

Exploring the role and expression pattern of WRKY transcription factor in the growth and development of *Bletilla striata* based on transcriptome

Qingqing Li

Zunyi Medical University

Ceyin Huang

Zunyi Medical University

Chaobo Liu

Zunyi Medical University

Xueyan Jia

Zunyi Medical University

Weie Wen

Zunyi Medical University

Lin Li

Zunyi Medical University

Yihuai He

the Affiliated Hospital of Zunyi Medical University

Delin Xu (✉ xudelin2000@163.com)

Zunyi Medical University

Research Article

Keywords: *Bletilla striata*, bioinformatics, functional exploring, SSR marker mining, WRKY gene family.

Posted Date: July 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1837954/v1>

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

[Read Full License](#)

Abstract

To explore the genetic mechanism of regulating growth and chemical biosynthesis is vital important for further studies and applications in medicinal herbs. WRKY type transcription factors (TFs) play crucial roles in the growth and development of plants. However, a comprehensive analysis of the *WRKY* family members in a valuable Chinese herbal orchid, *Bletilla striata*, or in other orchids, is limited. In this study, by applied bioinformatic programs, we screened 29 *WRKY* genes from former formed database and then identified their functions based on the sequences properties and explore their expression mode, and also identified some SSR molecular markers. Analysis showed that *BsWRKY* genes may work together to regulate plant growth and development. A single *BsWRKY* transcription factor may be involved in the regulation of several seemingly different processes in plants. *BsWRKY* genes play an important regulatory role in the synthesis of Coelonin, Hydroxybenzyl Alcohol, Dactylorhin A and Militarine. Taken together, these results provided a theoretical basis for further studies on the gene functions and regulatory mechanisms of what in *B. striata*.

Introduction

Transcription factors (TFs) are proteins that can bind to the specific sequence of gene, making gene express at a specific intensity at a specific period and tissue. WRKY transcription factors are important regulatory factors for plant growth and development in higher plants. They are involved in biological stress responses and abiotic stress responses such as drought, salt injury, pests and diseases, nutrient deficiency, high temperature and cold injury. For example, *WRKY* regulates fruit growth and development in jujube (Chen et al., 2019). *WRKY* combines with *NPR1* to regulate gene expression and then activate plant defense response (Wang et al., 2006). *WRKY13*, *WRKY45-2* and *WRKY42* formed regulatory cascades that were involved in rice blast resistance (Cheng et al., 2015). Xu used gene microarray technology to find that *SbWRKY14*, *SbWRKY32* and *SbWRKY39* were all involved in drought stress response of sorghum (Xu and Liu, 2021). *CIWRKY47* of lemon and *CsWRKY47* of sweet orange can be induced by high salinity, drought and low temperature stress (Shen et al., 2021). The *ZmWRKY102* in maize can improve the drought resistance of plants (Li, 2015). The *WRKY* of *Cinnamomum kanehirae* were involved in response to drought and low temperature stress (Zhao et al., 2020). *WRKY* gene family members were also involved in plant nutrient stress and plant hormone signal transduction (Jing et al., 2021; Wani et al., 2021).

The conserved WRKYGQK domain, which consists of about 60 amino acids in the DNA binding domain, was the most conserved structural feature of the *WRKY* gene family. The downstream of this family had a cis-acting element W-box (T) (T) TGAC (C/T) structure, which regulated the transcription of target genes and responds to adversity stress by combining with specific W-box (Wang, 2020). The C-terminal was an absolutely conserved domain composed of WRKYGQK, while the N-terminal contained a conserved domain composed of zinc finger protein. It was classified three types according to the number and structure of domains. The group Ⅱ typically contained two WRKY domains including a C₂H₂ (CX₄-5CX₂₂-23HXH) motif, while group Ⅰ and group Ⅲ were characterized by a single WRKY domain. Group Ⅲ also

contained a C₂H₂ zinc-finger motif which can be further divided into five subgroups (Ⅰ-a, Ⅰ-b, Ⅰ-c, Ⅰ-d and Ⅰ-e) based on the phylogeny of the WRKY domains, whereas group Ⅰ contained a zinc-finger motif ending with C₂HC (CX7CX23HCX), which was only found in higher plants (Chen et al., 2020).

Bletilla striata (Thunb.) Reich, f. is a perennial herb of the orchid family, which contains chemicals of phenolic acids, dihydrophenanthrene, bibenzyl and other components with medications of anti-tumor, hemostasis, antibacterial, anti-inflammatory, promoting wound healing and plasma substitutes (Tang et al., 2014). It is also used in food, cosmetics and other industries (Zhang et al., 2015; Zhou et al., 2020). Abundant secondary metabolites are the material basis of the pharmacological action of the *B. striata*. Therefore, from the perspective of gene family, it is of great significance to analyze the synthesis pathway of secondary metabolites of *B. striata* and explore functional genes, which helps to understand the regulatory mechanism of effective components synthesis, and finally enhance the producing of secondary metabolites by plants. WRKY TFs have been studied in many plants such as peach (Gu et al., 2017), apple (Gu et al., 2015), melon (Chen et al., 2021), chickpea (Waqas et al., 2019), *Hordeum vulgare* (Jiang et al., 2021), *Prunella vulgaris* (Zhu et al., 2020), etc., but little is known about *WRKY* gene in *B. striata*. Based on transcriptome data of *B. striata*, screening and identification *WRKY* genes, and then analyze their genetic information, conservative domain, evolutionary relationships and functions, expression pattern, etc. In addition, SSR molecular markers were mined to classify the function of *WRKY* genes in *B. striata*, so as to provide a reference for further exploration of its function in the regulation of secondary metabolite synthesis.

1. Materials And Methods

1.1 Materials

The *B. striata* capsules were collected from the *B. striata* Germplasm Garden of Zunyi Medical University, Xinpu District, Zunyi City of Guizhou Province, China (27°42'N, 107°01'E), and the seeds were induced for suspension culture for a total of 45 days (Pan et al., 2020). The samples were randomly sampled every 3 days since from the callus was induced (3 replicates were taken at each time point), a total of 10 time points, and the total RNA of each sample was extracted by liquid nitrogen grind. The RNA of each sample was mixed as one with equal amount to perform the subsequent transcriptome sequencing by using Iso-seq of PacBio platform (Li et al., 2020). The *WRKY* genes sequence of *B. striata* were screened out from the sequencing results for the following analysis.

1.2 RNA-Seq sequencing

The RNA samples were entrusted to Novogene Co., Ltd., for library construction and subsequent sequencing. First, after the total RNA samples passed the quality test in this experiment, the mRNA was enriched using the NEBNext R Poly (A) mRNA Magnetic Isolation Module, and a common transcriptome library was constructed using the NEBNext R mRNA Library Prep Master Mix Set for Illumina, the insert size was 250 bp. Subsequently, PacBio P6 Binding Kit and PacBio C4 Sequencing Kit library sequencing

reagents were used to load samples on the machine using magbeads, and data was collected for 1 × 240 minutes. After the qualified RNA (or polyA + RNA) is detected, the SMARTer PCR cDNA Synthesis Kit is used to synthesize cDNA strand. And the resulting data were de novo assembled by using Trinity software (trinityrnaseq_r2013-02-25 version) (Haas et al., 2013) to finally obtain the transcriptome data. The online software NCBI blast were employed to do the annotation. After sequencing on the PacBio Sequel sequencing platform and correcting the transcripts with the second-generation data, a total of 1015753 single molecule high-quality sequencing reads (polymer reads) were obtained. By filtering the original offline data, removing the connectors and the original offline data with a length of less than 50bp, 54.02 GB subreads were obtained. Finally, quality filtering was carried out on the Illumina platform, 589522461 clean reads were obtained, and finally 100276 unigenes were obtained. The name of the repository and accession number can be found in NCBI PRJNA451494. These polished sequences were upload to NCBI with the TSA submission accession number of SUB7210106.

1.3 *WRKY* gene family identification

The sequences annotated as *WRKY* genes were screened to identify the conserved domains. All obtained protein sequences were examined for the presence of *WRKY* (PF03106) domains by using the Hidden Markov Model of Pfam (<https://pfam.xfam.org/>) (Finn et al. 2008), SMART (http://smart.embl-heidelberg.de/smart/set_mode.cgi?GENOMIC=1) (Schultz et al., 1998) and InterPro (<http://www.ebi.ac.uk/interpro/>) (Blum et al., 2021) tools. After eliminating incomplete sequences, candidate sequences were obtained for the following analysis.

1.4 Physical and chemical properties exploring

The ORF Finder of NCBI (<https://www.ncbi.nlm.nih.gov/orffinder/>) was used to find out the open reading frame (ORF) of all *BsWRKY* candidate sequences. Online server ExPASy (<https://web.expasy.org/protparam/>) (Wilkins et al., 1999.) was used to characterize the physical and chemical properties of *WRKY* transcription factors, such as the protein molecule weight, amino acid size, isoelectric point, GRVAY index and so on. Protein secondary structure was predicted using the online server SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) (Geourjon et al., 1995).

1.5 Analysis of subcellular localization and conserved domains

The conserved domains of the detected *WRKY* members were analyzed by using the CD-search function of NCBI (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>). A conservative motif analysis was conducted through the MEME (<https://meme-suite.org/meme/>) (Timothy et al., 2015), in which the number of recognized motifs was set to 10. The WoLFPSORT (<https://wolfpsort.hgc.jp/>) (Paul et al., 2007) was used to predict the subcellular localization of the *WRKY* proteins.

1.6 Signal peptide, transmembrane structure prediction and promoter analysis

Signal P 5.0 Server (<https://services.healthtech.dtu.dk/service.php?SignalP-5.0>) (Almagro et al., 2019.) was used to analyze the signal peptide and TMHMM-2.0 (<https://services.healthtech.dtu.dk/service.php?TMHMM-2.0>) (Krogh et al., 2001) was used to predict the transmembrane domain. The cis-acting elements of the upstream sequence (2000 bp) of the *WRKY* genes' promoter were analyzed by using the online Plant CARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Rombauts et al., 1999), and the results were visualized by using TBTools (Chen et al., 2018).

1.7 Phylogenetic tree construction

The *Arabidopsis thaliana* *WRKY* sequences were obtained from the transcription factor database of Plant TFDB (<http://planttfdb.gao-lab.org/index.php?sp=Ath>) (Jin et al. 2014), *Dendrobium catenatum* *WRKY* sequences were obtained from NCBI (<https://www.ncbi.nlm.nih.gov/>) and the MEGA-X (Kumar et al., 2018) software was applied to construct a phylogenetic tree with the *WRKY* gene family members via the Neigh-Joining method (NJ), and the Bootstrap value was set to 1000. Then use iTOL V5 (<https://itol.embl.de/>) (Letunic et al., 2021) for beautification.

1.8 Functional analysis of the *BsWRKYs* genes

The Gene Ontology (GO) was performed to do GO functional classification. The biological pathways of the *WRKY* of *B. striata* were mapped to the reference pathways in KEGG. Based on the molecular functions and biological pathways of diseases in the KEGG database, the analysis results were used to mine biologically significant information. The differentially expressed cytoplasmic and nuclear proteins were matched with the KEGG pathway database to generate the predicted pathway (Ramdas et al., 2019).

1.9 SSR detection and verification

The MISA (<https://webblast.ipk-gatersleben.de/misa/>) (Beier et al., 2017) was used to detect the candidate SSR sites of *BsWRKYs*. Then, primers of each site were designed by using DNAMAN 6.0 program. The SSR candidates were verified on four landraces, which were respectively collected from Zheng'an County of Guizhou Province (28°55'N, 107°43'E), Xiuwen County of Guizhou Province (26°83'N, 106°58'E), Nanchuan District of Chongqing City (29°15'N, 107°07'E) and Chizhou County of Anhui Province (30°67'N, 117°49'E), by coupling conducts of PCR and PAGE. PCR reaction system was volume 10 µL in total, containing 1.5 µL DNA template with concentration of 50 ng/µL, 6 µL 2×PCR MIX, 0.75 µL primer each with concentration of 10 µmol/L, and 1 µL ddH₂O. Amplification conditions were set as: pre-denaturation at 95 °C for 5 min; denaturation at 95 °C for 30 s, annealing at 52 °C for 30 s, extension at 72 °C for 60 s, 30 cycles; extension at 72 °C for 5 min. A 10% polyacrylamide gel was used to separate the amplified products. The electrophoresis instrument was a PowerPac stabilized steady flow electrophoresis instrument. The electrophoresis buffer was 1×TBE, the constant voltage was 150 V, and the time was 90 minutes. After silver nitrate staining, the bands were observed and photographed.

1.10 qPCR Validation

In order to verify the accuracy of transcriptome data, differential expressed genes (DEGs) were selected for qPCR verification. 1 µL of the qualified 3 Dai, 9 Dai, 18 Dai, 27 Dai, 30 Dai, 33 Dai, 36 Dai, 39 Dai and 42 Dai total RNA was used as template, and the first strand of cDNA was synthesized by reverse transcription with TIANGEN kit. qPCR reaction was carried out on AGS 4800 real-time quantitative PCR instrument. qPCR reaction system was volume 10 µL in total, containing 1 µL cDNA template with concentration of 1-10 ng/µL, 5 µL SYBR Green qPCR Master Mix (Universal), 0.2 µL primer each with concentration of 10 µmol/L, and 3.6 µL ddH₂O. The reaction was performed using the following conditions: denaturation at 95 °C for 30 s, followed by 40 cycles of amplification (95 °C for 10 s, 60 °C for 30 s and 72 °C for 30 s), extension at 95 °C for 15 s. For each sample, three technical replicates of the qPCR assay were used with three biological replicates. Gene expression was evaluated using the $2^{-\Delta\Delta C_t}$ method (Schmittgen et al., 2008).

1.11 Expression pattern of *BsWRKY*

In order to explore the function of *BsWRKY*, on the basis of previous studies (Li et al., 2020), the expression of *BsWRKY* of *B. striata* suspension cultured cells at 10 time points (3 Dai, 18 Dai, 27 Dai, 30 Dai, 33 Dai, 36 Dai, 39 Dai and 42 Dai) were analyzed, and the Heatmap heat map was made by TBtools. The correlation between the expression of *BsWRKY* gene family members and the contents of Hydroxybenzyl Alcohol, Dactylorhin A, Militarine and Coelonin in *B. striata* were analyzed to explore the expression mode of gene. Correlation analysis was carried out in R language (Pearson correlation coefficient). The screening condition was $q < 0.05$, $|\text{Rho}| > 0.5$. Using R version 3.6.3 to draw heat map.

2. Results

2.1 Physical and chemical properties of WRKY protein

A total of 135 sequences annotated as *WRKY* gene were preliminarily obtained from the transcriptome data of *B. striata*. After deletions of the sequences without typical WRKY domain and incomplete sequences, 29 *WRKY* sequences were finally reserved and renamed as *BsWRKY* 1-29. Physical and chemical properties analysis showed that the protein size of the *BsWRKY* members was between 159-703 aa, and the molecular weight was between 17546.1-76820.1 Da. The theoretical isoelectric point of proteins ranged from 4.48 to 9.94, 11 of them were basic proteins with isoelectric point greater than 7.5, 13 of them were acidic proteins with isoelectric point less than 6.5, and 5 of them were neutral between 6.5 and 7.5. These results indicated that most proteins of *BsWRKYs* were acidic. The instability coefficients of the 29 WRKY proteins were all greater than 40, while fat index were less than 100, and GRVAY values were negative, indicated that the *WRKY* transcription factor family of *B. striata* was an unstable hydrophilic protein. The predicted secondary structure of the protein showed that α -helix, β -folding and elongation accounted for 21.07%, 4.07% and 11.84%, respectively, and random coil accounted for 63.02% (Table 1). Among the 29 proteins, the β -turn was normally more than the α -helix except in *BsWRKY20*.

2.2 Subcellular localization, signal peptide and transmembrane structure

The subcellular localization prediction showed that 24 WRKY proteins were all located in nucleus, except that BsWRKY16 was located in chloroplast, BsWRKY14 and BsWRKY29 were located in mitochondria, BsWRKY17 was located in vacuole, and BsWRKY28 was located in endoplasmic reticulum. Subcellular location determines its specific biological effects.

2.3 Conservative motif of WRKY protein

A total of 10 conserved motifs were obtained by using online MEME for motif analysis of *WRKY* transcription factors in *B. striata* (Fig. 2A, B, C). Among them, motif3 were contained in 28 members except BsWRKY7. Motif1 and motif2 existed in 26 members, motif8 was found in 6 members, motif4 was detected from 4 members. The CD-search analysis found that motif1, motif2, and motif4 belong to WRKY domains, motif8 was a zinc finger domain relates to WRKY. Interestingly, the motif3 had no function record in the database currently and no relevant references, which needs to be further studied.

2.4 Promoter cis-regulatory elements of *BsWRKY* genes

The upstream of these *BsWRKY* genes were detected for finding cis-regulatory elements, like promoter and other cis-acting elements related to hormone regulation and stress-response (Fig. 1). The result showed that the cis-regulatory elements of the promoters of *BsWRKY* genes were related to growth and development (meristem expression, specific to the endosperm, seed-specific regulation and regulates circadian rhythm), plant hormones (auxin, abscisic acid, methyl jasmonate (MeJA), gibberellin, and salicylic acid), and stress (drought, low temperature, oxygen specificity induced response element and anaerobic induced indispensable cis function adjustment). It also showed that all the 29 *BsWRKY* genes had light response elements (LRE), and 14 of them had the drought-inducibility response elements. However, the elements of cell cycle regulatory elements, elements involved in defense and stress response, flavonoid synthesis and seed germination were only existed in *BsWRKY28*, *BsWRKY7*, *BsWRKY16* and *BsWRKY24*, respectively. This not only indicated that *BsWRKY* genes are associated with plant growth, but also playing a vital role in drought stress regulatory networks. Collectively, these results indicated that *WRKY* family members participate in embryonic development, meristem growth and environmental stress regulation during the growth and development of *B. striata*.

2.5 Conserved domains identification and phylogenetic tree construction

Through the online server CD-search, the structure domains of the *WRKY* gene family members were analyzed for comparison. The results showed that 24 members of the 29 *BsWRKY* had typical WRKYGQK heptapeptide domain and W-box, but there was different degree of variation which mainly occurred in the N-terminal. Five transcription factors, i.e. *BsWRKY 7, 16, 17, 22* and *27*, had incomplete domains, like missing N or C terminus. It was speculated that the deletion may occur in the evolutionary process.

To place the evolution role and further identify the functions of *BsWRKY* genes, these 29 *WRKY* sequences from *B. striata*, 22 *WRKY* from *A. thaliana* and 20 *WRKY* from *D. catenatum* were used to

perform a phylogenetic analysis (Fig.3). The 29 *BsWRKY* transcription factors were grouped into three major groups. Group I included 6 members as *BsWRKY 4, 17, 19, 10, 25, 26 and 28*, which contains two WRKY conserved domains. Group III had members of *BsWRKY1, 15 and 22*, which containing only one WRKY domain with the zinc finger structure of type C₂HC. In addition, Group II had only one domain with the zinc finger structure of C₂H₂, which could be further divided into 5 subgroups. In which, subgroup II-a had only member of *BsWRKY12*, subgroup II-b had four as *BsWRKY 18, 23, 24 and 27*, subgroup II-c had *BsWRKY2, 5, 6, 7, 10 and 29*, subgroup II-d was consisted by six members i.e. *BsWRKY8, 9, 11, 13, 14 and 16*, subgroup II-e hold three members as *BsWRKY3, 20 and 21*.

According to the evolutionary tree, event of domain's gain or loss might occur in the process of evolution, and group I was closer the real ancestor. The evolutionary relationship of *WRKY* genes among *B. striata*, *A. thaliana* and *D. catenatum* indicated that the three may have similar roles in certain biological functions. The results showed that *BsWRKY5*, *BsWRKY12* and *BsWRKY3* had the highest similarity with *AtWRKY3*, *AtWRKY40* and *AtWRKY14*, respectively.

2.6 GO enrichment and KEGG functional cluster of *WRKY* genes

The GO and KEGG analysis results (Table S1) showed that most of *BsWRKY* genes were divided into three categories with functions of cellular components, biological processes and molecular functions, except *BsWRKY8, 14, 16, 20 and 22*. Among the biological processes, 19 genes were annotated for transcriptional regulation, *BsWRKY18* and *BsWRKY27* for cell cycle, *BsWRKY25* for DNA repair and *BsWRKY17* for meiosis prophase. In the molecular function category, 16 proteins were annotated for sequence specific DNA binding and 20 proteins for sequence specific DNA binding transcription factor activity. Among the large group of cellular components, only *BsWRKY12* was annotated into the troponin complex, and *BsWRKY25* had DNA ligase (NAD⁺) activity. KEGG analysis showed that *BsWRKY19, 20, 26 and BsWRKY28* were involved in environmental information processing, signal transduction, MAPK signal transduction pathway, environmental adaptation, plant pathogen interaction and biological system.

2.7 EST-SSR polymorphism of *WRKY* genes in different germplasms of *B. striata*

The EST-SSR markers have the advantages of high polymorphism and variability, high reproducibility, accurate and rapid detection (Li et al., 2015). A total of 10 among the 29 sequences were detected with SSR sites by NWISRL, of which 3 sequences were dinucleotide repeats and 7 sequences were trinucleotide repeats. The lowest number of replicates was 5, and the highest number was 18. The primer pairs of the 10 SSR sites (Table S2) were designed by DNAMAN software which could be amplified stably in all the four landraces, and the length of the amplified products ranged from 100 to 200 bp (Fig. 4). These results indicated that *WRKY* gene family was probably high-conserved in different *B. striata* germplasms. These newly found SSR primers could be used as molecular markers to identify the members of *BsWRKY* gene families in different germplasm (Zhong et al., 2021).

2.8 Expression analysis of *BsWRKY* gene family

In order to explore the function of *BsWRKY*, FPKM value were used to evaluate the expression pattern of 29 *BsWRKY* genes at 10 time points, and TBtools was used to make Heatmap heat map (Fig. 5). Except that *BsWRKY3* was not expressed, the other 28 *BsWRKY* genes were expressed during the growth period of suspension culture, and different genes not only had unique expression patterns, but also had expression commonalities. Among them, the expression of 10 *BsWRKY* genes were relatively stable within 42 days of suspension culture, including *BsWRKY4*, *BsWRKY5*, *BsWRKY6*, *BsWRKY7*, *BsWRKY8*, *BsWRKY10*, *BsWRKY22*, *BsWRKY24*, *BsWRKY25* and *BsWRKY26*. With the increase of culture time, most genes showed up-regulated expression trend. The *BsWRKY11*, *BsWRKY12* and *BsWRKY13* genes showed fluctuating changes at 10 time points. The expression trend of *BsWRKY19* gene was up-regulated first and then down regulated, the expression trend of *BsWRKY9*, *BsWRKY14* and *BsWRKY21* genes was up-regulated all the time, and the expression trend of *BsWRKY11* and *BsWRKY12* was down regulated. It was speculated that these genes may be directly or indirectly involved in the regulation of secondary metabolite synthesis. The expression patterns of the *BsWRKY* gene family members were also closely related to their functions, with *BsWRKY9* and *BsWRKY14* being higher at different developmental periods, presumably playing an important role in cell growth of *B. striata*.

2.9 qPCR verification

In order to verify the reliability of transcriptome data, randomly select 5 genes for validation in a sample of four periods, and the primer blast of NCBI website was used to design specific primers for gene qPCR reaction (Table S3). With β -*Actin* gene was used as reference gene. The results showed that the expression patterns of the five DEGs were consistent with the trend of transcriptome sequencing (Fig. 6), which confirmed the reliability of our transcriptome data.

2.10 Expression pattern of *BsWRKY* gene in secondary metabolites

The correlation analysis between the FPKM value of *BsWRKY* genes and the contents of Coelonin, Dactylorhin A, Hydroxybenzyl Alcohol and Militarine in *B. striata* with high contents were conducted, and the HeatMap heat map was plotted (Fig. 7). *BsWRKY16* and *BsWRKY21* were significantly positively correlated with the synthesis of Militarine ($r = 0.71$, $r = 0.84$), *BsWRKY21* was significantly positively correlated with the synthesis of Coelonin ($r = 0.70$), and *BsWRKY9*, *BsWRKY14*, *BsWRKY16*, *BsWRKY21*, *BsWRKY6* were significantly positively correlated with the synthesis of Hydroxybenzyl Alcohol ($r = 0.8$, $r = 0.81$, $r = 0.88$, $r = 0.74$, $r = 0.72$). *BsWRKY4* was significantly negatively correlated with the synthesis of Hydroxybenzyl Alcohol ($r = -0.86$), *BsWRKY4* was significantly positively correlated with the synthesis of Dactylorhin A ($r = 0.8$). *BsWRKY9*, *BsWRKY20*, *BsWRKY14*, *BsWRKY7* and *BsWRKY27* were negatively regulated the synthesis of Dactylorhin A ($r = -0.78$, $r = -0.79$, $r = 0.77$, $r = 0.86$, $r = 0.83$).

3. Discussion

WRKY transcription factors are important transcription factors involved in growth, development and response to stress in eukaryotes, and play an important role in plant survival. *WRKY* gene in longan (Xue and Jue, 2021), *Artemisia annua* (Li et al., 2021), rice (Fan et al., 2017), barley (Viana et al., 2021) and

other plants is involved in regulating the maturity and senescence of plants, promoting the accumulation of specific products, and responding to one or more abiotic stresses.

Based on bioinformatics analysis, the physicochemical properties, conserved motifs, cis elements, evolutionary relationships, functions and expression pattern of *WRKY* gene family members of *B. striata* were analyzed, and EST-SSR sites were detected and verified. Based on the transcriptome data of *B. striata*, a total of 29 members of *BsWRKY* gene family were screened. Among them, 24 members had typical WRKYGQK conserved structure and downstream W-box (T) (T) TGAC (C/T) homeostasis element. However, five members (*BsWRKY7*, 16, 17, 22 and 27) had different degrees of deletion at the C-terminal or N-terminal, which indicated that they might be self-regulation or cross regulation between *WRKY* genes. The loss and increase of domains had occurred in rice (Ross et al., 2007) and maize (Wei et al., 2012). Therefore, its function and binding specificity could be further investigated. *BsWRKY7*, *BsWRKY28* and *BsWRKY29* had multiple regulatory elements, which participate in important growth processes such as environmental adaptation, regulation of cell maturation and senescence, which indicating that they regulate the synthesis of secondary metabolites of flavonoids and terpenoids during the cell growth and development (Park et al., 2021). They participated in stress response, and belong to the category β -c subgroup in classification (Na et al., 2018). It was speculated that β -c subgroup played an important role in stress response, which was consistent with the research results in barley (Zheng et al., 2021).

BsWRKYs can form a net that contributes to various cytoplasmic and nuclear processes including signaling events from organelles or the cytoplasm to the nucleus (Bakshi and Oelmüller, 2014). Studies had shown that WRKY TFs on the ABAR-ABA complex in the downstream chloroplast envelope, regulates seed germination and other processes, and is one of the key nodes of abscisic acid signaling pathways (Rushton et al., 2012). These results indicated that *WRKY* genes might be involved in the regulation of plant growth and development and is an important node in metabolic regulation.

Former studies reported that *AtWRKY3* involved in biological stress response (Aboul-Maaty and Oraby, 2019), *AtWRKY40* participated in drought stress response in *A. thaliana* (Ju et al., 2019), *AtWRKY14* played a role in plant antiviral (Che et al., 2018). Therefore, we speculated that genes *BsWRKY5*, *BsWRKY12* and *BsWRKY3* might play a regulatory role in stress resistance in *B. striata*. EST-SSR polymorphism detection showed that the *WRKY* family in different *B. striata* varieties was conserved. This indicated that this gene family has genetic conservative. Therefore, the SSR sites can be used to identify the genetic diversity of germplasm resources, and provide some reference for the molecular-assisted breeding and genetic diversity analysis of plants, which is of great significance for the development and utilization of gene function and the evaluation of germplasm resources.

The expression pattern showed that *BsWRKY16*, *BsWRKY21* were associated with the synthesis of Militarine, *BsWRKY21* was associated with the synthesis of Coelonin, *BsWRKY9*, *BsWRKY14*, *BsWRKY16*, *BsWRKY21*, *BsWRKY6*, *BsWRKY4* was associated with the synthesis of Hydroxybenzyl Alcohol, *BsWRKY4*, *BsWRKY7*, *BsWRKY9*, *BsWRKY14*, *BsWRKY20* and *BsWRKY27* regulated the synthesis of Dactylorhin A. Among them, *BsWRKY9* and *BsWRKY14* showed a continuous high expression trend with

time and were important for the synthesis of Hydroxybenzyl Alcohol and Dactylorhin A. It was found that Militarine, Coelonin, Hydroxybenzyl Alcohol and Dactylorhin A were the components with high content in *B. striata*, which had obvious significant neuroprotective effects (Tian et al., 2021), anti-inflammatory and anti-pulmonary fibrosis (Jiang et al., 2019), inhibitory activity in experimental endometriosis models (Meresman et al., 2021) and treat gastrointestinal diseases (Liu et al., 2021). However, the synthesis pathway has not been resolved, and it is difficult to obtain and has few market applications. The study of expression patterns provides a basis for the analysis of the synthesis pathway of secondary metabolites, which is expected to help the application of *B. striata* resources.

Conclusion

In this study, we found that the *BsWRKY* gene family members were involved in many signal pathways and had crosstalk regulation, which affected the synthesis of secondary metabolites. The pathway of *WRKY* gene family members involved in the synthesis of secondary metabolites had also been verified in plants such as maize (Tang et al., 2021) and wheat (Ye et al., 2021). In conclusion, our results contribute to a comprehensive understanding of *BsWRKY* gene family and provide a theoretical basis for further study on the function and regulation mechanism of *BsWRKY* gene family. It is conducive to the synthesis of secondary metabolites and the study of related metabolic pathways in *B. striata*, and improve plant stress resistance and the content of secondary metabolites. In addition, we also identified some SSR molecular markers, which provided a reference for further research on the genetics and breeding of *B. striata*.

Declarations

Ethical Approval:

This declaration is “not applicable”.

Competing interests:

The authors declare that there is no conflict of interest that could be perceived as prejudicial to the impartiality of the reported research.

Authors' contributions:

DX and YH conceived, supervised and writing-reviewed the manuscript, QL, CH, CL and XJ originally wrote and writing-reviewed the draft, LL and WW cofounded and co-administrated the project. All authors read and approved the final version.

Funding:

This research was financially supported by the National Natural Science Foundation of China (31560079, 31960074), the Science and Technology Department Foundation of Guizhou Province of China (No.

[2017]5733-050, [2019]-027, [2019]5657), the Special Joint Bidding Project of Zunyi Sci & Tech Bureau and Zunyi Medical University (ZSKHHZ-2020-91), and the Postgraduate Research Fund Project of Zunyi Medical University (ZYK029).

Availability of data and materials:

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA451494.

References

1. Aboul-Maaty A F and Oraby A S. 2019. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bulletin of the National Research Centre*, 43(1): 2–10.
2. Almagro Armenteros, José Juan, Tsirigos K D, et al. 2019. Signal P 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*. 37(4):420–423.
3. Bakshi M and Oelmüller R. 2014. WRKY transcription factors: Jack of many trades in plants. *Plant Signal Behav.* 9(2), e27700.
4. Beier S, Thiel T, Münch T, et al., 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 33: 2583–2585.
5. Blum M, Chang H, Chuguransky S, et al., 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 49(D1): D344-D354.
6. Che Y M, Sun Y J, Song-Chong L U, et al. 2018. *AtWRKY40* functions in drought stress response in *Arabidopsis thaliana*. *Plant Physiology Journal*, 054(003), 456–464.
7. Chen X, Chen R, et al Y W, et al. 2019. Genome-Wide Identification of WRKY Transcription Factors in Chinese jujube (*Ziziphus jujuba* Mill.) and Their Involvement in Fruit Developing, Ripening, and Abiotic Stress. *Genes*, 10(5), 2–17.
8. Cheng H, Liu H, Deng Y, et al. 2015. The *WRKY45-2 WRKY13 WRKY42* transcriptional regulatory cascade is required for rice resistance to fungal pathogen. *Plant Physiology*, 167(3), 1087–99.
9. Chen C, Chen X, Han J, et al. 2020. Genome-wide analysis of the *WRKY* gene family in the cucumber genome and transcriptome-wide identification of WRKY transcription factors that respond to biotic and abiotic stresses. *BMC Plant Biology*. 20(1), 1–45.
10. Chen Y, Jing X, Wang S, et al. 2021. Genome-wide Analysis of WRKY Transcription Factor Family in Melon (*Cucumis Melo* L.) and Their Response to Powdery Mildew. *Plant Molecular Biology Reporter*, 1–14.
11. Chen C, Rui X, Hao C, et al. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. 2018.

12. Fan Zhongqi, Zhou Yuanfang, Fu Yali, et al. 2017. Molecular Characterization of a Transcriptional Factor, *MaWRKY31*, and Its Regulation on *MaACS1*, *MaACO1* and *MaSAG1*. Food Science, 38(06), 13 – 9.
13. Finn R D, Tate J, Mistry J, et al. 2008. The Pfam protein families database. Nucleic Acids Research, 32(1):D138.
14. Geourjon C, G Deléage. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci, 11(6):681–684.
15. Gu Yan-bing, Ji Zhi-rui, Chi Fu-mei, et al. 2015. Bioinformatics and Expression Analysis of the *WRKY* Gene Family in Apple. Scientia Agricultura Sinica, 48(16), 3221-38.
16. Gu YB, Ji ZR, Chi FM, et al. 2017. Genome-wide identification and expression analysis of the *WRKY* gene family in peach. Acta Agriculturae Boreali-Sinica, 38(3), 254–270.
17. Haas, B., Papanicolaou, A., Yassour, M. et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8, 1494–1512.
18. Jiang Liling, Chen Lihua, Nyima Tashi, et al. 2021. Bioinformatics Analyses of *WRKY* Family Members in *Hordeum vulgare* L. var. *nudum* Based on RNA-seq Data. Molecular Plant Breeding, 19(02), 370–379.
19. Jiang F, Li M, Wang H, et al. 2019. Coelonin, an Anti-Inflammation Active Component of *Bletilla striata* and Its Potential Mechanism. Int J Mol Sci., 20(18):4422.
20. Jin J, Zhang H, Kong L, et al. 2014. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Research, (D1):1182–1187.
21. Ju Ping, Li Li, Yu Lili, et al. 2019. Cloning and Expression Analysis of *CmWRKY4* Gene in Chrysanthemum. Molecular Plant Breeding, 17(01), 29–32.
22. Jing Yizhuo, Wei Yanqiu, Lu zhijing, et al. 2021. Research progress of *WRKY* transcription factor family in response to crop elemental stress. Plant Physiology, 57(2), 265–73.
23. Krogh A, Larsson B, Heijne G V, et al. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes - ScienceDirect. Journal of Molecular Biology, 305(3):567–580.
24. Kumar S, Stecher G, Li M, et al. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547–1549.
25. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49(W1):W293-W296.
26. Li C, Li D, Shao F, et al. 2015. Molecular cloning and expression analysis of *WRKY* transcription factor genes in *Salvia miltiorrhiza*. BMC Genomics, 16(1), 200.
27. Li L, Liu H, Wen W, et al. 2020. Full Transcriptome Analysis of Callus Suspension Culture System of *Bletilla striata*. Frontiers in Genetics, 11, 1–12.
28. Li Qi, Gao Xiao-yue, Zhang Lei, et al. 2021. Research on transcription factors related to artemisinin biosynthesis in *Artemisia annua*. Chinese Traditional and Herbal Drugs, 52(06), 1827–33.

29. Li Xiaoming. 2015. Cloning and analysis of Biological Functions of Transcription Factors *ZmWRKY102* from *Zea mays*. PhD thesis. Anhui Agricultural University.
30. Liu F, Ding H, Wang M, et al. 2021. A multi-evaluating strategy for Weikangling capsules: Chemical profiling, fingerprinting combined with quantitative analysis, quantity transfer, and dissolution curve. *J Pharm Biomed Anal.* 206:114347.
31. Meresman GF, Götte M, Laschke MW. 2021. Plants as source of new therapies for endometriosis: a review of preclinical and clinical studies. *Hum Reprod Update.*, 19;27(2):367–392.
32. Na Y, Li J, Han Z, et al. 2018. Research progress of jasmonate-responsive transcription factors in regulating plant secondary metabolism. *Zhongguo Zhong Yao Za Zhi*, 43(5), 897–903.
33. Park S H, Scheffler J A, Ray J D, et al. 2021. Identification of simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) that are associated with the nectariless trait of *Gossypium hirsutum* L. *Euphytica*, 217(4), 1–17.
34. Pan Y, Li L, Xiao S, et al. 2020. Callus growth kinetics and accumulation of secondary metabolites of *Bletilla striata* Rchb.f. using a callus suspension culture. *PLoS ONE*, 15(2), e0220084.
35. Paul H, Keun-Joon P, Takeshi O, et al. 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, 35:585–587.
36. Ramdas P, Radhakrishnan AK, Abdu Sani AA, et al. 2019. Advancing the Role of Gamma-Tocotrienol as Proteasomes Inhibitor: A Quantitative Proteomic Analysis of MDA-MB-231 Human Breast Cancer Cells. *Biomolecules*. 10(1), 19.
37. Rombauts, Dehais, Van, et al. 1999. PlantCARE, a plant cis-acting regulatory element database. *NUCL ACID RES*, 27(1):295–296.
38. Ross C.A., Liu Y., Shen Q.J. The *WRKY* gene family in rice (*Oryza sativa*). *J. Integr. Plant Biol.* 2007,49, 827–842.
39. Rushton DL, Tripathi P, Rabara RC, et al. 2012. WRKY transcription factors: key components in abscisic acid signalling. *Plant Biotechnol J.* 10(1), 2–11.
40. Shen Dan, Yang li, Hu wei, et al., 2021. Cloning and expression analysis of stress response gene *WRKY47* in citrus. *Jiangsu J. of Agr. Sci.*, 37(1), 129–138.
41. Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. *Nat. Methods* 3, 1101–1108.
42. Schultz J, Milpetz F, Bork P, et al. 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences*, 95(11):5857–5864.
43. Tang Yifei, Ruan chuanfen, Ying chen, et al., 2014. Research Progress on chemical constituents and pharmacological effects of *Bletilla striata*. *Chinese Traditional and Herbal Drugs*, 45(19), 2864–72.
44. Tang Y, Guo J, Zhang T, et al. 2021. Genome-Wide Analysis of *WRKY* Gene Family and the Dynamic Responses of Key *WRKY* Genes Involved in *Ostrinia furnacalis* Attack in *Zea mays*. *Int J Mol Sci.* 22(23):13045.

45. Tian SX, Cheng W, Lu JJ, et al. Role of Militarine in PM2.5-Induced BV-2 Cell Damage. *Neurochem Res.* 2021, 46(6):1423–1434.
46. Timothy L. Bailey, James Johnson, Charles E. Grant, et al., 2015. The MEME Suite. *Nucleic Acids Research*, 43(W1): W39-W49.
47. Viana VE, Carlos da Maia L, Busanello C, et al. 2021. When rice gets the chills: comparative transcriptome profiling at germination shows WRKY transcription factor responses. *Plant Biol (Stuttg)*. 1, 100–112.
48. Wani S H, Anand S, Singh B, et al. 2021. WRKY transcription factors and plant defense responses: latest discoveries and future prospects. *Plant Cell Reports*, 40(7):1071–1085.
49. Wang Congyu. 2020. Genome-wide analysis of the *WRKY* family response to salt stress and histidine kinase in *beta vulgaris*. PhD thesis. Harbin Institute of Technology.
50. Waqas M, Azhar MT, Rana IA, et al. 2019. Genome-wide identification and expression analyses of WRKY transcription factor family members from chickpea (*Cicer arietinum* L.) reveal their role in abiotic stress-responses. *Genes & Genomics*, 41(4):467–481.
51. Wang D, Amornsiripanitch N and Dong X. 2006. A genomic approach to identify regulatory nodes in the transcriptional network of systemic acquired resistance in plants. *PLoS Pathog*, 2(11), e123.
52. Wei K.F., Chen J., Chen Y.F., et al. 2012. Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Res*, 19, 153–164.
53. Wilkins MR, Gasteiger E, Bairoch A, et al.. 1999. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* 112:531–52.
54. Xu Lei and Liu yang. 2021. Expression Analysis of Group Ⅱ WRKY Transcription Factors in Sorghum Under Drought Stress. *Guangdong Agricultural Sciences*, 48(2), 11–6.
55. Xue Xin and Jue Dengwei. 2021. Cloning of Longan *DIWRKY25* Gene and Analysis of Its Function in Flowering Process. *Molecular Plant Breeding*, 1–13.
56. Ye H, Qiao L, Guo H, et al. 2021. Genome-Wide Identification of Wheat *WRKY* Gene Family Reveals That *TaWRKY75-A* Is Referred to Drought and Salt Resistances. *Front Plant Sci.* 12:663118.
57. Zhang Q, Wang N, Hu R, et al. 2015. Wet spinning of *Bletilla striata* polysaccharide/silk fibroin hybrid fibers. *Materials Letters*. 161(15), 576–579.
58. Zhu Yun-hao, Chen Zhi-heng, Dong Cheng-ming. 2020. Transcriptome-wide Identification and Expression Profiles of *WRKY* Transcription Factor in *Prunella vulgaris*. *Chinese Journal of Experimental Traditional Medical Formulae*, 26(20), 146–152.
59. Zhong H, Zhang F, Zhou X, et al. 2021. Genome-Wide Identification of Sequence Variations and SSR Marker Development in the Munake Grape Cultivar. *Frontiers in Ecology and Evolution*, 9, 1–11.
60. Zheng J, Zhang Z, Tong T, et al. 2021. Genome-Wide identification of *WRKY* gene family and expression analysis under abiotic stress in barley. *Agronomy*, 11(3), 521.
61. Zhou Jiaming, Ou Jingchun, Zhao Ai, et al. 2020. Research and development of *Panax notoginseng* whitening cosmetics. *Ginseng research*, 032(001), 40–42.

62. Zhao N, He M, Li L, et al. 2020. Identification and expression analysis of *WRKY* gene family under drought stress in peanut (*Arachis hypogaea* L.). PLoS ONE, 15(4), e0231396.

Tables

Table 1. Physical and chemical properties of WRKY protein

Gene name	AA	MW/Da	PI	Instability index	GRAVY	Alpha helix	Extended strand	Beta turn
<i>BsWRKY1</i>	307	34243	5.4	43.37	-0.651	24.76%	10.10%	3.58%
<i>BsWRKY2</i>	481	52166.7	7.01	46.65	-0.672	10.19%	14.35%	3.74%
<i>BsWRKY3</i>	406	44754.8	5.5	47.5	-0.695	25.62%	6.40%	4.43%
<i>BsWRKY4</i>	702	76105.4	5.67	48	-0.627	11.97%	11.11%	3.42%
<i>BsWRKY5</i>	500	54510.7	8.97	48.21	-0.752	12.00%	12.80%	3.60%
<i>BsWRKY6</i>	469	50789.1	6.98	50.9	-0.729	10.45%	14.07%	3.41%
<i>BsWRKY7</i>	185	21037.2	6.26	51.01	-0.97	24.86%	18.92%	10.27%
<i>BsWRKY8</i>	194	51538.4	9.59	52.71	-0.865	11.34%	13.92%	6.19%
<i>BsWRKY9</i>	335	35817.5	9.69	53.06	-0.484	21.49%	9.85%	6.87%
<i>BsWRKY10</i>	431	47054.3	6.57	54.84	-0.703	14.39%	12.30%	2.55%
<i>BsWRKY11</i>	341	36780.7	9.67	55.85	-0.484	18.18%	10.26%	6.45%
<i>BsWRKY12</i>	329	37045.9	8.57	56.08	-0.632	32.22%	10.33%	1.52%
<i>BsWRKY13</i>	347	38224.3	9.94	56.17	-0.688	25.07%	9.22%	5.48%
<i>BsWRKY14</i>	270	29189.3	5.64	56.66	-0.0743	17.04%	11.11%	3.70%
<i>BsWRKY15</i>	217	24804.6	5.91	57.12	-0.86	11.06%	15.21%	3.69%
<i>BsWRKY16</i>	233	25270	4.93	57.21	-0.642	27.90%	10.30%	3.43%
<i>BsWRKY17</i>	412	45995.3	8.62	57.49	-0.617	31.31%	8.98%	1.94%
<i>BsWRKY18</i>	595	64961.1	6.55	57.8	-0.65	28.57%	10.92%	1.51%
<i>BsWRKY19</i>	402	43491.8	8.58	59.9	-0.829	9.95%	11.44%	3.23%
<i>BsWRKY20</i>	159	17546.1	4.48	60.03	-0.82	40.25%	11.32%	8.81%
<i>BsWRKY21</i>	427	47131.5	5.25	60.59	-0.63	23.89%	6.56%	3.51%
<i>BsWRKY22</i>	174	19581.5	4.55	60.82	-0.506	21.84%	18.39%	5.17%
<i>BsWRKY23</i>	595	64961.1	6.55	62.18	-0.65	28.57%	10.92%	1.51%
<i>BsWRKY24</i>	335	35367.1	7.61	62.24	-0.354	25.37%	13.73%	3.88%
<i>BsWRKY25</i>	530	56563.4	6.36	62.43	-0.622	20.94%	10.94%	2.45%
<i>BsWRKY26</i>	703	76820.1	5.61	62.54	-0.648	13.23%	10.53%	3.27%
<i>BsWRKY27</i>	363	41768.4	8.52	63.93	-0.781	31.68%	13.50%	2.75%

<i>BsWRKY28</i>	526	57802.3	6.2	66.1	-0.657	13.50%	12.93%	3.42%
<i>BsWRKY29</i>	343	38673.7	9.76	66.63	-1.146	23.32%	12.83%	4.37%

Figures

Figure 1

Cis-responsive elements in the upstream region of the initiation codon of *BsWRKY* genes. Different colored boxes indicated different cis-responsive elements.

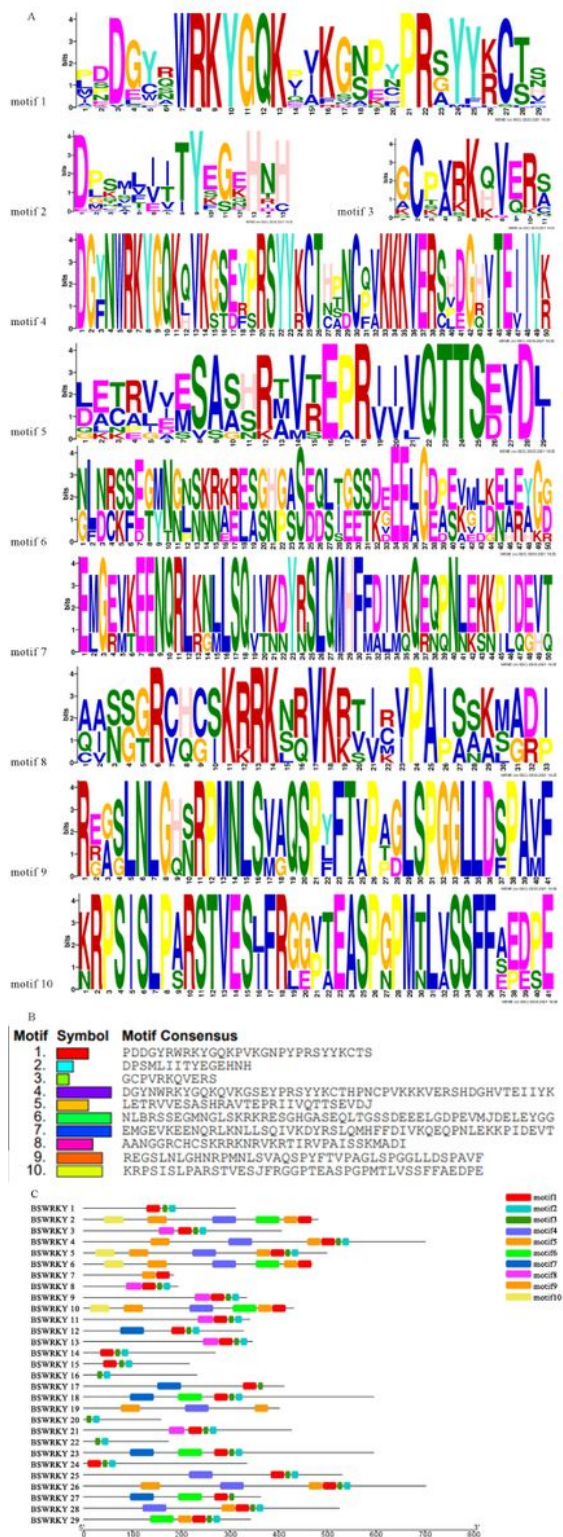


Figure 2

The size(A), sequence(B) and location(C) of the conserved motifs on WRKY protein family. The length and order of the boxes with different colors represent the actual size and location of each motif in protein sequence respectively.

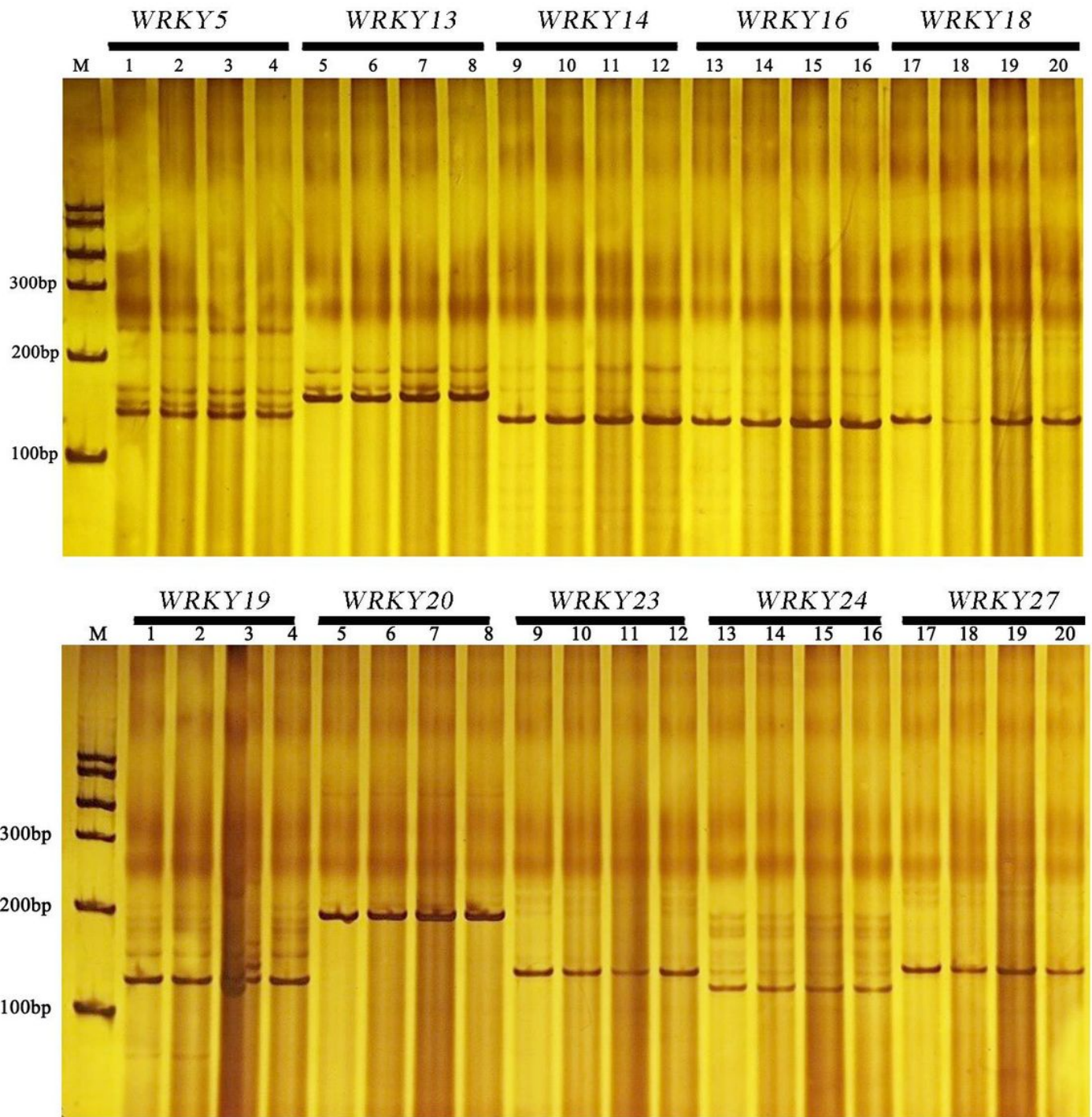


Figure 4

SSR profiles of *B. striata*. M is marker I (100-600 bp).

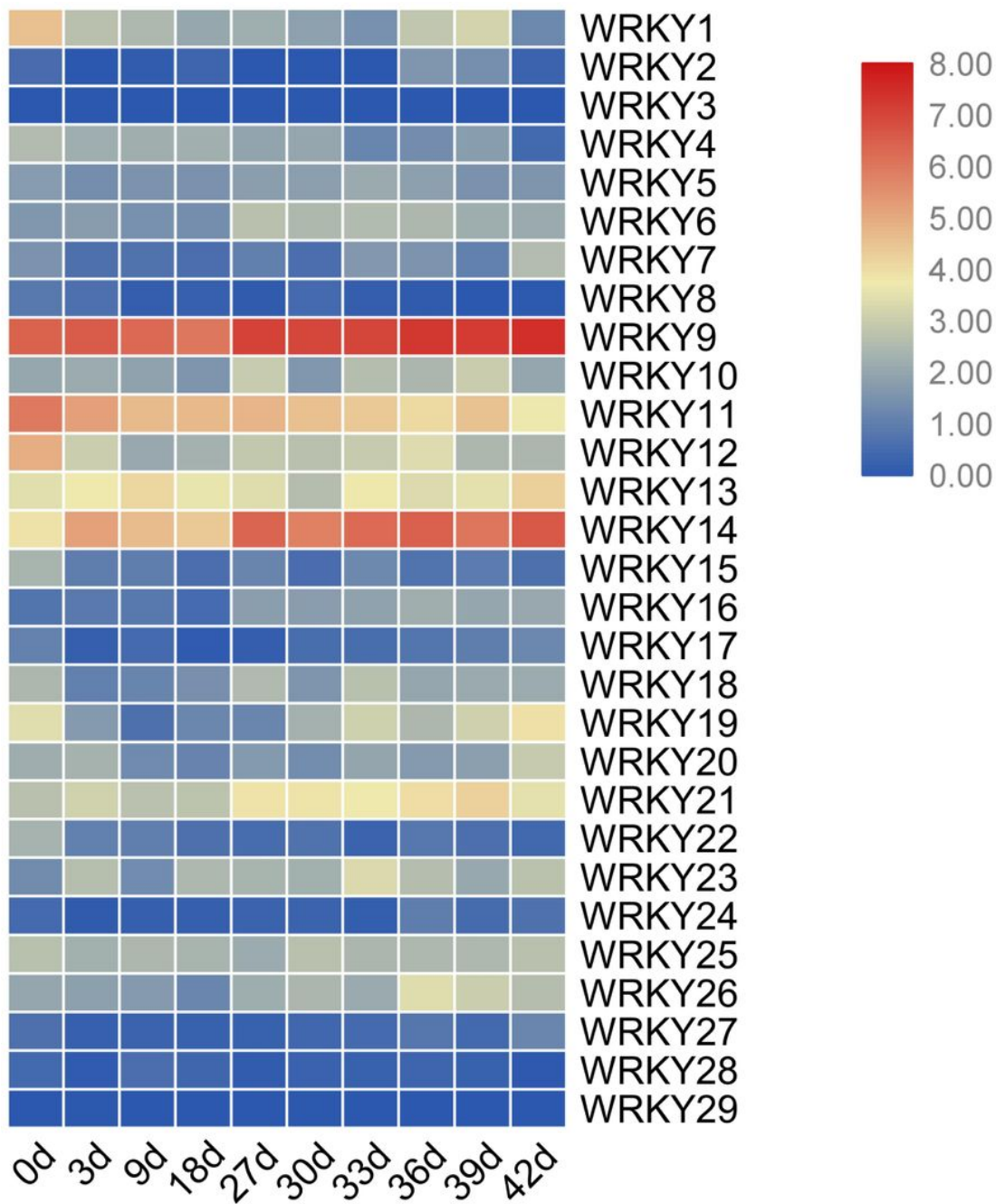


Figure 5

Expression patterns of *BsWRKY* gene family members in different stages of suspension culture. The abscissa represents the days of suspension culture, and the ordinate represents different *BsWRKY* genes.

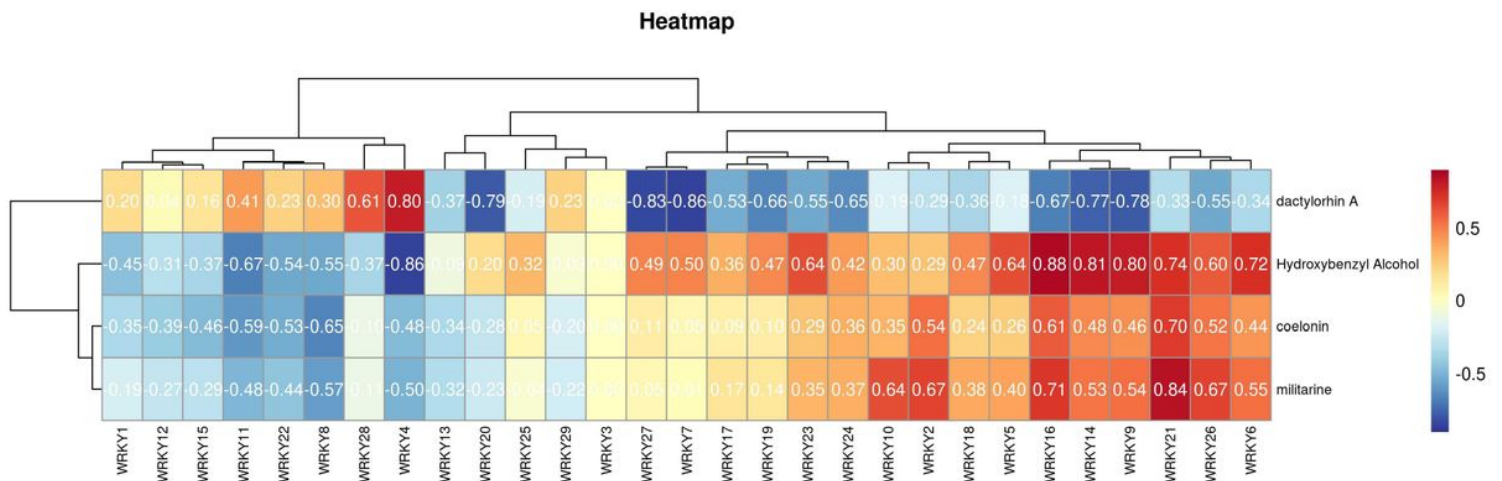


Figure 6

qPCR analysis of differential genes in *B. striata*. The ordinate corresponds to the average change of expression value, the abscissa represents different time, and the FPKM represents the value of transcriptome sequencing.

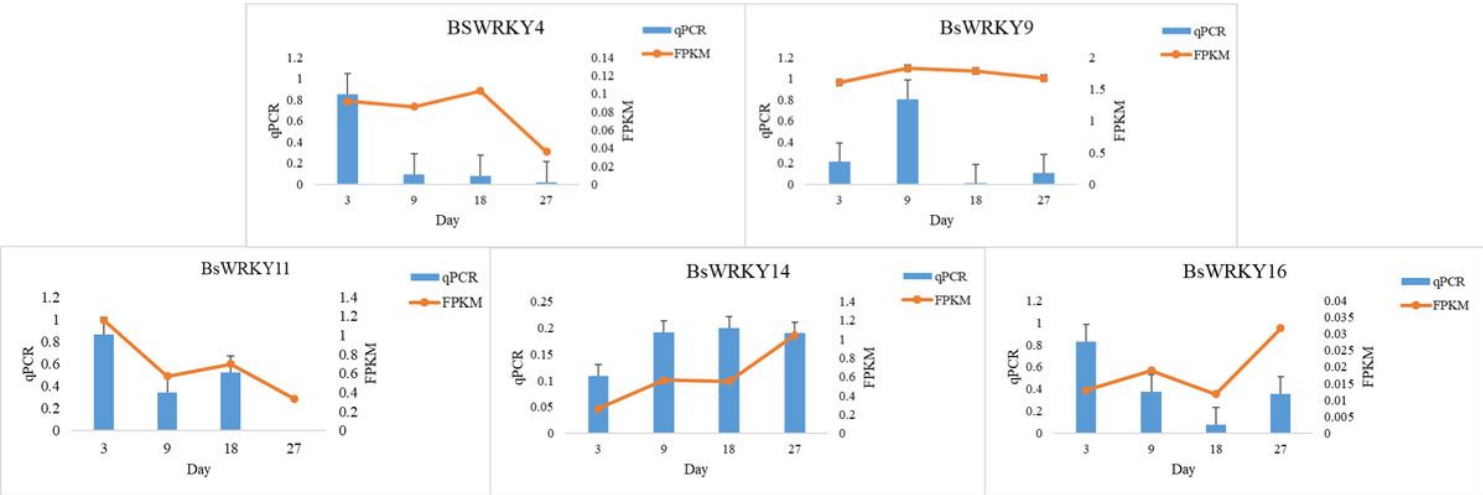


Figure 7

Heat map of correlation between *BsWRKY* genes and synthesis of four metabolites. The abscissa represents the gene, the ordinate represents the metabolite, and the numerical value represents the correlation coefficient.

Supplementary Files

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

- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.docx](#)