Genome-wide identification and expression analysis of the NF-Y family of transcription factors in tea plant

Yi Yue (✉ yyyue@ahau.edu.cn)

Anhui Agricultural University, Hefei, Anhui, P.P. China <https://orcid.org/0000-0002-6520-7171>

Di Zhang

Anhui Agricultural University

Huimin Dou

Anhui Agricultural University

Xinyi Liu

Anhui Agricultural University

Lewen Gao

Anhui Agricultural University

Yiqiong Chen

Anhui Agricultural University

Guohua Fan

Anhui Agricultural University

Youhua Zhang

Anhui Agricultural University

Yunzhi Wu

Anhui Agricultural University

Research article

Keywords: Tea plant, NF-Y transcription factor, Phylogenetic analysis, Expression patterns, Transcriptome sequencing

Posted Date: June 4th, 2019

DOI: <https://doi.org/10.21203/rs.2.10027/v1>

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

[Read Full License](#)

Abstract

Background Nuclear factor-Y(NF-Y) is a heterotrimeric transcription factor commonly found in higher eukaryotes. It usually consists of three subunits: NF-YA, NF-YB and NF-YC, which binds to CCAAT sequences to regulate the expression of target genes. Although NF-Y proteins have been reported in many plants like Arabidopsis and wheat, a comprehensive and systematic analysis of NF-Y genes in tea plant has not yet been reported. **Results** In this study, a total of 35 CsNF-Y genes, including 10 CsNF-YA, 16 CsNF-YB, and 9 CsNF-YC genes, were identified in the sequenced genomes of tea plant. The conserved domains analysis and multiple alignments suggested that domains were conserved within subgroups but differed between subgroups in the CsNF-Y family. Functional prediction and protein interaction analysis were performed by phylogenetic analyses with orthologous genes. Further, expression pattern analysis in various tissues of tea plant by transcriptome data indicated that these genes play different roles in tea plant development and many exhibit tissue-specific expression patterns. **Conclusions** In this study, we characterized 35 CsNF-Ys based on their distributions on tea plant genome, gene structures, phylogenetic relationship with Orthologous NF-Y genes, and their expression patterns. This study provides extensive evaluation of the tea plant NF-Y family and may contribute to the future studies about the CsNF-Y gene family.

Background

The CCAAT box is a cis-acting element that is widely found in 30% of eukaryotes promoters and plays a key role in regulating gene expression[1].The promoter element are usually located in the region 60 to 100 bp upstream of the transcription start site[2].Among homologous genes of different species, the CCAAT box is highly conserved, especially in the positional direction and nearby nucleotide sequences[3]. According to the regulatory role of transcription factors, genes under the control of a promoter containing a CCAAT sequence may be expressed universally, or may be expressed in different tissues and special stages[4]. Numerous experiments have demonstrated the importance of the biological function of the CCAAT box, whether in yeast ,mammals and plants[5-7].

Nuclear factor-Y (NF-Y), also known as CCAAT-binding factor (CBF) and heme activator protein (HAP) is a class of transcription factor which can be combined with CCAAT box in the promoter sequence[8]. Complex transcription factors have multiple subunits, and the spatial and temporal deployment of individual subunits (such as expression levels, post-translational modifications, and subcellular localization) has profound effects on their function[9]. The NF-Y was first founded in yeast[10], which includes at least three subunits: NF-YA (CBF-B or HAP2), NF-YB (CBF-A or HAP3), and NF-YC (CBF-C or HAP5). In addition, researcher only can found the fourth subunit so-called HAP4 in fungi where it provides the transcriptional activation domain[11]. All subunits contain evolutionarily conserved DNA binding and subunit interaction domains to form heterotrimeric complexes[12,13].

Each type of NF-Y protein in mammals and yeast is encoded by a single exon, but the three subunits of plants are encoded by multiple exons, indicating that NF-Y proteins have more complex regulation in

plants[14]. In yeast and animals, each of NF-Y protein is encoded by one or two exons, which act as a heterotrimer. In contrast, each NF-Y subunit has evolved into a polygene-encoded family with a large number of heterotrimeric combinations in plants[15,16]. For example, mice and humans encode only one copy of each subunit[17], while 10 NF-YAs, 13 NF-YBs, and 13 NF-YCs are encoded by the dicot model plant *Arabidopsis thaliana* and in *Triticum aestivum*, there are 10, 11 and 14 copies of NF-YA, NF-YB and NF-YC, respectively[18].

Numerous studies have shown that NF-Y proteins specifically bind to highly conserved CCAAT sequences in their target gene promoters, and each NF-Y subunit contains conserved domain associated with DNA binding and protein interactions[19,20]. The NF-Y subunits A is normally localized to the nucleus, and most NF-YA proteins bind to CCAAT cis elements in the promoter regions of the target gene, but have different functions[21,22]. It is known that NF-YA proteins have two conserved domains, one in the C-terminal region, which is related to DNA-binding, the other in the N-terminus, which is related to the interaction with the NF-YB/NF-YC heterodimer[23]. The NF-Y subunits B and NF-Y subunits C contain highly conserved Histone Folding Domains (HFDs) associated with histone H2B and H2A structures[24], respectively, which functions in protein-DNA and protein-protein interactions[25]. It is worth noting that the NF-YB protein has no nuclear localization signal (NLS), so it must interact with NF-YC in the cytoplasm to transfer to the nucleus, and combine the heterodimer with NF-YA to form the final heterotrimer[26,27]. The HFD domains of the NF-YB and NF-YC subunits assemble into a dimer in a head-to-tail fashion, resulting in DNA binding and structural scaffold interaction with the NF-YA subunit[28]. This research suggests that the regulation of various NF-Y proteins in plants is more diverse and complex than in other organisms. These numerous subunits implies the possibility of genetic redundancy and functional differentiation[29]. These subunits assemble into heterotrimers and bind to the promoter CCAAT box to regulate the expression of many genes[30].

In mammals and yeast, the NF-Y gene possesses many important biological functions, like regulation of energy metabolism[31], response to sugar signal[32], activation of mitochondria-related biosynthesis[33] and coordinating gene expression related to the cell cycle[34]. For example, the degree of NF-YA and CCAAT combination is closely related to human age[35,36]. In *Kluyveromyces lactis*, an additional subunit (HAP4) is proved that can provide powerful acidic activation surface to the complex by associates with the trimer[37]. Many studies have shown that the NF-Y gene family also play an important role in plants, like embryonic development[38], photosynthesis[39], controlling florescence[40] and stress response[41]. *Arabidopsis thaliana* LEAFY COTYLEDON1 (LEC1) and LEAFY COTYLEDON 1 LIKE (L1L) genes are the first nuclear transcription factors of NF-YB family reported in plants, which play an important role in regulating embryo development and embryo maturation[42,43]. In addition, the AtNF-YB2 gene can promote early flowering by activating the FLOWERING LOCUS (FT) and SUPPRESSOR OF OVEREXPRESSION OF CO1 (SOC1) genes[44,45]. Recent studies have shown that AtNF-YA5 expression is up-regulated under drought stress and abscisic acid (ABA) response, and AtNF-YA5 loss-of-function mutant is sensitive to drought stress, suggesting that AtNF-YA5 may enhance the expression of drought-resistance genes[46]. OsNF-YB7/L1L in rice not only plays a role in vegetative growth stage, but also participates in the regulation of flower meristems. Overexpressing OsNF-YB7/L1L plants are short,

densely flocculated, and the inflorescence axis is abnormally developed[47].In barley, overexpression of HvNF-YB1 leads to early flowering time, and NF-Ys regulate the flowering process by combining with the CCT domain in VRN2 (the vernalization pathway) and CO (photoperiod pathway) in wheat[48].

Recently, more and more plants have been discovered and studied in the NF-Y gene family, such as *Brachypodium distachyon*[49], *Zea mays*[50], *Solanum lycopersicum*[51], *Glycine max*[52], *Ricinus communis*[53], *Poplar*[54], *Physcomitrella patens*[55], and *Vitis vinifera* L.[56]. Although the NF-Y gene family has been extensively studied in the large number of plant species, there are no systematic investigations of this gene family have been reported in tea plant(*Camellia sinensis* L.).

Tea plant is an important evergreen, thermophilic, hydrophilic and woody perennial crop belonging to the *Camellia* family [58].The tea made from the young leaves of tea plants is the oldest in the world (from AD) and the most popular non-alcoholic caffeinated beverages since the first 3,000 years [61].In addition to its attractive aroma and pleasant taste, tea beverages also have many health and medicinal properties for humans due to many unique secondary metabolites in tea, such as polyphenols[62], Caffeine[63], and theanine[64]. During the growth of tea plant, the environment is a key factor affecting the growth and development of tea plant. For example, changes in temperature, soil moisture and salinity will affect the growth and harvesting of tea plant to varying degrees. Therefore, improving the stress tolerance of tea plant is essential for sustainable tea production. At present, many researchers have identified some stress-responsive tea plant genes, including genes heat shock proteins (HSPs)[65], HD-Zip[66] and SBP-Box[67]genes.The recently available whole-genome sequences offer a great opportunity for identification and characterization of the NF-Y family in tea plant, which could provide valuable information in understanding the functions of the NF-Y genes underlying the regulation of growth and development of tea plant.

In this study, we identified nine NF-YAs, sixteen NF-YBs and ten NF-YCs in *Camellia sinensis* using sequence information from *Arabidopsis thaliana*. Subsequently, we systematically analysis the physicochemical properties of NF-Y genes including gene structure, isoelectric point , hydrophobic property and protein stability. We also studied their conserved domains, phylogeny, divergence time and putative cis-regulatory elements in the promoters of these genes were conducted. Furthermore, expression difference of the CsNF-Y genes in different tissues and under cold and drought treatments were also detected by transcriptome analysis. The results obtained in this study will provide a reference for future studies of NF-Y family genes in tea plant growth and development.

Results

Identification of NF-Y genes in *Camellia sinensis*

All *Arabidopsis* NF-Y protein sequences were used as queries for tBLASTn program to search the CsNF-Y proteins. In total, there are 45 putative NF-Y genes were founded. After Pfam search for the conserved

domains of NF-Y proteins, a total of 35 NF-Y genes were identified and confirmed, including 10 NF-YA,16 NF-YB and 9 NF-YC genes (Table 1).

Table 1 NF-Y transcription factors in *Camellia sinensis*

Protein properties and sequence analysis of CsNF-Y genes

Characteristics of the 35 CsNF-Ys are shown in Table 2. The length of the amino acid sequence sequences ranged from 118 (CsNF-C2) to 881 (CsNF-YA10). The predicted molecular weights of the CsNF-Y genes products range from 12.86 kDa (CsNF-C2) to 98.50 kDa (CsNF-A10), with a mean value of 34.34 kDa. The isoelectric points (pIs) of the WKRY proteins ranged from 4.87 (CsNF-YB6) to 9.47 (CsNF-YA2), most values approximately 5.0 or 6.0. The protein subcellular localization prediction showed that all CsNF-Y proteins were located in the nucleus.

Table 2 The detailed information of CsNF-Y members

Gene structure and conserved motifs

In order to gain an insight into the diversity of NF-Y genes in tea plant, the exon/intron organization and conserved motifs were further analyzed. The GSDS2.0 online tool was used to determine the gene structure of each predicted CsNF-Y gene (Fig. 1). According to the statistics, 34.3% of genes contain no introns, 11.4% of genes containing 1 to 3 introns, and 54.3% of genes containing more than 3 introns. Using MEME web server, we predicted 10 conserved motif distributions with the full-length protein sequences of the three CsNF-Y subfamilies. Experiments show that all of the genes have functional motifs, and there are significant differences in the distribution of motifs before the different groups, that basically consistent with the predictions of the previous conserved domains. For 10 NF-YA proteins, all of them shared motif 3 and motif 4, and 9 members contained motif 8. In the case of 16 NF-YB members, two motifs (motif 1,2) were widely distributed. All of NF-YC members had the motif 1 and motif 3, and 5 NF-YC members contained motif 6, 7 and 9.

Fig.1 Exon-intron structure of the CsNF-Y family generated from GSDS2 online website. The exons and introns are represented by colored boxes and black lines, respectively.

Fig.2 Conserved motifs of the CsNF-Y proteins were displayed. The conserved motifs in the CsNF-Y proteins were identified using Multiple Em for Motif Elicitation (MEME). In total, 10 motifs were identified and are shown in different colors. Motif locations are also indicated.

Multiple alignments and phylogenetic analysis

We combined the tea plant and Arabidopsis NF-Y gene family with Clustal X software for multiple sequence alignments to show the specific location of the conserved domain. It can be seen from the

figure that each member of the NF-Y family contains a heterodimerization domain and a DNA-binding domain that recognizes the CCAAT site. In the CsNF-YA class protein sequence, there is a 20-amino acid alpha-helical structure called NF-YB/C interaction domain, located at the N-terminus of the core region, which is capable of interacting with the NF-YB and NF-YC subunits. The interaction plays a key role in the interaction with NF-YB and NF-YC subunits. Another conserved domain linked to the NF-YB/C interaction domain by Linker is called DNA binding, which consists of 21 amino acids. It primarily regulates specific binding to the CCAAT box. In the CsNF-YB class protein sequence, the central domain has 95 amino acids, and the conserved amino acid sequence of NF-YB interacts with NF-YC covers almost the entire NF-YB conserved domain. The conserved amino acid sequence of NF-YB interacts with NF-YA is the motif FIHYLSA. The central domain of CsNF-YC protein contains approximately 80 amino acid sequences with a conserved domain of interaction between the NF-YB and NF-YA subunits.

The experiment combines some orthologous genes that have been shown to have biological functions with CsNF-Y genes to construct a phylogenetic tree. From the phylogenetic tree of the three groups of NF-Y family, it can be found that the dicotyledono like tea palnt, *Arabidopsis thaliana* and monocotyledon such as rice and maize NF-Y members are staggered and do not form separate branches, which indicates CsNF-Y members are more conservative in the evolutionary system.

By observing the species evolution tree of NF-YA gene, we can find that the CsNF-YA1 gene is on the same branch as the AtNF-YA9 gene of *Arabidopsis thaliana*, and the AtNF-YA9 gene has been found by researchers in plant embryogenesis and seed development[68]. It plays an important role, and we predict that CsNF-YA1 gene is particularly important in the biological development of tea seed development. Similarly, we also predicted that CsNF-YA5 gene was involved in the tea plant maturation process based on the tomato SINP-YA8 and SINP-YA9 genes[51]. The CsNF-YA2 and CsNF-YA8 genes were predicted to participate in the drought-resistant process of tea plant according to the rice OsNF-YA7 gene[69]. At the same time, we observed that CsNF-YB12 gene maybe can regulate the flowering process of tea plant and the process of CsNF-YB14 involved in promoting the flowering of tea plant by observing the species evolution tree of NF-YB gene. Finally, by studying the species evolutionary tree of NF-YC gene, we predicted that CsNF-YC7 gene could promote the flowering process of tea plant.

Fig. 3 Multiple sequence alignments of CsNF-Y and AtNF-Y members. (a) Multiple alignment of CsNF-YA proteins. (b) Multiple alignment of CsNF-YB proteins. (c) Multiple alignments of CsNF-YC proteins. All members from *Arabidopsis thaliana* was included, respectively. Multiple alignments of NF-Y protein sequences were carried out with ClustalX and MEGA7.

Fig. 4 Phylogenetic tree of tea plant NF-Y members. Function prediction for tea plant genes was based on gene expression and homology to the referred characterized proteins. (a) Phylogenetic tree of CsNF-YA members. (b) Phylogenetic tree of CsNF-YB members. (c) Phylogenetic tree of CsNF-YC members.

Gene duplication and divergence time analysis

We first obtained 10 pairs of paralogous gene pairs based on the phylogenetic tree of the CsNF-Y gene family, and calculated the nonsynonymous(Ka) and synonymous(Ks) rate of each group. According to their ratio, which displayed from table3, we found that Ka/Ks were all less than 1, and we determined that the CsNF-Y gene family experienced a purifying selection.

In addition, we have obtained the replication events of these CsNF-Y genes based on the molecular evolution rate of dicotyledons, which occurred between 32.18 million and 3.9 million years ago. Then we chose kiwifruit as the near-source plant of tea plant, and we selected 11 pairs of orthologous genes based on the NF-Y gene family phylogenetic tree of two plants. It was found through calculation that the differentiation of tea plant and kiwifruit occurred between 49 and 13 million years ago. As shown in table 4, we found that the two pairs of genes CsNF-YB3_Ach28g200451.1 and CsNF-YB9_Ach24g427181.2 differentiated much earlier than other gene pairs. Excluding these two pairs of genomes, we found that the average time for tea tree and kiwifruit differentiation was 18 million years ago.

Table 3 The Ka/Ks rate of CsNF-Y replication genes

Table 4 The Ka/Ks rate of tea tree and kiwifruit NF-Y genes

GO and KEGG enrichment analysis

Through experimental analysis of Gene Ontology, it was found that there are 17 types of functional groups involved in the CsNF-Y gene family, including 3 molecular functions, 7 cell component functions and 7 biological processes. Through the analysis of the KEGG metabolic pathway combined with Figure 5, it was found that three genes in the CsNF-Y gene family are related to metabolism, including sputum metabolism and pyrimidine metabolism. There are also four genes related to the processing of genetic information, mainly based on basic excision and repair functions.

Fig. 5 Gene ontology (GO) and metabolic pathway (KEGG) enrichment information of CsNF-Y gene family. (a) GO enrichment information of CsNF-Y members. (b) KEGG enrichment information of CsNF-Y members. Different colors indicate different functions.

Putative promoter region analysis

Based on the analysis of the 1500 bp sequence upstream of the CsNF-Y gene sequences, we obtained a total of 96 different types of cis-elements in the promoter of CsNF-Y genes. Then, we divided it into six classes. Among them, there are some cis-elements (CAT-box, motif I, etc.) related to the growth and development of tea plant, some cis-elements (circadian, TC-rich repeats, etc.) related to the growth environment of tea plant, and another functional elements related to the response of tea plant to light.

After counting all the cis-elements as shown in Figure 6, we found that the functional components of light response of tea plant are the most widely distributed. We speculate that the CsNF-Y gene may regulate

the light conditions of tea plant. In addition, it was found that only CsNF-YA4 possesses the element motif I related to root growth, so we predict that CsNF-YA4 is closely related to the growth of tea roots. Moreover, we found that CsNF-YB6 has a functional element WUN-motif associated with plant damage and recovery function, and we predict that this gene is involved in the damaged recovery function of tea plant.

Fig. 6 Cis-element of CsNF-Y gene family, which based on the analysis of the 1500 bp sequence upstream of the CsNF-Y gene sequences. The figure shows that the functional components of light response of tea plant are the most widely distributed.

Gene expression analysis

In this experiment, the transcriptomes of different tissues of tea plant were analyzed to obtain the differential expression of CsNF-Y gene family in different tissues. According to the heat map as shown in Figure 7, we can clearly find that many genes in the CsNF-Y gene family are differentially expressed in different tissues, and some of them are significantly expressed in a single tissue.

Using detailed analysis of the heat map, we found that CsNF-YA6 is significantly expressed in young leaves, while CsNF-B10, CsNF-B13 and CsNF-C2 are specifically expressed in mature tea leaves. From this we conclude that these genes have an important relationship with the growth and development of tea leaves. At the same time, we also found that CsNF-B12 is differentially expressed in tea flower and CsNF-B4 and CsNF-B15 are abundantly expressed in tea fruit. Based on this, we conclude that these three genes are involved in the growth of tea flower and fruit. In the heat map, we can also find that CsNF-A8, CsNF-B9, CsNF-C4 and CsNF-C9 are only abundantly expressed in the roots of tea plant. We speculate that these genes play an active role in promoting the growth of tea root.

Fig. 7 Expression profile analysis of CsNF-Y genes in different tea plant tissues. Transcriptome data (FPKM) were used to measure the expression levels of CsNF-Y genes in apical bud, flower, fruit, young leaf, mature leaf, old leaf, stem and root . The colored scale for the different expression levels is shown.

Discussion

NUCLEAR FACTOR-Y (NF-Y), also called CCAAT-Binding Factor (CBF) or Heme Activator Protein (HAP), is one of the most widely found transcription factors found in several genes in yeast, animals and plants. It have been revealed to be important factors in multiple physiological processes in plants, like Arabidopsis, rice, maize, grape and tomato. However, there are no systematic investigations of this gene family have been reported in tea plant(*Camellia sinensis* L.). In this study, a comprehensive analysis of tea plant NF-Y gene family was performed, including analyses of sequence feature, phylogeny, gene duplication, promoter region , GO and KEGG enrichment and expression pattern.

In current study, a total of 35 NF-Y subunits, including 10 NF-YA, 16 NF-YB, and 9 NF-YC subunits, were identified based on the genome data of tea plant. In our study, we found that the CsNF-Y gene family have CBF conserved domain sequences, and Histone conserved sequences in both NF-YB and NF-YC gene sequences. These conserved sequences are intrinsic gene sequences of NF-Y gene family. We used the GSDS2.0 program to visualize the CsNF-Y gene structure. The results showed that compared with the relatively consistent intron/exon structure in the NF-YA subunit, most of the NF-YB and NF-YC gene structures were less, which were consistent with that observed in *Arabidopsis*, common bean and *Physcomitrella patens*. These structures of the CsNF-Y gene family are diverse and contain more genes in introns, and the proportion of genes containing more than 3 introns reached 54.3%.

In order to understand the phylogenetic relationship of the CsNF-Y gene family, many orthologous genes are constructed into phylogenetic trees after multiple sequence alignments.

From the phylogenetic tree of the three groups of CsNF-Y gene family, it can be found that NF-Y gene members of the dicotyledonous and monocotyledon are staggered and not form separate branches, which indicates CsNF-Y gene members are more conservative in the evolutionary system. By observing the phylogenetic tree, we predicted that CsNF-YA1 gene plays an important biological function in the seed development stage of tea plant. The CsNF-YA5 gene is predicted to be involved in the tea plant maturation process, and CsNF-YA2 and CsNF-YA8 genes are predicted to participate in the drought-resistant process of tea plant. At the same time, it is predicted that CsNF-YB12, CsNF-YB14 and CsNF-YC7 can regulate the flowering process of tea plants. And the ratio of the non-synonymous substitution rate (K_a) to the synonymous substitution rate (K_s) was less than 1, thereby determining that the CsNF-Y gene family experienced a purification selection. In addition, we have predicted the replication events of these CsNF-Y genes according to the molecular evolution rate of dicotyledons, which occurred between 32.18 million and 3.9 million years ago. The differentiation of tea plant and kiwifruit occurred between 49 and 13 million years ago.

The analysis of gene expression patterns in different tissues will help us to understand the dynamic gene expression of NF-Y genes in tea plant. Therefore, we displayed heat map of the gene expression patterns of 35 identified CsNF-Y genes using published transcriptome data of tea plant different tissues downloaded from NCBI database. According to the analysis of the heat map, our study found that many genes in the CsNF-Y gene family are differentially expressed in different tissues, and some of them are significantly expressed in a single tissue.

After further analysis, we can predict that CsNF-YA6, CsNF-YB10, CsNF-YB13 and CsNF-YC2 have an important relationship with the growth and development of tea leaves. At the same time, we also predicted that CsNF-YB12, CsNF-YB4 and CsNF-YB15 are involved in the growth of tea flower and fruit. In the heat map, we can also find that CsNF-YA8, CsNF-YB9, CsNF-YC4 and CsNF-YC9 are only abundantly expressed in the roots of tea plants. We speculate that these genes play an active role in promoting the growth of tea roots. These results might suggest that CsNF-Y genes had many putative functions in processes of tea plant growth and development.

Conclusions

In the present study, a comprehensive analysis of NF-Y gene family in tea plant was carried out for the first time. The conserved domain, gene structure, phylogenetic relationship, , gene duplication and expression pattern analysis of CsNF-Y genes were investigated. The CsNF-Y genes were divided into three clades: A (10 genes), B (16 genes), C (9 genes), which were supported by gene structural and conserved domain analysis. The biological function of the CsNF-Y proteins was predicted by phylogenetic analysis with orthologous proteins, and this provided a preliminary understood for the less-studied CsNF-Ys. The differential expression patterns of CsNF-Y genes in different tissues of tea plant showed that these genes play different roles in tea plant development and many exhibit tissue-specific expression patterns. This study is the first report to identify CsNF-Y genes in tea plant and may contribute to the future studies about the CsNF-Y gene family.

Materials And Methods

Identification of the NF-Y gene family in tea plant

The sequences of 36 Arabidopsis NF-Y proteins were downloaded from the TAIR database (<http://www.arabidopsis.org>) and PlantTFDB database (<http://planttfdb.cbi.pku.edu.cn>). The Camellia sinensis genome were obtained from the TPIA(<http://tpia.teaplant.org>). The TBLASTN program was applied with an E-value cut-off of $1.0e^{-5}$ to identify predicted CsNF-Y sequences in the tea plant genome. In addition, a Hidden Markov Model (HMM) profile for the NF-Y domain from the Pfam database (<http://pfam.janelia.org>) was also used as a query to confirm all candidate sequences by searching for NF-Y domain sequences. The identified NF-Ys were named according to their order in the scaffold from top to bottom.

Protein properties and sequence analysis

The protein molecular weights (MWs) and isoelectric points (pIs) were predicted using the ExPASy tools (http://web.expasy.org/compute_pi/) based on the deduced NF-Y sequences.

The subcellular localization of the NF-Y proteins were predicted using the TargetP 1.1 tool based on the N-terminal amino acid sequence. The MEME program was used to identify the conserved motifs in the CsNF-Y proteins (parameter setting: maximum number of motifs, 10; maximum width, 50). The gene exon/intron structure of the CsNF-Y genes were determined using the Gene Structure Display Server2.0 (GSDS2.0) [39] by comparing their predicted coding sequence (CDS) with their genomic sequences.

Multiple alignments and phylogenetic analysis

The multiple sequence alignments for CsNF-Y and AtNF-Y protein sequences were performed using CLUSTAL X software with manually adjustment where appropriate for the alignment of the NF-Y domain, then displayed using DNAMAN software. The phylogenetic trees were constructed based on the alignment with the Neighbor Joining method using MEGA7.0 software with the pairwise deletion option and bootstrap tests of 1,000 replicates. The combined phylogenetic tree of selected ZmNF-Y, SINP-Y and OsNF-Y, proteins was generated using the same method.

Gene duplication and divergence time analysis

The predicted CsNF-Y proteins were first aligned by ClustalW2 at MEGA7.0 software to a gene duplication analysis. The MCScanX software (<http://chibba.pgml.uga.edu/mcscan2/>) was used to identify the duplications of CsNF-Y proteins. The synonymous (Ks) and nonsynonymous substitution rate (Ka) of the different gene duplication pairs were estimated using DnaSP 5 software. The Ks values were used to predict the divergence time (T) occurred in tea plant, according the formula: $T = Ks / 2\lambda \times 10^{-6} \text{ Mya}$ ($\lambda = 6.5 \times 10^{-9}$).

GO and KEGG enrichment analysis

In this experiment, the GO and KEGG annotation files of all the genes of tea plant were used as background genes, and the CsNF-Y gene family was used as the foreground gene. The background gene and foreground genes of tea plant were analyzed by the clusterprofiler package in R language to obtain GO and KEGG enrichment analysis of the CsNF-Y gene family.

Putative promoter region analysis

For the promoter analysis, the 1,500 bp upstream sequences of the start codon of each CsNF-Y genes were extracted from the tea plant genome sequence as putative promoter region. The PLACE website was used to search for cis-acting regulatory elements in the putative promoter regions.

Gene expression analysis

To study gene expression levels of CsNF-Y genes in different tissues or development stages. The raw RNA-seq reads from different tea plant tissues were obtained from the NCBI SRA database, including apical bud, flower, fruit, young leaf, mature leaf, old leaf, stem and root .

There raw reads were trimmed to remove low quality adaptor sequences with Trimmomatic tool. Then, the paired clean reads were mapped to the *C. sinensis* reference genome using RSEM with defaults parameters. To visualize the expression patterns of the NF-Y genes in different tea plant tissues, the fragments per kilobase million (FPKM) and heat map were created with EdgeR software.

Abbreviations

NF-Y: Nuclear Factor- κ B; CDS: Coding sequence; CBF: CCAAT binding factor; FPKM: Fragments per kilobase of transcript per million fragments; HMM: Hidden Markov Model; NJ: Neighbor-joining

Declarations

Acknowledgments

We would particularly wish to Prof. Chaoling Wei and Prof. Yunshen Wan for laboratory assistance and feedback on the manuscript.

Funding

This research was supported by National Key Research and Development Project of China (No. 2017YFD0301303) and The Open Fund of State Key Laboratory of Tea Plant Biology and Utilization(KLTOF20150103).

Availability of data and materials

Clean data from transcriptome reads, genome assembly, and gene functional annotations were downloaded from pcsb.ahau.edu.cn:8080/CSS/.

Authors' contributions

DZ and YY designed and performed the experiments; DZ, HD, XL, and LG analyzed the data; YY, DZ, HD,GF and YC contributed reagents/materials/analysis tools; DZ wrote the paper. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹AnHui Provincial Engineering Laboratory of Beidou Precision Agriculture, Anhui Agricultural University, Hefei 230036, China. ²State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, China. ³College of Information and Computer, Anhui Agricultural University, Hefei 230036, China.

References

1. Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences[J]. *Journal of molecular biology*, 1990, 212(4): 563-578.
2. Edwards D, Murray J A H, Smith A G. Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis[J]. *Plant physiology*, 1998, 117(3): 1015-1022./ Mantovani R. A survey of 178 NF-Y binding CCAAT boxes[J]. *Nucleic acids research*, 1998, 26(5): 1135-1143.
3. Mantovani R. The molecular biology of the CCAAT-binding factor NF-Y[J]. *Gene*, 1999, 239(1): 15-27.
4. Stephenson T J, McIntyre C L, Collet C, et al. Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*[J]. *Plant molecular biology*, 2007, 65(1-2): 77-92.
5. Yang P, Taoka K, Nakayama T, et al. Structural and functional characterization of two wheat histone H2B promoters[J]. *Plant molecular biology*, 1995, 28(1): 155-172.
6. Rieping M, Schöffl F. Synergistic effect of upstream sequences, CCAAT box elements, and HSE sequences for enhanced expression of chimaeric heat shock genes in transgenic tobacco[J]. *Molecular and General Genetics MGG*, 1992, 231(2): 226-232.
7. Taoka K, Kaya H, Nakayama T, et al. Identification of three kinds of mutually related composite elements conferring S phase-specific transcriptional activation[J]. *The Plant Journal*, 1999, 18(6): 611-623.
8. Coustry F, Maity S N, Sinha S, et al. The transcriptional activity of the CCAAT-binding factor CBF is mediated by two distinct activation domains, one in the CBF-B subunit and the other in the CBF-C subunit[J]. *Journal of Biological Chemistry*, 1996, 271(24): 14485-14491.
9. Siefers N, Dang K K, Kumimoto R W, et al. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity[J]. *Plant physiology*, 2009, 149(2): 625-641.
10. McNabb D S, Tseng K A, Guarente L. The *Saccharomyces cerevisiae* Hap5p homolog from fission yeast reveals two conserved domains that are essential for assembly of heterotetrameric CCAAT-binding factor[J]. *Molecular and cellular biology*, 1997, 17(12): 7008-7018.
11. Forsburg S L, Guarente L. Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer[J]. *Genes & development*, 1989, 3(8): 1166-1178.
12. Sinha S, Maity S N, Lu J, et al. Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3[J]. *Proceedings of the National Academy of Sciences*, 1995, 92(5): 1624-1628.
13. Sinha S, Kim I S, Sohn K Y, et al. Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex[J]. *Molecular and Cellular Biology*, 1996, 16(1): 328-337.

14. Thirumurugan T, Ito Y, Kubo T, et al. Identification, characterization and interaction of HAP family genes in rice[J]. *Molecular Genetics and Genomics*, 2008, 279(3): 279-289.
15. Kahle J, Baake M, Doenecke D, et al. Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin β and importin 13[J]. *Molecular and cellular biology*, 2005, 25(13): 5339-5354.
16. Kumimoto R W, Adam L, Hymus G J, et al. The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*[J]. *Planta*, 2008, 228(5): 709-723.
17. Zhang Z, Li X, Zhang C, et al. Isolation, structural analysis, and expression characteristics of the maize nuclear factor Y gene families[J]. *Biochemical and biophysical research communications*, 2016, 478(2): 752-758.
18. Stephenson T J, McIntyre C L, Collet C, et al. Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*[J]. *Plant molecular biology*, 2007, 65(1-2): 77-92.
19. Romier C, Cocchiarella F, Mantovani R, et al. The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y[J]. *Journal of Biological Chemistry*, 2003, 278(2): 1336-1345.
20. Steidl S, Tüncher A, Goda H, et al. A single subunit of a heterotrimeric CCAAT-binding complex carries a nuclear localization signal: piggy back transport of the pre-assembled complex to the nucleus[J]. *Journal of molecular biology*, 2004, 342(2): 515-524.
21. Calvenzani V, Testoni B, Gusmaroli G, et al. Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits[J]. *PloS one*, 2012, 7(8): e42902.
22. Petroni K, Kumimoto R W, Gnesutta N, et al. The promiscuous life of plant NUCLEAR FACTOR Y transcription factors[J]. *The Plant Cell*, 2012, 24(12): 4777-4792.
23. Xing Y, Zhang S, Olesen J T, et al. Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short alpha-helix in HAP2[J]. *Proceedings of the National Academy of Sciences*, 1994, 91(8): 3009-3013.
24. Baxevanis A D, Arents G, Moudrianakis E N, et al. A variety of DNA-binding and multimeric proteins contain the histone fold motif[J]. *Nucleic acids research*, 1995, 23(14): 2685-2691.
25. Zhao H, Wu D, Kong F, et al. The *Arabidopsis thaliana* nuclear factor Y transcription factors[J]. *Frontiers in plant science*, 2017, 7: 2045.
26. Frontini M, Imbriano C, Manni I, et al. Cell-cycle regulation of NF-YC nuclear localization[J]. *Cell Cycle*, 2004, 3(2): 205-210.
27. Kahle J, Baake M, Doenecke D, et al. Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin β and importin 13[J]. *Molecular and cellular biology*, 2005, 25(13): 5339-5354.
28. Nardini M, Gnesutta N, Donati G, et al. Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination[J]. *Cell*, 2013, 152(1-2): 132-143.

29. Liang M, Hole D, Wu J, et al. Expression and functional analysis of NUCLEAR FACTOR-Y, subunit B genes in barley[J]. *Planta*, 2012, 235(4): 779-791.
30. Ceribelli M, Dolfini D, Merico D, et al. The histone-like NF-Y is a bifunctional transcription factor[J]. *Molecular and cellular biology*, 2008, 28(6): 2047-2058.
31. MAZÓN M J, GANCEDO J M, GANCEDO C. Phosphorylation and Inactivation of Yeast Fructose-Bisphosphatase in vivo by Glucose and by Proton Ionophores: A Possible Role for cAMP[J]. *European journal of biochemistry*, 1982, 127(3): 605-608.
32. Pinkham J L, Guarente L. Cloning and molecular analysis of the HAP2 locus: a global regulator of respiratory genes in *Saccharomyces cerevisiae*[J]. *Molecular and Cellular Biology*, 1985, 5(12): 3410-3416.
33. Dang V D, Bohn C, Bolotin-Fukuhara M, et al. The CCAAT box-binding factor stimulates ammonium assimilation in *Saccharomyces cerevisiae*, defining a new cross-pathway regulation between nitrogen and carbon metabolisms[J]. *Journal of bacteriology*, 1996, 178(7): 1842-1849.
34. Gancedo J M. Yeast carbon catabolite repression[J]. *Microbiol. Mol. Biol. Rev.*, 1998, 62(2): 334-361.
35. Chang Z F, Liu C J. Human thymidine kinase CCAAT-binding protein is NF-Y, whose A subunit expression is serum-dependent in human IMR-90 diploid fibroblasts[J]. *Journal of Biological Chemistry*, 1994, 269(27): 17893-17898.
36. Good L F, Chen K Y. Cell Cycle-and Age-Dependent Transcriptional Regulation of Human Thymidine Kinase Gene The Role of NF-Y in the CBP/tk Binding Complex[J]. *Neurosignals*, 1996, 5(3): 163-169.
37. Gusmaroli G, Tonelli C, Mantovani R. Regulation of the CCAAT-Binding NF-Y subunits in *Arabidopsis thaliana*[J]. *Gene*, 2001, 264(2): 173-185.
38. Kwong R W, Bui A Q, Lee H, et al. LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development[J]. *The Plant Cell*, 2003, 15(1): 5-18.
39. Kusnetsov V, Landsberger M, Meurer J, et al. The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids[J]. *Journal of Biological Chemistry*, 1999, 274(50): 36009-36014.
40. Kumimoto R W, Zhang Y, Siefers N, et al. NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*[J]. *The Plant Journal*, 2010, 63(3): 379-391.
41. Han X, Tang S, An Y, et al. Overexpression of the poplar NF-YB7 transcription factor confers drought tolerance and improves water-use efficiency in *Arabidopsis*[J]. *Journal of experimental botany*, 2013, 64(14): 4589-4601.
42. Lotan T, Ohto M, Yee K M, et al. *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells[J]. *Cell*, 1998, 93(7): 1195-1205.
43. Vicient C M, Bies-Etheve N, Delseny M. Changes in gene expression in the leafy cotyledon1 (*lec1*) and *fusca3* (*fus3*) mutants of *Arabidopsis thaliana* L[J]. *Journal of Experimental Botany*, 2000, 51(347): 995-1003.

44. Cai X, Ballif J, Endo S, et al. A putative CCAAT-binding transcription factor is a regulator of flowering timing in Arabidopsis[J]. *Plant Physiology*, 2007, 145(1): 98-105.
45. Chen N Z, Zhang X Q, Wei P C, et al. AtHAP3b plays a crucial role in the regulation of flowering time in Arabidopsis during osmotic stress[J]. *BMB Reports*, 2007, 40(6): 1083-1089.
46. Li W X, Oono Y, Zhu J, et al. The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and post transcriptionally to promote drought resistance[J]. *The Plant Cell*, 2008, 20(8): 2238-2251.
47. Ito Y, Thirumurugan T, Serizawa A, et al. Aberrant vegetative and reproductive development by overexpression and lethality by silencing of OsHAP3E in rice[J]. *Plant science*, 2011, 181(2): 105-110.
48. Li C, Distelfeld A, Comis A, et al. Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes[J]. *The Plant Journal*, 2011, 67(5): 763-773.
49. Cao S, Kumimoto R W, Siriwardana C L, et al. Identification and characterization of NF-Y transcription factor families in the monocot model plant *Brachypodium distachyon*[J]. *PloS one*, 2011, 6(6): e21805.
50. Zhang Z, Li X, Zhang C, et al. Isolation, structural analysis, and expression characteristics of the maize nuclear factor Y gene families[J]. *Biochemical and biophysical research communications*, 2016, 478(2): 752-758.
51. Li S, Li K, Ju Z, et al. Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening[J]. *BMC genomics*, 2016, 17(1): 36.
52. Quach T N, Nguyen H T M, Valliyodan B, et al. Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response[J]. *Molecular Genetics and Genomics*, 2015, 290(3): 1095-1115.
53. Wang Y, Xu W, Chen Z, et al. Gene structure, expression pattern and interaction of Nuclear Factor-Y family in castor bean (*Ricinus communis*)[J]. *Planta*, 2018, 247(3): 559-572.
54. Yan D H, Xia X, Yin W. NF-YB family genes identified in a poplar genome-wide analysis and expressed in *Populus euphratica* are responsive to drought stress[J]. *Plant molecular biology reporter*, 2013, 31(2): 363-370.
55. Zhang F, Han M, Lv Q, et al. Identification and expression profile analysis of NUCLEAR FACTOR-Y families in *Physcomitrella patens*[J]. *Frontiers in plant science*, 2015, 6: 642.
56. [56] Ren C, Zhang Z, Wang Y, et al. Genome-wide identification and characterization of the NF-Y gene family in grape (*Vitis vinifera* L.)[J]. *BMC genomics*, 2016, 17(1): 605.
57. [57] Liu S C, Yao M Z, Ma C L, et al. Physiological changes and differential gene expression of tea plant under dehydration and rehydration conditions[J]. *Scientia Horticulturae*, 2015, 184: 129-141.
58. [58] Banerjee, B. "Botanical classification of tea." *Tea*. Springer, Dordrecht, 1992. 25-51. Mondal, Tapan K., et al. "Recent advances of tea (*Camellia sinensis*) biotechnology." *Plant Cell, Tissue and Organ Culture* 76.3 (2004): 195-254.
59. Wang W, Zhou Y, Wu Y, et al. Insight into catechins metabolic pathways of *Camellia sinensis* based on genome and transcriptome analysis[J]. *Journal of agricultural and food chemistry*, 2018, 66(16):

4281-4293.

60. Xia E H, Zhang H B, Sheng J, et al. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis[J]. *Molecular plant*, 2017, 10(6): 866-877.
61. Nobre A C, Rao A, Owen G N. L-theanine, a natural constituent in tea, and its effect on mental state[J]. *Asia Pacific journal of clinical nutrition*, 2008, 17(S1): 167-168.
62. Das A, Das S, Mondal T K. Identification of differentially expressed gene profiles in young roots of tea [*Camellia sinensis* (L.) O. Kuntze] subjected to drought stress using suppression subtractive hybridization[J]. *Plant molecular biology reporter*, 2012, 30(5): 1088-1101.
63. Shen W, Li H, Teng R, et al. Genomic and transcriptomic analyses of HD-Zip family transcription factors and their responses to abiotic stress in tea plant (*Camellia sinensis*)[J]. *Genomics*, 2018.
64. Wang P, Chen D, Zheng Y, et al. Identification and Expression Analyses of SBP-Box Genes Reveal Their Involvement in Abiotic Stress and Hormone Response in Tea Plant (*Camellia sinensis*)[J]. *International journal of molecular sciences*, 2018, 19(11): 3404.
65. Lee H, Fischer R L, Goldberg R B, et al. Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor[J]. *Proceedings of the National Academy of Sciences*, 2003, 100(4): 2152-2156.
66. Lee D K, Kim H I, Jang G, et al. The NF-YA transcription factor OsNF-YA7 confers drought stress tolerance of rice in an abscisic acid independent manner[J]. *Plant Science*, 2015, 241: 199-210.

Tables

Due to technical limitations the tables could not be inserted. They can be found as image files in the supplemental files section.

Figures

Figure 1

Exon-intron structure of the CsNF-Y family generated from GSDS2 online website. The exons and introns are represented by colored boxes and black lines, respectively.

Figure 2

Conserved motifs of the CsNF-Y proteins were displayed. The conserved motifs in the CsNF-Y proteins were identified using Multiple Em for Motif Elicitation (MEME). In total, 10 motifs were identified and are shown in different colors. Motif locations are also indicated.

Figure 3

Multiple sequence alignments of CsNF-Y and AtNF-Y members. (a) Multiple alignment of CsNF-YA proteins. (b) Multiple alignment of CsNF-YB proteins. (c) Multiple alignments of CsNF-YC proteins. All members from *Arabidopsis thaliana* were included, respectively. Multiple alignments of NF-Y protein sequences were carried out with ClustalX and MEGA7.

Figure 4

Phylogenetic tree of tea plant NF-Y members. Function prediction for tea plant genes was based on gene expression and homology to the referred characterized proteins. (a) Phylogenetic tree of CsNF-YA members. (b) Phylogenetic tree of CsNF-YB members. (c) Phylogenetic tree of CsNF-YC members.

Figure 5

Gene ontology (GO) and metabolic pathway (KEGG) enrichment information of CsNF-Y gene family. (a) GO enrichment information of CsNF-Y members. (b) KEGG enrichment information of CsNF-Y members.

Different colors indicate different functions.

Figure 6

Cis-element of CsNF-Y gene family, which based on the analysis of the 1500 bp sequence upstream of the CsNF-Y gene sequences. The figure shows that the functional components of light response of tea plant are the most widely distributed.

Figure 7

Expression profile analysis of CsNF-Y genes in different tea plant tissues. Transcriptome data (FPKM) were used to measure the expression levels of CsNF-Y genes in apical bud, flower, fruit, young leaf, mature leaf, old leaf, stem and root . The colored scale for the different expression levels is shown.

Supplementary Files

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

- [table1.png](#)
- [table3.png](#)
- [table2.png](#)
- [table4.png](#)