

Genome-wide identification and expression profile analysis of the NF-Y transcription factor gene family in *Eucalyptus grandis*

Juan Li

l i . j u a n @ s x n u . e d u . c n

Shanxi Normal University

Chaoyan Gong

Shanxi University of Chinese Medicine

Li Zhuang

Shanxi Normal University

Guangyou Li

Chinese Academy of Forestry

Jianmin Xu

Chinese Academy of Forestry

Zhaohua Lu

Chinese Academy of Forestry

Research Article

Keywords: NF-Y, *Eucalyptus grandis*, Bioinformatics analysis, Expression analysis

Posted Date: July 31st, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4703272/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

The NF-Y (NUCLEAR FACTOR-Y) transcription factor in plants is composed of NF-YA, NF-YB and NF-YC subunits. It is known to play an important role in plant growth and development and response to stress. Although the NF-Y gene family has been systematically studied in many species, the understanding of the NF-Y gene family in *Eucalyptus* remains unknown.

Results

In this study, 31 (7 EgrNF-YA, 16 EgrNF-YB and 8 EgrNF-YC) *EgrNF-Y* genes were identified in *E. grandis* using Arabidopsis NF-Y protein sequences as queries and their structural characteristics were comprehensively analyzed. Phylogenetic, conserved domain and exon-intron structure analyzed that the closer relationship in each subfamily. Multiple alignments showed that all EgrNF-Y proteins had conserved core regions. Chromosomal localization of these genes revealed that they were randomly distributed across 11 chromosomes. *Cis*-element analysis of promoter indicated that *EgrNF-Y* gene was affected by various hormonal and abiotic stresses. Furthermore, tissue-specific expression showed that all 30 *EgrNF-Y* genes were widely expressed in various tissues and organs. Additionally, the stress response pattern of *EgrNF-Ys* was identified under phosphate-starved, and 12 genes and 3 genes were upregulated more than 2-fold in the leaves and roots, respectively.

Conclusion

Our studies have provided a general understanding of the conservation and characteristics of the EgrNF-Y genes family in *E. grandis*. And it has been demonstrated that members of the EgrNF-YB1 and EgrNF-YB11 may play important roles in the regulation of flowering in *E. grandis*. To provide reference for further study on the role of NF-Y gene in the regulation of flowering in *E. grandis*. In addition, our also established a theoretical basis for further functional studies on this family.

Background

Transcription factors (TFs), a protein that binds to a *cis*-acting element in the promoter region of a eukaryotic gene and enables the target gene to be expressed in a specific way, play vital roles in numerous in many plants growth, development and abiotic stress. Nuclear factor Y (NF-Y), also famous as heme activator protein (HAP) or CCAAT-binding factor (CBF), is one of the most widespread heterotrimeric transcription factors found in numerous genes in fungi, animal, and plants. In plants, NF-Y contains three unique subunits: NF-YA (HAP2 or CBF-B), NF-YB (HAP3 or CBF-A), and NF-YC (HAP5 or CBF-C) [1].

NF-Y is highly conserved in all higher eukaryotes [2]. In yeast and animals, each NF-Y subunit is encoded by a single gene. However, each NF-Y subunit gene is represented by multiple orthologs in plants, indicating that heterotrimer complexes composed of different NF-Y members play different regulatory roles [3]. To date, NF-Y transcription factors have been identified in several herb species, including 33 NF-Y member genes were identified in *Arabidopsis thaliana* [4], 28 NF-Y members in *Oryza sativa* [5], 68 NF-Y in *Glycine max* [6], 33 members in *Brassica napus*. L [7], 60 NF-Y member genes in *Medicago sativa* [8], 27 members *Petunia hybrida* [9]. Furthermore, NF-Y gene members have also been reported in many species of woody plants. For example, 46 members were identified in *Populus* [10], 28 members were identified in *Pinus tabulaeformis* [11] and 22 members in *Citrus* [12]. Recently, the NF-YB family of *Eucalyptus* has been analyzed and identified. Although many studies have reported the NF-Y gene family in various plant species, the *Eucalyptus* NF-Y gene family has not been systematically explored.

Many studies have shown that NF-Y genes are responsible for regulating diverse physiological processes related to plant growth and development. Among, AtNF-YB9, the first NF-Y gene cloned in a plant, plays a key role in Arabidopsis seed development[13]. The NF-Y has also been reported to play an important role in flowering time. Such as, NF-Y mediates the effect of photoperiod and GA signaling on *SOC1* expression partly through H3K27me3 demethylation [14]. Also, *CmNF-YB8* can influence flowering time through regulating the expression of cmo-MIR156 in the aging pathway in *Chrysanthemum* [15]. In addition, NF-Y plays an important role in the process of pollen tube growth [16], starch biosynthesis [17], root elongation[18], fruit ripening [19].

In addition to plant growth and development, NF-Y participates in stress response. Studies on NF-Y have focused mainly on drought resistance [20–22], salt stress[23] and cold [24]. Studies have also shown that NF-Y protein is a key factor in regulating root nodule formation and nutrient absorption in plants. It has been reported in the *Medicago truncatula* [25], *Lotus japonicus*[26], *Phaseolus vulgaris*[27], *Triticum aestivum*[28]. However, whether NF-Y plays a role in growth, development and nutrient absorption in *Eucalyptus* remains unclear.

Eucalyptus grandis, a species of *Eucalyptus* in Myrtaceae, is a significant fast-growing tree species. *E. grandis* is grown primarily in the southern and southwestern of China. It has the advantages of short growth cycle, strong stress resistance and strong toughness of wood and high economic value. It is considered one of the three fastest growing trees in the world along with poplar and pine. Because it is mainly distributed in southwest China and South China, the soil in these areas is generally lacking in phosphorus. Furthermore, the lack of nutrients affects the growth of *E. grandis*. Most previous studies only focused on the involvement of NF-Y transcription factors in plant root growth and development [18], but the mechanism of how NF-Y regulates root growth and development under nutrient stress is still limited. Therefore, it is of great significance to identify the NF-Y gene family members and clarify the effects of phosphorus deficiency on the growth of *E. grandis*. In this study, 31 NF-Y genes were identified from the *E. grandis* genome and their physicochemical properties, phylogenetic relationships, gene structure, and chromosome localization were comprehensively analyzed. Moreover, we also analyzed the expression levels of *EgrNF-Y* in six different tissues and organs and in the condition of phosphorus

deficiency by qRT-PCR, which is particularly important for identifying candidate genes involved in regulation of the growth and response to phosphorus stress of *E. grandis*. Overall, our results contribute to a more complete understanding of the function of the NF-Y genes in *E. grandis*, and the study of other woody plants also has high reference value.

Results

Isolation and identification of the NF-Y family members in *E. grandis*

To obtain information on the NF-Y genes of *E. grandis*, we used Arabidopsis NF-Y protein sequences as queries to search for the *E. grandis* NF-Y genes in Phytozome v13 [4]. Through comprehensive screening, including remove those with improper domains and redundant sequences, a total of 31 EgrNF-Y sequences were identified in the *E. grandis* genome, including 7 NF-YA, 16 NF-YB and 8 NF-YC genes. Furthermore, the physicochemical properties data of all gene members were estimated by ExPASy server (<http://WWW.expasy.org/>), and the characteristics of the *EgrNF-Y* sequences are listed in Table 1. Among them, the identified *EgrNF-Y* genes encoded peptides ranged from 94 to 339 aa. The molecular weights (MWs) and the isoelectric point (pI) values of these proteins ranged from 10.64 kDa to 37.21 kDa, and from 4.49 to 9.48, respectively (Table1).

Gene structure and conserved motifs analysis of the *EgrNF-Y* genes

The analysis of gene structure can understand the evolution of gene families. To investigated the evolutionary conservation and divergence of NF-Ys between *E. grandis* and Arabidopsis, we analyzed the exon-intron gene structure of 31 identified EgrNF-Ys using the GSDS website. Phylogenetic analysis revealed that the *EgrNF-Y* genes were divided into three groups: EgrNF-YA, EgrNF-YB and EgrNF-YC. Most of the EgrNF-YAs (except *EgrNF-YA5*) had five or six exons with similar distribution. More than two-thirds of *EgrNF-YBs* had no introns, and the results showed that members with similar numbers of exons and introns are distributed in the same clade. For the EgrNF-YC subfamily, EgrNF-YC1 and EgrNF-YC4 had two exons. EgrNF-YC5 and EgrNF-YC7 have six and five exons, respectively. In general, the gene structure of NF-Y members was positively correlated with their phylogenetic relationships (Fig. 1). Furthermore, the distributions of conserved motifs were assessed by MEME software. The results showed that all of the genes contained motif 1 except EgrNF-YA6. Interestingly, three EgrNF-Y subunits have a unique motif distribution (Fig. 2). For example, motifs 8 were only present in *EgrNF-YA*, motif 2 was only observed in *EgrNF-YB*, and motif 7 was unique to *EgrNF-YC*.

Chromosomal localization and collinearity analysis of *EgrNF-Ys*

The 31 *EgrNF-Y* gene members were widely distributed according to the positions of the annotated chromosomes in the *E. grandis* genome database (Table 1). To further analyze the distribution of EgrNF-Y family members on each chromosome, we constructed a location map using MapInspect software. The result showed that the distribution of NF-Y gene family on the chromosomes of *E. grandis* was uneven. Among them, the largest number of members distributed on chromosome 2, while only 1

member distributed on chromosome 3 and chromosome 9. Moreover, the number of genes was not positively correlated with chromosome length. For example, chromosome 3 had the largest length, but only one member is distributed (Fig. 3).

To further investigate the homologous genes and their evolutionary relationships between *E. grandis*, model plant *Arabidopsis* and woody model plant poplar, a multicomparative synteny map generated between three species. The 31 EgrNF-Ys were located at 11 scaffolds, 46 PtNF-Ys were distributed on 19 chromosomes, and 36 AtNF-Ys were distributed on 5 chromosomes, which have a collinearity relationship with EgrNF-Ys. Among them, the relationship between *E. grandis* and poplar is closer, which indicates that the NF-Y family is a close relationship between woody plants (Fig. 4).

Conserved regions and phylogenetic relationships of EgrNF-Ys

To further investigate the conserved regions of EgrNF-Ys, the protein sequences of the 31 members were analyzed using ClustalW 2.1 and Genedoc software. Multiple sequence alignment results suggested that each EgrNF-Y family member contains a heterodimerization domain and a DNA-binding domain that recognizes the CCAAT site. This core conserved regions of the EgrNF-YAs proteins were 53AAs, including two highly conserved domains: the NF-YB/C subdomain and the DNA binding, they were separated by a conserved linker with 21 AAs. As shown in Fig. 5 B/C, the central domain of EgrNF-YBs had 91AAs. Among EgrNF-YBs, EgrNF-YB2/4/16 had lower conserved domains. Fig.5C showed that EgrNF-YC subunits were also found to consist of a core histone-like sequence with a central domain about 79 AAs in length. Meanwhile, EgrNF-YC5 and EgrNF-YC7 were slightly different from other NF-YCs. This analysis also suggested that EgrNF-YAs is more evolutionarily conserved in the three subfamilies (Fig. 5).

To reveal the evolutionary relationship and potential function of EgrNF-Ys, an phylogenetic tree was constructed using the NF-Y protein sequences of *E. grandis*, *A. thaliana* and *P. trichocarpa* were created by MEGA7 software with the neighbor-joining (NJ) criteria. The phylogenetic analysis revealed that the 107 NF-Y proteins were clustered into three groups: NF-YA (yellow), NF-YB (green), and NF-YC (pinkish red). It is consistent with our subfamily classifications of the EgrNF-Ys (Table 1). Based on the phylogenetic relationship of the evolutionary tree and the reported functions of AtNF-Ys, the functions of EgrNF-Y members can be further predicted. In each group, we found that some pairs of paralogous NF-Y proteins were composed of one EgrNF-Y and one AtNF-Y, such as EgrNF-YA1 and AtNF-YA6, EgrNF-Y5 and AtNF-YA11, and this close evolutionary relationship generally suggested the similarity of their biological functions. We also found EgrNF-YB9, AtNF-YB6/9 and PtNF-YB3/5 clustered in a subgroup, belonging to LEC1 and its homolog LEC1-like. In addition, we also identified three pairs of analogues: EgrNF-YB4 and EgrNF-YB5, EgrNF-YB7 and EgrNF-YB14, EgrNF-YC2 and EgrNF-YC3, while most EgrNF-Ys share low homology with other members, suggesting that they have evolved in diversity.

Analysis of *cis*-elements in the promoter regions of *EgrNF-Y* genes

In order to further explore the potential function of the 31 *EgrNF-Y* at the transcriptional level, the distribution of *cis*-elements in the *EgrNF-Y* promoter regions (2000 bp) was scanned using the PlantCARE software. A total of 19 types of *cis*-elements were identified in the 31 *EgrNF-Y*, including light responsive, hormone responsive, stress responsive, growth regulation, and some common and core *cis*-elements, such as the TATA-box and CAAT-box. The detailed classification and sequence information of all the *cis*-elements are listed in Table S2. Meanwhile, the transcription regulatory *cis*-elements binding site are also shown in fig.7 except some common and core *cis*-elements. This result analysis showed that the promoter regions of the *EgrNF-Ys* contain several phytohormone response *cis*-elements, including gibberellin, salicylic acid, MeJA, auxin, ethylene, and abscisic acid responsive elements. The results indicated that *EgrNF-Ys* may play an important role in response to stress and growth regulation.

Expression profiles of the *EgrNF-Ys* in different tissues

To further investigate the possible functions of the *EgrNF-Ys* genes in the developmental processes of *E. grandis*, we examined their gene expression profiles using quantitative real-time PCR in six different tissues and organs (root, stem, young leaf, mature leaf, xylem and flower). All member genes except *EgrNF-YA5* were expressed. Therefore, the tissue-specific expression patterns of 30 *EgrNF-Ys* member genes were analyzed. The results demonstrated that 30 genes were widely expressed in various tissues and organs of *E. grandis*, but they exhibited different spatial and temporal expression patterns. For *EgrNF-YA* subfamilies, *EgrNF-YA1* and *EgrNF-YA7* were significantly expressed in flowers and roots, respectively. Expression was strongest for *EgrNF-YA4* and *EgrNF-YA6* in the leaves. For *EgrNF-YB* subfamilies, *EgrNF-YB1* and *EgrNF-YB11* was most highly expressed in the flowers, more than 90% of other *EgrNF-YB* members have high expression levels in young leaves. For *EgrNF-YC* subfamilies, although all members in *EgrNF-YC* subfamilies were expressed in almost all tissues, their expression levels were highest in young leaves (Fig. 8). The diversity of expression patterns of *EgrNF-Y* gene indicates that *EgrNF-Y* gene has different biological functions during the growth and development of eucalyptus, and has a wide range of biological applications.

Expression Profiles of the *EgrNF-Ys* under Low phosphorus environment

E. grandis is mainly distributed in southwest and South China where soil is generally deficient in phosphorus. Studies have shown that *NF-Y* is an important transcriptional regulator and plays an important role in plant stress and growth and development. In order to investigate the response of *EgrNF-Y* gene expression to phosphate starvation, *E. grandis* seedlings with the same growth were cultured in normal and phosphate-free nutrient solution for two weeks respectively, and compared their relative expression levels under the above two growth conditions by qRT-PCR. The results analysis showed that 12 genes (*EgrNF-YB3/B8/B11/B12/B13/B14/B15*, *EgrNF-YC1/C2/C3/C5/C7*) were upregulated more than 2-fold in the leaves after being phosphate-starved for 14 days. The other genes remained stable or were downregulated. In the root, only *EgrNF-YB6/B11/B13* was upregulated by more than a factor of 2 under low phosphate treatment, while most genes remained stable (Fig. 9). These

results suggested that these up-regulated genes may be involved in phosphate uptake when phosphate is limited in the soil.

Discussion

Multiple evidences showed that NF-Ys plays a variety of important roles in plant growth and response to environmental stress. Because of the duplication of genes, plants usually have large gene families, and have been broadly studied in many herbs. such as *A. thaliana* [4], *Glycine max* [6], *Oryza* [38], *Petunia hybrida* [9], *Medicago sativa* [8], *Brassica napus* [7]. In recent years, NF-Y family gene identification has been gradually identified in woody plants. Such as *Populus tomentosa* and *Pinus tabuliformis*. *Eucalyptus* is recognized by FAO as one of the three fastest growing trees in the world. However, the identification and analysis of the NF-Y gene family in *E. grandis* have not been reported. Here, we isolated and identified a total of 31 *EgrNF-Y* genes. Compared with the numbers of NF-Ys in other species, such as 36 in *A. thaliana* [4], 28 in *Oryza* [38], 46 in *Populus trichocarpa* [10] and 28 in *P. tabuliformis* [11], *Eucalyptus* harbored a comparable number of genes. In addition, their gene structures, conserved domains, phylogenetic relationship and expression patterns were systematically analyzed. These findings provide valuable information for a subsequent functional analysis and precision plant breeding of a single *EgrNF-Y* gene.

To explore the function of these *EgrNF-Ys*, we performed exon-intron structure and motif analysis (Fig. 1, Fig. 2). The number and distribution of exon-intron structures and motifs are similar to previous results [11, 39], suggesting that the function of *EgrNF-Ys* may be similar to that of homologous genes in other species.

Studies have shown that Arabidopsis NF-YB subunits can be divided into LEC1 and non-LEC1 categories. Among them, aspartic acid at D55 is considered to be a key protein interaction site in the AtNF-YB subfamily to distinguish LEC1 from non-LEC1 types. Study found that *LEC1* plays an important role in plant embryogenesis and seed development. For example, Arabidopsis [4], castor [23]. In this study, *EgrNF-YB9* changed from Lys (K) to Asp (D) at this binding site. Phylogenetic analysis also showed that *EgrNF-YB9* was clustered with AtNF-YB9 (AtLEC1) and AtNF-YB6 (AtLEC1). Therefore, it can be speculated that they play an important role in regulating seed development and embryogenesis.

In this study, we constructed a phylogenetic tree to analyze the NF-Y proteins of *E. grandis*, poplar and Arabidopsis. It was previously reported that the NF-Y family members of Arabidopsis may not contain AtNF-YB11/12/13 and AtNF-YC10/11/13, because they do not have the corresponding structure [40]. Interestingly, the analysis of evolutionary tree results in this study showed that *EgrNF-YB2/B16* and *EgrNF-YC5/C7* also had distant evolutionary relationships with other *EgrNF-YA/B/C* gene clusters. Similar to AtNF-YB11/12/13/C11 (Fig. 6). In addition, multiple comparison results and our phylogenetic tree support this view, respectively (Fig. 5).

It has been reported that NF-Y is an important transcriptional regulatory factor, which is widely expressed in the whole life cycle of vegetative and reproductive growth of plants. In this study, the relative

expression levels of *EgrNF-Y* in roots, stems, young leaves, mature leaves, xylem and flowers of *E. grandis* were analyzed by qRT-PCR. In this study, we found that they are widely expressed in various tissues and organs at different levels. However, the results of this study showed that *EgrNF-YC* subfamily had high expression levels in young leaves, indicating that they play an important role in leaves. Studies have shown that NF-YC subunits can interact with CONSTANS proteins to regulate transcription levels of the key integrons FT, leading to early flowering in plants. Such as, NF-YC3, NF-YC4, and NF-YC9 are necessary for proper photoperiod-dependent induction of flowering in Arabidopsis, and the function of CO in FT transcriptional activation requires these NF-YCs, whereas synthetic FT proteins are only present in leaves [41, 42]. FT is also an important integron regulating plant flowering [41]. Therefore, high expression of *EgrNF-YC* subfamily genes in leaves may help activate FT synthesis and promote flowering.

In addition to NF-YC members participating in flowering regulation through photoperiodic pathways, studies have shown that NF-YB subfamily members can also participate in flowering regulation through age pathways. For example, chrysanthemum *CmNF-YB8* can enhance *SPLs* transcript accumulation by regulating *miR156* expression, and finally regulates the flowering time of chrysanthemum [15]. Interestingly, in the process of analyzing the tissue-specific expression levels of *EgrNF-YB* subfamily members, the expression levels of *EgrNF-YB1* and *EgrNF-YB11* were the highest in flower (Fig. 8). Meanwhile, we constructed an evolutionary tree between *EgrNF-YB* subfamily members and *CmNF-YB8* for phylogenetic analysis. The results showed that, *EgrNF-YB1*, *EgrNF-YB11* and *CmNF-YB8* are clustered into a subclade (Fig. S2). It is further speculated that *EgrNF-YB1* and *EgrNF-YB11* may have similar functions to *CmNF-YB8* in participating in the regulation of flowering. Due to the existence of multiple promoters in the upstream sequence of *EgrNF-YB* gene, which usually function as dimers or trimers, the mechanism of age-regulation regulation is not clear. Therefore, the exact functional mechanism of the *EgrNF-Y* gene still needs further investigation.

NF-Y is an important transcriptional regulatory factor that plays an important role in response to plant stress. At present, most studies focus on abiotic stress. Such as drought [43], salt [8], cold and heat [9, 24]. However, there are few studies on how *NF-Y* affects plant growth and development under nutrient stress conditions. In this study, we found that different subunits of the *EgrNF-Y* family showed different expression patterns under low phosphorus environment. Among them, the expressions of 12 genes of *NF-YBs* and *NF-YCs* in leaves were up-regulated by more than 2 times after 14 days of phosphate starvation treatment. In the root, only three genes were upregulated more than two-fold, while most remained stable. These results suggest that these up-regulated genes may be involved in phosphate uptake when phosphate is limited in the soil.

Conclusions

In this study, 31 *EgrNF-Y* genes were identified, including 7 *EgrNF-YAs*, 16 *EgrNF-YBs* and 8 *EgrNF-YCs*, and their conserved domains, gene structure, phylogenetic relationships, cis-elements and expression patterns were analyzed in the *E. grandis* reference genome. This provides a theoretical basis for

researchers to further understand this gene family, and can be used as a candidate functional gene for further study of NF-Y in the growth, development and phosphorus tolerance of *E. grandis*. In conclusion, our results can provide valuable information for further elucidating the biological function of EgrNF-Ys.

Materials and methods

Identification of NF-Y family members in *E. grandis*

The NF-Y protein sequences (10 NF-YA, 13 NF-YB and 13 NF-YC genes) in *Arabidopsis thaliana* from the Arabidopsis Information Resource (TAIR) (<http://www.Arabidopsis.org/>). These sequences were used to search *E. grandis* in Phytozome v13 (<https://phytozome-next.jgi.doe.gov/>) using the BLAST program and selected all sequences cut-off was set as $e\text{-value} < 10^{-10}$. In addition, we used the Pfam and SMART tools to verify whether each candidate gene had NF-Y protein conserved domains [29]. Further, all protein sequences obtained by BLAST and SMART retrievers were manually screened to remove incomplete and redundant sequences. Finally, 31 candidate members of *E. grandis* NF-Y were obtained for subsequent analysis. Meanwhile, ProtParam (<http://web.expasy.org/protparam/>) was used to analyze the information about each *EgrNF-Y* gene, including protein sequence, molecular weight (MW), isoelectric point (pI).

Gene structure, conserved motifs and cis-element analysis of the NF-Y gene family in *E. grandis*

For gene structure analysis, the online service of the Gene Structure Display Server (GSDS) (<http://gsds.cbi.pku.edu.cn/>) to visualize the number and distribution of exons and introns *EgrNF-Y* genes [30]. For conserved motifs analysis, MEME software was used with the default parameters (the maximum number of motifs set to 10) to analyze the conserved motifs in the EgrNF-Y proteins [31]. For *cis*-element analysis, the promoter sequences (length, 2 kb) of *EgrNF-Y*s were collected from the genome of *E. grandis* as the regulatory promoter region, and were analyzed using the online software PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [32]. Meanwhile, *cis*-elements present in the promoter regions were also visualized by TBtools [33].

Chromosome position and cross-species collinearity analysis of the EgrNF-Y genes

The chromosomal locations of the *EgrNF-Y* genes using the MapInspect software [34]. TBtools was used to extract the location information of NF-Y genes from *A. thaliana*, *P. trichocarpa*, *E. grandis* genome files and gene annotation files to construct the physical map of NF-Ys of three plants on chromosomes. The MCscanX command in TBtools was used to analyze the collinear relationship among the three species and to identify the collinear blocks of NF-Y genes in *A. thaliana*, *P. trichocarpa* and *E. grandis* genome files [35].

Multiple alignments and phylogenetic analysis of EgrNF-Y proteins

The protein sequences of the *EgrNF-Y* members were obtained using Phytozome v12.1. Multiple sequence alignment analysis of the identified *EgrNF-Y*s were constructed using ClustalX2.1 with the default parameters [36]. For phylogenetic relationship analyse, the phylogenetic tree was constructed using MEGA7.0 with the complete *NF-Y* protein sequences. The evolutionary relationships were estimated with 1000 bootstrap replications [37]. The phylogenetic tree constructed by MEGA7 was uploaded to iTOL (<http://itol.embl.de/>) for further editing.

Plant material, growth conditions and abiotic treatment

All the seedlings were grown in the Institute of Tropical Forestry, Chinese Academy of Forestry (Guangzhou, China) in this study. The greenhouse was kept at $23 \pm 1^\circ\text{C}$, 60–70% humidity, and 16/ 8 h light/dark photoperiod. In this study, for the phosphate treatments, we designed normal growth condition (KH_2PO_4 concentrations of 1mM) and phosphorus deficient growth conditions (KH_2PO_4 concentrations of 0M) for the experiments. Control group and treatment group were replaced with a new nutrient solution every 5 days (800 mL/pot). Here, we used seedling samples grown under normal conditions as controls. Control group and treatment group was applied to 18 plants (with three replicate trials) for one month. After P treatments, the roots (the newest and the second lateral roots) and leaves were harvested, respectively. Roots, stems, and leaves tissue were obtained from 6-month-old cultured plantlets. In addition, mature leaves, xylem and flowers were collected from Jiangmen City, Guangdong Province. All samples are immediately frozen in liquid nitrogen after collection and stored at -80°C until use.

RNA extraction and quantitative real-time PCR

Total RNAs were isolated from different tissues and organs of *E.grandis*, as well as from the different phosphate treatment samples. Total RNA from the samples was extracted using the OMEGA plant RNA Kit (Shanghai, China) and following the operating instruction. These RNAs were assessed by agarose gel and NanoDrop 2000 spectrophotometer (Implen, Inc., Westlake Village, CA, USA). cDNA was synthesized using total RNA as the template with the Takara's PrimeScript Synthesis 1st Strand cDNA Synthesis Kit (Takara, Beijing, China). The obtained cDNA was diluted 10-fold with ddH₂O and used as the template for qRT-PCR.

To measure the expression levels of *EgrNF-Y* genes, qRT-PCR analysis was performed using SYBR® Premix Ex Taq™ (TaKaRa). The qRT-PCR primers were listed in Table S1, and *EgrEF2* was used as the internal control. PCR was performed on an Light Cycler 96 Real-Time PCR system following the manufacturer's instructions. The PCR program was 94°C for 30 s, followed by 40 cycles of 94°C for 5 s and 60°C for 10 s, and a final elongation step of 72°C for 5 min. The program was completed with a melting curve from 70 to 95°C . The relative expression levels of these genes were analyzed using the $2^{-\Delta\Delta\text{CT}}$ method.

Statistical analyses

For all experiments, we all performed three independent biological replicates and three technical replicates. Data are statistically described as mean and standard deviation (mean $\pm$ SD). SPSS19.0 software was used for data analysis (IBM SPSS, Chicago, IL, USA). Error bars were calculated according to Tukey's multiple range test, and $^{*}(p < 0.05)$ and $^{**}(p < 0.01)$ being used to indicate statistically significant effects.

Declarations

Acknowledgements

We appreciate the contributions and support from the lab members. We extend our sincere gratitude to the editor and reviewers for their thorough evaluation of this manuscript and for providing valuable feedback for its enhancement.

Funding

This research has been supported by the Young Scientists Fund of the National Natural Science Foundation of China (32201524), the National Key Research and Development Program of China during the 14th five-year plan Period (2022YFD2200203), the Fundamental Research Funds for the Central Non-profit Research Institution of CAF (CAFYBB2022SY017) and the Guangzhou Science and technology plan project (2023A04J0711).

Author contribution

JL designed and supervised the whole experiments, and wrote the manuscript. JL and GL collected plant materials. JL, LZ and CG analyzed the data and edited manuscript. JL, ZL and JX revised manuscript. All authors approved the manuscript before submission.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Bucher P, Trifonov EN. CCAAT box revisited: bidirectionality, location and context. J Biomol Struct Dyn. 1988;5(6):1231–6.

2. Crombrugghe SNMaBd. Role of the CCAAT-binding protein CBF/NF-Y in transcription. Trends Biochem Sci. 1998;23:174–8.
3. Giuliana Gusmaroli CT, Roberto M. Regulation of novel members of the Arabidopsis thaliana CCAATbinding nuclear factor Y subunits. Genes. 2002;283:41–8.
4. Siefers N, Dang KK, Kumimoto RW, Bynum WE, Tayrose G, Holt BF 3. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant Physiol. 2009;149(2):625–41.
5. Thirumurugan T, Ito Y, Kubo T, Serizawa A, Kurata N. Identification, characterization and interaction of HAP family genes in rice. Mol Genet Genomics. 2008;279:279–89.
6. Quach TN, Nguyen HT, Valliyodan B, Joshi T, Xu D, Nguyen HT. Genome-wide expression analysis of soybean *NF-Y* genes reveals potential function in development and drought response. Mol Genet Genomics. 2015;290(3):1095–115.
7. Wang J, Jin Z, Zhou M, Yu Y, Liang M. Characterization of NF-Y transcription factor families in industrial rapeseed (*Brassica napus* L.) and identification of *BnNF-YA3*, which functions in the abiotic stress response. Ind Crops Prod. 2020; 148.
8. An Y, Suo X, Niu Q, Yin S, Chen L. Genome-wide identification and analysis of the NF-Y transcription factor family reveal its potential roles in salt stress in alfalfa (*Medicago sativa* L.). Int J Mol Sci. 2022; 23(12).
9. Wei Q, Wen S, Lan C, Yu Y, Chen G. Genome-wide identification and expression profile analysis of the NF-Y transcription factor gene family in *Petunia hybrida*. Plants (Basel). 2020; 9(3).
10. Li J, Gao K, Khan WU, Yang X, Yang X, Zhao T, et al. Genome-wide analysis of the poplar NF-Y gene family and its expression in floral bud development of *Populus tomentosa*. Trees. 2019;34(1):285–96.
11. Guo Y, Niu S, El-Kassaby YA, Li W. Transcriptome-wide isolation and expression of NF-Y gene family in male cone development and hormonal treatment of *Pinus tabuliformis*. Physiol Plant. 2021;171(1):34–47.
12. Pereira SLS, Martins CPS, Sousa AO, Camillo LR, Araujo CP, Alcantara GM et al. Genome-wide characterization and expression analysis of citrus NUCLEAR FACTOR-Y (NF-Y) transcription factors identified a novel NF-YA gene involved in drought-stress response and tolerance. PLoS ONE. 2018; 13(6), e0199187.
13. Braybrook SA, Harada JJ. LECs go crazy in embryo development. Trends Plant Sci. 2008;13(6):624–30.
14. Hou X, Zhou J, Liu C, Liu L, Shen L, Yu H. Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. Nat Commun. 2014;5:4601.
15. Wei Q, Ma C, Xu Y, Wang T, Chen Y, Lu J, et al. Control of chrysanthemum flowering through integration with an aging pathway. Nat Commun. 2017;8(1):829.
16. YL YZ, Huang GX, Meng ZD, et al. PwHAP5, a CCAAT-binding transcription factor, interacts with *PwFKBP12* and plays a role in pollen tube growth orientation in *Picea wilsonii*. J EXP BOT.

- 2011;62(14):4805–17.
17. Bai AN, Lu XD, Li DQ, Liu JX, Liu CM. NF-YB1-regulated expression of sucrose transporters in aleurone facilitates sugar loading to rice endosperm. *Cell Res.* 2015.
 18. Zhou Y, Zhang Y, Wang X, Han X, An Y, Lin S, et al. Root-specific NF-Y family transcription factor, *PdNF-YB21*, positively regulates root growth and drought resistance by abscisic acid-mediated indoleacetic acid transport in *Populus*. *New Phytol.* 2020;227(2):407–26.
 19. Li S, Li K, Ju Z, Cao D, Fu D, Zhu H, et al. Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening. *BMC Genomics.* 2016;17:36.
 20. Li WX, Oono Y, Zhu J, He XJ, Wu JM, Iida K, et al. The Arabidopsis NF-YA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell.* 2008;20(8):2238–51.
 21. Ma XJ, Yu TF, Li XH, Cao XY, Xu ZS. Overexpression of *GmNF-YA5* confers drought tolerance to transgenic Arabidopsis and soybean plants. *BioMed Central.* 2020; (1).
 22. Sato H, Suzuki T, Takahashi F, Shinozaki K, Yamaguchi-Shinozaki K. NF-YB2 and NF-YB3 Have Functionally diverged and differentially induce drought and heat stress-specific genes. *Plant Physiol.* 2019;180(3):1677–90.
 23. Wang Y, Xu W, Chen Z, Han B, Haque ME, Liu A. Gene structure, expression pattern and interaction of Nuclear Factor-Y family in castor bean (*Ricinus communis*). Volume 247. *Planta*; 2017. pp. 559–72. 3.
 24. Shi H, Ye T, Zhong B, Liu X, Chan Z. AtHAP5A modulates freezing stress resistance in Arabidopsis through binding to CCAAT motif of *AtXTH21*. *New Phytol.* 2014;9(2):554–67.
 25. Mendes A, Kelly AA, Van Erp H, Shaw E, Powers SJ, Kurup S, et al. bZIP67 regulates the omega-3 fatty acid content of Arabidopsis seed oil by activating fatty acid desaturase3. *Plant Cell.* 2013;25(8):3104–16.
 26. Soyano T, Kouchi H, Hirota A, Hayashi M. NODULE INCEPTION Directly targets NF-Y subunit genes to regulate essential processes of root nodule development in *Lotus japonicus*. *PLoS Genet.* 2013; 9(3), e1003352.
 27. Zanetti M, Eugenia, Blanco, Flavio A, et al. A C Subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean\\u2013 rhizobium etli symbiosis. *Plant Cell.* 2010;22(12):4142–57.
 28. Baoyuan, Qu, Xue, He, Jing W et al. A wheat CCAAT box-binding transcription factor increases the grain yield of wheat with less fertilizer input. *Plant Physiol.* 2015.
 29. Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. *Nucleic Acids Res.* 2009;37:D229–232.
 30. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31(8):1296–7.

31. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;34:W369–373.
32. Magali L, Patrice D, Gert T, Kathleen M, Yves M, Yves VDP et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002; (1) 1.
33. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13(8):1194–202.
34. Zhang S, Tong Y, Li Y, Cheng ZM, Zhong Y. Genome-wide identification of the *HKT* genes in five Rosaceae species and expression analysis of *HKT* genes in response to salt-stress in *Fragaria vesca*. *Genes Genomics.* 2019;41(3):325–36.
35. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012; 40(7), e49.
36. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947–8.
37. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33(7):1870–4.
38. Thirumurugan T, Ito Y, Kubo T, Serizawa A, Kurata N. Identification, characterization and interaction of HAP family genes in rice. *Mol Genet Genomics.* 2008;279(3):279–89.
39. Yang JY, Yongxue. Genome-wide identification and expression analysis of NF-Y transcription factor Families in watermelon (*Citrullus lanatus*). *HortTechnology.* 2017; 27(4).
40. Gusmaroli G, Tonelli C, Mantovani R. Regulation of novel members of the *Arabidopsis thaliana* CCAAT-binding nuclear factor Y subunits. *Gene.* 2002;283(1–2):41–8.
41. Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science.* 2006;312:1040–3.
42. Turck F, Fornara F, Coupland G. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Annu Rev Plant Biol.* 2008;59(1):573–94.
43. Quach TN, Nguyen HTM, Valliyodan B, Joshi T, Nguyen HT. Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. *Mol Genet Genomics.* 2015;290(3):1095.

Tables

Table 1 is available in the Supplementary Files section.

Figures

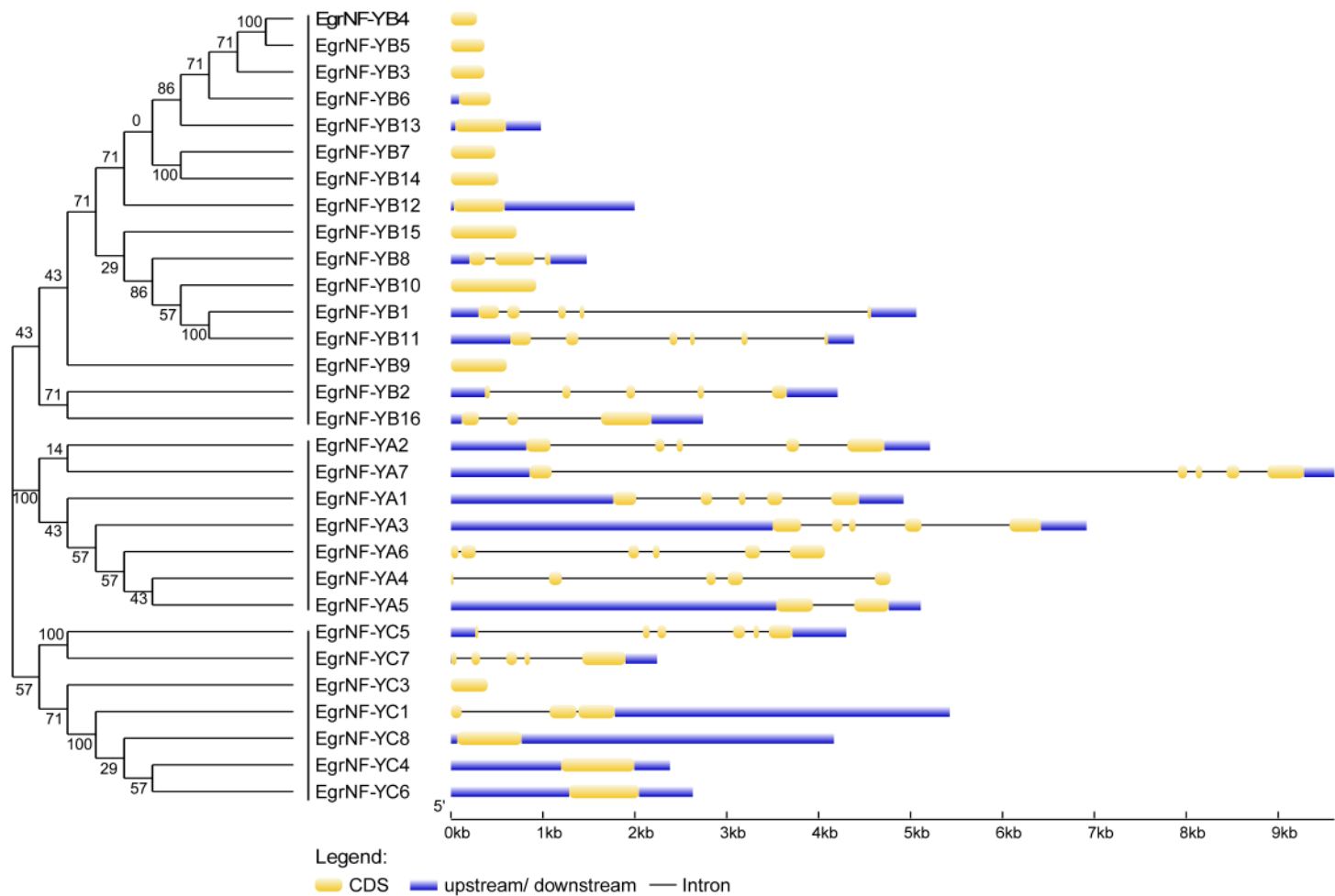


Figure 1

Phylogenetic and gene structure analyses of the *EgrNF-Y* genes. The intron is represented by a line and exons are represented by yellow boxes.

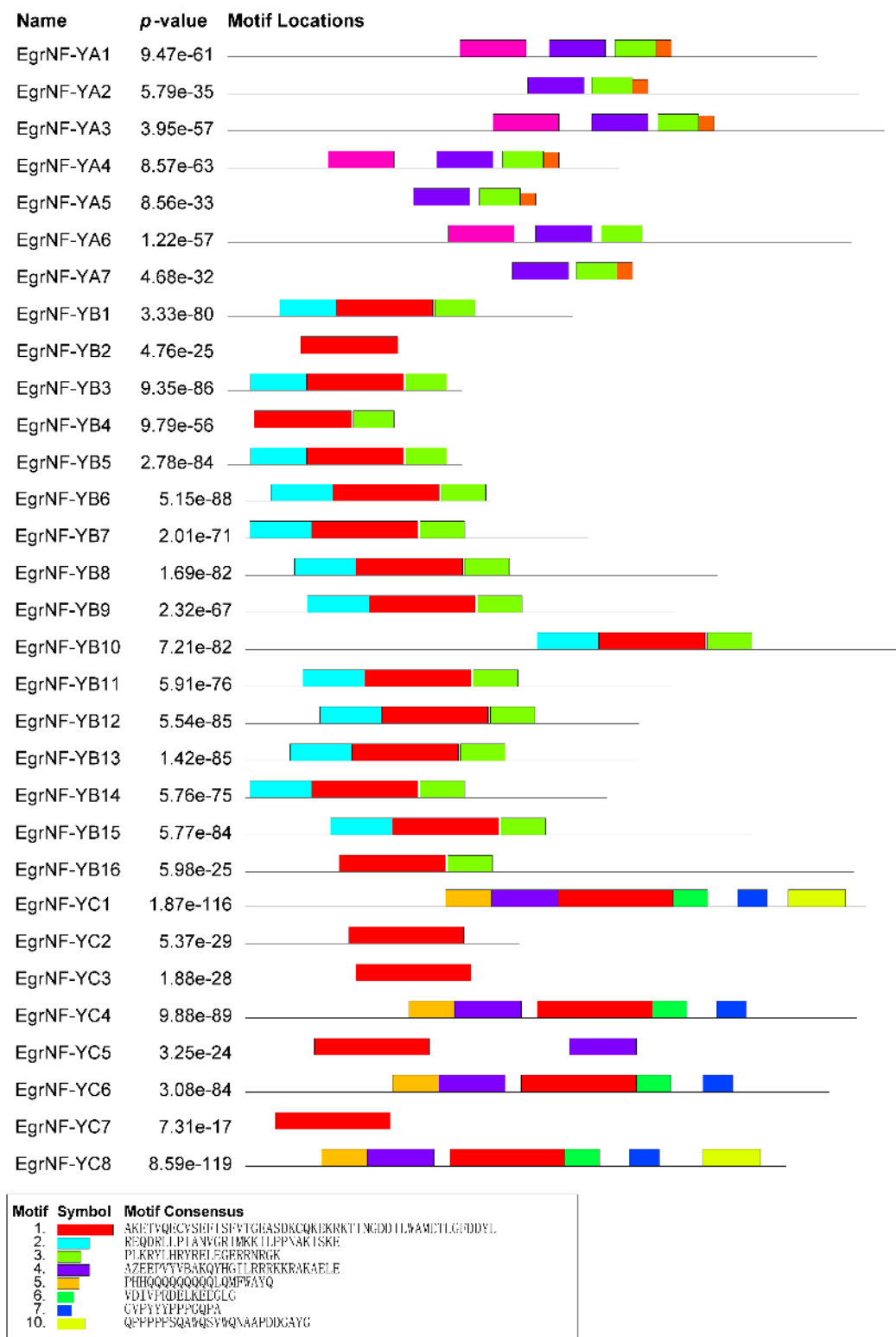


Figure 2

Conserved motif analyses of *E. grandis* and *Arabidopsis* NF-Ys. Conserved motifs (1-10) are represented by different colored boxes, and nonconserved sequences are indicated by gray line (sequences of conserved motifs are given in Figure S1).

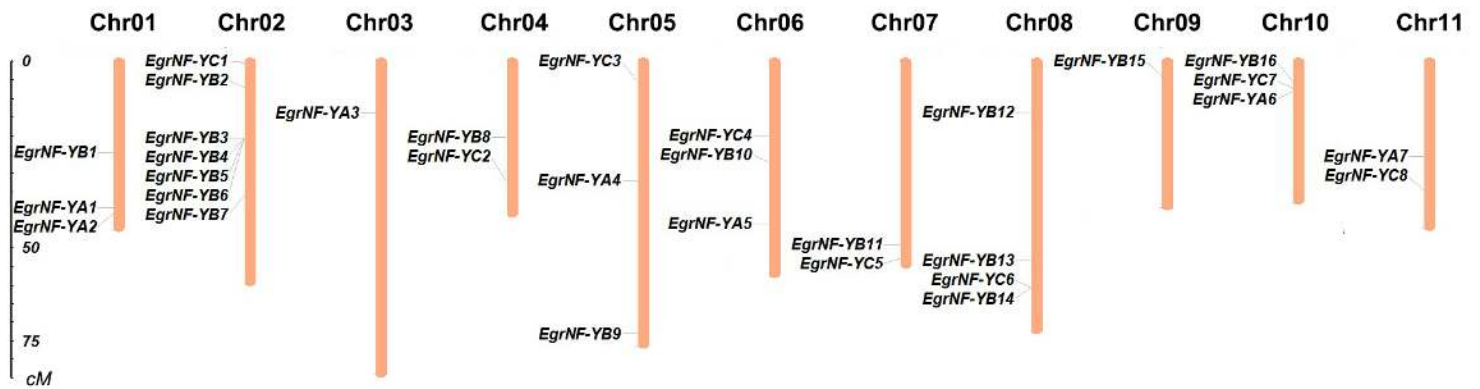


Figure 3

Chromosomal location and gene duplication of NF-Y proteins in *E. grandis*. Chromosome numbers are shown at the top of each bar. Chromosome size is indicated by the vertical scale.

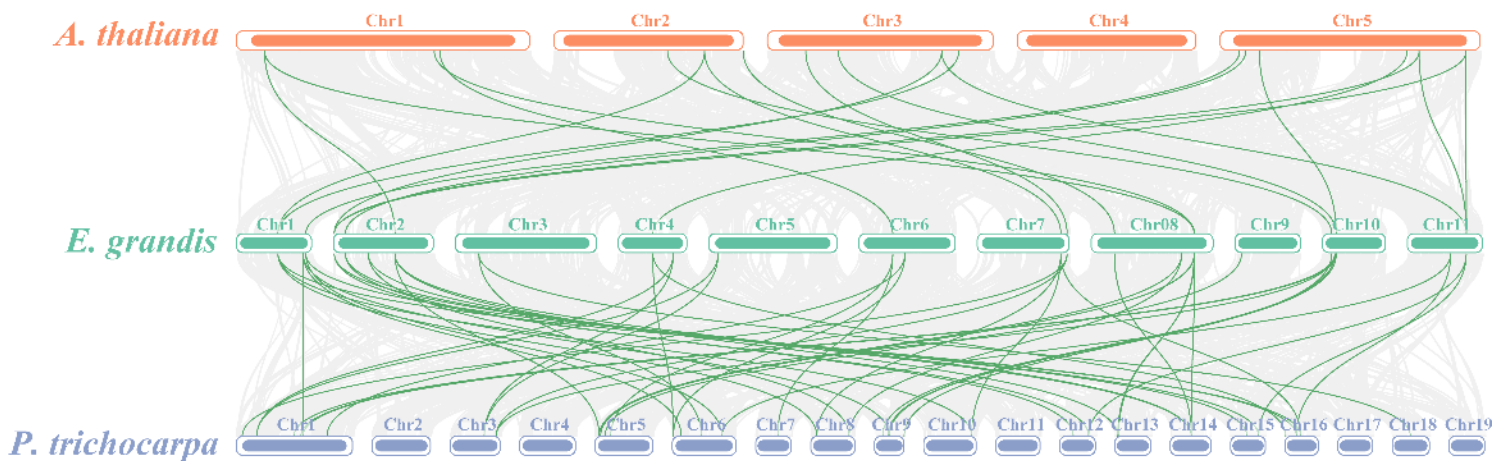


Figure 4

Synteny analyses of NF-Y genes between *E. grandis* and two model plant (*A. thaliana* and *P. trichocarpa*). Green lines indicate the collinear blocks with in *E. grandis* and other two model plant genomes.

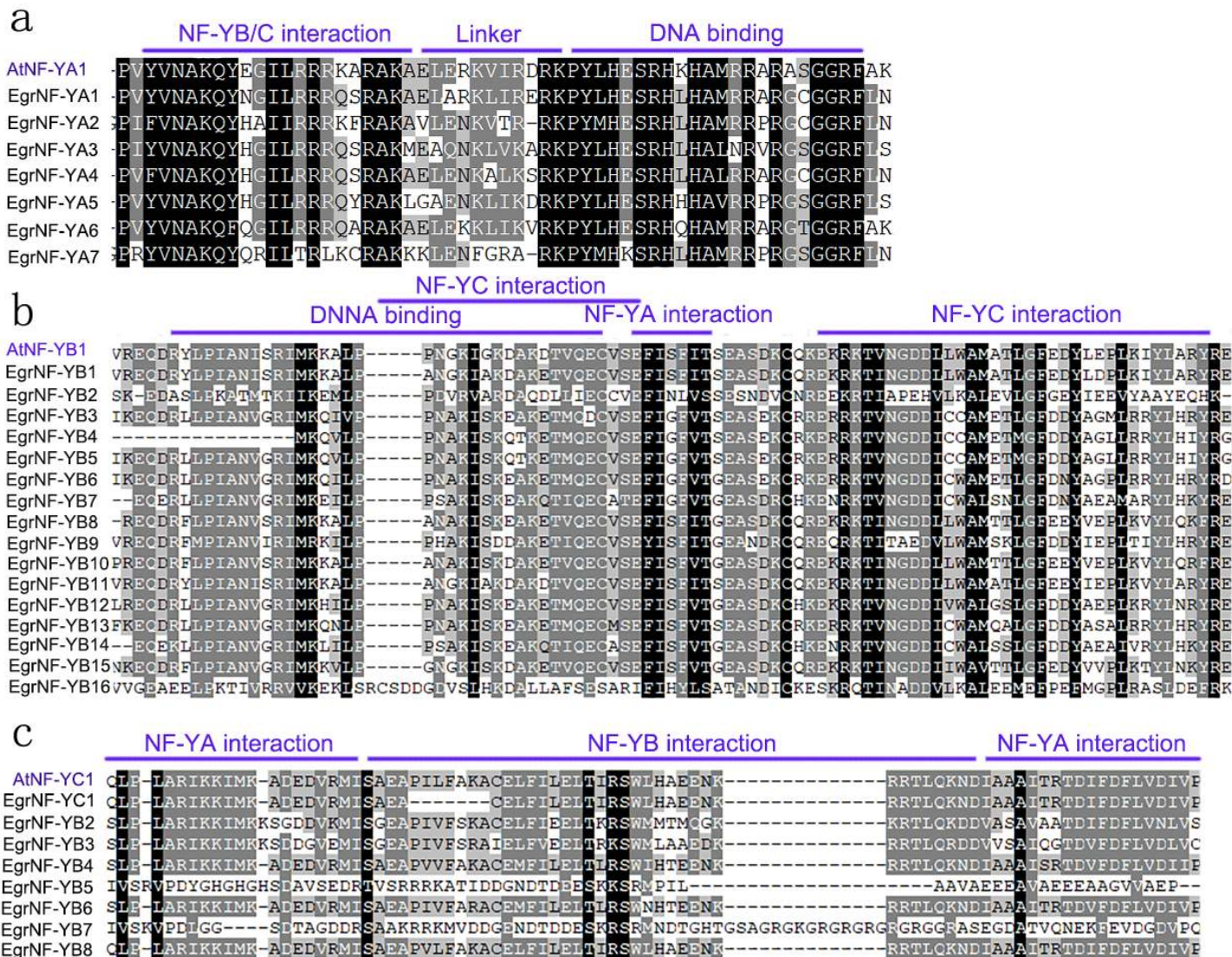


Figure 5

Multiple sequence alignments of *E. grandis* NF-Y TF members. a. EgrNF-YA proteins. **b.** EgrNF-YB proteins. **c.** EgrNF-YC proteins. The AA numbers of the conserved domains and the functional regions are marked.

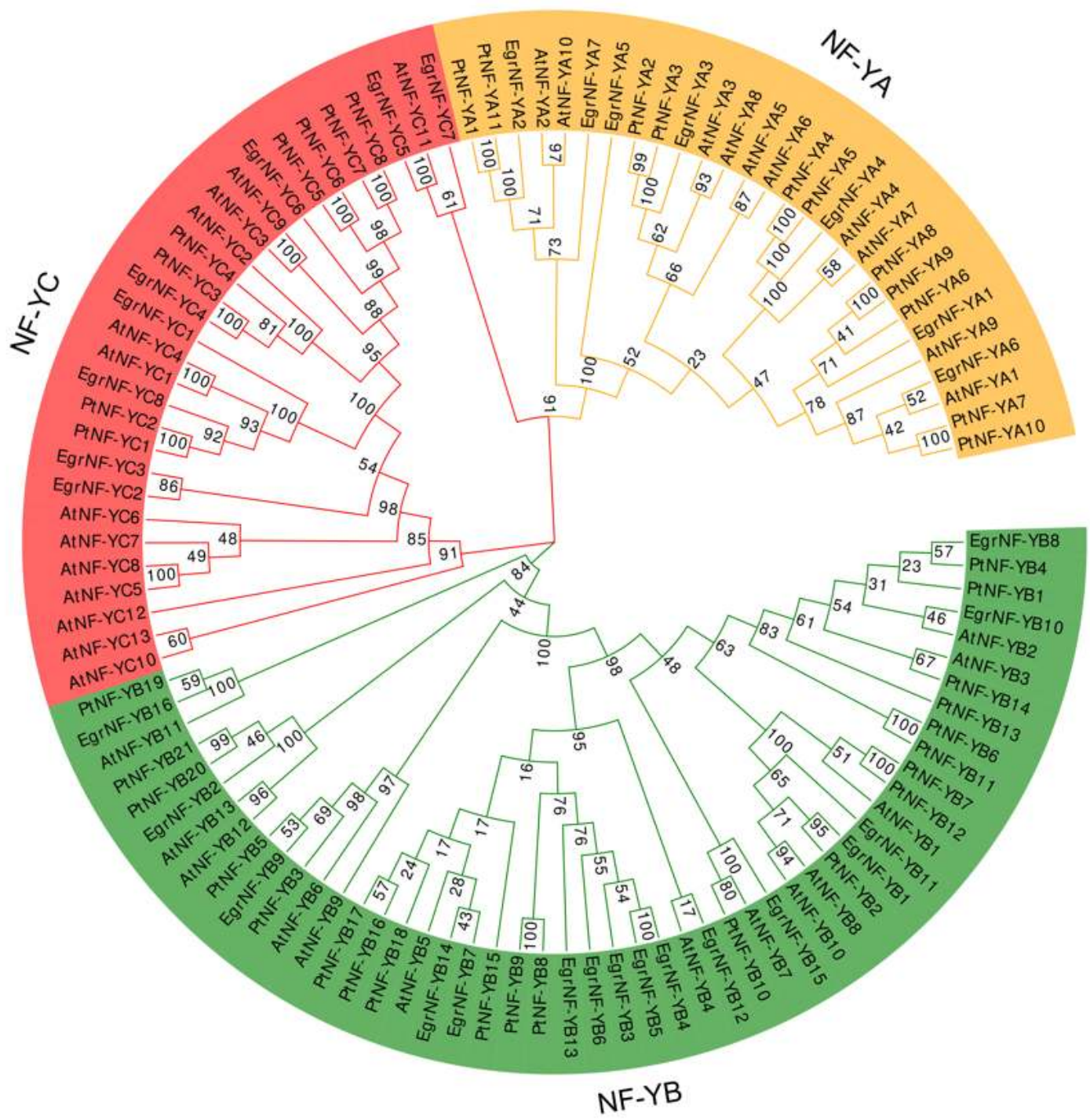


Figure 6

Phylogenetic analysis of *E. grandis*, *Arabidopsis* and poplar NF-Y proteins. Three branches were formed corresponding to genes with different subunits. Pinkish red represents branches of NF-YC, green represents NF-YB and yellow represents NF-YA.

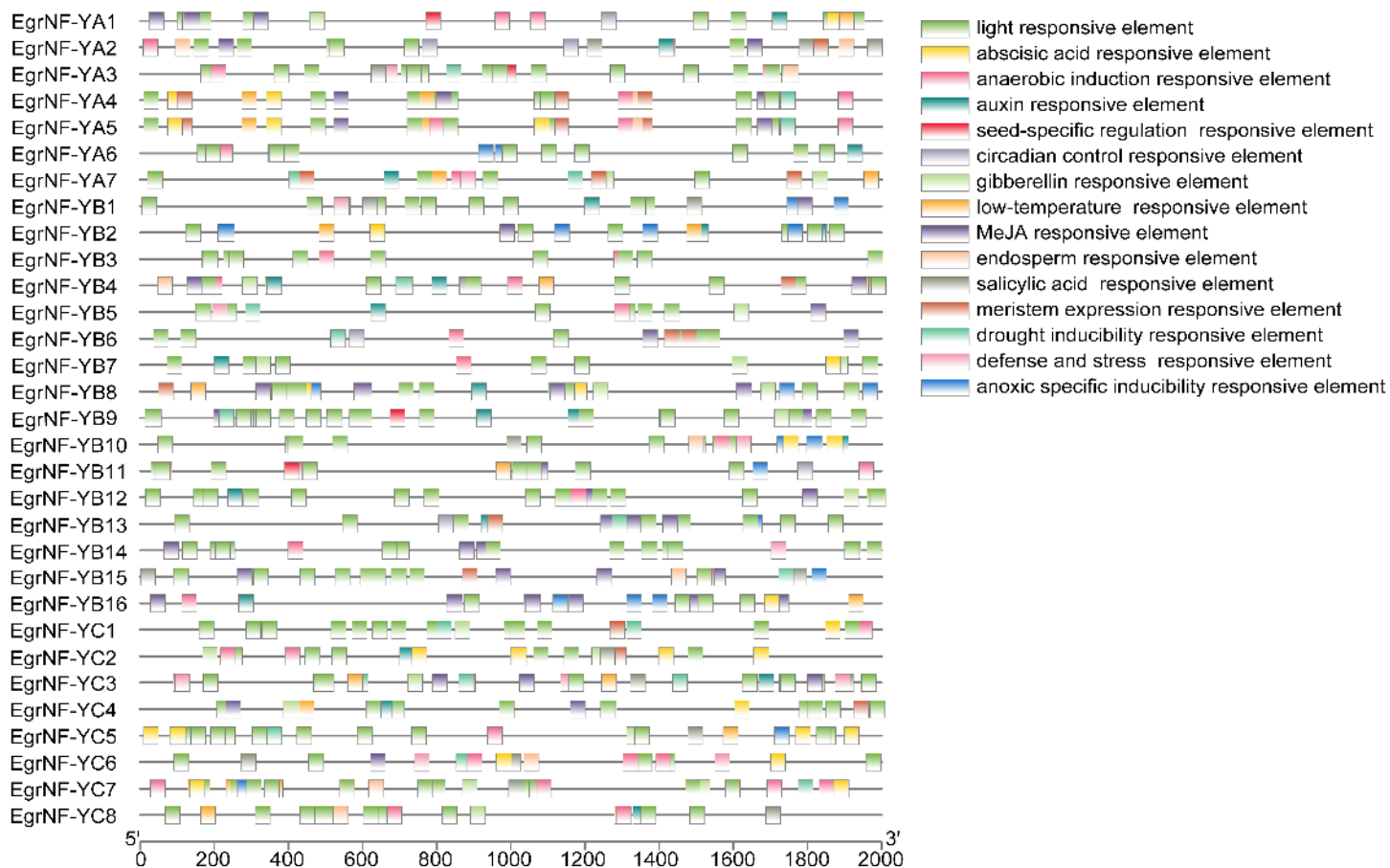


Figure 7

Analysis of *cis*-regulatory elements in the promoter region of *EgrNF-Y* gene in *E. grandis*. The *cis*-elements are marked with different-colored boxes in their respective positions.

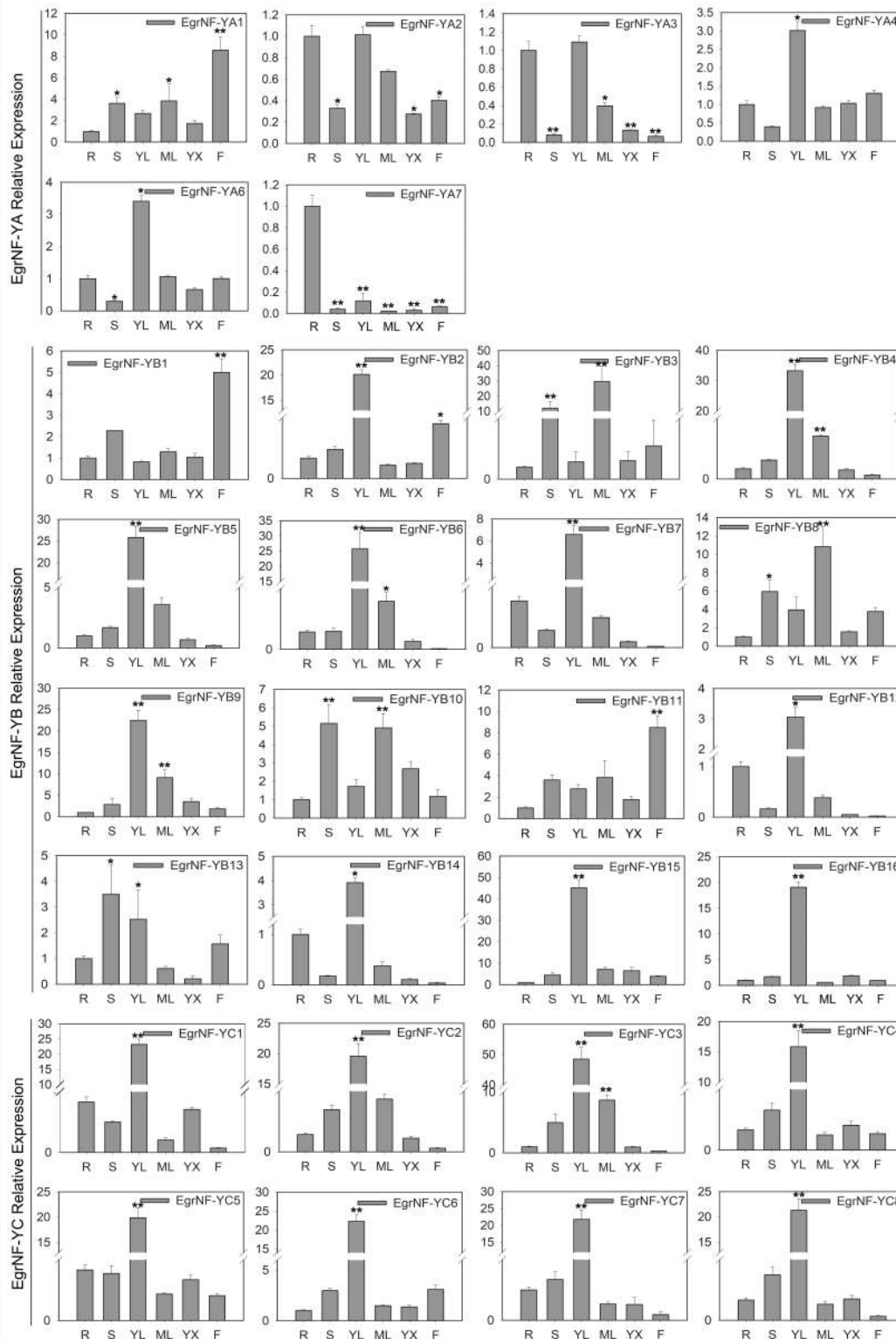


Figure 8

Tissue-specific expression patterns of *EgrNF-Y* in *E. grandis*. R, S, YL, ML, YX and F represent roots, stems, young leaves, mature leaves, xylem and flowers, respectively. Data are mean $\pm$ SE of three biological replications. Asterisks on top of the bars indicate statistically significant differences between the stress and counterpart controls (* $p < 0.05$, ** $p < 0.01$)

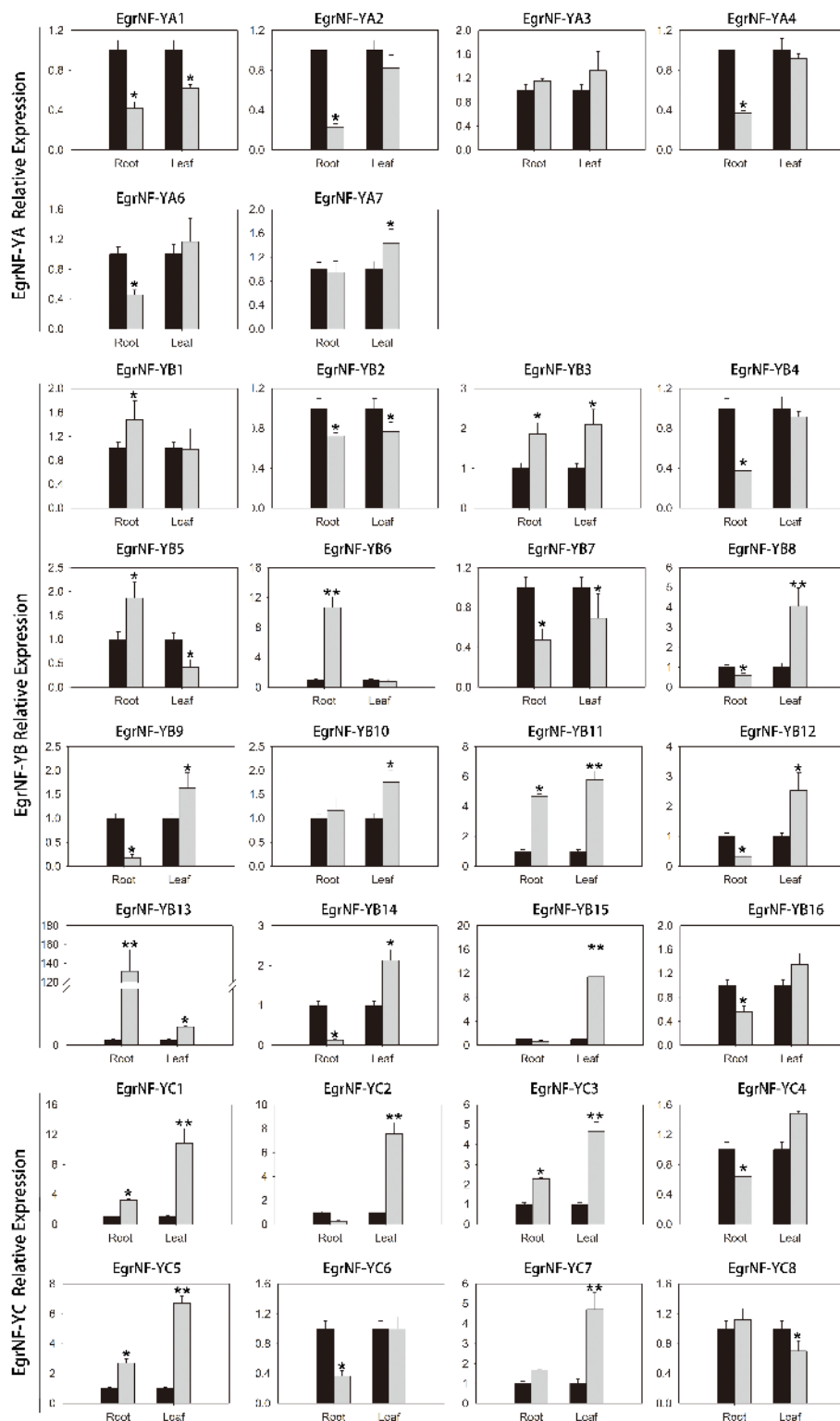


Figure 9

Expression pattern analysis of *EgrNF-Y* gene in *E. grandis* seedlings under phosphorus deficiency condition. The black column represents normal growth condition (KH_2PO_4 concentrations of 1mM), and the gray column represents phosphorus deficient growth conditions (KH_2PO_4 concentrations of 0M). The roots and leaves were collected for gene expression analysis, respectively. Error bars show the standard

deviation of three biological replicates. Single and double asterisks represent significant differences from the control sample at the 0.05 and 0.01 levels (*t-test*), respectively.

Supplementary Files

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

- [Table1.docx](#)
- [SupplementaryFile.zip](#)