a detailed functional characterization of *StNAC* genes to elucidate their molecular mechanisms in alkaloid and flavonoid biosynthesis in *S. tonkinensis*. These findings provides novel insights for advancing the understanding of *StNAC* gene functions in secondary metabolism.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06564-0.

Supplementary Material 1: Fig. S1. Alignment of multiple NAC protein sequences in S. tonkinensis. Some incomplete sequences or sequences that do not contain an N-terminal structural domain have been omitted from the above comparisons. No color indicates less than 35% similarity between amino acids of NAC gene family members at this locus. Fig. S2. Sequence logo representing conserved motifs. The height of each amino acid reflected the relative frequency of amino acids in position. Fig. S3. Prediction of the transmembrane structure for the NAC protein StNAC23. Fig. S4. Secondary structure of the NAC family in S. tonkinensis. The blue line, red line, green line and purple line mean Alpha helix, Extended stand, Beta turn and Random coil respectively.The number under the box indicated the sequence length of the protein. Fig. S5. Tertiary structure of NAC proteins in S. tonkinensis. Table S1. Characteristic features of the NAC gene family in S. tonkinensis. Table S2. Identification of NAM domains in S. tonkinensis. Table S3. Gene structure of the NAC gene family in S. tonkinensis. Table S4. Cis-elements in the promoter regions of the 85 StNAC genes. Table S5. Secondary structure analysis of NAC proteins from S. tonkinensis. Table S6. Ka/Ks ratios for tandem duplicate and collinear gene pairs. Table S7. Number of homologous gene pairs between S. tonkinensis and A. thaliana. Table S8. Primers used in the study.

Supplementary Material 2

Acknowledgements
We thank the editor and reviewers for their helpful comments and their input.

Author contributions
LX analyzed data and wrote the paper. SW, FW and YL designed the project. YL, SQ and FW performed experiments. SQ, GW and XG helped with the data analysis and examined the results. All authors contributed to the article and approved the submitted version.

Funding
This study was supported by the National Natural Science Foundation of China (82260747), Guangxi Natural Science Foundation (2025GXNSFAA069971), Guangxi Qihuang Scholars Training Program (GXQH202402), Guangxi elite team of medicinal plant conservation.

Data availability
The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 27 November 2024 / Accepted: 15 April 2025
Published online: 25 April 2025

References

1. 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. Plant Cell. 1997;9(6):841–57.
2. Souer E, van Houwelingen 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):159–70.
3. Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Matsusoka M, Hirano HY. Molecular analysis of the NAC gene family in rice. Mol Gen Genet. 2000;262(6):1047–51.
4. Olsen AN, Ernst HA, Lo Leggio L, Skriver K. NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci. 2005;10(2):79–87.
5. Luo G, Yang Y, Zhou M, Ye Q, Liu Y, Gu J, et al. Novel 2-arylbenzofuran dimers and polyisoprenylated flavanones from *Sophora tonkinensis*. Fitoterapia. 2014;99:21–7.
6. Song SS, Ran WX, Gao LH, Wang YC, Lv WY, Tao Y, et al. A functional study reveals *CsNAC086* regulated the biosynthesis of flavonols in *Camellia sinensis*. Planta. 2024;259(6):147.
7. Hao X, Wang C, Zhou W, Ruan Q, Xie C, Yang Y, et al. OpNAC1 transcription factor regulates the biosynthesis of the anticancer drug camptothecin by targeting Loganic acid O-methyltransferase in *Ophiorrhiza pumila*. J Int Plant Biol. 2023;65(1):133–49.
8. Zhang Y, Fu W, Pu Q, He Z, Li Z, et al. The white clover single-copy nuclear gene TrNAC002 promotes growth and confers drought resistance in plants through flavonoid synthesis. Plants (Basel). 2024;14(1):31.
9. Fu Y, Guo H, Cheng Z, Wang R, Li G, Huo G, et al. NtNAC-R1, a novel NAC transcription factor gene in tobacco roots, responds to mechanical damage of shoot meristem. Plant Physiol Biochem. 2013;69:74–81.
10. Wang R, Wang M, Wang S, Yang K, Zhou P, Xie X, et al. An Integrated characterization of contractile, electrophysiological, and structural cardiotoxicity of *Sophora tonkinensis* Gapnep. In human pluripotent stem cell-derived cardiomyocytes. Stem Cell Res Ther. 2019;10(1):20.
11. Liang JJ, Zhang PP, Zhang W, Song D, Wei X, Yin X, et al. Biological activities and secondary metabolites from *Sophora tonkinensis* and its endophytic fungi. Molecules. 2022;27(17):5562.
12. Zou J, Zhao L, Yi P, An Q, He L, Li Y, et al. Quinolizidine alkaloids with antiviral and insecticidal activities from the seeds of *Sophora tonkinensis* Gagnep. J Agr Food Chem. 2020;68(50):15015–26.
13. He LJ, Liu JS, Luo D, Zheng YR, Zhang YB, Wang GC, et al. Quinolizidine alkaloids from *Sophora tonkinensis* and their anti-inflammatory activities. Fitoterapia. 2019;139:104391.
14. Bunsupa S, Yamazaki M, Saito K. Lysine-derived alkaloids: overview and update on biosynthesis and medicinal applications with emphasis on Quinolizidine alkaloids. Mini-rev Med Chem. 2017;17(12):1002–12.

15. Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MS, Wang L, et al. The phenylpropanoid pathway and plant defence-a genomics perspective. Mo Plant Pathol. 2002;3(5):371–90.
16. Nabavi SM, Samec D, Tomczyk M, Milella L, Russo D, Habtemariam S, et al. Flavonoid biosynthetic pathways in plants: versatile targets for metabolic engineering. Biotechnol Adv. 2020;38:107316.
17. Weisshaar B, Jenkins GI. Phenylpropanoid biosynthesis and its regulation. Cur Opin Plant Biol. 1998;1(3):251–7.
18. Xie N, Guo Q, Li H, Yuan G, Gui Q, Xiao Y, et al. Integrated transcriptomic and WGCNA analyses reveal candidate genes regulating mainly flavonoid biosynthesis in *Litsea Coreana* Var. Sinensis. BMC Plant Biol. 2024;24(1):231.
19. Fu H, Xiong L, Ma M, Lu B. The potential function of chalcone isomerase (CHI) gene on flavonoid accumulation in Amomum tsao-ko fruit by transcriptome and metabolome. Int J Biol Macromol. 2025;297:139897.
20. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, et al. The Pfam protein families database. Nucleic Acids Res. 2002;30(1):276–80.
21. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222–30.
22. Cao Y, Feng X, Ding B, Huo H, Abdullah M, et al. Gap-free genome assemblies of two Pyrus bretschneideri cultivars and GWAS analyses identify a CCCH zinc finger protein as a key regulator of stone cell formation in Pear fruit. Plant Commun. 2023;6(3):101238.
23. National Pharmacopoeia Commission. Pharmacopoeia of the People.'s Republic of China-I (2020 edition). China Pharmaceutical Science and Technology Press of Beijing. 2020; 28.
24. Dorjee T, Cui Y, Zhang Y, Liu Q, Li X, Sumbur B, et al. Characterization of NAC gene family in *Ammopiptanthus mongolicus* and functional analysis of *AmNAC24*, an osmotic and cold-stress-induced NAC gene. Biomolecules. 2024;14(2):182.
25. Kuzmin E, Taylor JS, Boone C. Retention of duplicated genes in evolution. Trends Genet. 2022;38(1):59–72.
26. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. Plant Physiol. 2016;171(4):2294–316.
27. Foresti C, Orduna L, Matus JT, Vandelle E, Danzi D, Bellon O, et al. NAC61 regulates late- and post-ripening osmotic, oxidative, and biotic stress responses in grapevine. J Exp Bot. 2023;75(8):2330–50.
28. Bi Y, Wang H, Yuan X, Yan Y, Li D, Song F. The NAC transcription factor ONAC083 negatively regulates rice immunity against Magnaporthe *Oryzae* by directly activating transcription of the RING-H2 gene *OsRFPH2-6*. J Integra Plant Bio. 2023;65(3):854–75.
29. Liu X, Xing Q, Liu X, Muller-Xing R. Expression of the populus orthologues of AtYY1,YIN and YANG activates the floral identity genes AGAMOUS and SEPALLATA3 accelerating floral transition in Arabidopsis thaliana. Int J Mol Sci. 2023;24(8):7639.
30. Briesemeister S, Rahnenfuhrer J, Kohlbacher O. Going from where to why–interpretable prediction of protein subcellular localization. Bioinformatics. 2010;26(9):1232–8.
31. Zecha J, Gabriel W, Spallek R, Chang YC, Mergner J, et al. Linking post-translational modifications and protein turnover by site-resolved protein turnover profiling. Nat Commun. 2022;13(1):165.
32. Ramazi S, Zahiri J. Posttranslational modifications in proteins: resources, tools and prediction methods. Database (Oxford). 2021; baab012.
33. Xiao H, Zou Y, Wang J, et al. A review for artificial intelligence based protein subcellular localization. Biomolecules. 2024;14(4):409.
34. Fu C, Liu M. Genome-wide identification and molecular evolution of NAC gene family in *Dendrobium nobile*. Front Plant Sci. 2023;14:1232804.
35. Zhang H, Xu J, Chen H, Jin W, Liang Z. Characterization of NAC family genes in *Salvia miltiorrhiza* and NAC2 potentially involved in the biosynthesis of Tanshinones. Phytochemistry. 2021;191:112932.
36. Fuertes-Aguilar J, Matilla AJ. Transcriptional control of seed life: new insights into the role of the NAC family. Int J Mol Sci. 2024;25(10):5369.
37. Ray S, Basnet A, Bhattacharya S, Banerjee A, Biswas K, et al. A comprehensive analysis of NAC gene family in Oryza sativa Japonica: a structural and functional genomics approach. J Biomol Struct Dyn. 2023;41(3):856–70.
38. Ernst HA, Olsen AN, Larsen S, Lo Leggio L. Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. EMBO Rep. 2004;5(3):297–303.
39. Puranik S, Sahu PP, Srivastava PS, Prasad M. NAC proteins: regulation and role in stress tolerance. Trends Plant Sci. 2012;17(6):369–81.
40. Xie Q, Frugis G, Colgan D, Chua NH. *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. 2000;14(23):3024–36.
41. Ren Y, Huang Z, Jiang H, Wang Z, Wu F. A heat stress responsive NAC transcription factor heterodimer plays key roles in rice grain filling. J Exp Bot. 2021;72(8):2947–64.
42. Liu C, Ma H, Zhou J, Li Z, Peng Z. TsHD1 and TsNAC1 cooperatively play roles in plant growth and abiotic stress resistance of *Thellungiella halophile*. Plant J. 2019;99(1):81–97.
43. Tao Y, Wan JX, Liu YS, Yang XZ, Shen RF. The NAC transcription factor ANAC017 regulates aluminum tolerance by regulating the cell wall-modifying genes. Plant Physiol. 2022;189(4):2517–34.
44. Xu Y, Cheng J, Hu H, Yan L, Jia J, et al. Genome-wide identification of NAC family genes in oat and functional characterization of *AsNAC109* in abiotic stress tolerance. Plants (Basel). 2024;13(7):1017.
45. Zhang S, Xu T, Ren Y, Song L, Liu Z, Kang X, et al. The NAC transcription factor family in *Eucommia ulmoides*: Genome-wide identification, characterization, and network analysis in relation to the rubber biosynthetic genes. Front Plant Sci. 2023;14:1030298.
46. Huang X, Qiu X, Wang Y, Abubakar AS, Chen P, et al. Genome-wide investigation of the NAC transcription factor family in *Apocynum venetum* revealed their synergistic roles in abiotic stress response and Trehalose metabolism. Int J Mol Sci. 2023;24(5):4578.
47. Birchler JA, Yang H. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. Plant Cell. 2022;34(7):2466–74.
48. Niu Z, Li G, Hu H, Lv J, Zheng Q, et al. A gene that underwent adaptive evolution, LAC2 (LACCASE), in *Populus euphratica* improves drought tolerance by improving water transport capacity. Hortic Res. 2021;8(1):88.
49. Hu R, Qi G, Kong Y, Kong D, Gao Q, et al. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biol. 2010;10:145.
50. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005;39:121–52.
51. Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol. 2009;60:433–53.
52. Sun Q, Jiang S, Zhang T, Xu H, Fang H, Zhang J, et al. Apple NAC transcription factor MdNAC52 regulates biosynthesis of anthocyanin and Proanthocyanidin through MdMYB9 and MdMYB11. Plant Sci. 2019;289:110286.
53. Singh S, Koyama H, Bhati KK, Alok A. The biotechnological importance of the plant-specific NAC transcription factor family in crop improvement. J Plant Res. 2021;134(3):475–95.
54. Xu N, Meng L, Song L, Li X, Du S, Hu F, et al. Identification and characterization of secondary wall-associated NAC genes and their involvement in hormonal responses in tobacco (*Nicotiana tabacum*). Front Plant Sc. 2021;12:712254.
55. Wang HL, Zhang Y, Wang T, Yang Q, Yang Y, Li Z, et al. An alternative splicing variant of PtRD26 delays leaf senescence by regulating multiple NAC transcription factors in *Populus*. Plant Cell. 2021;33(5):1594–614.
56. Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Seki M, et al. Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. Plant Cell. 2004;16(9):2481–98.
57. Yamaguchi-Shinozaki K, Shinozaki K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci. 2005;10(2):88–94.
58. Yang YT, Yu YL, Yang GD, Zhang JD, Zheng CC. Tissue-specific expression of the PNZIP promoter is mediated by combinatorial interaction of different cis-elements and a novel transcriptional factor. Nucleic Acids Res. 2009;37(8):2630–44.
59. He H, Li Q, Fang L, Yang W, Xu F, et al. Comprehensive analysis of NAC transcription factors in *Scutellaria baicalensis* and their response to exogenous ABA and GA(3). Int J Biol Macromol. 2023;244:125290.
60. Liu X, Zong X, Wu X, Liu H, Han J, Yao Z, et al. Ectopic expression of NAC transcription factor HaNAC3 from *Haloxylon ammodendron* increased abiotic stress resistance in tobacco. Planta. 2022;256(6):105.
61. Shang X, Yu Y, Zhu L, Liu H, Chai Q, Guo W. A cotton NAC transcription factor GhirNAC2 plays positive roles in drought tolerance via regulating ABA biosynthesis. Plant Sci. 2020;296:110498.
62. Feng X, Meng Q, Zeng J, Yu Q, Xu D, et al. Genome-wide identification of sucrose non-fermenting-1-related protein kinase genes in maize and their responses to abiotic stresses. Front Plant Sci. 2022;13:1087839.

63. Yang F, Mitra P, Zhang L, Zhao D, Tao J, et al. Engineering secondary cell wall deposition in plants. Plant Biotechnol J. 2013;11(3):325–35.
64. Yu S, Wu M, Wang X, Li M, Gao X, et al. Common bean (*Phaseolus vulgaris* L.) NAC transcriptional factor PvNAC52 enhances Transgenic *Arabidopsis* resistance to salt, alkali, osmotic, and ABA stress by upregulating stress-responsive genes. Int J Mol Sci. 2024;25(11):5818.
65. Mei C, Yang J, Mei Q, Jia D, Yan P, Feng B, et al. *MdNAC104* positively regulates Apple cold tolerance via CBF-dependent and CBF-independent pathways. Plant Biotechnol J. 2023;21(10):2057–73.
66. Dalman K, Wind JJ, Nemesio-Gorriz M, Hammerbacher A, Lunden K, Ezcurra I, et al. Overexpression of PaNAC03, a stress induced NAC gene family transcription factor in Norway Spruce leads to reduced flavonol biosynthesis and aberrant embryo development. BMC Plant Biol. 2017;17(1):6.
67. Mahmood K, Xu Z, El-Kereamy A, Casaretto JA, Rothstein SJ, et al. The *Arabidopsis* transcription factor ANAC032 represses anthocyanin biosynthesis in response to high sucrose and oxidative and abiotic stresses. Front Plant Sci. 2016;7:1548.
68. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
69. Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther. 2012;22(4):271–4.
70. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12(2):87–98.
71. Udvardi MK, Czechowski T, Scheible WR. Eleven golden rules of quantitative RT-PCR. Plant Cell. 2008;20(7):1736–7.
72. Nygaard V, Rodland EA, Hovig E. Methods that remove batch effects while retaining group differences May lead to exaggerated confidence in downstream analyses. Biostatistics. 2016;17(1):29–39.
73. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11(10):733–9.
74. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
75. Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, et al. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun. 2004;313(4):856–62.
76. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3(7):RESEARCH34.
77. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010;11:94.
78. Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, et al. *Arabidopsis* NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. Plant Cell Physiol. 2009;50(12):2210–22.
79. Zhang S, Chen Y, Zhao L, Li C, Yu J, Li T, et al. A novel NAC transcription factor, MdNAC42, regulates anthocyanin accumulation in red-fleshed Apple by interacting with MdMYB10. Tree Physiol. 2020;40(3):413–23.
80. Gong L, Zhang H, Liu X, Gan X, Nie F, et al. Ectopic expression of HaNAC1, an ATAF transcription factor from *Haloxylon ammodendron*, improves growth and drought tolerance in Transgenic *Arabidopsis*. Plant Physiol Biochem. 2020;151:535–44.
81. Liu F, Ma D, Yu J, Meng R, Wang Z, et al. Overexpression of an ART1-interacting gene OsNAC016 improves al tolerance in rice. Int J Mol Sci. 2023;24(23):17036.
82. Qu W, Huang W, Chen C, Chen J, Zhao L, et al. AdNAC20 regulates lignin and coumarin biosynthesis in the roots of *Angelica Dahurica* Var. Formosana. Int J Mol Sci. 2024;25(14):7998.
83. Wang T, Ma X, Chen Y, Wang C, Xia Z, et al. *SlNAC3* suppresses cold tolerance in tomatoes by enhancing ethylene biosynthesis. Plant Cell Environ. 2024;47(8):3132–46.
84. Huang Y, Liang Z, Lu J, Zhang M, Cao X, et al. The transcription factor EjNAC5 regulates Loquat fruit chilling lignification. J Exp Bot. 2024;75(20):6625–43.
85. Huang C, Toyokura K, Murakami EI, Ishiwata A, Kurotani KI, et al. *Nicotiana benthamiana* VASCULAR-RELATED NAC-DOMAIN7-2 (NbVND7-2) has a role in xylem formation during interfamily grafting. J Exp Bot. 2025;23:eraf074.

Publisher's note