Transcriptome-wide identification and screening of WRKY factors involved in the regulation of taxol biosynthesis in *Taxus chinensis*

Meng Zhang1,2, Ying Chen1,2, Lin Nie1,2, Xiaofei Jin1,2, Weifang Liao1,2, Shengying Zhao1,2, Chunhua Fu1,2 & Longjiang Yu1,2

WRKY, a plant-specific transcription factor family, plays important roles in pathogen defense, abiotic cues, phytohormone signaling, and regulation of plant secondary metabolism. However, little is known about the roles, functions, and mechanisms of WRKY in taxane biosynthesis in *Taxus* spp. In this study, 61 transcripts were identified from *Taxus chinensis* transcriptome datasets by using hidden Markov model search. All of these transcripts encoded proteins containing WRKY domains, which were designated as TcWRKY1–61. After phylogenetic analysis of the WRKY domains of TcWRKYs and AtWRKYs, 16, 8, 10, 14, 5, 7, and 1 TcWRKYs were cladded into Group I, IIa–IIe, and III, respectively. Then, six representative TcWRKYs were selected to classify their effects on taxol biosynthesis. After MeJA (methyl jasmonate acid) and SA (salicylic acid) treatments, all of the six TcWRKYs were upregulated by MeJA treatment. TcWRKY44 (IId) and TcWRKY47 (IIa) were upregulated, whereas TcWRKY8 (IIc), TcWRKY20 (III), TcWRKY26 (I), TcWRKY41 (IIe), and TcWRKY52 (IIb) were downregulated by SA treatment. Overexpression experiments showed that the six selected TcWRKYs exerted different effects on taxol biosynthesis. In specific, TcWRKY8 and TcWRKY47 significantly improved the expression levels of taxol-biosynthesis-related genes. Transcriptome-wide identification of WRKY factors in *Taxus* not only enhances our understanding of plant WRKY factors but also identifies candidate regulators of taxol biosynthesis.

Taxol is the most effective chemotherapy medication used to treat many cancers, including ovarian, breast, lung, Kaposi sarcoma, cervical, and pancreatic1. Taxol biosynthesis is complicated and involves approximately 19–20 steps of enzyme reaction catalyzed from geranylgeranyl pyrophosphate2–4. Most of the biosynthesis genes were isolated and their functions investigated in the past decades. However, only a few studies focused on the regulation mechanisms underlying the biosynthesis of these secondary metabolites. Recently, 5’ flanking sequences of several enzyme genes, including 10-deacetylbaccatin III-10-β-O-acetyltransferase (*DBAT*), taxadiene synthase (*TS*), phenylpropanoyl transferase (*BAPT*), taxoid 5α-hydroxylase (*T5H*), taxoid 7α-hydroxylase (*T7H*), and taxoid 10α-hydroxylase (*T10H*), have been submitted to public databases. Notably, W-box, the specific cis-element binding with WRKY proteins, was detected in all of these promoter regions, indicating that WRKY plays important roles in the regulation of taxol biosynthesis5,6.

WRKY, which is named after the conserved WRKYGQK motif in the WRKY domain, constitutes a dominant family of transcription factors in plants. All of the known WRKY factors could be divided into Group I to III according to the numbers of WRKY domains and their types of zinc finger motifs7. Group I has two WRKY domains, whereas Group II and III contain only one WRKY domain. Furthermore, Group II and III are distinguished because of their different zinc finger motifs, i.e., C2H2 (C-X4,5-C-X22,23-H-X1-H) type in Group II WRKYs, whereas C2HC (C-X7-C-X23-H-X1-C) type in Group III. Commonly, Group II WRKY factors could...
be further divided into five subgroups (i.e., Groups IIa–IIe) based on the specific characteristics of the WRKY domains. Recently, five novel WRKY proteins have been observed to contain three or four WRKY domains from 43 plants by genome-wide identification, which is rare in plants.

Currently, WRKY transcription factors have become the most pivotal transcription factors in plants because of their indispensable roles in regulating various physiological processes, such as biotic and abiotic stress, signal molecule delivery, and plant senescence and biosynthesis of secondary metabolites. For example, GaWRKY1 regulates the sesquiterpene synthase CAD1 (+)-6-cadinene synthase-A gene, which is involved in gossypol pathway regulation in *Gossypium arboretum*. AaWRKY1 regulates amorpha-4,11-diene synthase, which is involved in artemisinin biosynthesis in *Artemisia annua*. We also observed that TcWRKY1 could specifically interact with the promoter of the *DBAT* gene. All of these results indicate that WRKY factors play important roles in taxol biosynthesis.

Nowadays, high-throughput screening of regulation factors from various omics datasets has become economical and effective for researchers. In wheat (*Triticum aestivum* L.), 48 putative drought-induced WRKY genes were initially identified from a transcriptome, and TaWRKY33 was found to serve excellent functions in enhancing the drought tolerance of wheat. In Arabidopsis thaliana, analysis of RNA sequencing data revealed that AtWRKY46, AtWRKY54, and AtWRKY70 play global roles in promoting brassinosteroid-mediated gene expression and inhibiting drought-responsive genes. In summary, omics analysis has become an effective method to screen important transcription factors.

In this study, 61 TcWRKYs were identified by transcriptome-wide identification from Taxis chinensis. Multiple sequence alignment, motif analysis, and phylogenetic analysis were conducted to analyze the evolutionary relationship of WRKY factors between *Taxis* and angiosperms. Thus, six TcWRKYs were selected for functional studies to identify their relationships with taxol biosynthesis. Our work enhances the understanding of WRKY factors in gymnosperm and identifies several effective candidate regulators of taxol biosynthesis.

**Materials and Methods**

**Transcriptome-wide identification of TcWRKYS in Taxis chinensis.** Previously, three pairs of samples, MeJA- (Methyl Jasmonate acid), GA-treated and NA/CA cells (Accession Numbers: SRR1343578, SRR1339474, and GSE28539), were high-throughout sequenced. In MeJA-treated and NA/CA cells, taxol and taxanes contents were significantly higher than control cells, while there was no difference between GA-treated and control cells. Thus, these datasets would help us analyze the relationships of expression patterns of WRKY factors with taxol biosynthesis. To improve the efficiency of gene screening, all reads of these transcriptomes were re-assembled by Trinity, totally 34 Gbp (2 Gbp for MJ-, 16 Gbp for GA- and NA/CA cells respectively). To improve the efficiency of gene screening, all reads of these transcriptomes were re-assembled by Trinity, totally 34 Gbp (2 Gbp for MJ-, 16 Gbp for GA- and NA/CA cells respectively).

Finally, 67,147 unigenes were obtained with N50 value of 1,552 bp. Then, the HMM model of WRKY (PF03106) was retrieved from the Pfam database (http://pfam.sanger.ac.uk). After redundant sequences were removed, the HMMER program was used to identify the WRKYs, with an e-value cutoff of 1e-5. These unigenes containing two WRKY domains were separated as Group I, the others with only WRKY domain need further analysis to divide. Moreover, the WRKYs of *Arabidopsis thaliana* were downloaded from PlantTFDB database (http://planttfdb.cbi.pku.edu.cn/).

**Classification and phylogenetic analysis of conserved sequence of TcWRKY genes.** The AtWRKYs protein sequences were downloaded from TAIR (http://www.arabidopsis.org/), and Pfam database was downloaded at http://pfam.xfam.org/. Hmmscan programe of HMMER package was used to identify the conserved domains of AtWRKYS and TcWRKYS with E cut-off 1e-5. MEME was used to generate the motif logo of AtWRKYS and TcWRKYS. Motif LXXLL (or LXLXLX) and HARF (RTGHARFRR (A/G) P) were found manually.

The conserved sequences of *A. thaliana* were selected to build the phylogenetic tree. Multiple alignment was conducted by ClustalW with identity protein weight matrix. Phylogenetic analysis was performed with a neighbor-joining (NJ) method by using bootstrapping with 1000 repeats and Possion Correction model with 1000 resamplings in MEGA 5.0. Phylogenetic tree was modified by FigTree V1.4.2.

**Plant hormones treatment.** In vitro long-term subcultured cells of *Taxis chinensis* were maintained on 62* medium containing 0.5 mg/L 6-BA, 0.5 mg/L 6BA, and 0.5 mg/L 2,4-D under two-day conditions. Then, 6 g from 43 plants by genome-wide identification, which is rare in plants.

**Gene cloning and construction of TcWRKY Overexpression Vectors.** The total RNA of *Taxis chinensis* cells was reverse-transcribed to cDNA by reverse transcription kit (Thermo Scientific, USA). Specific primers were designed based on our transcriptome data (Supplementary Table S1). PCR procedures were as following: 96 °C for 5 min; 94 °C for 40 s, 52 °C for 40 s, 72 °C for 30 s, 30 cycles; 72 °C for 10 min, 16 °C for 10 min. The PCR products were subcloned into pMD18-T (TaKaRa, Japan) for sequencing.

**Quantificational real-time polymerase chain reaction.** The overexpression of TcWRKYS was analyzed by qRT-PCR with SYBR Green II method. The reaction system contained 5 μl SYBR Premix buffer, 0.5 μl each of the primers and 1 μl template and 3 μl dDH2O. The thermal profile for qRT-PCR was as follows: holding
stage: 95 °C for 5 min; cycling stage: 95 °C for 10 s, 52 °C for 10 s, 72 °C for 15 s, 40 cycles; melting stage: 95 °C for 1 min, 65 °C to 95 °C 0.3 °C increase per cycle for 15 s. Each reaction was run in triplicate to obtain the average value and $2^{-\Delta\Delta C_t}$ method was applied for the analysis gene expression.

**The transformation assay in Transgenic Taxus chinensis cells.** 6 g Taxus chinensis CA cells were suspended in 50 mL fresh 62# medium. Then the Agrobacterium tumefaciens strain LBA4404 containing the TcWRKY-overexpressing constructs were added to ensure the value of absorbance were 0.6 in these mediums. Finally, the concentration of 1 mol/L AS (Acetosyringone, Biofroxx, German) was added into the liquid medium. They were shaked at 25 °C with 125 rpm for 48 h dark period. LBA4404 with the pBI121 or pCAM-BIA1303-overexpressing constructs were as the control group.

**Results**

**Identification of TcWRKYs from T. chinensis transcriptome datasets.** For the identification of WRKY genes from T. chinensis, all known WRKY of A. thaliana and Oryza sativa were used as queries to perform local BLASTP search on the T. chinensis transcriptome datasets. Then, the obtained sequences were submitted to the NCBI-CDD web server (http://www.ncbi.nlm.nih.gov/cdd/) to analyze their conserved protein domain. Finally, a total of 61 unique TcWRKY genes encoding conserved WRKY domains were identified.

**Phylogenetic analysis of WRKY domains.** Phylogenetic analysis was performed using all putative 61 TcWRKY proteins in T. chinensis and 71 AtWRKY proteins in A. thaliana to categorize and investigate the evolutionary relationships of TcWRKY genes.
Results of phylogenetic analysis (Fig. 1) showed that all selected WRKYs, including TcWRKY proteins, could be categorized into three groups, i.e., Group I, II, and III. Among these TcWRKYs, 16 were categorized as Group I. Meanwhile, only 5 of 16 were full length after online BLAST. However, eight TcWRKYs encoded two WRKY domains (called Group IN and IC) and C2H2 type of zinc finger motif. Then, 44 TcWRKYs were assigned to Group II, which contains a single WRKY domain and C2H2 type of zinc finger motif. Furthermore, these 44 Group II TcWRKYs were classified into five subgroups, i.e., 8 in Group IIA, 10 in Group IIB, 14 in Group IIC, 5 in Group IID, and 7 in Group IIE. Notably, these five subgroups could be summed up to three branches, i.e., IIA + B, IIC, and IID + E; the same results were reported in many other plants. Finally, only TcWRKY20 was categorized as Group III, which had one WRKY domain and C2HC type of zinc finger motif. These results are in accordance with the results of Arabidopsis and other plants, indicating that WRKY differentiated completely before evolution-induced bifurcation of gymnosperm and angiosperm.

AtWRKY19 (Group I), AtWRKY16 (Group II), and AtWRKY52 (Group II) appear to be unique. These results are similar to previous reports because only three AtWRKYs were R protein-WRKYs (both resistance (R) proteins and WRKY transcription factors) in Arabidopsis. Moreover, AtWRKY16 and AtWRKY52 could be separated as Group IIg WRKYs, which is rare in plants, although AtWRKY16 contains two WRKY domains. According to our phylogenetic analysis, no Group I IIg WRKY was detected in T. chinensis.

Sequence alignment of WRKY domains of TcWRKYs. The WRKY domain, which determines the molecular structures and functions of all WRKY proteins, consisted of the WRKYGQK and zinc finger motifs. The WRKYGQK motif was conserved in almost all TcWRKY factors, except for several subgroup IIC WRKYs, which was WRKYGK instead in TcWRKY34, TcWRKY49, TcWRKY51, and TcWRKY55 (Fig. 2). The WRKYGKK sequence is the most common variant present in Group IIc not only in Taxus but also in soybean, Solanum lycopersicum, Lotus japonicus, and Brassica oleracea var. capitata.

Conserved domains of TcWRKY proteins. In addition to the conserved domains/motifs, WRKY proteins contain many diverse conserved domains, such as B3 domain, NAC (named after NAM, ATAF1/2, and CUC2 proteins) domain, zinc finger, SQUAMOSA promoter binding domain, Toll–interleukin receptor (TIR), leucine-rich repeat (LRR), paired amphipathic helix repeat (PAH), cystathionine-β-synthase (CBS), and kinase domain.

Aside from the Ca2+-dependent CaM-binding domain (CaMBD), all of Group IId TcWRKYs and AtWRKYs also contain the plant zinc finger cluster (Plant_zf_clust, PF10533) domain (Fig. 3). This result is in accordance with previous reports. Group Ile and IId are commonly considered one subgroup, although Group Ile has no CaMBD. According to our results, Group Ile contains a peptide that is highly similar to the Plant_zf_clust domain of Group IId, such that they share similar functions and relationships (Fig. 3). TcWRKY57 also contains an HSF-type DNA-binding (HSF_DNA-bind, pfam00447) domain at its N-terminal (44 aa-133 aa) (Fig. 3).

TcWRKY24 (Group IId) contains the HARF (RTGHARFRR [A/G] P) motif, whose function has not been clearly determined, in its CaMBD (Fig. 3). Several TcWRKYs also contain the LxxLL motif, which participates in many protein–protein interactions associated with different aspects of transcriptional regulation, i.e., TcWRKY12 (LSQLL, Group I), TcWRKY20 (LYQLL, Group III), TcWRKY24 (LYQLL, Group IId), TcWRKY32 (LYQLL, Group I), TcWRKY33 (LSPLL, Group I), TcWRKY40 (LQLL, Group IId), and TcWRKY48 (LSPLL, Group I) (Fig. 3, supplement S2). The NCBI BLASTP results showed that TcWRKY3 and TcWRKY60 contain a leucine zipper (pfam15294) domain at the N-terminal (93 aa-147 aa), which is also detected in AtWRKY6 (IId) (Fig. 4b).

Overall, some novel conserved domains were detected in TcWRKY proteins. However, current conserved domains detected in WRKY proteins, such as LRR, were not identified in TcWRKYs. Moreover, completed sequences need to be further identified to clarify the molecular structure of WRKYs in Taxus.

Group IIB TcWRKY proteins contain EAR motifs. Notably, several Group IIB TcWRKYs contain the ERF-associated amphiphilic repression (EAR) motif, which is a strong repression domain present in various repressors. TcWRKY4/27/52 has the LKLDLY-type EAR motif, whereas TcWRKY3/16/60 has the LKLAL-type EAR motif. TcWRKY3/60 also has an LSLGLN-type EAR motif at the C-terminal of the LKLAL-type EAR motif (Fig. 4).

Several AtWRKYs also contain the EAR motif, i.e., AtWRKY14 (IId) and AtWRKY18 (IId) had the DLNxp-type EAR motif, whereas AtWRKY9 (IId) has two LSLAL-type EAR motifs. According to our results, AtWRKY36 (IId) contains the LKLALS-type EAR motif (Fig. 4). Notably, no Group Ila or Ile TcWRKYs contains any EAR motif.

Expression profiles of TcWRKY genes. A hierarchical cluster analysis was performed using the three experimental datasets to determine the potential roles of the 61 TcWRKY genes in plant responses to various environmental stresses (Fig. 5). NA denotes the newly isolated cells, whereas CA denotes the 10 year long-term subcultured cells; the taxol content between these two samples was highly different. Methyl Jasmonate acid (MeJA) and gibberellin (GA) are plant endogenous hormones.

Most TcWRKYs changed significantly in the three experimental datasets (Fig. 5). Of 61 TcWRKYs, 50 were changed significantly, 43 were downregulated after MeJA treatment for 15 h, and only TcWRKY3/60 (Group IIB) and TcWRKY17/18/49/55/56 (Group IIC) were increased. Of 55 differentially expressed (DE) TcWRKYs, 45 were downregulated after GA treatment for 3 h. TcWRKY4/27/52 (Group IId), TcWRKY7/8 (Group IId), TcWRKY22/37 (Group I), TcWRKY44 (Group IId), and TcWRKY25/36 (Group Ila) were induced by GA.
Of 57 DE TcWRKYs, 42 were downregulated in NA than in CA. TcWRKY7/8/11/17/46/56 (Group IIc), TcWRKY10/25/50/36 (Group IIa), TcWRKY14/19/43/41 (Group IIe), and TcWRKY44 (Group IId) were upregulated in NA.

Figure 2. Sequence alignment of WRKY domain of 61 TcWRKY proteins. WRKYGQK motif was underlined as WRKY domain, and C(H/H/C) of zinc-finger motif was designated by star (*). TcWRKY45 was uncomplete so that its HxH residues were not found. All TcWRKYS were grouped according to phylogenetic analysis. The names with N or C represented for N- and C-terminal WRKY domain of Group I respectively. Several Group I TcWRKYs didn't find their C-terminal WRKY domain. Sequence alignment was conducted by ClustalW, the figure was generated by DNAMAN 6.0. Different colors indicate sequence similarities: black indicated 100%, purple is 75%, green is 50%.
Expression patterns of selected TcWRKYs induced by SA and MeJA. A total of seven TcWRKYs, one from each subgroup, i.e., TcWRKY26 (Group I), TcWRKY47 (Group IIa), TcWRKY52 (Group IIb), TcWRKY8 (Group IIc), TcWRKY44 (Group IId), TcWRKY41 (Group IIe), and TcWRKY20 (Group III), were selected. The expression patterns of these TcWRKY genes responding to MeJA and SA treatments were determined. Results showed that the seven TcWRKYs had different response patterns to MeJA and SA treatments.

After MeJA treatment, TcWRKY8, TcWRKY20, TcWRKY26, TcWRKY41, TcWRKY44, and TcWRKY47 significantly increased at 3 h, reaching 12.1, 11.0, 2.6, 11.2, 16.5, and 10.1 times, respectively. Meanwhile, at 6 h, their expression levels decreased in varying degrees, with only TcWRKY20 and TcWRKY47 remaining >2 times higher than the control. Moreover, TcWRKY52 was insensitive to MeJA treatment (Fig. 6a). TcWRKY8, TcWRKY20, TcWRKY26, TcWRKY41, and TcWRKY52 were significantly inhibited at 6 h after SA treatment. By contrast,
although TcWRKY44 and TcWRKY47 were reduced at 3 h, they were significantly enhanced at 6 h by 3.3 and 4.8 times, respectively (Fig. 6b). Overall, only TcWRKY44 and TcWRKY47 were induced by MeJA and SA treatments.

Overexpression of TcWRKYs. According to previous reports, Group IId and IIe were grouped together as Group IId + e and verified to function redundantly in other plants13. Therefore, TcWRKY44 was selected for further studies to test the functions of Group IId + e WRKYs on taxol biosynthesis. Then, six TcWRKYs were overexpressed in the T. chinensis cell lines. Our previous research verified W-box as the key cis-element of promoters of DBAT and T5H. TcERF12 and TcERF15, which encode regulators of taxol biosynthesis, also contain W-boxes in their promoters (unpublished work). When these TcWRKYs are overexpressed, the expression levels of DBAT, T5H, TcERF12, and TcERF15 change in varying degrees (Fig. 7).

TcWRKY8, TcWRKY20, TcWRKY47, and TcWRKY26 could increase the expression levels of the DBAT (3.1, 2.7, 5.4, and 2.2 times), T5H (4.6, 2.8, 4.1, and 1.5 times), and TcERF15 (3.2, 4.7, 5.5, and 1.4 times) genes. However, only TcWRKY20 and TcWRKY26 could also increase TcERF12 by 2.8 times. By contrast, TcWRKY47 and TcWRKY8 inhibit TcERF12 by 0.4 and 0.8 times, respectively.

Meanwhile, TcWRKY44 and TcWRKY52 inhibit the expression levels of the DBAT (0.7 and 0.2 times), T5H (0.4 and 0.3 times), and TcERF12 (0.5 and 0.4 times) genes significantly. TcERF15 is also downregulated by 0.3 times in TcWRKY44-overexpressing cell lines but upregulated by 2.0 times in TcWRKY52-overexpressing cell lines. The results show that the overexpression of different TcWRKYs exerts different effects on the expression of taxol-biosynthesis-related genes.

Discussion
WRKY, the most pivotal transcription factors in plants, could regulate the expression of downstream genes by combining with cis-acting element W-box39. WRKY plays indispensable roles in regulating various physiological processes, including biotic and abiotic stress responses, signal molecule delivery, plant senescence, and synthesis of secondary metabolites. Taxol is a precious secondary metabolite initially isolated from Taxus spp. and widely used as an anticancer drug. Previous studies on promoters of taxol biosynthesis genes suggested that WRKY transcription factors play important regulatory roles in taxol biosynthesis38,39. Therefore, we conducted a systematic research on WRKY transcription factors of T. chinensis.

In the present study, 61 TcWRKYs were identified from the T. chinensis transcriptome datasets. Meanwhile, 75 TcWRKYs were identified in Arabidopsis, 109 in O. sativa, 83 in Pinus monticola, and 62 in Picea abies. Identifying all WRKYs from the transcriptome datasets is difficult because of the limited genome information of Taxus. For
instance, *P. abies*, whose genome has been sequenced, is also a conifer and considered the most relative species of *Taxus* spp. Therefore, 61 TcWRKYs should comprise almost the entire WRKY factor family of *Taxus* compared with the 62 WRKYs of *P. abies*. All of these results indicated that we have identified almost all WRKYs of *T. chinensis*, such that the phylogenetic and functional analyses were highly representative.

On the basis of our results, all WRKY proteins from *T. chinensis* and *A. thaliana* were cladded into eight subgroups (i.e., Group I, IIa-IIe, IIg, and III) by their WRKY domains. Group IIf and IIg are not widespread WRKY proteins, such that Group IIf does not exist in *Arabidopsis* and cotton, whereas Group IIg does not exist in physic nut and rice; however, these WRKYs contain an additional MAP/ERK kinase kinase domain or TIR-NBS-LRR (NBS-LRR short for nucleotide-binding site leucine-rich repeat) domain. Currently, some WRKY proteins were determined to contain three or four WRKY domains, whereas no such WRKYs were detected in *T. chinensis* in our work.

In general, WRKY proteins could be classified by a particular WRKY domain, which mainly consists of 60 highly conserved amino acid sequences and has a WRKYGQK motif in the N-terminus. Sometimes, the core sequence can be mutated into WRKYGKK, which is the most common variant in soybean, *S. lycopersicum*, *L. japonicus*, and *B. oleracea* var. *capitata* and has the highest frequency in Group IIc. In the present study, TcWRKY34, TcWRKY49, TcWRKY51, and TcWRKY54, all of which belong to Group IIc, contain the

Figure 5. Expression profiles of 61 TcWRKYs. Expression profiles of 61 TcWRKYs were analyzed according to three transcriptome datasets. NA was newly isolated *Taxus* cells while CA was 10-years long-term subcultured cells, the taxol content between the two samples was highly different (Zhang *et al*., 2015). MeJA (methyl jasmonate acid)-treated (Li *et al*., 2012) and GA (gibberellin)-treated (unpublished work) mean the cells were treated by MeJA for 15 h and GA for 6 h respectively. The heatmap was generated by Heml 1.0 software (http://hemi.biocuckoo.org/). The black boxes indicated that these TcWRKYs were not found in the dataset.
Figure 6. Expression patterns of selected TcWRKYs in response to MeJA and SA. Seven TcWRKYs, one from each subgroup, were selected to clarify the expression patterns in response to (a) MeJA (Methyl Jasmonate acid) and (b) SA (Salicylic acid). Expression levels of TcWRKY8 (Group IIC), TcWRKY20 (III), TcWRKY26 (I), TcWRKY41 (IIe), TcWRKY44 (IId), TcWRKY47 (IIa) and TcWRKY52 (IIb) were determined by qRT-PCR. Actin was used as reference gene, and each experiment was conducted three repeats.

Figure 7. Expression profiles of taxol-biosynthesis-related genes in six overexpression cell lines. Expression of DBAT and TSH genes, which encoded taxol biosynthesis enzymes, were certificated to be controlled by W-boxes in their promoters. Promoter of TcERF12 and TcERF15, which were regulators of taxol biosynthesis, also contained W-boxes. The expression levels of taxol-biosynthesis-related genes were determined by qRT-PCR. Actin was used as reference gene, and each experiment was conducted three repeats.
WRKYGKK motif. In tobacco, the WRKYGKK sequence could bind specifically to the WK-box (TTTCCAC), which is significantly different from the consensus sequence of W-box, indicating that the four TcWRKY proteins have similar functions in Taxus. In addition to the WRKYGKK sequence, other heptapeptide variants in the WRKY domains exist in many plants. For example, WRKYGEX, WRKYGKR, WRKYEDK, WRKYGQK, and WHQYGKL variants were detected in the WRKY domains of Glycine max var. Williams 82. According to our results and prior knowledge, no such variants have been identified in either Taxus or P. abies, indicating that these variants should be specific in angiosperm plants.

HARF, Leu-zipper, LxxLL, and EAR motifs were commonly detected in WRKY proteins, whereas TIR, LRR, PAH, and CBS domains were rarely identified. In Taxus, WRKY proteins contain HARF, Leu-zipper, LxxLL, and EAR motifs, indicating that these motifs of TcWRKYs serve similar functions to AtWRKYs. The HSF domain was also identified in TcWRKY proteins, and the influence of these domains needs to be further clarified. Moreover, verifying the details of the molecular structural information of TcWRKY proteins are difficult because of the limited genome information of Taxus spp.

WRKY proteins from different subgroups play either positive or negative roles; some of them even play dual roles in regulating downstream gene expression levels. Thus, six TcWRKYs, i.e., TcWRKY8 (Group IIc), TcWRKY20 (Group III), TcWRKY26 (Group I), TcWRKY44 (Group II), TcWRKY47 (Group Ia), and TcWRKY52 (Group Ib), were selected to verify their own functions on taxol biosynthesis. Then, these six TcWRKYs were overexpressed in T. chinensis cells, and DBAT, TSH, and TcERF15, and TcERF12, all of which are taxol-biosynthesis-related genes, were analyzed by qRT-PCR.

TcWRKY44 (Group IIa) and TcWRKY52 (Group Ib) suppressed the expression of the four genes, indicating that they are putative negative regulators of taxol biosynthesis. To our knowledge, AtWRKY7/11/17, which are all Group IIb WRKYs, serve as negative defense regulators. Meanwhile, TcWRKY6 (Group Ib) could play dual roles, i.e., a negative regulator in low Pi stress and a positive regulator in low B stress. TcWRKY52 contains an EAR motif, resulting in its negative roles in regulating taxol biosynthesis.

By contrast, TcWRKY26 (Group I) and TcWRKY20 (Group III) improve the expression of four taxol-biosynthesis-related genes. Actually, Group III WRKYs always play positive roles in regulating the biosynthesis of secondary metabolites. For instance, CrWRKY1 and AaWRKY1, both of which belong to Group III, could improve the biosynthesis of vinblastine in Catharanthus roseus and artemisinin in A. annua, respectively. Moreover, AtWRKY25, AtWRKY26, and AtWRKY33, which are Group I WRKys in Arabidopsis, could upregulate the expression of HsfA2, HsfB1, Hsp101, and MFR1. In Capsicum annuum, CaWRKY58 acts as a transcriptional activator of negative regulators in the resistance of pepper to Ralstonia solanacearum infection. All of these results indicate that TcWRKY26 and TcWRKY20 are positive regulators in T. chinensis.

Our results showed that TcWRKY8 (Group IIc) and TcWRKY47 (Group Ia) upregulate the expression of the DBAT, TSH, and TcERF15 genes but downregulate the expression of TcERF12. TcERF12 is a negative regulator of the TASY gene, which encodes one of the key taxol biosynthesis enzymes. TcWRKY8 is a Group Ic WRKY protein with the typical WRKYGQK motif, whereas TcWRKY47 is a Group Ia WRKY protein. WRKY proteins from these two subtypes are reported as positive or negative regulators. For instance, GaWRKY1 and OsWRKY62 are Group IIa WRKys, but GaWRKY1 positively regulates the sesquiterpene synthase gene (+)-β-cadinene synthase- A in cotton, whereas OsWRKY62 negatively regulates the basal and Xa21-mediated defense against Xanthomonas oryzae. In summary, the 61 WRKY proteins from the T. chinensis transcriptome datasets were highly representative and adequate for further research. Phylogenetic analysis of the WRKY domains showed that the TcWRKYs of T. chinensis could be divided into Group I, Ia–IIe, and III as well as those of Arabidopsis, indicating that the WRKY transcription factors exhibit species divergence. Further overexpression of TcWRKY8/20/26/44/47/52 indicated the diverse functions of TcWRKYs in Taxus and identified candidate regulators of taxol biosynthesis.

Data availability. All the protein sequences of TcWRKYs were included in Supplementary file 2. The AtWRKYs were obtained from TAIR (http://www.arabidopsis.org/browse/genefamily/WRKY-Som.jsp).

References
1. Schiff, P. B., Fant, J. & Horwitz, S. B. Promotion of microtubule assembly in vitro by taxol. Nature 277, 665–667 (1979).
2. Jennenwein, S. & Croteau, S. A.: Taxol: biosynthesis, molecular genetics, and biotechnological applications. Appl Microbiol Biotechnol 57, 13–19 (2001).
3. Croteau, R., Ketchum, R. E. B., Long, R. M. & Kaspera, R. & Wildung, M. R. Taxol Biosynthesis and Molecular Genetics. Phytochemistry Reviews 5, 75–97, https://doi.org/10.1007/s11101-005-3748-2 (2006).
4. Tang, P. & Wang, F. P. Recent Progresses in the Synthesis of Taxol. Phytochemistry 57, 458–468, https://doi.org/10.1016/S0031-9422(00)00478-9 (2000).
5. Ciolkowski, I., Wanke, D., Birkenbihl, R. P. & Somssich, I. E. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. Plant Molecular Biology 68, 81–92, https://doi.org/10.1007/s11103-008-9335-1 (2008).
6. Li, S., Zhang, P., Zhang, M., Fu, C. & Yu, L. Functional analysis of a WRKY transcription factor involved in transcriptional activation of the DBAT gene in Taxus chinensis. Plant science 195, 19–26, https://doi.org/10.1111/j.1365-311X.2008.01261.x (2013).
7. Fulgen, T., Rushon, P. J., Robatzeek, S. & Somssich, I. E. The WRKY superfamily of plant transcription factors. Trends Plant Sci 5, 199–206 (2000).
8. Wu, K. L., Guo, Z. I., Wang, H. H. & Li, J. The WRKY family of transcription factors in rice and Arabidopsis and their origins. DNA evolution 12, 9–26, https://doi.org/10.1093/dnares/12.1.9 (2005).
9. Zhang, Y. & Wang, L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. BMC evolutionary biology 5, 1, https://doi.org/10.1186/1471-2148-5-1 (2005).
10. Rushon, P. J. et al. Tobacco transcription factors: Novel insights into transcriptional regulation in the Solanaceae. Plant Physiology 147, 280–295, https://doi.org/10.1104/pp.107.114041 (2008).
20. Li, S. T.
21. Zhang, M.
23. Zhang, M.
24. Mangelsen, E.
25. Zhou, L.
26. Rinerson, C. I., Rabara, R. C., Tripathi, P. J., Shen, Q. X. J. & Rushton, P. J. The evolution of WRKY transcription factors. 
27. Rinerson, C. I., Rabara, R. C., Tripathi, P. J., Shen, Q. X. J. & Rushton, P. J. The evolution of WRKY transcription factors. 
28. Wang, N. N. et al. Genome-wide identification of WRKY family genes and their response to cold stress in Vitis vinifera. Bmc Plant Biology 14, doi:10.1186/s12870-012-0015-2 (2014).
29. Zhou, L. et al. Molecular characterization of 26 cotton WRKY genes that are expressed differentially in tissues and are induced in seedlings under high salinity and osmotic stress. Plant Cell Tissue And Organ Culture 119, 141–156, https://doi.org/10.1007/s12229-012-0155-y (2014).
30. Song, H. et al. Global Analysis of WRKY Genes and Their Response to Dehydration and Salt Stress in Soybean. Front Plant Sci 7, Arts 910.3389/Fpls.2016.00009 (2016).
31. Park, C. Y. et al. WRKY group IId transcription factors interact with calmodulin. FEBS Lett 579, 1545–1550, https://doi.org/10.1016/j.febslet.2005.01.057 (2005).
32. Ayadi, M. et al. The WRKY Transcription Factor Family in Citrus: Valuable and Useful Candidate Genes for Citrus Breeding. Applied biochemistry and biotechnology 180, 516–543, https://doi.org/10.1007/s12010-012-0456-y (2015).
33. Huang, S. et al. Genome-wide analysis of WRKY transcription factors in Solanum lycopersicum. Molecular genetics and genomics: MGG 287, 495–513, https://doi.org/10.1007/s00438-012-0969-6 (2012).
34. Song, H., Wang, P., Nan, Z. & Wang, X. The WRKY Transcription Factor Genes in Lotus japonicus. International journal of genomics 2014, 420128, https://doi.org/10.1155/2014/420128 (2014).
35. Yao, Q. X., Xie, E. H., Liu, F. H. & Gao, L. Z. Genome-wide identification and comparative expression analysis reveal a rapid expansion and functional divergence of duplicated genes in the WRKY gene family of cabbage, Brassica oleracea var. capitata. Gene 557, 35–42, https://doi.org/10.1016/j.gene.2014.12.005 (2015).
36. Song, H. et al. Genome-wide analysis of WRKY transcription factors in cotton. BMC genomics 15, 116, https://doi.org/10.1186/s12870-016-0806-4 (2016).
37. Ayadi, M. et al. Genome and transcriptome analysis of the grapevine (Vitis vinifera L.) WRKY gene family. Horticulture research 3, 116, https://doi.org/10.1038/s41438-016-0002-7 (2016).
38. Plevin, R. et al. The LxxLLL motif: a multifunctional binding sequence in transcriptional regulation. Trends in biochemical sciences 30, 66–69, https://doi.org/10.1016/j.tibs.2004.12.001 (2005).
39. Robatzek, S. & Somssich, I. E. A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. Plant Journal 28, 123–133, https://doi.org/10.1046/j.1365-313X.2001.01131.x (2001).
40. Hiratsu, K., Matsu, K., Koyama, T. & Ohme-Takagi, M. Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. FEBS Journal 273, 733–739, https://doi.org/10.1111/j.1742-4658.2006.05455.x (2006).
41. Xiong, W. et al. Genome-wide analysis of the WRKY gene family in physic nut (Jatropha curcas L.). Gene 254, 128–132, https://doi.org/10.1016/j.plante.2013.04.047 (2013).
42. Xie, Z. et al. Annotations and functional analysis of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiol 137, 176–189, https://doi.org/10.1104/pp.109.151704 (2010).
43. Oh, S. K. et al. Capsicum annuum WRKY protein CaWRKY1 is a negative regulator of pathogen defense. New Phyto 177, 977–989, https://doi.org/10.1111/j.1469-8137.2007.02310.x (2008).
44. van Verk, M. C., Pappaioannou, D., Neeleman, L., Bol, J. F. & Linthorst, H. J. M. A novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. Plant Physiology 146, 1983–1995, https://doi.org/10.1104/pp.107.112789 (2008).
45. Journot Catalino, N., Somssich, I. E., Roby, D. & Kroj, T. The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in Arabidopsis thaliana. The Plant cell 18, 3289–3302, https://doi.org/10.1105/tpc.2010.020606 (2010).
46. Eulgem, T. & Somssich, I. E. Networks of WRKY transcription factors in defense signaling. Current opinion in plant biology 10, 366–371, https://doi.org/10.1016/j.pbi.2007.04.020 (2007).

References
11. Mohanta, T. K., Park, Y. H. & Bae, H. Novel Genomic and Evolutionary Insight of WRKY Transcription Factors in Plant Lineage. Scientific reports 6, 37309, https://doi.org/10.1038/srep37309 (2016).
12. Banerjee, A. & Roychoudhury, A. WRKY proteins: signaling and regulation of expression during abiotic stress responses. TheScientificWorldJournal 2015, 8073560, https://doi.org/10.1155/2015/807560 (2015).
13. Rushton, P. J., Somssich, I. E., Ringer, P. S. & Shen, Q. J. WRKY transcription factors. Trends in Plant Science 15, 247–258, https://doi. org/10.1016/j.tplants.2010.02.006 (2010).
14. Vom Endt, D., Kijne, J. W. & Memelink, J. Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry 61, 107–114, https://doi.org/10.1007/s10303-002-00185-1 (2002).
15. Bakshi, M. & Oelmuller, R. WRKY transcription factors: Jack of many trades in plants. Plant Signal Behav 9, et7700 (2014).
16. Xu, Y. H., Wang, J. W., Wang, S. W., Wang, J. Y. & Chen, X. Y. Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-delta-cadinene synthase-A. Plant Physiology 135, 507–515, https://doi.org/10.1104/pp.104.038612 (2004).
17. Ma, D. M. et al. Isolation and Characterization of AaWRKY1, an Artemisia annua Transcript Factor that regulates the Amorphophallus ceboi diene Synthase Gene, a Key Gene of Artemisinin Biosynthesis. Plant And Cell Physiology 50, 2146–2161, https://doi.org/10.1093/pcp/pcp149 (2009).
18. He, G.-H. et al. Drought-responsive WRKY transcription factor gene TaWRKY63 and TaWRKY33 from wheat confer drought and/or heat resistance in. Arabidopsis. BMC Plant Biology 16, 116, https://doi.org/10.1186/s12870-016-0806-4 (2016).
19. Chen, J. et al. Arabidopsis WRKY46, WRKY54, and WRKY70 Transcription Factors Are Involved in Brassinosteroid-Regulated Plant Growth and Drought Responses. The Plant cell 29, 1425–1439, https://doi.org/10.1105/tpc.17.00364 (2017).
20. Li, S. T. et al. Transcriptional profile of Taxus chinensis cells in response to methyl jasmonate. BMC genomics 13, 295, https://doi.org/10.1186/s12866-014-2355-z (2012).
21. Zhang, M. et al. High-throughput sequencing reveals mRNA effects on the primary and secondary production properties in long-term subcultured Taxus cells. Front Plant Sci 6, 604, https://doi.org/10.3389/fpls.2015.00604 (2015).
47. Kasajima, I. & Fujiwara, T. Micorrarray analysis of B nutrient response: Identification of several high-B inducible genes and roles of WRKY6 in low-B response. *Plant And Cell Physiology* **48**, S46–S46 (2007).
48. Chen, Y. F. et al. The WRKY6 transcription factor modulates PHOSPHATE1 expression in response to low Pi stress in *Arabidopsis*. *The Plant cell* **21**, 3554–3566, https://doi.org/10.1105/tpc.108.064980 (2009).
49. Chen, L. et al. The role of WRKY transcription factors in plant abiotic stresses. *Biochim Biophys Acta* **1819**, 120–128, https://doi.org/10.1016/j.bbabm.2011.09.002 (2012).
50. Suttipanta, N. et al. The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol* **157**, 2081–2093, https://doi.org/10.1104/pp.111.181834 (2011).
51. Li, S., Fu, Q., Chen, L., Huang, W. & Yu, D. Arabidopsis thaliana WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. *Planta* **233**, 1237–1252, https://doi.org/10.1007/s00425-011-1375-2 (2011).
52. Wang, Y. et al. CaWRKY58, encoding a group I WRKY transcription factor of *Capsicum annuum*, negatively regulates resistance to *Ralstonia solanacearum* infection. *Mol Plant Pathol* **14**, 131–144, https://doi.org/10.1111/j.1364-3703.2012.00836.x (2013).
53. Peng, Y. et al. OsWRKY62 is a negative regulator of basal and *Xa21*-mediated defense against *Xanthomonas oryzae* pv. *oryzae* in rice. *Mol Plant* **1**, 446–458, https://doi.org/10.1093/mp/ssn024 (2008).

**Acknowledgements**
This work was supported by the National Natural Science Foundation of China (31270342 and 31670293). And the authors thank the Analytical and Testing Center in Huazhong University of Science & Technology for Real-time PCR analysis.

**Author Contributions**
Yu M.Z., C.F. and L.Y. conceived and designed research. C.Y., L.N., X.J., S.Z. and L.W. conducted experiments. M.Z. and C.Y. analyzed data and wrote the manuscript. All authors read and approved the manuscript.

**Additional Information**
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-23558-1.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018