Genome-wide characterization of WRKY genes involved in flavonoid biosynthesis in Erigeron breviscapus

Wan-ling Song
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Shuang-yan Zhang
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Qi Li
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Gui-sheng Xiang
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Yan Zhao
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Sheng-chao Yang
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Wei Fan
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Bing Hao (Bing.Hao@hotmail.com)
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Zhang Guang-Hui
National & Local Joint Engineering Research Center on Germplasms Innovation & Utilization of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University

Research Article

Keywords: flavonoid, Erigeron breviscapus, WRKY, pharmaceutical, structure gene
Abstract

Background

The transcription factors of WRKY genes play essential roles in plant growth, stress responses, and metabolite biosynthesis. *Erigeron breviscapus*, a traditional Chinese herb, is abundant in flavonoids and has been used for centuries to treat cardiovascular and cerebrovascular diseases. However, the WRKY transcription factors that regulate flavonoid biosynthesis in *E. breviscapus* remain unknown.

Results

In this study, genome-wide characterization of WRKY genes in *E. breviscapus* was conducted to predict 75 *EbWRKY* transcription factors using phylogenetic, gene structure, and conserved motif analyses. In addition, the chromosomal location of each *EbWRKY* gene was analyzed. RNA sequencing showed that several *EbWRKY* genes transiently responded to exogenous abscisic acid (ABA), salicylic acid (SA), and gibberellin 3 (GA3) after 4h of treatment. In contrast, the expression of key structural genes involved in flavonoid biosynthesis increased after 4h in GA3 treatment. However, the content of flavonoid metabolites in leaves significantly increased at 12h. The quantitative real-time PCR (qRT-PCR) results showed that the expression patterns of *EbWRKY11*, *EbWRKY30*, *EbWRKY31*, *EbWRKY36*, and *EbWRKY44* transcription factors were similar to those of the 11 structural genes involved in flavonoid biosynthesis.

Conclusions

This study provides comprehensive information on the regulatory control network of flavonoid accumulation mechanisms, which could contribute to improving the pharmaceutical value of *E. breviscapus*.

1 Background

WRKY transcription factors play important roles in plant growth and development, defense regulation, stress, and synthesis of secondary metabolites[1–4]. The typical structure of WRKY is the N-terminus, which contains a WRKY DNA-binding domain, whereas the C-terminus contains a zinc finger motif (C2H2 or C2HC). According to the number of WRKY domains and the type of zinc finger structure, WRKY can be divided into three categories: Group I contains two WRKY domains, whereas Groups II and III have a single WRKY domain. Groups I and II WRKY had the same zinc finger structure, C2H2, at the C-terminus, and Group III WRKY had the same zinc finger structure, C2HC. The two WRKY domains of Group I WRKY transcription factors exhibit distinct functions. The C-terminal WRKY domain has DNA-binding activity, and the N-terminal WRKY domain can bind to DNA with the assistance of the C-terminal WRKY domain, enhancing the affinity and specificity of the DNA-binding activity[5]. According to the amino acid sequence differences of the WRKY domain, Group II WRKY can be divided into five subgroups: Ila, including a, Ilc, IId, and Ile.
Since the first WRKY gene (SPL7) was cloned from sweet potatoes, the identification and functional analysis of WRKY genes have developed rapidly in plants, especially in crops, fruits, and medicinal plants[6]. Several WRKY transcription factors have been identified in *Arabidopsis thaliana*, *Oryza sativa*, *Glycine max*, *Phyllostachys heterocyla*, and *Vitis vinifera* [5, 7–10]. A total of 14549 WRKY genes were recorded in the Plant Transcription Factor Database (PlantTFDB)[11]. Recently, plant hormones were shown to induce WRKY expression. 

*ZmWRKY106* is significantly induced by exogenous abscisic acid (ABA) [12]. *Arabidopsis AtWRKY70* positively regulates the salicylic acid (SA)-mediated defense signaling pathway[13]. *AtWRKY53* responds to and participates in regulating the plant senescence gene[14]. SA-induced transient expression of *AtWRKY29* in leaves can alleviate its disease symptoms[15]. *AtWRKY49* in *Arabidopsis* is regulated by SA or pathogenic bacteria and participates in plant disease resistance responses[16]. *Arabidopsis* mutant knockouts *WRKY18, WRKY40*, and *WRKY60* are highly sensitive to ABA, causing growth arrest in mutant plants[17]. Most of the reported WRKY transcription factors are related to plant disease resistance, drought resistance, and other adverse stresses in model plants and crops. However, only a few studies report regulating WRKY in medicinal plants using hormones on metabolism.

*Erigeron brevisapus* is a traditional medicinal plant of the genus *Erigeron* in the Composite family and is mainly distributed in Southwest China. In 2020, the sales revenue of traditional Chinese medicine preparations with *E. breviscapus* as a raw material in China reached 3 billion RMB (http://yn.chinadaily.com.cn/a/202007/27/WS5f1e9bb2a310a859d09da571.html). The main active components of flavonoids are phenylalanine and malonyl COA. Scutellarin, extracted from *E. brevisapus* leaves, has been used in prescription injections and as a drug for treating cardiovascular diseases[18, 19]. Genes and transcription factors that regulate scutellarin (R2R3-MYB) and anthocyanin (bHLH) biosynthesis have also been reported[20, 21]. However, the WRKY transcription factor family of *E. breviscapus*, which responds to exogenous substances and regulates the flavonoid pathway, has not been identified.

Only a few relevant studies on the WRKYs involved in flavonoid biosynthesis exist. WRKY transcription factors are closely associated with plant anthocyanin formation. Previous studies have shown that proanthocyanidin biosynthesis is enhanced by the overexpression of grape *VvWRKY26* in *Petunia hybrida*[22]. *AtWRKY41* is a repressor of anthocyanin biosynthesis, and mutations could significantly increase anthocyanin levels in *Arabidopsis*[23]. *AtWRKY44 (TTG2)* transcription factors are involved in flavonoid biosynthesis. However, WRKY transcription factors related to early flavonoid compound biosynthesis has only been reported in *Arabidopsis thaliana*. *AtWRKY44 (TTG2)* transcription factors are involved in flavonoid biosynthesis, whereas *AtWRKY23* transcription factors regulate flavonol accumulation, auxin transport, root growth, and development [24, 25].

In this study, the physical and chemical properties, phylogeny, classification, gene structure, chromosomal location, gene expression profile, and function of WRKY genes were identified based on the whole genome of *E. breviscapus*. In addition, integrated metabolomic and transcriptomic analyses were
performed to study the expression patterns of WRKY genes and flavonoid metabolites in response to exogenous hormone treatments. This study uncovered the regulatory mechanisms and functional identification of WRKY transcription factors involved in the growth, development, and flavonoid biosynthesis of *E. breviscapus*.

2 Results

2.1 Identification and physicochemical properties of WRKY genes in *E. breviscapus*

A total of 75 putative *EbWRKYs* were identified from *E. breviscapus* genomic data, which were designated as *EbWRKY1* to *EbWRKY75*. The number of amino acids ranged from 144 (*EbWRKY51*) to 756 (*EbWRKY19*), and the isoelectric points ranged from 4.97 (*EbWRKY29*/47) to 10.12 (*EbWRKY73*), including 45 acidic and 30 basic amino acids. In addition, the relative molecular weights ranged from 19.99 (*EbWRKY68*) to 80.89 kDa (*EbWRKY25*). The 75 *EbWRKY* proteins were located in the nucleus, indicating that these transcription factors might play regulatory roles in the nucleus. In addition, *EbWRKY19* is located on the cell membrane and may be involved in the expression and regulation of genes related to membrane transport (Supplementary Table S1).

2.2 Evolution and sequence analysis of WRKY transcription factors

To better understand the overall mechanisms of plant biodiversity and the function of WRKY genes in *E. breviscapus* network regulation, an evolutionary analysis of WRKY transcription factors was performed. Phylogenetic analyses of 75 *E. breviscapus* and 72 *Arabidopsis* WRKY transcription factors were performed (Fig. 1a). A total of 147 WRKYs were divided into seven branches. *E. breviscapus* and *Arabidopsis* WRKY proteins with the same classifications were classified into classes I, II, and III. Class II can be divided into five subtypes according to the different zinc-finger structural sites: IIa, IIb, IIC, IId, and IIe. There were five *EbWRKYs* in Group Ila, eight *EbWRKYs* in Group IIB, 11 *EbWRKYs* in Group IIC, 10 *EbWRKYs* in Group IId, nine *EbWRKYs* in Group IIe, and 15 *EbWRKYs* in Group III. The three major classes and five subclasses in the phylogenetic tree contained both the WRKY genes from *E. breviscapus* and *Arabidopsis*, indicating that the WRKY families of *Arabidopsis* and *E. breviscapus* are highly similar at the evolutionary level. Furthermore, the WRKY transcription factor branching relationships of the three major classes of *E. breviscapus* were relatively close, indicating that the WRKY gene family of *E. breviscapus* became more similar during evolution.

In addition to WRKYGQK, the core motif of *E. breviscapus* WRKY heptapeptide contained five variants: WRKYGKK, WKKYGQK, WKKYGDK, WKKYGEK, and WSKYGQK (Fig. 1b). Sequence comparison results showed that each WRKY protein, except *EbWRKY8*, contained a typical WRKY conserved domain at the N-terminus and a complete zinc finger structure at the C-terminus (CX4-5CX22-23HXH/C), which is an important feature for identifying WRKY transcription factors. All WRKY sequences of *E. breviscapus*
showed a high similarity and conservation of the WRKY domain. Group I *EbWRKYs* contained the same heptapeptide core motif WRKYGQK at the N-and C-termini and the zinc finger structure C2H2 behind the WRKY structure at the C-terminus. Group II and Group III had a WRKY domain at the N-terminus, but the zinc finger structure at the C-terminus differed (C2HX). However, *EbWRKY8* was highly similar to the other *EbWRKYs*, and evolutionary analysis clustered it into Group III. The domain sequence was lost, and the C-terminus retained a zinc-finger structure similar to other sequences.

### 2.3 Gene structure and conserved motif analysis of WRKY proteins

A total of 75 WRKY protein motifs were analyzed using MEME and TBtools. The heptapeptide core sequence of the WRKY conserved domain exists in motifs 1 and 3, and the zinc-finger structure exists in motif 2. Almost all *EbWRKY* proteins contain two WRKY motifs, motifs 1 and 2. The results showed that different *EbWRKYs* had different motif structures. The motif structures of each branch clustered by the same type of *EbWRKYs* had similar motifs. However, each motif was different. Motif 15 only existed in *EbWRKY49*, *EbWRKY54*, *EbWRKY59*, *EbWRKY60*, and *EbWRKY74* transcription factor genes in Group I. Motif 16 was the characteristic motif of *EbWRKYs* in Group IIb, and motif 4 was the characteristic motif of Groups I and II (b, c). Motif 11 was only found in eight *EbWRKYs* of Group III, and motif 12 was present in the *EbWRKY* sequences of Groups III and d (Fig. 2A; Supplementary Table S2). Phylogenetic trees of WRKY proteins were established, and the three groups were clustered according to their sequence similarity. WRKY sequences with similar structures in the evolutionary tree clustered into a single branch, indicating that these WRKY proteins may have similar functions. TBtools were used to analyze the number and distribution of exons of CDS sequences of the 75 WRKYs. The results showed a significant difference in the number of introns and exons in the WRKY gene family of *E. breviscapus*. The numbers of introns and exons were 1–6 and 2–7, respectively. (Fig. 2B).

### 2.4 Chromosomal mapping and collinearity analysis of WRKY genes

The 75 *EbWRKY* genes were irregularly distributed on all nine *E. breviscapus* chromosomes (Fig. 3a). A total of 21 *EbWRKY* genes were localized to chromosome 1, accounting for 28% of the *EbWRKY* gene family. Ten tandem duplications occurred on each of the six chromosomes. Eight WRKY members were collinear on chromosomes 1, 3, 4, and 6, respectively (Fig. 3b). To further predict the potential evolutionary patterns of the *EbWRKY* gene family, we constructed five comparative syntenic maps of *E. breviscapus* associated with five representative species, *Arabidopsis*, *carrot*, *sunflower*, *tomato*, and *potato*, which belong to the *Brassicaceae*, *Umbelliferae*, *Asteraceae*, and *Solanaceae* families, respectively (Fig. 3c). The orthologous gene pairs between *E. breviscapus* and *Arabidopsis*, *sunflower*, *carrot*, *tomato*, and *potato* were 21, 37, 33, 34, and 36, respectively. These results revealed that the identified orthologous events of *EbWRKY-HaWRKY* were considerably higher than those of other WRKY species based on their close evolutionary relationship. An extensive level of syntenic conservation and an increased number of
orthologous events in *EbWRKY-HaWRKY* indicated that *EbWRKY* genes in *E. breviscapus* shared a similar structure and function with *HaWRKY* genes.

### 2.5 Expression pattern of WRKY genes

The RPKM values of the roots, stems, leaves, and flowers were extracted from the genomic database to clarify the expression of *EbWRKY* family genes. The seven *EbWRKY* genes were not expressed in any tissue. However, 26, 31, three, and nine *EbWRKY* genes were specifically expressed in the leaves, roots, stems, flowers and leaves, respectively (Supplementary Table S3; Fig. 4a). In this study, exogenous SA, GA3, and ABA were sprayed onto Eucalyptus leaf *E. breviscapus* to increase their scutellarin (SE) content. The results showed that spraying ABA significantly increased the scutellarin content in leaves. Sprayed leaves were collected at 4, 12, and 24h for transcriptome analysis to investigate the mechanism of hormone-induced expression of *EbWRKY*. The results showed that the expression levels of *EbWRKY* significantly changed after 4h of hormone treatment. In the ABA treatment group, the expression levels of *EbWRKY69* and *EbWRKY30* were significantly up-regulated after 4h. In contrast, in the SA treatment assays, the gene expression levels of *EbWRKY8*, *EbWRKY17*, *EbWRKY51*, *EbWRKY67*, *EbWRKY18*, *EbWRKY66*, and *EbWRKY64* were significantly up-regulated after 12h of treatment, and *EbWRKY52* and *EbWRKY57* were significantly up-regulated after 24h. In the GA treatment, the expression levels of *EbWRKY3*, *EbWRKY41*, *EbWRKY47*, *EbWRKY2*, and *EbWRKY39* genes were significantly increased after 4h of treatment. Expression levels in the leaves gradually decreased over time. (Fig. 4b).

### 2.6 Flavonoid metabolites and structural gene expression analyzed in leaves

Ultra-high-performance liquid chromatography (UPLC) and tandem mass spectrometry (MS/MS) were used to determine dynamic changes in flavonoid metabolites in nine *E. breviscapuses* treated with the three hormones. Pearson correlation coefficients of the QC samples were calculated based on the relative quantitative values of the metabolites. The R² values of all the samples were close to 1, indicating better stability of the entire detection process and higher data quality (Supplementary Table S4; Fig. 5a).

Scutellarin biosynthesis begins with phenylalanine, followed by catalysis (*PAL*), cinnamate 4-hydroxylase (*C4H*), coumaric acid coenzyme A ligase (*4CL*), chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavone synthase II (*FSII*), flavonoid 7-O-glucuronosyltransferase (*F7GAT*), and flavone-6 hydroxylase (*F6H*) to form scutellarin[31]. In addition, other flavonoids, such as kaempferol, quercetin hesperidin, and Luteolin, were biosynthesized by *F3'H*, *F3H*, and *FLS* [32, 33].

### 2.7 Hormone-induced expression profiling of WRKY genes in *E. breviscapus*

The expression patterns of 11 key enzyme genes involved in flavonoid biosynthesis were analyzed after ABA, SA, and GA3 treatments. The results showed that the expression of 11 genes in cells treated with the three hormones was up-regulated at 4h. The downstream genes regulating flavonol biosynthesis, *F3H* and *F3'H*, were significantly up-regulated in ABA treatment at 24h but down-regulated in GA3 treatment.
Gene expression of GA3 was the highest at 4h, and the response to SA was not obvious. The expression levels of FLS, C4H, and F6H were significantly up-regulated after 4h of GA3 treatment (Fig. 5b).

A total of 159 flavonoids were identified, and three biological replicates were set for each sample (Supplementary Table S5). Among the 159 metabolites, flavones and flavonols (57.8%) accounted for a large proportion, followed by flavanones (15.7%), isoflavones (10.6%), and anthocyanins (8.1%). Whereas chalcones, dihydrochalcones (4.4%), and flavonoids (3.1%) were the least abundant (Fig. 5b). A total of 92 flavone and flavonol metabolites were identified, among which scutellarin had the highest content, followed by Apigenin7-O-beta-D-glucuronide. After the 92 metabolites were treated with ABA, SA, and GA3, the contents of 54 metabolites were differentially expressed (Fig. 5c), and the responses of flavonol and flavonoid compounds significantly increased after 12h of GA3 treatment, with 14 compounds significantly increased compared to other levels. In addition, at 24h of SA treatment, the content of pectinarigenin was significantly increased, and after 24h of GA3 treatment, the content of Oroxylin A significantly increased compared to other levels.

2.8 Integrated analysis of WRKYS involved in flavonoid metabolism

Transcriptome and metabolome data were integrated and analyzed to construct a co-expression network of key genes involved in flavonoid metabolism pathways. Through Pearson analysis of gene expression amount and compound content, relevant pairs were screened to construct a co-expression network (PCC ≥ 0.6, p ≤ 0.05). A total of 231 related pairs were identified and visualized using Cytoscape software (version 3.3.0). The network showed that 102 nodes were connected by 231 edges, including 10 key enzyme genes and 45 EbWRKYS, as well as 47 flavonoid metabolites that contained 26 flavone and flavonol metabolites. In addition, 143 pairs showed a positive correlation, and 88 pairs were negatively correlated (Fig. 6). Five candidate EbWRKYS were identified in the flavonoid metabolic pathway, and five positive candidate EbWRKYS were found, including EbWRKY11, EbWRKY30, EbWRKY31, EbWRKY36, and EbWRKY44. EbWRKY11 was connected to four key genes (EbF6H, F7GAT, FLS, and CHI). EbWRKY30 related to 4CL, EbWRKY31, and F3’H were a positive correlation. EbWRKY36 and EbWRKY44 were associated with PAL, EbF6H, C4H, 4CL, F7GAT, CHI, CHS, and FLS, while EbWRKY44 also connected with FSII as well (Fig. 6).

Moreover, EbWRKY11 was associated with 17 flavonoid metabolites, including nine flavone and flavonol metabolites. Eight pairs positively correlated, and five negatively correlated with the 13 EbWRKYS. EbWRKY31 was associated with six metabolites, including three flavones and three flavonols. Seven pairs were positively correlated, and eight pairs were negatively correlated with 15 EbWRKYS. However, EbWRKY30 was not directly connected with metabolites but was related to 17 EbWRKYS. EbWRKY36 was connected to 18 flavonoid metabolites and 12 EbWRKYS, of which 11 pairs were positively correlated, and one pair was negatively correlated. Twelve flavonoid metabolites and 16 EbWRKYS were connected to EbWRKY44, of which 13 pairs were positively correlated, and three pairs were negatively correlated.
Overall, the co-expression analysis of the selected EbWRKY genes revealed that these genes might play an essential role in flavonoid synthesis.

### 2.9 Quantitative real-time PCR (qRT-PCR) profiling characterization of genes under exogenous hormone treatment

To investigate the expression pattern responses of genes under ABA, SA, and GA3 exogenous hormones in *E. breviscapus*, 11 structural genes of the flavonoid biosynthesis pathway and five genes (*EbWRKY11, EbWRKY30, EbWRKY31, EbWRKY36, and EbWRKY44*) were selected for qRT-PCR analysis after exogenous hormone treatment. Eleven genes in the flavonoid synthesis pathway were significantly up-regulated, suggesting that these genes may function in the process (Fig. 7). The relative expression of the selected key genes peaked at different times under different treatments. *FLS* exhibited the highest expression level, showing a 20-fold increase after 4h of GA3 treatment and a 10-fold increase after 4h of ABA and SA treatment, which decreased at 12h. *C4H, F6H,* and *F3H* showed similar expression patterns after 4h of treatment, indicating that these genes are sensitive to GA3, SA, and ABA. *CHI* and *PAL* showed relatively high expression levels after 4h of ABA treatment (> 3.5-fold). *CHS* exhibited the highest expression level, with an 8-fold expression at 4h of ABA treatment and a 6.8-fold expression with GA3 treatment. *FS* showed a 2.9-fold higher expression after 4h of SA treatment.

The expression patterns of *EbWRKY11, EbWRKY30, EbWRKY31, EbWRKY36,* and *EbWRKY44* transcription factors related to the structural genes involved in flavonoid biosynthesis varied under different hormone treatments. *EbWRKY11* showed the highest expression level after 4h of hormone treatment and gradually decreased after 12h, indicating that *EbWRKY11* was sensitive to ABA, SA, and GA3. *EbWRKY30* was sensitive to ABA and had the highest expression level at 4h with a 4.3-fold increase. In the SA treatment, an increase was followed by a decrease in the volatility of *EbWRKY30. EbWRKY31* showed a significant response to ABA treatment. Its expression level gradually increased over time, with a 3.9-fold change at 24h. *EbWRKY36* exhibited a > 2-fold change in expression after 4h of hormone treatments. *EbWRKY44* was more sensitive to SA than to ABA and GA3, with a 2.2-fold expression at 4h. However, the gene expression level was highest at 12h and 24h of ABA treatment.

### 3 Discussion

With the development of omics, mining and identification of WRKY gene families in plants have become convenient and feasible. In this study, 75 members of EbWRKYs were identified from the genome of *E. breviscapus*. According to the phylogenetic analysis of *AtWRKYs*, *EbWRKYs* were divided into two large branches: Group I was separated into one branch, and the remaining *EbWRKYs* were composed of two complex branches, which contained the single branches Group IIC, Group IIA + IIB, Group IID + IIe, and Group III. WRKYs originated from prokaryotes, and Group I WRKYs were relatively primitive and evolved into Group II, which included Group IIA + IIB, Group c, and Group d + IIe. Group III is closely related to Group IID and Group e[34]. Phylogenetic analysis showed that the evolutionary classification of *E.*
*E. breviscapus* is consistent with previous studies. However, the genetic relationship between Group IIa + IIb and Group III was closer than between Group IId + Ile. Therefore, it is believed that the evolution of *EbWRKYs* in Group IIa + IIb was more closely related to that in Group III.

In this study, the conserved domains of the *EbWRKY* protein were evaluated and divided into three categories and five subtypes according to domain characteristics. Except for *EbWRKY8*, all the WRKY protein sequences contained a complete WRKY conserved domain and zinc finger structure. The sequence alignment was consistent with the results of the motif structure and gene family phylogenetic analyses. Multiple sequence alignments indicated that, regardless of the common WRKYGQK heptapeptide sequence, there were other variations, mainly distributed in Group IIC (*EbWRKY68/69*) and Group IId (*EbWRKY2/72/73*). In addition, Group III (*EbWRKY40*) and three variations were observed in Group IId. These results further confirmed that the evolution of Group I WRKYs was more conservative than that of other types. The differences in the conserved domain of the WRKY protein may be caused by variations in the WRKYGQK heptapeptide sequence and zinc finger structure during evolution or the deletion mutation of amino acid residues [34]. Mutations in the conserved domain reflect the diversity of the evolution of the plant WRKY gene family. Similar minor variations have been observed in *Citrus* and rice [35, 36]. Variations in WRKYGQKs affect its affinity for the W-box and further influence its function. Therefore, the interactions between *EbWRKY* proteins with these variations and downstream target genes, and their binding preferences with cis-acting W-box elements, should be further investigated.

WRKY transcription factors play key roles in plant growth, development, and secondary metabolite biosynthesis [37]. In recent years, many studies have reported the function of WRKY transcription factors in regulating plant disease resistance, stress resistance, and metabolite accumulation [38–40]. The expression of WRKY genes in four different tissues of *E. breviscapus* was analyzed, and the results showed that most (76%) of the *EbWRKY* genes were expressed in the roots and leaves. These *EbWRKYs* may be involved in stress defense and metabolite biosynthesis. In addition, WRKY transcription factors can mediate the signal transduction pathways of plant endogenous and exogenous hormones, such as SA, ABA, GA3, and MeJA, which can mediate the expression of WRKYs and related genes to affect plant abiotic stress and the accumulation of secondary metabolites [41, 42]. After ABA treatment, *OsWRKY71* could bind the W-box of Amy32b- α-amylase promoter element to control α-Amylase production in rice [43]. In this study, *EbWRKY* genes were differentially expressed at 4h, 12h, and 24h after GA3, SA, and ABA treatments of leaves, indicating that these *EbWRKY* transcription factors are involved in the exogenous hormone response.

Flavonoids are a class of important secondary metabolites in plants with various biological functions, such as attracting pollen, diffusing seeds, and preventing ultraviolet rays and pathogens. Applying exogenous plant hormones can regulate plant growth and secondary metabolism, promote plant growth, improve plant resistance, and, most importantly for medicinal plants, increase the pharmaceutical metabolites [34]. In a previous study, exogenous SA treatment significantly promoted the biosynthesis and accumulation of flavonoids, oleanolic acid, and ursolic acid in *Glechoma longituba*, a traditional Chinese herb used to promote blood circulation and reduce swelling [44]. The ABA treatment of *Ginkgo biloba*
leaves promotes the biosynthesis of flavonoids, the main active ingredients for treating Alzheimer's disease[45]. The SA and GA3 treatments of Fagopyrum esculentum and Citrus aurantium significantly increase the content of total flavonoids[46, 47]. In this study, 54 flavone and flavonol compounds showed spatiotemporal accumulation in the leaves of E. breviscapus after SA, ABA, and GA3 treatments, and GA3 significantly improved the accumulation of flavones and flavonols in the leaves.

Our previous results showed that PAL, EbF6H, C4H, 4CL, F7GAT, CHI, CHS, FLS, FSII, F3H, and F3’H were key structural genes regulating flavonoid biosynthesis in the leaves of E. breviscapus[20, 21]. The expression of these structural genes significantly increased after 4h of exogenous hormone treatment. Transcription factors can enhance or repress the initiation of gene transcription, indicating that genes begin to be transcribed in specific tissues or developmental stages. In addition, these genes are also related to transcription induced by the gene responses to the external environment via the identification of the upstream promoters of functional genes to regulate the level of gene transcription that participates in plant growth and development, defense regulation, stress response, and secondary metabolite accumulation[48–50]. According to the co-expression network analysis, EbWRKY11, EbWRKY30, EbWRKY31, EbWRKY36, and EbWRKY44 were identified as key structural genes. They were specifically overexpressed in leaves. qRT-PCR analysis further verified the expression patterns of the structural genes involved in the flavonoid biosynthesis pathway and EbWRKYs. The results showed that the expression patterns of EbWRKY transcription factors were consistent with those of the associated structural genes. Therefore, these five EbWRKY transcription factors may participate in the transcription of key structural genes that regulate flavonoid metabolite accumulation.

The phylogenetic analysis predicted the function of genes in different branches at the same node based on the distance of the phylogenetic relationships of the evolutionary branches. The phylogenetic analysis revealed that EbWRKY11, EbWRKY30, EbWRKY31, EbWRKY36, and EbWRKY44 were distributed across all three WRKY groups. The homologous genes of EbWRKY44 in Group IId were AtWRKY7/11/15/17 in Arabidopsis. In Group Iib, AtWRKY6, AtWRKY31, and AtWRKY42 were co-orthologous to EbWRKY31. EbWRKY30 in Group III was homologous to AtWRKY30/46/41/53. EbWRKY36 and EbWRKY11 were orthologs of AtWRKY1 and AtWRKY31, which belonged to different nodes in the same branch of Group I. Previous studies have shown that WRKY transcription factors in Arabidopsis are widely involved in the regulation of biotic and abiotic stress, plant growth, and development. AtWRKY15 regulates Arabidopsis growth and the salt stress response, whereas AtWRKY7/11/17 are negative regulators of the PAMP immune system, which could enhance plant resistance to pathogens[51, 52]. In addition, AtWRKY7 contains a CaM-binding domain, a new CaM-binding transcription factor that regulates plant growth and development and plays an important role in Ca^{2+} signal transduction[53]. Overexpression of AtWRKY30 enhances abiotic stress tolerance in A. thaliana at the early growth stage[54]. AtWRKY1 and AtWRKY41 can resist Pseudomonas syringae[55]. AtWRKY42 regulates Pi homeostasis to adapt to environmental changes, and AtWRKY6 participates in Pi transportation[56, 57]. AtWRKY53 regulates stomatal movement and negatively regulates drought resistance, whereas AtWRKY46 is involved in the sensitivity of Arabidopsis to drought and salt stress[58, 59]. Based on the evolutionary analysis results of EbWRKYs
and AtWRKYs, it was speculated that EbWRKYs homologous to various branches of Arabidopsis might participate in plant growth, development, and responses to biological and abiotic stresses with other members in the branch.

4 Conclusions

In this study, a total of 75 EbWRKY transcription factors were predicted from genome of E. breviscapus. The amino acid number, molecular weight, predicted isoelectric point (PI) value, chromosome position, domain pattern and conservative motif of EbWRKYs were revealed by bioinformatics-based analyses. The specificity of EbWRKYs gene expression in different tissues and their expression pattern under hormone treatment were determined based on RNA sequencing. Combining metabolome and transcriptome result revealed the regulatory mechanism between WRKY transcription factor and key genes involved in flavonoid biosynthesis. The expression patterns of EbWRKY11, EbWRKY30, EbWRKY31, EbWRKY36, and EbWRKY44 transcription factors were similar to those of the 11 key structural genes involved in flavonoid biosynthesis. We provided comprehensive information about the WRKY gene family of E. breviscapus and mechanism of EbWRKY genes involved in flavonoids metabolism regulation.

5 Methods

5.1 Plant treatment

In the present study, the E. breviscapus two-month seedlings were germinated and grown in a growth chamber at 22 °C with a 16/8h light/dark photoperiod. In addition, 200 mL ABA, SA, and gibberellin 3 (GA3) at a concentration of 200 µmol/L were used to treat the leaves of E. breviscapus. Leaf samples were taken at 4h, 12h, and 24h and subsequently kept frozen in liquid nitrogen and stored in a -80 °C refrigerator[26].

5.2 Identification and physicochemical properties of WRKY proteins

PfamScan v1.6 was performed to annotate the domains of the whole genome protein sequence of E. breviscapus based on the Pfam 35.0 database. Sequences with E-values less than Le-5 and containing the PF03106 domain were screened, and the atypical characteristics of the WRKY genes were manually removed. The ProtParam tool (https://web.expasy.org/protparam/) was used to predict the molecular weights (MWs), isoelectric points (pIs), number of amino acids, open reading frame (ORF) lengths, and locations of EbWRKY proteins.

5.3 Protein domain and phylogenetic evolution analysis

Multiple sequence alignments were performed using MAFFT v7.490 to understand the evolutionary relationship between E. breviscapus and A. thaliana. To investigate the relationship between E. breviscapus WRKY proteins, a phylogenetic tree of e E. breviscapus and A. thaliana WRKY proteins were
constructed. Phylogenetic analysis was performed using Phylipv3.698 following the neighbor-joining method (1000 repeats). Next, the evolutionary tree software was used for EvolView(https://evolgenius.info/evolview/#/) [27]. Lastly, the WRKY sequence of A. thaliana was downloaded from the TAIR database (https://www.arabidopsis.org/).

5.4 Comprehensive analysis of WRKY genes

The intron-exon structures of the EbWRKY genes were determined using the gene structure display server (http://gsds.cbi.pku.edu.cn/). The conserved motifs of WRKY proteins were predicted using MutipleEm for Motif Elicitation (http://alternate.meme-suite.org/tools/meme). The parameter configuration was set with the following optimized parameters: any number of repetitions, 20; minimum width, 10; maximum width, 80. Gene structure and chromosome mapping of WRKY family members were performed using TBtools v1.098691. The one-step MCScanX tool was used to analyze the collinearity of the WRKY gene family between E. breviscapus and Arabidopsis, carrot, sunflower, tomato, and potato. Collinearity within E. breviscapus was displayed using Advanced Circos software.

5.5 Expression profiling analysis of EbWRKY genes in various tissues

The expression of candidate EbWRKYs in the roots, stems, leaves, and flowers of E. breviscapus was extracted from genome and transcriptome data. Three biological replicates of the different plant leaves were used per sample. The RPKM value was used to homogenize the gene expression analysis, and clustering results and heat maps were generated using TBtools v1.098691 [28].

5.6 Metabolites analysis

Root, stem, leaf, and flower samples (100 mg) were ground in liquid nitrogen, and the homogenate was resuspended in pre-chilled 80% methanol and 0.1% formic acid by vortexing. The samples were incubated on ice for 5 min and then centrifuged at 15,000 g at 4°C for 20 min. The supernatant was diluted to a final methanol concentration of 53% for the LC-MS/MS analysis [29]. Samples were injected onto an Xselect HSS T3 column (2.1×150 mm, 2.5 µm) with a 20-min linear gradient at a 0.4 mL/min flow rate for the positive/negative polarity mode. The eluents used were eluent A (0.1% formic acid water) and eluent B (0.1% formic acid-acetonitrile). The solvent gradient was set as follows: 2% B, 2 min; 2–100% B, 15.0 min; 100% B, 17.0 min; 100–2% B, 17.1 min; 2% B, 20 min [30]. The data files generated by HPLC-MS/MS were processed using SCIEX OS Version.

5.7 Co-expression network analysis

The expression of candidate EbWRKYs and key genes involved in flavonoid biosynthesis in the roots, stems, leaves, and flowers of E. breviscapus was extracted from the transcriptome data. First, statistically significant correlations between differential metabolites were calculated using the R language. Statistical significance was set at \( p < 0.05 \). The expression of the gene and the relative content of metabolites were collected to screen the correlation pairs with Pearson product-moment correlation coefficient (PCC) \( \geq 0.6 \) and \( p\)-value \( \leq 0.05 \). The filtered genes were then used to construct a correlation network. The co-
expression network was visualized using Cytoscape version 3.3.0 software (https://www.cytoscape.org) (Fig. 6).

5.8 Quantitative real-time PCR (qRT-PCR) analysis

Leaves were collected, and total RNA was extracted from hormone-treated samples using a HiPure HP Plant RNA Mini Kit (R4165-02). cDNA was synthesized using a PrimeScript RT Reagent Kit with gDNA Eraser (Takara, Japan). Gene-specific primers for qRT-PCR reactions were designed using Primer3 web version 4.1.0 (https://primer3.ut.ee/) (Supplementary Table S6). A Quantstudio 5 Flex Real-Time PCR System (Thermo Fisher Scientific, USA) was used to analyze three technical replicates. The expression levels of genes from different treatments were normalized to EbACTIN2. Lastly, the relative expression levels were calculated by the $2^{-\Delta\Delta Ct}$ method and visualized with GraphPad Prism 8.0.2.

Abbreviations

ABA
abscisic acid
SA
salicylic acid
GA3
gibberellin 3
AtWRKY
Arabidopsis thaliana WRKY Transcription factor
EbWRKY
Erigeron breviscapus (Vant.) Hand.-Mazz WRKY Transcription factor
qRT-PCR
Quantitative RT-PCR
PAL
Phenylalanine ammonia lyase
C4H
Cinnamate 4-hydroxylase
4CL
4-coumarate:CoA ligase
CHS
Chalcone synthase
CHI
Chalcone isomerase
FSII
Flavone synthase
EbF6H
Flavone 6-hydroase
F7GAT
Flavonoid 7-O-glucuronosyltransferase
F3H
Flavanone 3-hydroxylase
F3’H
Flavonoid 3’-hydroxylase
FLS
Flavonol synthase
UPLC
Ultra-high-performance liquid chromatography
SE
Scutellarin.

Declarations

7.1 Ethics approval and consent to participate

Experimental research and field studies on plants are comply with relevant institutional, national, and international guidelines and legislation, and all methods were performed in accordance with the relevant guidelines and regulations. The cultivated *E. breviscapus* were collected with official permissions of Yunnan Hongling Biological Technology Co., LTD. Honghe, China.

7.2 Consent for publication

Not Applicable.

7.3 Availability of data and materials

The raw data of *E. breviscapus* transcriptome in the current study are available in the National Center for Biotechnology Information (NCBI) database under project number PRJNA971382 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA971382). The genomic date of *E. breviscapus* is downloaded from Medicinal Plants multi-Omics Database (http://medicinalplants.ynau.edu.cn/genome/detail/68).

7.4 Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7.5 Funding

This work was supported by Fundamental Research Project of Yunnan (202101AS070037), Science and Technology Innovation team of Yunnan (202105AE160011), The Major Science and Technique Programs
in Yunnan Province (202102AE090042), Yunnan Characteristic Plant Extraction Laboratory (2022YKZY001), the First Projects of Science and Technology Plan in the Biomedical field in 2021 (202102AA310048).

7.6 Authors' contributions

BH, GHZ, and SCY conceived the study. WLS, SYZ and QL performed the experiments. GHZ, WF, and BH the designed experiments. WLS and GSX analyzed the data. BH and WLS drafted the manuscript. BH, GHZ, and SCY reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

7.7 Acknowledgments

The authors thank to lab members for assistance.

References


Figures
Figure 1

Phylogenetic and WRKY protein domain sequence analysis of *E. breviscapus*: (a) Phylogenetic analysis of *E. breviscapus* and *Arabidopsis* WRKY transcription factor; (b) WRKY protein domain sequence analysis in *E. breviscapus*. 
Figure 2

The evolutionary relationship, genetic structures, and motifs of WRKY genes in *E.breviscapus*: (A) Phylogenetic evolution and motif structure of *E.breviscapus* WRKY genes; (B) Gene structure of *E.breviscapus* WRKY genes (green box represents the exon).
Figure 3

The Synteny analysis and chromosome location of WRKY genes in *E. breviscapus*: (a) Chromosomal location of WRKY genes; (b) The internal collinearity circle diagram of the *E. breviscapus* genome (the black line represents the position of genes on chromosomes; the arc represents the collinearity relationship between genes; the WRKY gene pair is highlighted with a red line; the gene name color represents different subgroups); (c) WRKY gene pairs are highlighted with red lines according to the analysis of collinearity between *E. breviscapus* and different plant genomes.
Figure 4

The expression profiles of WRKY genes in *E. breviscapusa*:

(a) Expression profiles of WRKY genes in different tissues of *E. breviscapusa*; (b) Expression profiles of *E. breviscapusa* WRKY under different hormone treatments at different times. The color scale on the right of each diagram represents RPKM expression values: green indicates higher levels and blue indicates lower levels of transcript abundance.
Figure 5

Quantitative and structural gene expression analysis of flavonoid pathway metabolites in the leaves of *E. breviscapus*: (a) Pearson correlation coefficient analysis of the QC of samples; (b) Structural gene expression analysis; (c) Contents of 54 flavone and flavonol metabolites in *E. breviscapus* under abscisic acid (ABA), salicylic acid (SA), and gibberellin 3 (GA3) hormone treatments at different times.
**Figure 6**

Co-expression analysis of structural genes involved in the flavonoid biosynthesis pathway and *EbWRKYS* in the leaves of *E. breviscapus*. Yellowish-green nodes represent genes; Blackish-green nodes represent WRKY TFs; Lavender nodes represent other flavonoid metabolites; Green nodes represent flavone and flavonol metabolites. The size of the circle is associated with the number of *EbWRKY* genes. Black circles outside the genes are associated with the number of metabolites.
Figure 7

Relative expression of selected *Eb* genes in response to exogenous hormone treatment. Genes expression was analyzed by RT-qPCR. Blue was used as the untreated control (expression= 1); Red, green and purple represent 4h, 12h, and 24h. Error bars represent standard errors. Data were calculated using the $2^{-\Delta\Delta Ct}$ method.

Supplementary Files

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

- 03Supplementarytablesforthisstudy.xlsx