

RESEARCH ARTICLE

Open Access



New different origins and evolutionary processes of AP2/EREBP transcription factors in *Taxus chinensis*

Meng Zhang^{1,2}, Ying Chen^{1,2}, Xiaofei Jin^{1,2}, Yuxin Cai^{1,2}, Yuanyuan Yuan^{1,2}, Chunhua Fu^{1,2*} and Longjiang Yu^{1,2}

Abstract

Background: *Taxus* spp. produces the anticancer drug, taxol, and hence is planted as an industrial crop in China. APETALA2/ethylene response element binding proteins (AP2/EREBPs) are the key regulators of plant development, growth, and stress responses. Several homologues control taxol biosynthesis. Identifying the AP2/EREBP proteins from *Taxus* is important to increase breeding and production and clarify their evolutionary processes.

Results: Among the 90 genes from multi *Taxus chinensis* transcriptome datasets, 81 encoded full-length AP2-containing proteins. A domain structure highly similar to that of angiosperm AP2/EREBPs was found in 2 AP2, 2 ANT, 1 RAV, 28 dehydration-responsive element-binding proteins, and 47 ethylene-responsive factors contained, indicating that they have extremely conservative evolution processes. A new subgroup protein, TcA3Bz1, contains three conserved AP2 domains and, a new domain structure of AP2/EREBPs that is different from that of known proteins. The new subtype AP2 proteins were also present in several gymnosperms (*Ginkgo biloba*) and bryophytes (*Marchantia polymorpha*). However, no homologue was found in *Selaginella moellendorffii*, indicating unknown evolutionary processes accompanying this plant's evolution. Moreover, the structures of the new subgroup AP2/EREBPs have different conserved domains, such as B3, zf-C3Hc3H, and agent domains, indicating their divergent evolution in bryophytes and gymnosperms. Interestingly, three repeats of AP2 domains have separately evolved from mosses to gymnosperms for most of the new proteins, but the AP2 domain of Gb_11937 has been replicated.

Conclusion: The new subtype AP2/EREBPs have different origins and would enrich our knowledge of the molecular structure, origin, and evolutionary processes of AP2/EREBP transcription factors in plants.

Keywords: *Taxus chinensis*, APETALA2/ethylene response element binding protein transcription factors, TcA3Bz1, Evolutionary processes

Background

Taxus spp. is an ancient gymnosperm that could produce a secondary metabolite, taxol (generic Paclitaxel), the most used clinical anticancer drugs [1–3]. *Taxus* spp. is currently the main source of taxol and its precursors and thus is widely planted as an industrial crop in many provinces of China. However, the taxol content in *Taxus* spp. is extremely low, and the biosynthesis

pathway is highly complex and requires 19–20 enzymatic steps, leading to a taxol supply shortage.

Regulating secondary metabolite biosynthesis is a promising way to guide breeding, genic manipulation, and planting of industrial crops [4]. Our previous work identified two AP2/EREBPs proteins, TcERF12 and TcERF15, that could positively and negatively regulate the expression of *TASY* gene, which encodes the first committed step enzyme in the taxol biosynthesis pathway [5]. Therefore, identifying AP2/EREBP transcription factors in *Taxus* spp. would facilitate the screening of candidate regulators for taxol biosynthesis.

AP2/EREBP transcription factors play various roles in plants. AP2/EREBP superfamily proteins are divided into six subgroups, namely, APETELA2 (AP2), AINTEGUMENTA

* Correspondence: fuch2003@126.com

¹Department of Biotechnology, Institute of Resource Biology and Biotechnology, College of Life Science and Technology, Huazhong University of Science and Technology, No.1037 Luoyu Road, Wuhan 430074, People's Republic of China

²Key Laboratory of Molecular Biophysics Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, No.1037 Luoyu Road, Wuhan 430074, People's Republic of China



(ANT), related to ABI3/VP1 (RAV), dehydration-responsive element-binding protein (DREB), ethylene-responsive factor (ERF), and soloist in most plants [6, 7]. AP2 and ANT proteins mainly function as key developmental regulators in reproductive and vegetative organs and lateral organ development [8, 9]. As negative regulators, RAV proteins mediate plant defense during abiotic and biotic stress [10–12]. DREBs usually function in cold-, drought- and heat-stress responses, and ERFs are often linked to biotic stress responses, such as pathogen attack and methyl jasmonate and ethylene application [13, 14]. Thus, the identification of AP2/EREBP transcription factors is meaningful for further cultural practices.

AP2/EREBP is a highly conserved superfamily in angiosperms, but its transcription factors in gymnosperm plants have not been systematically studied. Gymnosperms are more ancient but as important as angiosperms and often provide new understandings on the evolutionary processes of genes [15]. Thus, *Taxus* spp. would be a valuable material to study the evolution processes of AP2/EREBP proteins in gymnosperms.

We integrated multi transcriptome datasets of *T. chinensis* related to taxol biosynthesis for the identification of AP2/EREBP transcription factors [2, 15–17]. More than 100 Gb sequenced data were used to ensure that most AP2/EREBP proteins were obtained by hidden Markov model (HMM) search. Only the genes encoded with at least a complete AP2 domain were employed for further analysis. AP2/EREBP proteins were also obtained from angiosperms (*Arabidopsis thaliana*), gymnosperms (*Picea abies*, *Ginkgo biloba*, *Pinus tadae*, and *Gnetum montanum*), and bryophytes (*Physcomitrella patens* and *Selaginella moellendorffii*) to clarify the evolutionary differences. Finally, the expression patterns of these TcAP2/EREBP proteins were analyzed and clustered with taxol biosynthesis genes. Results showed that the evolution of AP2/EREBP transcription factors is highly divergent among higher plants and provided interesting findings on the evolution of AP2/EREBP proteins.

Methods

Plant materials and transcriptome datasets

The *T. chinensis* cells were induced from 2-years-old Chinese yew (*Taxus chinensis* (Pilger) Rehd.) at May 2005 and subcultures with 62# medium [18]. A series of transcriptome datasets was sequenced by BGI (Shenzhen, China); some of these data were previously reported, such as NA/CA [16] and MJ treatment [17]. Two unpublished datasets with an individual size of 36 Gb and the re-assembled transcriptome dataset from public reports were also used [15, 19].

Identification of AP2/EREBP proteins in *T. chinensis*

The transcriptome data in the present study originated from the database of our previous report that was expanded by adding several new datasets such as miR5298OE and WRKY47OE (two genes, miR5298 and WRKY47 involve in taxol biosynthesis were over-expressed in *Taxus* cells and their transcriptome data were high-throughput sequenced; unpublished results) [2, 16, 17, 19]. A large dataset is helpful to obtain numerous full-length genes.

The AP2/EREBP proteins of *T. chinensis* were identified by HMM search in the HMM v3.1b2 (<http://hmmer.org/>) by `hmmsearch` in the HMM v3.1b2 (<http://pfam.xfam.org/>) using the Hidden Markov Model (HMM) of AP2 domain (PF00847) with a cut-off score of $1e-5$ [20]. All nucleotide sequences obtained from various transcriptome datasets were further reassembled by CAP3 (<http://doua.prabi.fr/software/cap3>) [21], and redundancy was removed by CD-HIT-est with sequence identity cut-off of 0.98 (<http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi?cmd=cd-hit-est>) [22]. The sequences that could not be extended were aligned with the NR database to identify their conserved domains (AP2 domain, B3 domain, and others) by using online Blastx with default parameters (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). ORF finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) was employed to obtain the full-length deduced protein sequences. For the genes without an intact ORF, the aligned sequences that contained AP2 domains were utilized for further bioinformatics analysis. All nucleotide and amino acid sequences could be downloaded as supplementary files.

Obtaining AP2/EREBP proteins from *A. thaliana*, *P. patens* and *S. moellendorffii*, *P. abies*, and *G. biloba*

Arabidopsis AP2/EREBP proteins were searched from the plnTFdb database (<http://plntfdb.bio.uni-potsdam.de/v3.0/>) [23]. The protein databases of *P. patens* and *S. moellendorffii* were downloaded from JGI (<https://phytozome.jgi.doe.gov/pz/portal.html#>). The protein sequences of *P. abies* and *P. tadae* were obtained from ConGenIE (<http://congenie.org/>), those of *G. biloba* were downloaded from Giga (<http://gigadb.org/dataset/100209>), and those of *G. montanum* were acquired from DRYAD (<https://datadryad.org/resource/doi:10.5061/dryad.0vm37.2>). HMM search was used to screen AP2/EREBP proteins that contain an AP2 domain with a cut-off score of $1e-5$. The HMM file (PF02362, <http://pfam.xfam.org/family/B3>) was used to identify RAV proteins containing a B3 domain in addition to the AP2 domain from AP2/EREBP proteins.

Sequence alignment and phylogenetic and motif analysis of conserved AP2 domains

The proteins were aligned with MAFFT v7.312 (<http://mafft.cbrc.jp/alignment/software/>) [24] and online ClustalW

with default parameters and colored by EsPript 3.0 (<http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>) [25]. MEGA 5.0 was used to construct a neighbor-joining tree with these sequences of AP2 domain based on the JTT model after 1000 bootstrap resampling [26]. The tree was validated by phyML 3.1 (<http://www.atgc-montpellier.fr/phyml>) [27]. When the ML and NJ trees were similar, they were used as the phylogenetic tree and colored by Fig-Tree V1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). The specific logo of AP2 domain from various plants was generated using MEME Suite V5.0.2 webtools in the classic motif discovery mode (<http://meme-suite.org/tools/meme>) [28].

Phylogenetic analysis of full-length ERF transcription factors

ERF transcription factors are AP2/EREBP proteins that contain only an AP2 domain and constitute the biggest subfamily of the AP2/EREBP family. The full-length ERF proteins of *T. chinensis* and other plants were phylogenetic analyzed to clarify the differences of ERFs in *Taxus* and other plants. MAFFT was used for alignment, and MEGA 5.0 was used to construct the tree based on the JTT model with 1000 bootstrap resampling. The tree was validated by phyML 3.1 (<http://www.atgc-montpellier.fr/phyml>) [27]. When the ML and NJ trees were similar, they were used as the phylogenetic tree and colored by FigTree V1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). B3 ERFs containing EDLL motif were aligned using ClustalW with protein weight matrix “Identity”. phyML and FigTree were used to validate and color the tree.

Motif analysis of ERF transcription factors

The full-length amino acid sequences of *T. chinensis* AP2/EREBP transcription factors, which have intact ORFs and other AP2/EREBP proteins, were analyzed by MEME software with default parameters to identify new and specific motifs in various plants. Domain and motif patterns were illustrated by TBtools (<http://cj-chen.github.io/tbtools>) [29] and WebLogo (<http://weblogo.berkeley.edu/logo.cgi>).

Expression patterns of *T. chinensis* AP2/EREBP transcription factors

Taxus spp. is famous for its anticancer metabolite, taxol. Two ERF factors, namely, TcERF12 and TcERF15, regulate taxol biosynthesis. Thus, the known taxol biosynthesis genes were added into the heatmap to clarify the potential regulating roles of AP2/EREBP factors on taxol biosynthesis. Five transcriptome data were obtained from the samples with remarkably different taxol contents. For instance, NA comprised newly induced and accumulating *Taxus* cells, whereas CA comprised long-term subculture cells [16]. Gene expression was

calculated by FPKM (NA/CA, GA-treated), TPM (miR5298bOE and WRKY47OE); two genes, miR5298 and WRKY47 involve in taxol biosynthesis were over-expressed in *Taxus* cells and their transcriptome data were high-throughput sequenced), and RPKM (MeJA-treated). Values of log₂ (Expression ratio) were used to generate and cluster the heatmap using Morpheus with default parameters (<https://software.broadinstitute.org/morpheus>). Co-expression coefficient values were calculated using the CORREL formula in Excel.

Results

Ninety AP2/EREBP genes were identified in *T. chinensis* transcriptome datasets

Although many unigenes were annotated as AP2/EREBP transcription factors according to the Blast results, only the genes that could encode a full-length AP2 domain were considered. After redundancy removal, sequence assembly, and elongation, 93 genes with a length ranging from 206 bp to 6951 bp could encode at least one full-length AP2 domain (Additional files 1 and 4). In recent studies on AP2/EREBP, nearly 90 genes were found using only transcriptome data [30, 31]. A few *Taxus* AP2/EREBPs were not found, indicating that the expression of several AP2/EREBPs is strictly regulated.

AP2-, ANT-, and RAV-type AP2/EREBP proteins were minorities in *T. chinensis*

Among the 90 genes, 81 unigenes could encode full-length AP2/EREBP proteins ranging from 146 aa to 1955 aa (Additional file 1).

Three unigenes named as TcAP2a-c could encode unique proteins with two AP2 domains (Fig. 1a). The R1 of TcAP2c is similar to that of ANT subgroups called TcANT2c (Fig. 1a) [9]. In addition, *TcRAVI* encodes a RAV protein with an extra B3 domain, and 76 proteins contain only one AP2 domain. TcERF41, also named as TcWRI1, is highly conserved with WRI1 proteins and contains only one AP2 domain but has a divergent C-terminal AP2 domain-like motif, which lacks (Y/H/W)LG- and RAYD-motif (Fig. 1b) [32]. AtWRI1/3/4 has two conserved AP2 domains, but the homologues have a highly varied C-terminal AP2 domain (R2) in *Larix gmelinii var. olgensis*, *Larix kaempferi*, and *Persea americana* (Fig. 1b) [29]. However, the R2 of WRI homologues is closely related to that of ANTs, indicating that WRI homologues have special evolutionary processes in plants [33] (Fig. 1c).

New AP2/EREBP with three AP2 domains found in several gymnosperms and bryophytes

T. chinensis has a unigene designated as TcA3Bz1 that encodes a protein containing three AP2, one B3, and one zf-C3Hc3H domains (pfam13891). Different from

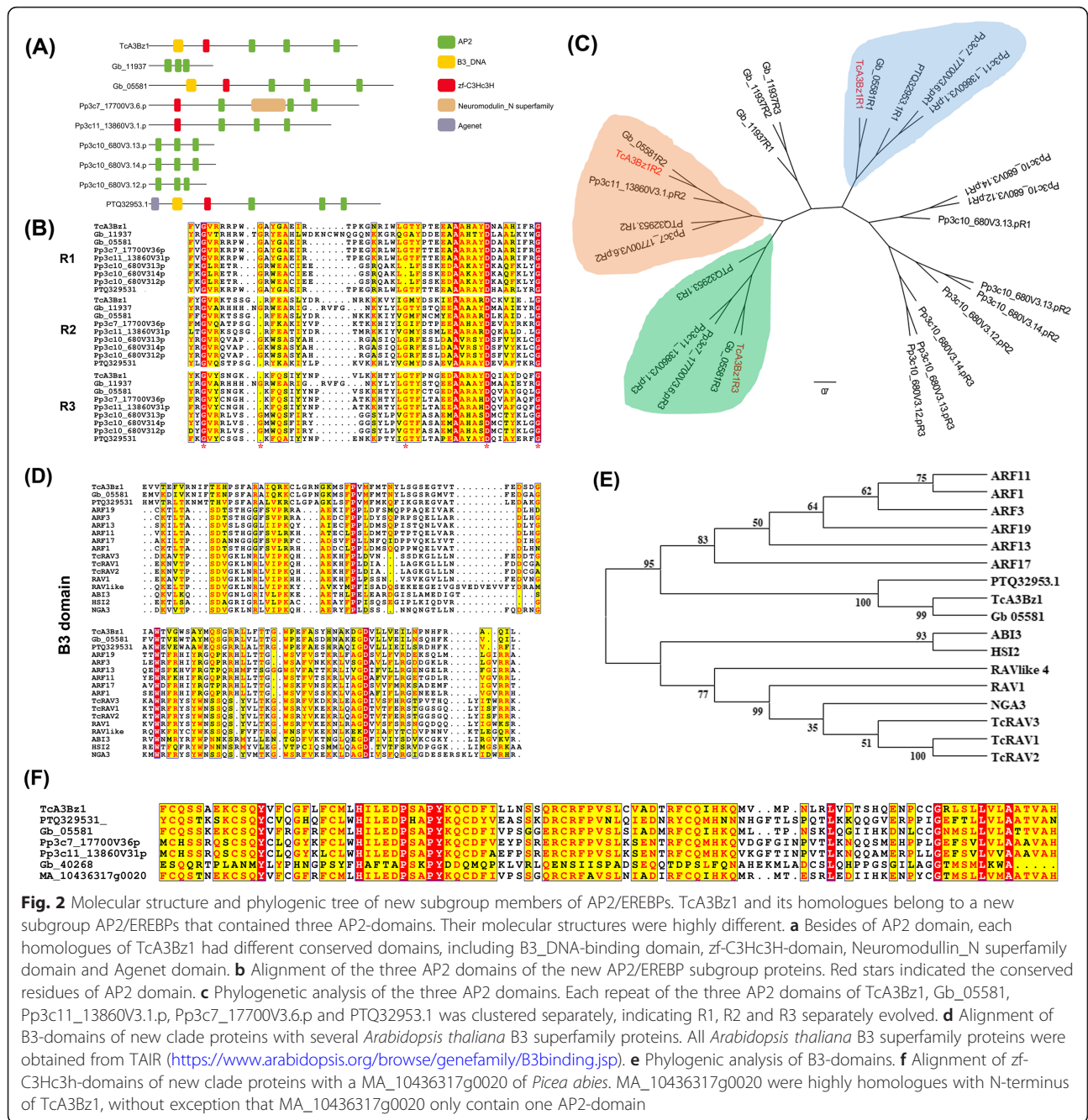


Fig. 2 Molecular structure and phylogenetic tree of new subgroup members of AP2/EREBPs. TcA3Bz1 and its homologues belong to a new subgroup AP2/EREBPs that contained three AP2-domains. Their molecular structures were highly different. **a** Besides of AP2 domain, each homologue of TcA3Bz1 had different conserved domains, including B3_DNA-binding domain, zf-C3Hc3H-domain, Neuromodulin_N superfamily domain and Agenet domain. **b** Alignment of the three AP2 domains of the new AP2/EREBP subgroup proteins. Red stars indicated the conserved residues of AP2 domain. **c** Phylogenetic analysis of the three AP2 domains. Each repeat of the three AP2 domains of TcA3Bz1, Gb_05581, Pp3c11_13860V3.1.p, Pp3c7_17700V3.6.p and PTQ32953.1 was clustered separately, indicating R1, R2 and R3 separately evolved. **d** Alignment of B3-domains of new clade proteins with several *Arabidopsis thaliana* B3 superfamily proteins. All *Arabidopsis thaliana* B3 superfamily proteins were obtained from TAIR (<https://www.arabidopsis.org/browse/genefamily/B3binding.jsp>). **e** Phylogenetic analysis of B3-domains. **f** Alignment of zf-C3Hc3H-domains of new clade proteins with a MA_10436317g0020 of *Picea abies*. MA_10436317g0020 were highly homologues with N-terminus of TcA3Bz1, without exception that MA_10436317g0020 only contain one AP2-domain

Pc3Pv3.12-14 (Fig. 2c). The nucleotide sequences of the R2 and R3 repeat of Gb11937 are completely identical, indicating a duplication that leads to three AP2 domains containing a protein (Fig. 2b, Additional file 2).

B3 and zf-C3Hc3H domains are highly conserved in these new clade proteins but not homologous to any reported proteins according to the search within NR database (version 2019/6/30) (E-value >1e-5) (Fig. 2d and f). B3 domain can also be found in RAV proteins but are different from those of the new clade proteins

(Fig. 2d). B3-containing proteins are divided into four classes, LAV (Leafy cotyledon2 [LEC2]-abscisic acid insensitive3 [ABI3]-VAL), RAV, ARF (Auxin response factor), and REM (Reproductive meristem), leading to three other classes. The B3 of new clade proteins with three AP2 domains shares low similarities with these four B3 repeats (Fig. 2d). However, phylogenetic analysis suggested that the B3 of new clade proteins and ARFs

have a common ancestor though they share low similarities, suggesting that TcA3Bz1 and its homologues could bind similar cis-elements with ARF (Fig. 2e).

All these results showed that AP2/EREBP transcription factors have complicated evolutionary processes. To the best of our knowledge, this study first found the AP2/EREBP proteins with three or more AP2 domains and extra domains such as the zf-C3Hc3H domain, which was first identified in AP2/EREBP proteins. These findings could be interesting study points on the functions of AP2/EREBP transcription factors.

AP2 domains of DREB and ERF proteins have evolved conservatively in *T. chinensis*

The EREBP subfamily proteins that contain only one AP2 domain in *A. thaliana* could be divided into 13 clades, namely, two subgroups, DREBs (A1–A6), ERFs (B1–B6), and soloists [7]. Many plants only have 1–2 soloists [11, 35–37]. According to our results, no soloist ERFs, which have a conserved HLG-motif in AP2 domain, exist in *T. chinensis* and *P. abies*. However, this protein is highly conserved in *G. biloba*. Hence, the evolutionary process of soloist ERFs in gymnosperms should be further clarified.

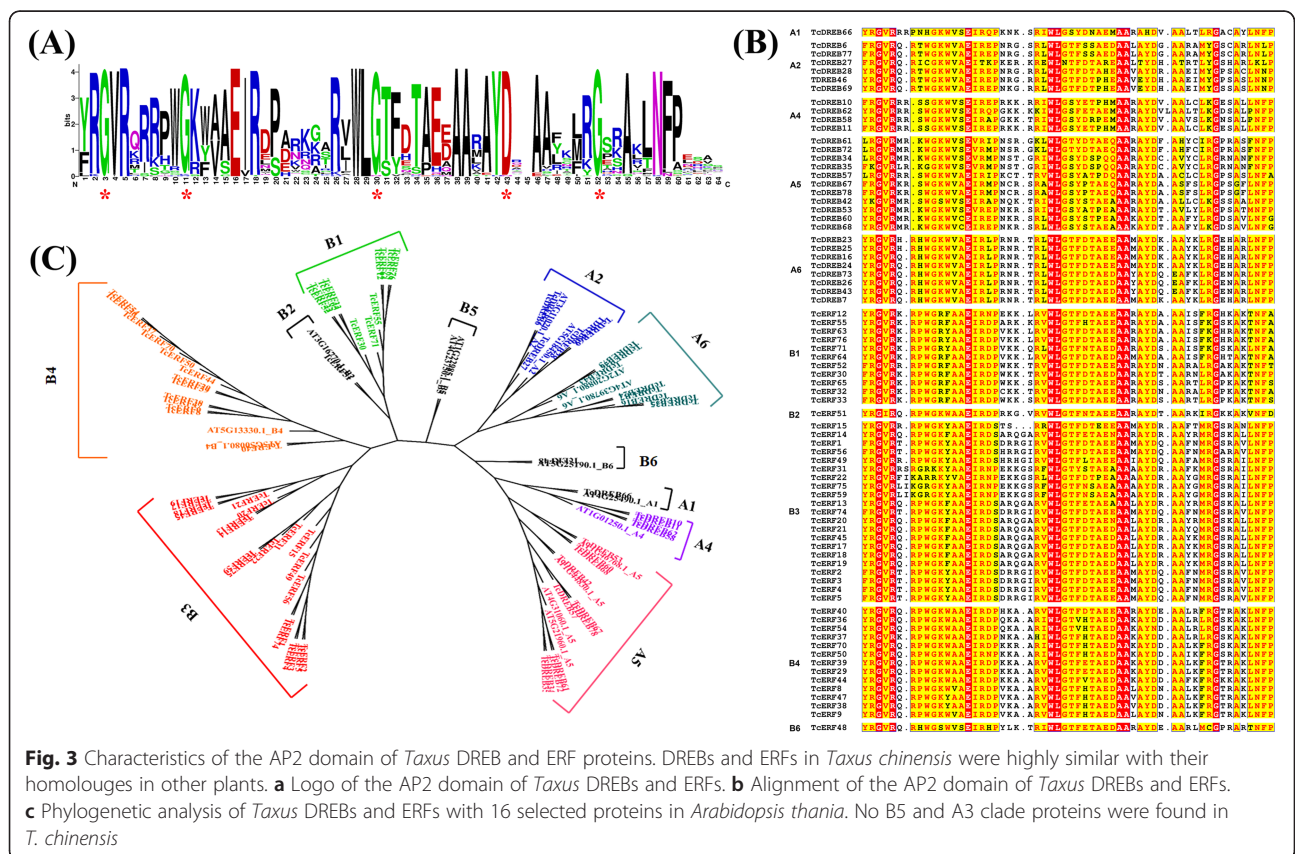
The AP2 domain of 76 full-length EREBP proteins of *T. chinensis* was phylogenetically analyzed with 16

representative homologues in *A. thaliana* (Fig. 3c). According to our results, the EREBP subfamily proteins of *T. chinensis* could be divided into 10 clades, and no A3 and B5 members exist in *T. chinensis* [30, 38]. Only one protein belongs to B2, B6, and A1 clades. The B3 clade in *T. chinensis* has 21 ERFs, whereas B4, B1, and A5 groups contain 13, 11, and 10 ERFs, respectively (Fig. 3b). In many plants, B3 is the dominant clade with most members in the ERF subfamily [31, 39–41].

Conserved motifs in ERF and DREB proteins in *T. chinensis*

Almost all *Taxus* ERFs and DREBs contain the WLG-motif in their AP2 domain, except for TcERF27 that converts it into WLN (Fig. 3b). WLG-motif is 100% identical in all DREBs and ERFs, even in RAVs, but is converted to YLG in AP2 subfamily proteins [31, 42]. The substitution of WLN-motif might infer an interesting biological importance.

All B1 TcERF proteins have a (L/F)DLNL/(F/X)P-type EAR-motif in their C-terminal, which is commonly present in B1 ERFs [43] and could be considered as an enhanced EAR-motif combination of LxLxLx- and DLNxxP-type that allows physical interaction by TOPLESS co-repressor (Fig. 4a and c) [44–46]. Moreover, a DCDSSS-motif is highly conserved and present at the N-



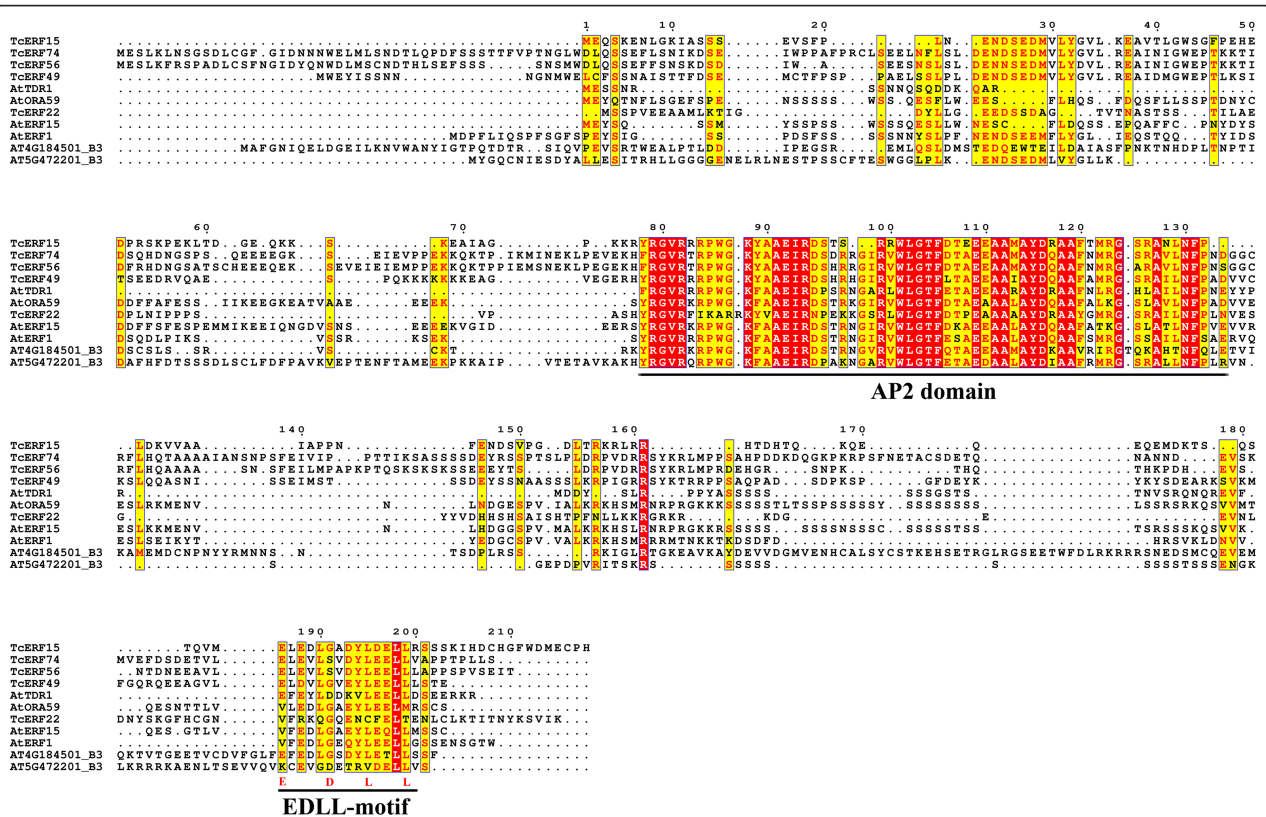


Fig. 5 Domain structure of several B3 ERFs that containing EDLL-motif. Sequence alignment of several B3 ERFs that contained EDLL-motif in *Taxus* and *Arabidopsis*. EDLL-motif was not a conserved motif, even the four signature residues were not conserved. Moreover, many B3 ERFs had no EDLL-motif

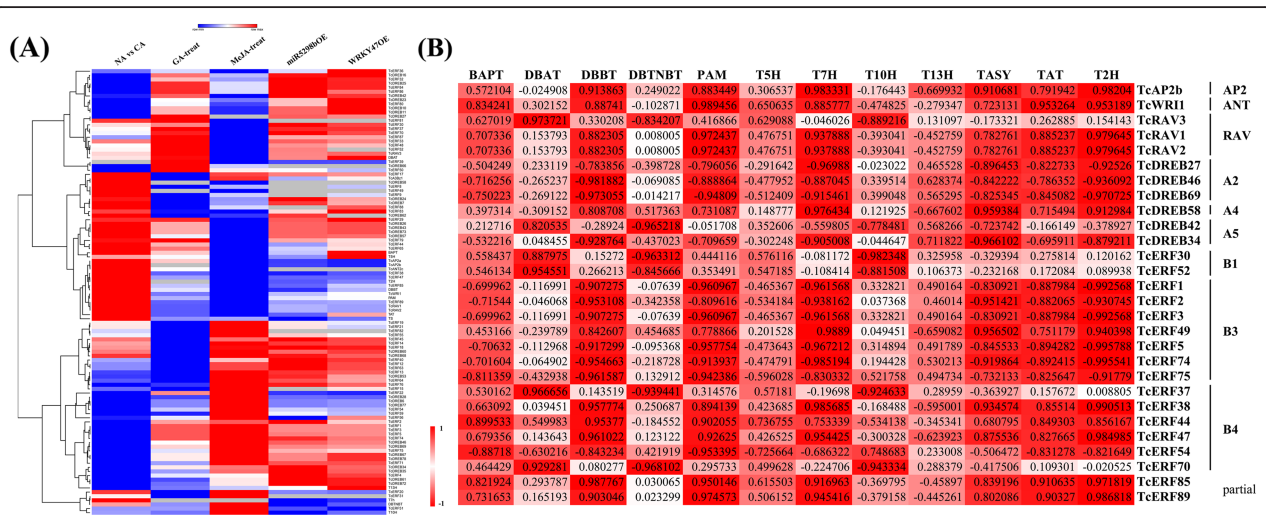


Fig. 6 Expression patterns of *Taxus* AP2/EREBP proteins. **a** Expression heatmap of *Taxus* AP2/EREBP proteins. NA means the new induced cells, GA- and MeJA- treated means the cells were treated with related hormones for 24 h and 16 h respectively. And, miR5298bOE and WRKY47OE indicated the samples were transformed into two genes, miR5298 and WRKY47 which were certificated to involve in taxol biosynthesis. **b** Co-expression coefficient values of selected *Taxus* AP2/EREBPs with taxol biosynthesis genes. Only those *Taxus* AP2/EREBPs were chosen which had a coefficient value more than 0.95 with taxol biosynthesis genes. The subgroups of those *Taxus* AP2/EREBPs were indicated, partial means that the proteins were not full-length

entire plant life cycle and are important parts of gene regulatory networks that integrate metabolic, hormonal, and environmental signals in stress acclimation and retrograde signaling in various biological plant activities [4, 5, 49, 51, 52]. AP2/EREBP superfamily proteins were once considered as plant-specific transcription factors. However, several HNH endonucleases were found to contain an AP2 domain, which selectively recognizes the stretches of poly(G)/poly(C) as binding sites in cyanobacterium (*Trichodesmium erythraeum*), ciliate (*Tetrahymena thermophile*), and viruses (*Enterobacteria phage Rb49* and *Bacteriophage Felix 01*) [53]. In plants, the binding sites of the AP2 domain are also G/C-rich and highly similar to poly(G)/poly(C) stretches [53]. The DNA binding affinity of AP2/EREBP transcription factors indicated the conservative evolution of AP2/EREBP proteins from virus to plants.

In plants, the AP2/EREBP superfamily proteins are commonly divided into AP2, ANT, ERF, DREB, RAV, and soloist subgroups. AP2 and ANT subgroups contain two conserved AP2 domains characterized by a YLG-motif. ERF and DREB have only one WLG-motif containing an AP2 domain. RAV has a B3 domain in addition to the AP2 domain. Soloists have only one AP2 domain with an HLG-motif instead of YLG- and WLG-motifs and are the smallest group in AP2/EREBP proteins, whereas DREB and ERF constitute the largest groups in nearly all plants [8, 11, 30, 35, 39]. These domain structures of AP2/EREBP proteins have evolved conservatively in many plant species, although most of them are angiosperms. However, our results indicated that AP2/EREBP proteins have highly different evolutionary processes and domain structures in gymnosperms and bryophytes.

A new clade of AP2/EREBP proteins was discovered to have three conserved AP2 domains in *T. chinensis*, *G. biloba*, *P. patens*, and *M. polymorpha* (Fig. 2). However, fern *S. moellendorffii* has no such subgroup members. This kind of AP2/EREBP proteins is not found in all gymnosperms, such as *P. abies*, *P. tadae*, and *G. montanum*. Moreover, the proteins of the new subgroup have different domain structures constituting other conserved domains, such as B3 and zf-C3Hc3H domains. After identification via transcriptome datasets, *T. chinensis* has only one protein, named as TcA3Bz1, which contains three AP2, one B3, and one zf-C3Hc3H domains, and this homologue of TcA3Bz1 was only found in *M. polymorpha* and *G. biloba*. B3 domain was found at the N-terminal of the three AP2 domains, whereas it is always located at the C-terminal in RAV proteins [54]. In *G. biloba* and *P. patens*, most proteins belong to the new clade with only three AP2 domains. Two other proteins of *P. patens* contain an extra zf-C3Hc3H domain at the N-terminus of the three AP2 domains. In summary, the new subgroup of AP2/EREBP proteins only exists in

Taxus spp., *G. biloba*, and bryophytes, and these proteins have diverse function domains and thus have evolved differently in higher plants.

B3 domain is an essential element for DNA-binding, but different classes bind various cis-elements; for instance, ARF binds to 5'-TGTCTC-3', and RAV1 recognizes 5'-CACCTG-3' [30]. Although the B3 of TcA3Bz1 has a common ancestor with ARF, they share few similarities, resulting in their different binding properties [30]. In particular, the RGQP(K/R)R-motif in ARF B3 domain, which is essential for ARF to recognize related cis-elements, is absent in the TcA3Bz1 B3 domain (Fig. 2d), suggesting that the latter domain binds an unknown cis-element [30]. zf-C3Hc3H domains are highly conserved in almost all new clade proteins, but minimal information is available on this domain, except that it is DNA binding [51, 55]. In summary, these new clade proteins, such as TcA3Bz1, might play novel functionalities in plants.

Each subgroup of AP2/EREBP proteins is highly conserved in *T. chinensis* and angiosperms, and only the soloist type is not present. The AP2 domain of each subgroup bind differently with various DNA cis-elements and functions [53]. The YRG- and RAYD-motifs of the AP2 domain are responsible for DNA binding and protein-protein interaction, respectively [52]. The first residues Y and R would sometimes differentiate into F and M/L, respectively, in a few types of proteins [42, 56]. In addition, (H/Y/W)LG-motif is the characteristic conserved motif that distinguishes these AP2/EREBP subgroups, YLG in AP2 and ANT, HLG in soloist, and WLG in DREB, ERF, and RAV [11, 31, 57]. In *T. chinensis*, YRG-, (Y/W/H)LG-, and RAYD-motifs are also highly conserved in common subtype proteins. The TcAP2a&b of AP2 clade and TcAP2c of ANT clade both contain the YLG-motif that is conserved in all AP2 domains of AP2 and ANT proteins [9, 31, 57]. Most TcDREBs, TcERFs, and TcRAVs have the WLG-motif.

One difference is that the first residue of YRG- and RAYD-motif varies between *T. chinensis* and angiosperms, especially in RAYD-motif where R is substituted by L/M/V/K/E/Q/I/Y/H (Fig. 3). Determining the influence of highly variable RAYD-motif in *T. chinensis* is difficult due to the limited information about the molecular basis of the interactions of RAYD-motif and related proteins. (Y/W/H)LG-motif exhibits better conservatism than YRG- and RAYD-motifs in *T. chinensis*, and only two proteins have different motifs, including the WLN- of TcERF27 and NTG-motif of R2 (C-terminal AP2 domain) of TcWRI1. All these results indicate that the (Y/W/H)LG-motif has an important role in AP2 domain functions (Fig. 1) [57].

The 14th and 19th residues differ between DREB (V14, E19) and ERF (A14, D19), though the 19th residue is not highly conserved sometimes [5, 35, 58]. The

appearance of TcDREBs and TcERFs is also in accordance with these results, and only TcERF8&9 (B4) and TcERF48 (B6) have a V14 residue similar to DREBs (Fig. 3). Further verification might help clarify the effects of V substitution in TcERF8&9&48.

Despite the differences observed, the AP2 domain evolution is highly conserved in *T. chinensis* and angiosperm plants [6, 11, 30, 31, 35, 36, 38–41]. In conclusion, four Gs (two Gs from YRG- and WLG-motif, two Gs from G11 and G52, Fig. 3a) and one D residue from RAYD-motif are extremely conserved and thus are essential for AP2 domains (Fig. 3). However, in these new subgroup proteins, the three conserve motifs and the four residues of AP2 domain are all variable (Fig. 2). The existence and diverse domain structure of the new AP2/EREBP factors with three AP2 domains indicated the complexity of the evolutionary processes of AP2/EREBP proteins in angiosperm plants and other higher plants.

The EAR- and EDLL motifs of B1 and B3 ERFs, respectively, are also highly conserved in *T. chinensis*.

With the presence of EAR-motif, B1 ERF functions as negative regulators in plants. In *T. chinensis*, all B1 TcERFs have the (L/F)DLNL/F(X)P-type EAR-motif, which is highly conserved in all B1 ERFs from other plants [43]. The EAR-motif generally has two types, LxLxL- and DLNxxP-types, such as JAZ, AUX/IAA, and NINJA [41, 46]. (L/F)DLNL/F(X)P-type could be considered as an enhanced EAR-motif combination of the two types and could strengthen the negative regulation functions. TcERF12, a B1 ERF transcription factor, has a (L/F)DLNL/F(X)P-type EAR-motif and was previously confirmed as a negative regulator of *TASY* gene in taxol biosynthesis (Fig. 4) [5].

EDLL motif, a short motif characterized by four discrete residues, has a strong activation domain and was first found in AtERF98 [47, 48]. Four B3 TcERFs contain a highly similar EDLL motif with AtTDR1 at their C-terminus, including TcERF15, a positive regulator of taxol biosynthesis (Fig. 5) [5]. Further studies revealed that the two L residues, especially the L228 of ORA59, are essential for the activation by binding with MED25, whereas the E and D residues are not conserved [48, 59]. Although not all B3 ERFs, including TcERFs, have conserved EDLL motif, most of them function as positive regulators; for instance, CrORCA3 does not contain an C-terminus EDLL motif but could activate the expression of *STR* gene (Additional file 3) [4]. The mechanism of these B3 ERFs without an EDLL motif should be further clarified.

Many reports characterized the AP2/EREBP transcription factors in various plants; the AP2 domain for DNA-binding is highly conservative across plant kingdoms [11, 39, 41]. Moreover, nearly all subtype proteins of AP2/EREBP factors, including AP2, ANT, RAV, DREB

and ERF, are present from *Physcomitrella patens* to angiosperms. The high number of classifications concluded the same evolutionary process of AP2/EREBP transcription factors, but the identification of new clade proteins reveals a different evolution process of AP2/EREBPs in higher plants. This finding might be an important guidance for us to re-recognize the speciation in plant kingdoms.

In addition to the three conserved AP2 domains, two DNA-binding domains, a new B3 domain and a new zf-C3Hc3H domain, showed extreme conservation in all new subtype AP2/EREBP proteins and were first identified in this study. Additional DNA-binding domains may provide novel functionalities for these new clade proteins by regulating downstream genes. However, identifying the roles of the new clade proteins in plants is difficult due to the uncharacterized B3- and zf-C3Hc3H domains, which were highly different from known related domains. Given its absence in angiosperms, the new clade AP2/EREBPs may play specific roles in *Marchantia polymorpha* and several gymnosperms. Such mutant lines and genetic manipulations of *Marchantia polymorpha* and several gymnosperms are required to clarify the functions of these AP2/EREBPs with newly domain structures.

Conclusions

Identification and evolutionary analysis revealed that AP2/EREBP transcription factors have complex evolutionary processes during plant divergence. The newly found proteins, which have three conserved AP2 domains, exhibited irregularity among higher plants. To our knowledge, no such report was found for angiosperms and pteridophyte *S. moellendorffii*. Not all gymnosperm plants have new AP2 clade proteins; *T. chinensis* and *G. biloba* contain this protein, whereas *Picea abies*, *Pinus tadae* and *G. montanum* do not. For bryophytes, *P. patens* and *M. polymorpha* have these new proteins. Moreover, the domain structure of these new AP2 proteins varies in these plants. These results indicate that the proteins of this new AP2 clade must have a special evolutionary process in plants, but the other clades of AP2/EREBPs have evolved conservatively in *T. chinensis* and other plants. The new subgroup proteins might have separately evolved from the common groups, and angiosperms and several gymnosperms might have abandoned the new-type AP2/EREBPs while diverging.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12870-019-2044-z>.

Additional file 1. Protein sequences of *Taxus* AP2/EREBPs.

Additional file 2. Sequence alignment of all B3 ERFs. *Arabidopsis* B3 ERF were obtained from TAIR (<https://www.arabidopsis.org/>), while CrORCA3

downloaded from UniprotKB under accession number Q9LDB6. This CrORCA3, lacked of the EDLL-motif, was certificated to up-regulate biosynthesis of terpenoid indole alkaloids in *Catharanthus roseus*.

Additional file 3. Partial nucleotide sequence of Gb_11937. The nucleotide sequence of Gb_11937 encoded R2 and R3 of AP2 domain were completely identical. The red characters encoded the R2 and R3 repeat of AP2 domain, the yellow ones were also completely identical, indicating a duplication.

Additional file 4. Nucleotide sequences of *Taxus* AP2/EREBPs.

Abbreviations

ANT: AINTEGUMENTA; AP2: APETELA2; DREB: Dehydration responsive element binding protein; EREBP: Ethylene-responsive element binding protein; ERF: Ethylene-responsive factor; GA: Gibberellin acid; HMM: Hidden Markov Model; MeJA: Methyl jasmonate; RAV: related to ABI3/VP1; TASY: Taxadiene synthase

Acknowledgements

Not applicable.

Authors' contributions

MZ, CF and LY conceived the work, YC, XJ, YC and YY involved in the data analysis, including sequence reassembling and intact Open Reading Frame identification. MZ and CF wrote the manuscript. All authors read and approved the final manuscript.

Funding

The work was funded by the National Natural Science Foundation of China [grant numbers: 31670293]. The funding bodies did not play a role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript, but just provide the financial support.

Availability of data and materials

Plant materials, *Taxus chinensis* (Pilger) Rehd. (deposited the voucher specimen in Herbarium of Wuhan Botanical Garden of CAS under accession number HIB0087305 and identified by Shuxia Fu), obtained from the Huawei Seedling Farmers' Professional Cooperative (Xianning, China), who provided permission to use the seedlings for our scientific research. The *Taxus* cells were cultivated in our lab. Raw sequencing data were submitted in GEO database (accession numbers: GSE28539) and SRA databases (accession numbers: SRR1339463, SRR1339474, SRR1343578, and SRR8083193 to SRR8083198). All the proteins in *Taxus chinensis* we used could obtained in Additional file 1. The others were get from public databases, PTQ32953.1 (accession number) from GenBank (<https://www.ncbi.nlm.nih.gov/protein/PTQ32953.1>), the genomes of *Physcomitrella patens*, *Selaginella moellendorffii* were from JGI (<https://phytozome.jgi.doe.gov/pz/portal.html#>). *Arabidopsis* AP2/EREBP proteins were searched from plnTFdb database (<http://plntfdb.bio.uni-potsdam.de/v3.0/>), *Picea abies* and *Pinus tadae* were obtained from ConGenIE (<http://congenie.org/>), *Ginkgo biloba* was from Giga (<http://gigadb.org/dataset/100209>) and *G. montanum* was downloaded from DRYAD (<https://datadryad.org/resource/doi:10.5061/dryad.0vm37.2>).

Ethics approval and consent to participate

The authorities responsible for the *Taxus* resources are the Mount Jiugong National Reserve, who provided permission to collect the samples; and the Huawei Seedling Farmers' Professional Cooperative (Xianning, China), provided permission to use the seedlings of *T. chinensis* for our scientific research.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 30 January 2019 Accepted: 20 September 2019

References

- Zhang M, Jin X, Chen Y, Wei M, Liao W, Zhao S, Fu C, Yu L. TcMYC2a, a basic helix-loop-helix transcription factor, Transduces JA-Signals and Regulates Taxol Biosynthesis in *Taxus chinensis*. *Front Plant Sci.* 2018;9:863.
- Zhang M, Chen Y, Nie L, Jin X, Liao W, Zhao S, Fu C, Yu L. Transcriptome-wide identification and screening of WRKY factors involved in the regulation of taxol biosynthesis in *Taxus chinensis*. *Sci Rep.* 2018;8(1):5197.
- Weaver BA. How Taxol/paclitaxel kills cancer cells. *Mol Biol Cell.* 2014;25(18):2677–81.
- van der Fits L, Memelink J. ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science.* 2000;289(5477):295–7.
- Zhang M, Li ST, Nie L, Chen QP, Xu XP, Yu LJ, Fu CH. Two jasmonate-responsive factors, TcERF12 and TcERF15, respectively act as repressor and activator of tasy gene of taxol biosynthesis in *Taxus chinensis*. *Plant Mol Biol.* 2015;89(4-5):463–73.
- Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol.* 2006;140(2):411–32.
- Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun.* 2002;290(3):998–1009.
- Zeng L, Yin Y, You C, Pan Q, Xu D, Jin T, Zhang B, Ma H. Evolution and protein interactions of AP2 proteins in *Brassicaceae*: evidence linking development and environmental responses. *J Integr Plant Biol.* 2016;58(6):549–63.
- Shigyo M, Hasebe M, Ito M. Molecular evolution of the AP2 subfamily. *Gene.* 2006;366(2):256–65.
- Fu M, Kang HK, Son SH, Kim SK, Nam KH. A subset of *Arabidopsis* RAV transcription factors modulates drought and salt stress responses independent of ABA. *Plant Cell Physiol.* 2014;55(11):1892–904.
- Song X, Li Y, Hou X. Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa ssp. pekinensis*). *BMC Genomics.* 2013;14:573.
- Romanel EAC, Schrago CG, Counago RM, Russo CAM, Alves-Ferreira M. Evolution of the B3 DNA Binding Superfamily: New Insights into REM Family Gene Diversification. *PLoS One.* 2009;4(6). <https://doi.org/10.1371/journal.pone.0005791>
- Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim Biophys Acta.* 2012;1819(2):86–96.
- Dietz KJ, Vogel MO, Viehhauser A. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma.* 2010;245(1-4):3–14.
- Zhang M, Chen Y, Nie L, Jin X, Fu C, Yu L. Molecular, structural, and phylogenetic analyses of *Taxus chinensis* JAZs. *Gene.* 2017;620:66–74.
- Zhang M, Dong Y, Nie L, Lu M, Fu C, Yu L. High-throughput sequencing reveals miRNA effects on the primary and secondary production properties in long-term subcultured *Taxus* cells. *Front Plant Sci.* 2015;6:604.
- Li ST, Zhang P, Zhang M, Fu CH, Zhao CF, Dong YS, Guo AY, Yu LJ. Transcriptional profile of *Taxus chinensis* cells in response to methyl jasmonate. *BMC Genomics.* 2012;13:295.
- Zhang CH, Mei XG, Liu L, Yu LJ. Enhanced paclitaxel production induced by the combination of elicitors in cell suspension cultures of *Taxus chinensis*. *Biotechnol Lett.* 2000;22(19):1561–4.
- Liao W, Zhao S, Zhang M, Dong K, Chen Y, Fu C, Yu L. Transcriptome assembly and systematic identification of novel cytochrome P450s in *Taxus chinensis*. *Front Plant Sci.* 2017;8:1468.
- Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7(10):e1002195.
- Huang X, Madan A. CAP3: a DNA sequence assembly program. *Genome Res.* 1999;9(9):868–77.
- Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;26(5):680–2.
- Perez-Rodriguez P, Riano-Pachon DM, Correa LG, Rensing SA, Kersten B, Mueller-Roeber B. PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 2010;38(Database issue):D822–7.

24. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30(4):772–80.
25. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 2014;42(Web Server issue):W320–4.
26. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28(10):2731–9.
27. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–21.
28. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–8.
29. Chen C, Xia R, Chen H, He Y. TBtools, a toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface; 2018.
30. Li X, Zhang D, Gao B, Liang Y, Yang H, Wang Y, Wood AJ. Transcriptome-wide identification, classification, and characterization of AP2/ERF family genes in the desert Moss *Syntrichia caninervis*. *Front Plant Sci.* 2017;8:262.
31. Wu ZJ, Li XH, Liu ZW, Li H, Wang YX, Zhuang J. Transcriptome-based discovery of AP2/ERF transcription factors related to temperature stress in tea plant (*Camellia sinensis*). *Funct Integr Genomics.* 2015;15(6):741–52.
32. Cernac A, Benning C. WRINKLED1 encodes an AP2/ERF domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. *Plant J.* 2004;40(4):575–85.
33. Tang T, Du C, Song H, Aziz U, Wang L, Zhao C, Zhang M. Genome-wide analysis reveals the evolution and structural features of WRINKLED1 in plants. *MGG: Molecular genetics and genomics;* 2018.
34. Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. *Trends Plant Sci.* 2008;13(12):647–55.
35. Du DL, Hao RJ, Cheng TR, Pan HT, Yang WR, Wang J, Zhang QX. Genome-wide analysis of the AP2/ERF gene family in *Prunus mume*. *Plant Mol Biol Rep.* 2013;31(3):741–50.
36. Zhuang J, Peng RH, Cheng ZM, Zhang J, Cai B, Zhang Z, Gao F, Zhu B, Fu XY, Jin XF, et al. Genome-wide analysis of the putative AP2/ERF family genes in *Vitis vinifera*. *Sci Hortic.* 2009;123(1):73–81.
37. Zhuang J, Cai B, Peng RH, Zhu B, Jin XF, Xue Y, Gao F, Fu XY, Tian YS, Zhao W, et al. Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa*. *Biochem Biophys Res Commun.* 2008;371(3):468–74.
38. Licausi F, Giorgi FM, Zenoni S, Osti F, Pezzotti M, Perata P. Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. *BMC Genomics.* 2010;11:719.
39. Li X, Tao S, Wei S, Ming M, Huang X, Zhang S, Wu J. The mining and evolutionary investigation of AP2/ERF genes in pear (*Pyrus*). *BMC Plant Biol.* 2018;18(1):46.
40. Du C, Hu K, Xian S, Liu C, Fan J, Tu J, Fu T. Dynamic transcriptome analysis reveals AP2/ERF transcription factors responsible for cold stress in rapeseed (*Brassica napus* L.). *Mol Genet Genomics.* 2016;291(3):1053–67.
41. Rao G, Sui J, Zeng Y, He C, Zhang J. Genome-wide analysis of the AP2/ERF gene family in *Salix arbutifolia*. *FEBS Open Bio.* 2015;5:132–7.
42. Zhang JY, Broeckling CD, Blancaflor EB, Sledge MK, Sumner LW, Wang ZY. Overexpression of WXP1, a putative *Medicago truncatula* AP2 domain-containing transcription factor gene, increases cuticular wax accumulation and enhances drought tolerance in transgenic alfalfa (*Medicago sativa*). *Plant J.* 2005;42(5):689–707.
43. Dong CJ, Liu JY. The *Arabidopsis* EAR-motif-containing protein RAP2.1 functions as an active transcriptional repressor to keep stress responses under tight control. *BMC Plant Biol.* 2010;10:47.
44. Tao SB, Estelle M. Mutational studies of the aux/IAA proteins in *Physcomitrella* reveal novel insights into their function. *New Phytol.* 2018; 218(4):1534–42.
45. Pauwels L, Barbero GF, Geerinck J, Tilleman S, Grunewald W, Perez AC, Chico JM, Bossche RV, Sewell J, Gil E, et al. NINJA connects the co-repressor TOPLESS to jasmonate signalling. *Nature.* 2010;464(7289):788–91.
46. Kazan K. Negative regulation of defence and stress genes by EAR-motif-containing repressors. *Trends Plant Sci.* 2006;11(3):109–12.
47. Tiwari SB, Belachew A, Ma SF, Young M, Ade J, Shen Y, Marion CM, Holtan HE, Bailey A, Stone JK, et al. The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J.* 2012;70(5):855–65.
48. Cevik V, Kidd BN, Zhang PJ, Hill C, Kiddle S, Denby KJ, Holub EB, Cahill DM, Manners JM, Schenk PM, et al. MEDIATOR25 acts as an integrative hub for the regulation of Jasmonate-responsive gene expression in *Arabidopsis*. *Plant Physiol.* 2012;160(1):541–55.
49. Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. *Biol Chem.* 1998;379(6):633–54.
50. Wang SJ, Li C, Wang HJ, Zhong XM, Zhao J, Tong YY, Zhou YJ. Effect of elicitors, precursors and metabolic inhibitors on paclitaxel production by *Taxus cuspidata* cell culture. *J Forestry Res.* 2016;27(6):1257–63.
51. Li CW, Su RC, Cheng CP, Sanjaya YSJ, Hsieh TH, Chao TC, Chan MT. Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. *Plant Physiol.* 2011;156(1):213–27.
52. Kizis D, Lumberras V, Pages M. Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. *FEBS Lett.* 2001;498(2-3):187–9.
53. Magnani E, Sjolander K, Hake S. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *Plant Cell.* 2004;16(9):2265–77.
54. Liu Y, Yuan J, Ma H, Song J, Wang L, Weng Q. Characterization and functional analysis of a B3 domain factor from *Zea mays*. *J Appl Genet.* 2015;56(4):427–38.
55. Malhotra S, Sowdhamini R. Collation and analyses of DNA-binding protein domain families from sequence and structural databanks. *Mol BioSyst.* 2015; 11(4):1110–8.
56. Yang Y, Dong C, Li X, Du J, Qian M, Sun X, Yang Y. A novel Ap2/ERF transcription factor from *Stipa purpurea* leads to enhanced drought tolerance in *Arabidopsis thaliana*. *Plant Cell Rep.* 2016;35(11):2227–39.
57. Kavas M, Kizildogan A, Gokdemir G, Baloglu MC. Genome-wide investigation and expression analysis of AP2-ERF gene family in salt tolerant common bean. *EXCLI J.* 2015;14:1187–206.
58. Sun ZM, Zhou ML, Xiao XG, Tang YX, Wu YM. Genome-wide analysis of AP2/ERF family genes from *Lotus corniculatus* shows LcERF054 enhances salt tolerance. *Funct Integr Genomics.* 2014;14(3):453–66.
59. Li JB, Zhang KX, Meng Y, Hu JP, Ding MQ, Bian JH, Yan ML, Han JM, Zhou ML. Jasmonic acid/ethylene signaling coordinates hydroxycinnamic acid amides biosynthesis through ORA59 transcription factor. *Plant J.* 2018;95(3):444–57.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

