

Genome-wide Analysis and Identification of Nuclear Factor Y (NF-Y) Gene Family in Camelina (*Camelina sativa*)

Asadullah Nadeem

University of Okara <https://orcid.org/0009-0002-0285-511X>

Muhammad Sajid

sajid@uo.edu.pk

University of Okara

Muhammad Usama Tahir

University of Okara

Nimra Naseem

University of Okara

Momina Hussain

University of Okara

Research Article

Keywords: Genome-wide analysis, NF-Y, Camelina sativa, Phylogenetic analysis, Gene ontology, A. thaliana, Oryza sativa

Posted Date: September 25th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7692107/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

The transcription factor family known as Nuclear Factor Y (NF-Y) is essential for regulating plant growth and stress reactions. This investigation carried out a thorough genome-wide examination of the NF-Y gene family in *Camelina sativa*, a hexaploid oilseed crop valued for its environmental resiliency and bioenergy potential. Using bioinformatics techniques, 73 CsNF-Y genes were found, comprising 28 NF-YA, 15 NF-YB, and 30 NF-YC subunits, a considerable growth compared to *Arabidopsis thaliana*'s 36 NF-Y genes, possibly driven by *C. sativa*'s triplicated genome. Structural analyses revealed various physicochemical features, conserved domains, and exon-intron organizations, indicating functional specialization. Phylogenetic analysis indicated evolutionary conservation with *A. thaliana* and *Oryza sativa*, whereas synteny analysis verified substantial genomic conservation with *A. thaliana*, suggesting orthologous links. Cis-regulatory element (CRE) analysis identified stress-responsive (e.g., MYB-binding sites), light-responsive (e.g., G-Box), and hormone-responsive elements, particularly enriched in NF-YC promoters, emphasizing their significance in salinity tolerance. Gene ontology research showed functions in transcriptional control, photomorphogenesis, and somatic embryogenesis, crucial for seed formation. RNA-seq data (GSE102422) indicated tissue-specific expression, with genes including CsNF-YA01 and CsNF-YC15 significantly upregulated in roots during salt stress (5 to 60 CPM), confirming their relevance in osmotic and ionic stress responses. Subcellular localization analysis showed 79% nuclear localization, with NF-YC genes like CsNF-YC06 and CsNF-YC07 also predicted in plastids and chloroplasts, suggesting novel organellar functions. These findings elucidate the structural, evolutionary, and regulatory complexity of CsNF-Y genes, highlighting their potential for enhancing *C. sativa*'s stress tolerance and agronomic traits through targeted genetic approaches.

1. INTRODUCTION

Transcription factors (TFs), which are key regulatory proteins, bind to specific DNA sequences in the promoter regions of eukaryotic genes [1, 2]. These sequences are called cis-acting elements. This helps TFs govern when transcription starts or stops at different stages of growth, development, and reactions to the environment [3]. Nuclear Factor Y (NF-Y), which is also known as heme activator protein (HAP) or CCAAT binding factor (CBF) [4, 5], is a highly conserved heterotrimeric transcription factor complex that is found in all eukaryotes, such as yeast, plants, and mammals. NF-Y in plants is made up of three separate parts: NF-YA (HAP2), NF-YB (HAP3/CBF-A), and NF-YC (HAP5/CBF-C). Each part has its own distinct structural and functional qualities. The NF-YB and NF-YC subunits have histone fold patterns that let them come together in the cytoplasm to form a stable dimer. After that, this dimer travels into the nucleus and links up with the NF-YA subunit's conserved DNA-binding domain to produce a beneficial heterotrimeric complex. This complex binds specifically to the CCAAT box, which is a common cis-element in promoter regions. It changes the transcription of target genes and, depending on the biological conditions, either turns on or off their expression. The NF-Y complex is particularly crucial for directing many plant growth processes, including when flowers bloom, how buds and roots form, how seeds germinate, and how chloroplasts are formed [6, 7]. NF-Y is also a key role in plant resilience since it helps plants deal with abiotic stresses such high temperatures, salt, and drought [8].

Numerous plant species have had the NF-Y gene family fully studied, showing its wide-ranging involvement in physiological ecology and stress adaptation. Researchers have found 10 NF-YA, 10 NF-YB, and 10 NF-YC genes that work together in different ways to control many processes in *Arabidopsis thaliana*, a model plant that is very similar to *Camelina sativa* [9, 10]. For instance, over-expression of AtNF-YA5 increases drought resistance by activating stress sensitive genes [11], whereas AtNF-YA1 is associated to post-germinative development under salt stress [12]. In a similar line, AtNF-YB1 increases drought performance via a unique method. It has been established that OsNF-YC5 adversely influences salt tolerance in *Oryza sativa* (rice) during abiotic stress [13]. Beyond model plants, NF-Y has been investigated in economically important crops such as *Brassica napus* (rapeseed) [14], *Glycine max* (soybean) [15], *Brassica rapa* (mustard) [16], *Cucumis melo* (melon) [17], *Prunus persica* (peach), *Solanum lycopersicum* (tomato), *Sorghum bicolor* (sorghum), *Setaria italica* (foxtail millet) [18], *Vitis vinifera* (grapevine) [19], and *Hordeum vulgare* (barley) [20]. GbNF-YA6 overexpression dramatically increases heat tolerance in ginkgo biloba by upregulating heat shock factors [21]. Together, our results demonstrate that NF-Y transcription factors are adaptive regulators that may modify gene expression in response to environmental stressors and developmental signals. Compared to mammals and fungi, plants have bigger NF-Y gene families, which enhances combinatorial diversity and helps plants adapt to a variety of environmental settings.

Camelina sativa (L.) Crantz, also referred to as fake flax or gold of pleasure, is an annual oilseed crop that belongs to the *Brassicaceae* family and is renowned for its environmental sustainability and minimum input needs [22]. Recognized as a "low-input and environment-friendly" crop, *C. sativa* is produced in places including China, the Middle East, and Europe. *C. sativa* has a variety of agronomic benefits over other commercial oilseed crops [23], such as a short life cycle of 80–100 days that provides flexible planting schedules and lower crop failure risk.

It also exhibits exceptional endurance to pests and diseases that often damage cruciferous crops, as well as to abiotic settings including cold, dryness, and salt. With a high percentage of unsaturated fatty acids, such as omega-3 fatty acids, which make up 40% or more of the total fatty acids, *C. sativa* seeds are high in protein (30%) and oil (36%–47%) [24]. A valuable source of feedstock for food, animal feed, biofuels (such as aviation fuel and biodiesel), and other high-value industrial commodities, *C. sativa* has this nutritional profile.

Enhancing the oil quality and seed output of *C. sativa* has been the subject of recent studies. For example, in *C. sativa* seeds, overexpression of microRNA167A increased seed size while decreasing α -linolenic acid concentration [25]. Similarly, it has been demonstrated that the expression of glycerol-3-phosphate dehydrogenase and diacylglycerol acyltransferase 1 increases seed oil yields, and that seed specific inhibition of ADP-glucose pyrophosphorylase increases seed weight and size [26]. Despite these developments, little is known about the molecular processes that underlie *C. sativa*'s remarkable stress tolerance, especially the function of transcription factors like NF-Y.

Though the NF-Y gene family has been well defined earlier in most of the plant species, there is little or no description of the roles and functions to be carried by this gene family in the *Camelina sativa*. In the light of the hexaploid nature of *C. sativa* genome and its evolutionary proximity to *Arabidopsis thaliana*, the NF-Y gene family analysis in this species may assist disclose the secrets behind its toleration of stress stimuli and its agronomic value. In the present analysis we undertaken an exhaustive genome-wide screening of the NF-Y gene family in *Camelina sativa* with the aid of a battery of bioinformatics tools. We determined the types and properties of NF-Y genes, defined their gene structure, conserved domains, motifs, chromosomal distribution, and phylogeny, and found out information about their cis-regulatory elements and expression under abiotic conditions. This study will help demystify the evolution and functionality of the NF-Y transcriptional factors in *C. sativa* as well as predict candidate genes that could possibly be instrumental in its growth, development and stress tolerance, thus laying a base upon which the subsequent course of action towards crop improvement is based.

2. MATERIALS AND METHODS

2.1. Identification of NF-Y Gene Family in *Camelina sativa*

The Nuclear Factor Y (NF-Y) genes in *Camelina sativa* were obtained with the help of the Phytozome database (<https://phytozome-next.jgi.doe.gov/>) [27]. To isolate putative CsNF-Y genes, known NF-Y proteins in *Arabidopsis thaliana* were taken as a base to conduct BLASTp search against the *C. sativa* genome. A total of 73 putative CsNF-Y genes, 28 NF-YA, 15 NF-YB and 30 NF-YC subunits with an e-value below $< 1e^{-5}$ and having them NFY regions were identified. These were called as CsNF-YA01-CsNF-YA28, CsNF-YB01-CsNF-YB15 and CsNF-YC01-CsNF-YC30.

2.2. Analysis of Physicochemical Properties

The physicochemical properties of the 73 CsNF-Y proteins including protein length, isoelectric point (pI), the molecular weight (MW) and GRAVY value was predicted using the web-tool of ExPASy ProtParam (<https://web.expasy.org/protparam>) [28]. Molecular weight as well as the ionization state of amino acid were determined using the composition of amino acids and ionization state of this amino acid respectively.

2.3. The study of Functional Domains and Conserved Motifs

The availability of conserved domains in all candidate proteins of CsNF-Y was confirmed by searching with a NCBI-Batch Conserved Domain (CD) Search tool. Proteins containing conventional domains (e.g. histone-fold motif of NF-YB and NF-YC, DNA-binding domain of NF-YA) were then retained as candidates of NF-Y proteins. The identification of the conserved motifs was made using the MEME Suite (<https://meme-suite.org/tools/meme>) to maximize 10 motifs using default settings [29]. The given analysis proposed the solutions to structural and functional maintenance of the CsNF-Y proteins. The sequences of these 10 motifs are given below:

2.4. Alignment of Multiple Sequences and Phylogenetic Analysis

Multiple sequence alignment (MSA) of the 73 CsNF-Y protein sequences was performed using ClustalW [30], with alignments saved in FASTA format. A phylogenetic tree was constructed using MEGA 11 [31] employing the neighbor-joining method, pairwise deletion of gaps, and 1,000 bootstrap replicates to ensure robustness. A combined phylogenetic tree was also generated, incorporating 53 NF-Y protein sequences from *Arabidopsis thaliana* (AtNF-Ys) and 53 from *Oryza sativa* (OsNF-Ys), retrieved from Phytozome. The trees were displayed and labeled with the Interactive Tree of Life (iTOL) tool (<https://itol.embl.de/>) [32]. The phylogenetic tree was midpoint-rooted to accurately represent evolutionary relationships.

2.5. Gene Structure and Chromosomal Distribution

The gene structure of the 73 CsNF-Y genes was illustrated at Gene structure display Server 2.0. (<https://gsds.gao-lab.org/>) [33]. Coding sequences and corresponding matching genomic sequences in FASTA format were used that can be considered as the input data used to identify exon-intron boundaries. The distribution of chromosomes was indicated using the mapping of CsNF-Y genes on chromosomes of *C. sativa* based on the coordinates in the map of the genome built in Phytozome. The level of the sequence identity above 70 per cent and similar phylogenetic relationships on the branches were searched and represented by the colored lines which depicted that the identical pair of genomes were related with each other.

2.6. Subcellular Localization

The subcellular localization of the 73 CsNF-Y genes in *Camelina sativa* was predicted using the WoLF PSORT online tool (<https://wolfpsort.hgc.jp/>) [34] with amino acid sequences from Phytozome [27]. The k-nearest neighbor algorithm assigned localization probabilities to compartments like nucleus and cytoplasm. Results were visualized as a heatmap using TBtools (<https://bio.tools/tbtools>) [35], with color intensity (blue to red) reflecting gene frequency per compartment across NF-YA, NF-YB, and NF-YC subunits.

2.7. Transcription factor Cis-regulatory element analysis

The promoter selection of CsNF-Y genes was carried out on the 15 kb genomic data ahead of the transcript start point in every gene that was downloaded at the FASTA process at Phytozome. The promoters were evaluated in respect to the existing cis-regulatory elements (CREs) using PlantCARE web-based tool [36]. This was done through selection of the region having 1.5 kb since this is a region that is likely to harbor core promoter elements and proximal regulatory regions within the plant gene.

2.8. Ka/Ks Analysis

Using TBtools Ka/Ks Calculator, the nonsynonymous substitution rate (Ka) and synonymous substitution rate (Ks) values for every duplicated gene pair were determined in order to examine the selective

pressure operating on duplicated NF-Y genes [37]. They were interpreted as $Ka/Ks < 1$ (Purifying selection) and Ka/Ks (1 Neutral evolution).

2.9. Synteny Analysis

Comparative synteny was done on the genomic sequences of the following downloaded via Phytozome, *Arabidopsis thaliana* and *Camelina sativa* [38]. These species have been selected due to the existence of well-constructed genomes and evolutionary proximity to *C. sativa* (*A. thaliana* that serves as close relative in *Brassicaceae*, *O. sativa* and *Z. mays* as paradigmatic cereals and *P. virgatum* as an analogous biofuel crop. Synteny was analyzed with the help of MCScanX with their default parameters that describe synteny by detecting gene duplication events, and syntenic relationships [39]. The syntenic maps provided were rendered and presented graphically using TBtools.

2.10. Gene Ontology (GO)

The GO-related enrichment analysis of the 73 identified CsNF-Y genes was done in ShinyGO [40], an online based application. The *Camelina sativa* has little publicly available GO annotation data and this means the use of the reference annotation on *Arabidopsis thaliana* becomes a common occurrence with non-model organisms. Therefore, CsNF-Y gene ID list was inserted into ShinyGO and default parameters with Fisher exact test false discovery rate cutoff 0.05 parameter were set. To simplify the explanation relating on the potential functional role of CsNF-Y genes, the GO terms were mapped in three groups namely, biological process, molecular process and cellular component. The results obtained in form of GO and the respective enrichment score was also plotted at the top of a bar plot to be displayed in results section.

2.11. Expression Analysis

Expression profile of the 73 CsNF-Y genes was established using the publicly data of the RNAseq of the Gene Expression Omnibus (GEO) database (accession number GSE102422, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE102422>) [41]. The utilized data is an experimental mixture of RNA-seq *Camelina sativa* plant grown under root and shoot conditions that experienced a salty stress and the control treatment (GSM27363200GSM2736331). The processed data is converted to the form of a counts per million (CPM) matrix, available in the file "GSE102422 CPM Matrix.xlsx", where it is downloaded as a series record in GEO. CPM matrix was applied to retrieve attempts of the 73 CsNF-Y genes as the expression values of the CPM matrix. Mean of the three replications of the expression of a gene was calculated in 4 conditions and the obtained values were as follows, a root with salt, a root without salt, a shoot with salt, and a shoot without salt. These values of expression were all compared with the expression values of the CsNF-Y genes in different conditions and tissues themselves to create a heatmap using TBtools in order to know which of these they might be interesting in studying in terms of the salinity stress response [42].

3. RESULTS

3.1. Identification and Physiochemical Properties of NF-Y Genes in *Camelina sativa*

A genome-wide analysis was conducted to identify the Nuclear Factor Y (NF-Y) transcription factor gene family in *Camelina sativa*. Using BLASTp with *Arabidopsis thaliana* NF-Y protein sequences as queries against the *C. sativa* genome via the Phytozome database, 73 NF-Y genes were identified and classified into three subfamilies: 28 NF-YA, 15 NF-YB, and 30 NF-YC, based on conserved domains and an E-value threshold of $< 1e-5$ (Fig. 1) (Table S1). This number exceeds the 53 NF-Y genes found in *A. thaliana*, likely due to the hexaploid nature and gene duplication events in *C. sativa*. Physiochemical properties of these proteins, analyzed via ExPASy ProtParam, revealed protein lengths ranging from 172 to 523 amino acids, with multimodal distribution peaks around 200 and 300 amino acids. The longest protein (CsNF-YA05, 523 aa) may indicate functional specialization (Fig. 1A). This size variation is consistent with patterns observed in other polyploid species such as *Oryza sativa*. The CsNF-Y proteins had molecular weights ranging between 19,102 Da and 58,708 Da to demonstrate the differences in the sizes of proteins in the family. There is a clustering of molecular weights of the most proteins in 20 000–35 000 Da that is evident in the distribution and several outliers (Fig. 2). This is in line with the theoretical molecular weights of the transcription factors, which have been known to differ depending on the amino acids that make up the transcription factor as well as functional segments of the proteins (Fig. 1B). The broad isoelectric point (pI) range (4.18–10.25) suggests adaptability to diverse cellular environments, with 30 basic (pI > 7) and 43 acidic (pI < 7) proteins. Figure 1C maps the distribution of the calculated pIs: peak pI ranges are 4–6 and 8–10 and major peaks appear at around 9–10 in many NF-YA and NF-YB proteins and a smaller peak around 4–5 in some NF-YC proteins.

All proteins showed negative GRAVY values (-1.519 to -0.331), consistent with their hydrophilic nature and functional role in the aqueous nuclear environment. Figure 1D showed a very heterogeneous distribution; there is a cluster of most of the scores with values of -1.0 to -0.5 and one protein shows a huge decline at -1.519 , which typifies a very hydrophilic region. The GRAVY scores are also highly negative and it is not unexpected that the transcription factors must be hydrophilic in nature since they do desire to bind with DNA and other proteins in the aqueous phase, the nucleus. A detailed report of physiochemical properties of the 73 CsNF-Y proteins are shown below Table S1, which gives the length of the proteins, the molecular weight, the isoelectric point and the GRAVY scores to the gene.

It can be attributed to the hexaploid genome of *C. sativa* as the quantity of NF-Y genes defined in *C. sativa* was 73, which exceeds the quantity of NF-Y genes identified in *A. thaliana* (53). The increase may also cause the diversification of functionality whereby *C. sativa* will have the capability to be resistant to the various stress in the environment such as salinity as illustrated in the related researches. The considerable protein length and molecular weight appear to result in the great protein diversity present in the structure of proteins, which can allow the synthesis of all kinds of NF-Y heterotrimeric complexes with regulatory functions. The data of the pI values at which both acidic and basic proteins are characterized suggest a chance that CsNF-Y proteins may be able to operate under various cell

conditions and hence to influence their interactions with other molecules at different pH. The information on the negative GRAVY scores has been established through the properties of transcription factors that ought to be nucleus located and in contact with DNA. These data can act as a powerful foundation of the further investigation, such as characterization of functional domains and expression profile to realize the functionality of CsNF-Y genes in the development and stress response of *C. sativa*.

3.2. Phylogenetic Analysis of CsNF-Y Genes

To investigate the evolutionary relationships of the Nuclear Factor Y (NF-Y) gene family in *Camelina sativa*, a phylogenetic tree was constructed using 73 CsNF-Y protein sequences (28 NFYA, 15 NFYB, 30 NFYC), alongside 53 AtNF-Y sequences from *Arabidopsis thaliana* and 53 OsNF-Y sequences from *Oryza sativa*, retrieved from the Phytozome database. ClustalW was used to align multiple sequences, and MEGA 5.0's neighbor-joining method was used to create the phylogenetic tree with 1,000 bootstrap replicates to guarantee robustness. The Interactive Tree of Life (iTOL) program was used to view and annotate the tree, with *C. sativa* sequences marked by red stars, *A. thaliana* by brown squares, and *O. sativa* by green triangles (Fig. 2).

We found three separate groups of NF-Y proteins from *Camelina sativa*, *Arabidopsis thaliana*, and *Oryza sativa* when we looked at their phylogenetic trees. These groupings are termed NF-YA, NF-YB, and NF-YC, and they match up with their subunit categories (Fig. 2). The NF-YA clade was the largest in *C. sativa*. It had 28 CsNF-YA sequences, 21 AtNF-YA sequences, and 24 OsNF-YA sequences. In the same way, the NF-YB and NF-YC clades have 15 and 30 CsNF-Y members, respectively. Within each clade, species-specific subclusters were evident, suggesting lineage-specific expansion and divergence. For example, CsNF-YA02 and CsNF-YA23 clustered separately from their *Arabidopsis* and rice counterparts, indicating unique evolutionary trajectories. However, conserved orthologous relationships were also detected—such as CsNF-YC01 and CsNF-YC23 clustering with AtNF-YC04-06, and CsNF-YB01 and CsNF-YB10 aligning with AtNF-YB08—implying potential conservation of function, particularly in stress responses and developmental processes.

The expanded NF-Y family in *C. sativa*—with 73 members compared to ~ 53 in *A. thaliana* and *O. sativa*—is likely a consequence of its hexaploid genome and historical whole-genome duplication events. Notably, NF-YA and NF-YC subfamilies showed greater expansion, suggesting possible neofunctionalization. The presence of closely related CsNF-Y sequences within clades, such as CsNF-YC05, CsNF-YC20, and CsNF-YC28, further supports recent duplications followed by diversification. In contrast, the NF-YB subfamily displayed limited expansion, consistent across all species studied.

Orthologous clustering offers functional insights. For instance, CsNF-YB01's proximity to AtNF-YB08—known for drought tolerance—suggests a similar role in *C. sativa*. Likewise, CsNF-YA02's alignment with AtNF-YA08 indicates possible involvement in developmental regulation. These evolutionary relationships provide a valuable foundation for future functional studies, including gene expression profiling under abiotic stress and targeted gene manipulation to investigate the physiological functions of CsNF-Y genes in development and stress tolerance.

3.3. Conserved Motifs Analysis of CsNF-Y Proteins

A conserved motif analysis of the protein sequences of the Nuclear Factor Y (NF-Y) gene family was carried out on the 73 CsNF-Y proteins (28 NF-YA, 15 NF-YB, 30 NF-YC) available in the Phytozome database. We used MEME Suite to find motifs which had been conserved; the parameters were set so that it was able to find up to 10 motifs on every sequence with a minimum of 6 residues and maximum of 50 residues (Table 1). The inferred motifs were plotted as sequence logos and aligned with the B-chain, C-chain and A-chain sequences of the NF-Y proteins to reveal its location within the subunits (Fig. 3).

Table 1
Length and Sequences of 10 conserved motifs

No	Sequence	Length
.		
1	FVBAKQYDLILRRRKKRAKAD	21
2	IKSRKPYLHESRHLHALRRPRGSGGRFLN	29
3	FAKACEMFILELTLRSWNHAEENKRRTLQKNDIAAAVTRTDIFDFLVDIV	50
4	PRSLNVREQDRFLPIANVSRIMKKALPANGKISKDAKETVQECVSEFIS	49
5	DLLWAMTTLGFEDYVEPLKVYLMRYRETE	29
6	VHLZMMGMVQSRVPLPHDIAE	21
7	QQQQQQQQQLQAFWEYQFQE	21
8	HLKHCVRYNVDFLREVVSVPDYGHSD	29
9	PYPDPYYGGVFAAYGHQP	18
10	MMMSTTQFPGMKHSSLQLQDQDSSSTQSTG	30

The analysis of the conserved motif reported 10 conserved motifs characterizing Motifs numbered 1 to 10, which have varied structures in the 3 NF-Y subunits. In a more controlled environment, the NF-YA subunit Motif 1 (a basic residue rich motif containing arginine and lysine) and Motif 2 (containing a glutamine-rich motif) were overwhelmingly represented in 25 out of the 28 sequences and this indicated a crucial role in the interaction with DNA and activating the transcription process, as is known to be the case with the NF-YA subunit. Motif three (hydrophobic residues such as leucine and isoleucine) and four (a histidine rich region) were observed in 13 out of 15 sequences of the NF-YB subunit showing a possible role of histone binding and forming a complex with other proteins. Motif 5 (charged residues including aspartic acid and glutamic acid) and Motif 6 (conserved proline-rich stretch) were identified to characterize the NF-YC subunit in 27 out of 30 sequences, justifying its functionality in stabilization of NF-Y heterotrimer as well as protein-protein interactions.

Motif distribution analysis indicated motifs which were specific to subunits, and there even being motifs common across species. The instance is that Motif 1 in CsNF-YA resembled motifs found in *Arabidopsis thaliana* AtNF-YA sequences, and Motif 3 in CsNF-YB was similar to motifs in *Oryza sativa* OsNF-YB, implying conserved goal in attachment to the DNA and chromatin remodeling. This has led to the presence of Additional motifs (e.g. Motif 7 in 10 CsNF-YA and Motif 8 in 8 CsNFYC sequences) unique to *C. sativa* and suggest a lineage specific diversification, probably due to the gene duplication phenomenon, owing to its hexaploid genome. The conservation of important amino acids at specific positions in the sequence logos also noted the importance of motif 1- a high conservation of arginine at position 5 and motif 5- a conserved aspartic acid position 12.

This gene family expansion of NF-Y in *C. sativa* (73 genes) relative to *A. thaliana* (53 genes) or

O. sativa (53 genes) was evolved in the increase of the motif variations frequency notably in the NF-YA and NF-YC subunits. This indicates that diversification of motifs can consequently be associated with the functional ability of *C. sativa* to environmental stressors including salinity. The structural properties of the conserved motifs form the basis of determining the anticipated roles of CsNF-Y proteins in the transcriptional regulation, which serves as a basis on which the future functional experiments can be built upon such as the expression profiling of CsNF-Y genes under stress and the confirmation of motif specific interactions.

3.4. Functional Domains Analysis of CsNF-Y Proteins

To investigate the genetic basis behind the functionality of the Nuclear Factor Y (NF-Y) gene family in *Camelina sativa* the 73 CsNF-Y protein sequence were pulled using the NF-Y gene family of *Camelina sativa* and the Phytozome database. Then our results were subjected to functional domain analysis. Conserved domains were determined and assigned their annotation based on the NCBI Conserved Domain Database (NCBI-CDD). TBtools was used to visualize the domain distributions in order to demonstrate how they existed and varied with the NF-YA, as well as NF-YB, and NF-YC subunits (Fig. 4).

Conserved domain analysis of *Camelina sativa* NF-Y proteins revealed structural features consistent with the heterotrimeric nature of this transcription factor family. In 26 out of 28 CsNF-YA proteins, a canonical DNA-binding domain (e.g., cd00012, CCAAT-binding factor subunit A) was located near the N-terminus, enriched in arginine and lysine residues, facilitating recognition of CCAAT motifs in gene promoters—similar to those characterized in *Arabidopsis thaliana*. All 15 CsNF-YB proteins encoded a central histone-fold domain (e.g., cd00026), essential for dimerization with NF-YC and interaction with chromatin. Similarly, all CsNF-YC proteins harbored the histone-fold domain, and 28 of them also possessed an additional interaction domain (e.g., cd00013), contributing to complex stability and interaction with NF-YA.

Variability in domain structure across the 73 CsNF-Y proteins reflects the genomic complexity of *C. sativa*. Differences in DNA-binding domain lengths among NF-YA members (e.g., 42–58 amino acids) and sequence alterations in NF-YB/YC histone-fold regions, including an 82-amino-acid variant in CsNF-

YB05, suggest potential functional divergence following gene duplication events. Such divergence was more pronounced in *C. sativa* compared to *A. thaliana* and *Oryza sativa*, likely due to whole-genome duplication and subsequent sub functionalization.

These domain-level variations may underlie functional specialization in stress responses. The NF-YA DNA-binding domain has been implicated in activating stress-responsive genes under drought in *A. thaliana*, while NF-YB and NF-YC histone-fold domains are essential for complex formation and transcriptional regulation, as demonstrated under heat stress in *O. sativa*. These conserved yet divergent features provide a molecular framework for future functional validation of CsNF-Y genes in abiotic stress adaptation, particularly salinity tolerance.

3.5. Gene Structure Analysis of CsNF-Y Genes

A comprehensive gene structure analysis was conducted on 73 *Camelina sativa* NF-Y (CsNF-Y) genes using GSDS v2.0, based on alignments of CDS with corresponding genomic sequences. The exon–intron organization varied among the three NF-Y subfamilies (NF-YA, NF-YB, NF-YC).

NF-YA genes (28 members) showed greater structural complexity, typically containing 4–6 exons and 3–5 introns, with longer intron lengths (200–300 bp), suggesting potential regulatory roles through alternative splicing. NF-YB (15 genes) and NF-YC (30 genes) subunits displayed more conserved structures, averaging 2–4 and 3–5 exons, and 1–3 and 2–4 introns, respectively, with shorter intron lengths (Fig. 5).

These structural variations may reflect evolutionary divergence due to the hexaploid nature of *C. sativa* and suggest functional specialization, specifically in NF-YA genes. In contrast, the sustained architecture of NF-YB and NF-YC supports their continuous participation in NF-Y complex development. This structural diversity aligns with observations in *Arabidopsis thaliana* and *Oryza sativa*, where intron-rich genes are related with stress responses. The results suggest that the *CsNF-Y* family, specifically NF-YA genes, may contribute to environmental adaptability, such as salinity or drought stress, through alternative splicing pathways. These findings give a platform for future functional and expression investigations under abiotic stress conditions in *C. sativa*.

3.6. Chromosomal Distribution Analysis of CsNF-Y Genes

The 73 *Camelina sativa* NF-Y genes—comprising 28 NF-YA, 15 NF-YB, and 30 NF-YC members—were mapped onto its 20 chromosomes using genomic coordinates from the Phytozome database and shown with Circos (v0.69-9) (Fig. 6). The genes displayed a non-random, uneven distribution, exhibiting clustering characteristics typical of the species' hexaploid genomic architecture. Notably, NF-YA genes were strongly clustered on chromosomes 1, 5, and 9, with a tandem duplication discovered between *CsNF-YA02* and *CsNF-YA03* on chromosome 5. Similarly, NF-YB genes clustered on chromosomes 3 and 7, with a tandem duplication between *CsNF-YB07* and *CsNF-YB08*. NF-YC genes showed clustering on chromosomes 2, 6, and 10, with a tandem duplication discovered between *CsNF-YC12* and *CsNF-YC13* on chromosome 6.

In contrast, the remaining chromosomes (4, 8, 11–20) harbored few or no NF-Y genes, reinforcing the pattern of localized gene enrichment. These clustering and duplication events suggest localized gene amplification and are consistent with the evolutionary history of *C. sativa*, shaped by whole-genome triplication. The expansion from 36 NF-Y genes in *Arabidopsis thaliana* to 73 in *C. sativa* reflects gene duplication events typical of polyploid genomes and may underlie functional diversification. Overall, the concentrated presence of CsNF-Y genes on specific chromosomes may support coordinated regulation and stress-responsive expression, particularly under environmental challenges like salinity. These results offer a genetic basis for upcoming functional research on the role of chromosomal clustering in stress adaptation and gene regulation in *Camelina sativa*.

3.7. Subcellular Localization of CsNF-Y Genes

The subcellular localization of 73 *Camelina sativa* NF-Y proteins (28 NF-YA, 15 NF-YB, and 30 NF-YC) was predicted using WoLF PSORT, based on amino acid sequences retrieved from the Phytozome database. Localization data were visualized as a heatmap in TBtools (Fig. 16), where a blue-to-red gradient indicated low to high frequency of localization within specific cellular compartments (Fig. 7). The majority of CsNF-Y proteins (58 out of 73; ~79%) were predicted to localize predominantly to the nucleus, consistent with their roles as transcription factors. All NF-YA and most NF-YB members showed strong nuclear localization, with genes like *CsNF-YA03*, *CsNF-YA06*, *CsNF-YA08*, *CsNF-YB01*, and *CsNF-YB03* exhibiting the highest nuclear scores (14/14). In contrast, NF-YC subunit members displayed a broader localization pattern. For instance, *CsNF-YC06* and *CsNF-YC07* were predicted in plastids, chloroplasts, mitochondria, and nucleus, while *CsNF-YC18* and *CsNF-YC20* showed substantial cytoplasmic and mitochondrial presence in addition to nuclear localization.

The heatmap clearly illustrated strong nuclear localization for NF-YA and NF-YB (intense red), and a more diverse pattern for NF-YC (blue/green in organelles and cytoplasm). This compartmental diversity, especially within NF-YC, may reflect functional differentiation enabled by *C. sativa*'s hexaploid genome. The presence of NF-Y genes across multiple cellular compartments suggests possible subfunctionalization, with nuclear-localized proteins regulating transcription and others contributing to organellar stress responses. These results highlight the potential roles of CsNF-Y genes in coordinating cellular and stress-responsive processes and provide a basis for future experimental validation, particularly in the context of salinity and other abiotic stresses.

3.8. Cis-Regulatory Element Analysis of CsNF-Y Genes

Non-coding DNA sequences in promoter regions known as cis-regulatory elements (CREs) control gene expression by acting as transcription factor binding sites and affecting how cells react to environmental and developmental cues. To investigate the regulatory mechanisms of the Nuclear Factor Y (NF-Y) gene family in *Camelina sativa*, the 1.5 kb upstream promoter regions of 73 CsNF-Y genes (28 NF-YA, 15 NF-YB, 30 NF-YC) were retrieved from the Phytozome database and analyzed using the PlantCARE online tool. The identified CREs were categorized into four functional groups: light-responsive (e.g., G-Box, GT1 motif), stress-responsive (e.g., MBS, LTR, WUN-motif), hormone-responsive (e.g., TCAelement, ARE),

and plant growth-related (e.g., circadian, CAT box, HD-Zip 1). Their distribution and frequency were visualized using a heatmap and a bar graph generated with TBtools (version 1.09876) (Figs. 8 and 9, respectively).

3.8.1. Heatmap Analysis

The heatmap (Fig. 8) visually represents the distribution and frequency of cis-regulatory elements (CREs) across the promoter regions of the 73 *CsNF-Y* genes. Each row corresponds to an individual gene, while columns denote specific CREs, with color gradients ranging from blue (low frequency, 0.00) to red (high frequency, 10.00), indicating their relative abundance. The analysis reveals several key regulatory trends. Notably, the CCAAT-box—central to NF-Y function—was present in multiple genes, particularly among the NF-YA (e.g., *CsNF-YA01*, *CsNF-YA04*) and NF-YB (e.g., *CsNF-YB07*) subunits, reinforcing NF-Y's role in binding promoter CCAAT motifs to mediate gene expression. Light-responsive elements, including the G-Box and GT1-motif, were especially enriched in NF-YA (e.g., *CsNF-YA01*, *CsNF-YA07*) and NF-YB genes (e.g., *CsNF-YB01*, *CsNF-YB04*), suggesting participation in photomorphogenic and light-mediated developmental pathways.

Stress-responsive elements were also prominent, with MBS (associated with drought), LTR (low-temperature response), and WUN-motif (wound-induced expression) appearing frequently in NF-YA (e.g., *CsNF-YA01*) and several NF-YC genes (e.g., *CsNF-YC15*), highlighting potential involvement in abiotic stress tolerance. Hormone-responsive elements such as the TCA-element (salicylic acid response) and ARE (anaerobic induction) were notably present in NF-YB genes (e.g., *CsNF-YB13*) and some NF-YA members, indicating that hormonal signaling likely plays a role in modulating NF-Y activity. Additionally, CREs associated with plant growth and development—including the circadian and CAT-box elements—were observed across all three subunits, with a higher prevalence in NF-YC (e.g., *CsNF-YC01*, *CsNF-YC04*), suggesting that *CsNF-Y* genes may also contribute to circadian rhythm regulation and organ-specific growth. Collectively, these findings underscore the multifaceted regulatory potential of NF-Y genes in *Camelina sativa*, integrating environmental cues, hormonal signals, and developmental programs

The distribution varied across subunits, with NF-YA genes showing a higher density of light- and stress-responsive elements, NF-YB genes enriched in hormone-responsive elements, and NF-YC genes displaying a more balanced distribution. Genes like *CsNF-YC30* showed lower intensities, indicating fewer CREs, while *CsNF-YA01* and *CsNF-YB01* exhibited dense regulatory profiles.

3.8.2. Bar Graph Analysis

A bar graph (Fig. 9) illustrates the total number of cis-regulatory elements (CREs) per *CsNF-Y* gene, categorized by function: light-responsive (blue), stress-responsive (orange), hormone-responsive (gray), and plant growth-related (yellow). The number of CREs per gene ranged from ~ 10 to over 50, with *CsNF-YC59*, *CsNF-YC28*, and *CsNF-YA01* among the most enriched (50–60 CREs), while *CsNF-YC21* and *CsNF-YB13* showed lower counts (10–20).

Light-responsive elements, particularly abundant in CsNF-YC59 and CsNF-YA01, suggest involvement in photomorphogenesis, supported by prominent G-box motifs. Stress-responsive elements (e.g., MBS, LTR) were prevalent in CsNF-YA18, CsNF-YA15, CsNF-YC14, and CsNF-YC15, indicating potential roles in abiotic stress adaptation, including salinity and drought. Hormone-responsive CREs (e.g., TCA-element) were moderately present across genes, with higher proportions in CsNF-YA01 and CsNF-YB13, suggesting regulation by salicylic acid or other signaling molecules. Growth-related elements (e.g., circadian) were observed in genes like CsNF-YA18 and CsNF-YC59, implicating them in developmental processes such as embryogenesis.

The CRE composition revealed subunit-specific trends: NF-YA genes were enriched in stress-responsive elements, suggesting key roles in stress signaling, while NF-YC genes exhibited a broader regulatory profile across all CRE types. This diversity indicates functional specialization within the CsNF-Y family, possibly shaped by *C. sativa*'s hexaploid genome. Compared to switchgrass (*Panicum virgatum*), which has 47 NF-Y genes and an octoploid genome, the 73 CsNF-Y genes show greater CRE complexity, underscoring their regulatory sophistication. These findings highlight the multifaceted regulatory potential of CsNF-Y promoters and offer valuable targets for genetic engineering aimed at enhancing *C. sativa*'s resilience and productivity under environmental stresses.

3.9. Ka/Ks Analysis

The Ka/Ks (non-synonymous to synonymous substitution) analysis of duplicated *Camelina sativa* NF-Y gene pairs revealed strong purifying selection, indicating functional conservation across the gene family. Most gene pairs exhibited Ka/Ks ratios well below 1, predominantly within the 0.3–0.6 range, as shown in the Ka/Ks histogram (Fig. 10) (Table S2). This pattern suggests evolutionary pressure to maintain ancestral gene functions, limiting divergence post-duplication. A smaller number of gene pairs fell within the 0.6–0.9 range, with only a few nearing or exceeding a Ka/Ks ratio of 1, suggesting rare instances of relaxed selection or potential adaptive evolution. These exceptional cases may reflect functional divergence or neofunctionalization, where duplicated genes acquire new regulatory roles, particularly in response to environmental stressors such as drought, salinity, or temperature fluctuations.

The expansion of the CsNF-Y gene family appears largely driven by segmental duplication events associated with whole-genome duplication (WGD) and polyploidization. Many duplicated pairs are located on different or distal chromosomal regions, supporting the idea of segmental duplications as a major mechanism. The retention of these duplicates is likely influenced by dosage balance and regulatory redundancy, crucial in the context of *C. sativa*'s allotetraploid genome structure. Overall, the Ka/Ks analysis underscores the evolutionary stability of CsNF-Y genes, with limited positive selection suggesting their conserved roles in transcriptional regulation. However, the few gene pairs with elevated Ka/Ks ratios represent promising candidates for further functional investigation, particularly in relation to stress adaptation.

3.10. Synteny Analysis of CsNF-Y Genes

To explore the evolutionary relationships and genomic organization of the NF-Y gene family, a synteny analysis was conducted between *Camelina sativa* and *Arabidopsis thaliana*, both members of the Brassicaceae family. A total of 73 CsNF-Y genes (28 NF-YA, 15 NF-YB, 30 NF-YC) and 36 AtNF-Y genes (10 NF-YA, 13 NF-YB, 13 NF-YC) were identified from the Phytozome database (accessed 03 April 2024). Syntenic blocks were identified using MCScanX and visualized via a dual synteny Circos plot in TBtools (Fig. 11).

The results revealed strong syntenic conservation, particularly between *C. sativa* chromosomes 1–5 and *A. thaliana* chromosomes 1–5, suggesting ancestral genomic collinearity. For instance, *C. sativa* chromosome 1 displayed extensive synteny with *A. thaliana* chromosome 1, with similar patterns seen for chromosomes 2 through 5. This implies the presence of orthologous NF-Y genes conserved across both species. Beyond chromosomes 1–5, dispersed syntenic relationships were observed between *C. sativa* chromosomes 6–20 and multiple *A. thaliana* chromosomes, indicating more complex evolutionary events involving rearrangements and duplications.

The syntenic associations are consistent with the hexaploid nature of *C. sativa* and its history of whole-genome triplication, which likely contributed to the expansion of the NF-Y gene family. Several *A. thaliana* chromosomes showed syntenic links with multiple *C. sativa* chromosomes—for example, *A. thaliana* chromosome 1 aligned with *C. sativa* chromosomes 1, 6, and 11—suggesting that NF-Y genes on these *C. sativa* chromosomes may be duplicated orthologs of those in *A. thaliana*.

Although the Circos plot does not show the specific positions of individual NF-Y genes, the syntenic blocks provide a framework for identifying orthologous gene relationships. For example, NF-YA genes involved in stress responses in *A. thaliana* may have functional orthologs in *C. sativa* chromosomes 1, 6, or 11. Similarly, conserved NF-YB and NF-YC genes likely retain roles in heterotrimer assembly and developmental regulation. The duplication of these syntenic regions in *C. sativa* may have facilitated neofunctionalization or sub functionalization, contributing to enhanced adaptability. In conclusion, this synteny analysis underscores the conserved genomic structure between *C. sativa* and *A. thaliana*, while highlighting gene family expansion through duplication events. These findings support the functional diversification of CsNF-Y genes and provide a valuable foundation for future studies on gene expression, stress tolerance, and agronomic improvement in *C. sativa*.

3.11. Gene Ontology Analysis of CsNF-Y Genes

Gene Ontology (GO) enrichment analysis was performed to elucidate the functional roles of the 73 Nuclear Factor Y (NF-Y) genes identified in *Camelina sativa* (28 NF-YA, 15 NF-YB, 30 NF-YC). The analysis utilized the ShinyGO online tool with *Arabidopsis thaliana* serving as the reference genome due to the limited GO annotations available for *C. sativa*. The list of CsNF-Y gene IDs was submitted, and enrichment was assessed using Fisher's exact test with a false discovery rate (FDR) cutoff of 0.05, as outlined in the methodology. The enriched GO terms were categorized into biological process, molecular function, and cellular component, and the results were visualized as a barplot using ShinyGO (Fig. 12). The barplot displays the number of genes associated with each GO term on the x-axis, the terms on the

y-axis, and statistical significance as $-\log_{10}(\text{FDR})$ values, with color intensity ranging from blue (low significance, $-\log_{10}(\text{FDR}) \sim 25$) to dark red (high significance, $-\log_{10}(\text{FDR}) \sim 90$). The biological process category revealed significant enrichment in terms related to transcriptional regulation and developmental responses. The term "positive regulation of transcription from RNA polymerase II promoter" was highly enriched, with approximately 40 genes and a $-\log_{10}(\text{FDR})$ of 85, indicating a primary role in controlling gene expression through RNA polymerase II. Light signaling was also prominent, with "regulation of photomorphogenesis" involving around 25 genes and a $-\log_{10}(\text{FDR})$ of 70, suggesting a key function in light-dependent developmental processes such as photomorphogenesis.

Developmental terms like "somatic embryogenesis" and "regulation of flower development" each encompassed about 20 genes, with $-\log_{10}(\text{FDR})$ values of 55 and 50 respectively, pointing to involvement in embryo formation and flowering regulation, which may support *C. sativa*'s reproductive success.

The term that arose as most significant within the molecular function category was that of transcription factor activity, sequence-specific DNA binding, with a number of about 35 genes and a $-\log_{10}(\text{FDR})$ of 80, meaning that the NF-Y genes have a transcription control activity that DNA sequence specific. The protein heterodimerization activity term, containing approximately 30 genes, and a $-\log_{10}(\text{FDR})$ of 75, points at their engagement in construction of the heterotrimeric NF-Y complex, which is basic in transcriptional activation. Other terms including the DNA binding transcription factor activity that has 25 genes with $-\log_{10}(\text{FDR})$ DNA-binding transcription factor activity of 65 further points to their regulatory capabilities in DNA-binding and the modification of gene expression.

Cellular component category presented the highest enrichment, with the term of transcription factor complex representing the most significant term, approximately 45 genes have the $-\log_{10}(\text{FDR})$ of 90. This emphasizes the nuclear localization and complex formation of NF-Y proteins which is in tandem with their role as transcription factors. The fact that the most significantly enriched terms include nucleus, around 40 genes and a $-\log_{10}(\text{FDR})$ of 85, and CCAAT-binding factor complex, 35 genes and a $-\log_{10}(\text{FDR})$ of 80, corroborates their nuclear functionality, and of being a part of the NF-Y family, which registers to CCAAT motifs in order to regulate gene transcription.

Overall analysis of the enriched GO terms implies that CsNF-Y genes have a wide variety of and significant impact in *Camelina sativa*, especially in the regulation of transcription, light signal pathways and development. The great importance of the term's relative to transcription factor complexes and DNA binding point to the fact that, these are genes, which are key to making regulatory units that regulate expression of genes, a role that is probably doubled by the hexaploid genome of the *C. sativa*, which has amplified the number of NF-Y genes (36 in *A. thaliana* transcriptome). The high frequency of light-related words and developmental-related terms indicates that CsNF-Y genes are involved in photomorphogenesis and embryo growth, which might be associated with the plasticity of *C. sativa* to the changing environmental conditions, including stress factors, such as salt. The proximity of *A. thaliana* to *C. sativa* in phylogeny under the same family (Brassicaceae) was used as a reference

because it presented a sound foundation of annotation which elucidated that the functional similarities can be utilized to enhance the agronomic properties of *C. sativa*.

3.12. Differential gene expression of CsNF-Y genes across Salinity Stress

The expression profiles of the 73 Nuclear Factor Y (NF-Y) genes identified in *Camelina sativa* were analyzed using publicly available RNA-seq data from the Gene Expression Omnibus (GEO) database under accession number GSE102422 (Table 2). This dataset encompasses RNA-seq profiles of *C. sativa* plants exposed to salinity stress and control conditions, with samples collected from root and shoot tissues, each represented by three biological replicates (GSM2736320–GSM2736331). The processed data, provided as a count per million (CPM) matrices in the file "GSE102422_CPM_Matrix.xlsx," was downloaded from the GEO series record. Expression values for the 73 CsNF-Y genes were extracted from this matrix, and mean expression levels were calculated across the three replicates for each of the four conditions: root with salt, root without salt, shoot with salt, and shoot without salt. These mean CPM values were utilized to generate a heatmap using TBtools, which visualized the expression patterns across tissues and conditions to assess the potential roles of CsNF-Y genes in the salinity stress response (Fig. 13).

Table 2
Description of samples used for salinity stress expression analysis in *Camelina Sativa*. Expression data were retrieved from the NCBI Gene Expression Omnibus (GEO) under accession GSE102422.

Description	Accession No.
Root after salt treatment 1	GSM2736320
Root after salt treatment 2	GSM2736321
Root after salt treatment 3	GSM2736322
Root without salt treatment 1	GSM2736323
Root without salt treatment 2	GSM2736324
Root without salt treatment 3	GSM2736325
Shoot after salt treatment 1	GSM2736326
Shoot after salt treatment 2	GSM2736327
Shoot after salt treatment 3	GSM2736328
Shoot without salt treatment 1	GSM2736329
Shoot without salt treatment 2	GSM2736330
Shoot without salt treatment 3	GSM2736331

The heatmap exhibited various expression patterns across the CsNF-Y genes, with color intensity ranging from blue (low expression, CPM ~ 0–10) to red (high expression, CPM ~ 50–100), reflecting the degree of transcript abundance under different conditions. Several genes, particularly within the NF-YA and NF-YC subunits, demonstrated considerably greater expression in root tissues under salt stress compared to normal conditions. For instance, genes such as CsNF-YA01 and CsNF-YC15 showed a large increase in expression, with mean CPM values rising from roughly 5 under root without salt to 60 under root with salt, demonstrating a robust induction in response to salinity. This increase in roots implies that these genes may play a substantial role in mediating stress tolerance mechanisms, such as osmotic adjustment or ion homeostasis, which are crucial for *C. sativa*'s survival under saline conditions. In contrast, shoot tissues indicated more mild changes, with genes including CsNF-YB07 showing a slight spike from 10 CPM (shoot without salt) to 20 CPM (shoot with salt), reflecting a less substantial response or a tissue-specific regulatory role.

There was also evidence of tissue specific expression with certain genes having lower baseline expression in shoots had controls. As a specific example, CsNF-YC28 showed an average CPM of about 30 in the shoots in the absence of salt and a value decreased to 15 under salt stress and this may indicate a part of normal developmental processes that are repressed during stress. The heatmap also revealed a group of such genes with low expression (CPM = 510) that are not differentially expressed in any of the conditions and conversely are expressed in all tissues (such as CsNF-YA18). On the whole, the differential expression and especially its upregulation to salt stress in roots confirms the hypothesis that CsNF-Y genes provide *C. sativa* with adaptability to environmental stress with root as the major stress response site.

Diversity in expression profiles is probably supported by the extended NF-Y gene family of *C. sativa*, because of its hexaploid genome. The up-regulation of some genes that occurs under the salinity condition is in agreement with previous reports on the resistance of *C. sativa* to salinity conditions, hence a specific CsNF-Y gene may be deemed as resource to improve the tolerance to salinity condition. The heatmap gives us the visual context within which we can use to select our candidate genes that need to be further assayed function-wise such as CsNF-YA01 and CsNFYC15 that had the most downstream induction. These findings provide the foundation to decipher the molecular architecture of *C. sativa* on how it responds to stress and how it could potentially be an economically viable stress tolerant crop.

4. DISCUSSION

The extensive genome-wide research of the Nuclear Factor Y (NF-Y) gene family in *Camelina sativa* offers crucial insights into the structural, functional, and evolutionary complexity of this transcription factor family in a hexaploid context [43]. A total of 73 CsNF-Y genes were identified—28 NF-YA, 15 NF-YB, and 30 NF-YC—suggesting significant gene family expansion, likely due to the whole-genome triplication that characterizes *C. sativa*. Compared to diploid species like *Arabidopsis thaliana*, this growth suggests a diversification of regulatory mechanisms and better resilience to environmental stressors [44]. The

structural variety of NF-Y subunits reflects specialized roles in creating heterotrimeric complexes that regulate gene expression via the CCAAT box [45]. NF-YA subunits, with their DNA-binding domains, are critical for targeting certain promoter regions, while NF-YB and NF-YC are principally engaged in protein-protein interactions and complex stability [46]. The higher number of NF-YC genes suggests an enlarged capability for transcriptional control, possibly permitting combinatorial specificity in response to various physiological conditions.

Polyploidy in *C. sativa* has not only resulted in gene duplication but also contributed to functional diversification through sub functionalization or neofunctionalization [47]. Phylogenetic study demonstrates evolutionary conservation with *A. thaliana* [48, 49], yet also highlights species-specific divergence that may underpin unique stress adaptation mechanisms. Gene retention post duplication is likely to have enabled *C. sativa* to sustain critical developmental processes while also acquiring additional stress response roles—especially under abiotic stresses like drought and salinity [50].

Physicochemical study of CsNF-Y proteins demonstrated significant variation in molecular weight, isoelectric points, and hydropathy, demonstrating their adaptability across diverse cellular settings. Subcellular localization analysis suggests that most NF-Y proteins localize to the nucleus, consistent with their activity as transcription factors. However, numerous NF-YC subunits display localization to plastids or chloroplasts, suggesting possible non-nuclear activities such as engagement in organellar signaling or stress mitigation at the subcellular level.

The non-uniform chromosomal distribution and clustering of CsNF-Y genes—especially on chromosomes like Chr11 and Chr2—points toward localized duplication events and co-regulated gene groups. Such genomic architectures may allow coordinated expression during stress or development. Synteny analysis with *A. thaliana* indicates substantial conservation, specifically among chromosomes 1–5, underlining the shared ancestry and possible orthologs of multiple NF-Y genes.

Cis-regulatory element (CRE) research further stresses the functional breadth of CsNF-Y genes. The presence of light-responsive (G-box), stress-responsive (MBS), and hormone-related (TCA-element) motifs throughout promoter regions shows that NF-Y genes are vital to integrating environmental and hormonal inputs [8]. Particularly, NF-YA and NF-YC subunits exhibit promoter enrichment for light and stress factors, tying them to photomorphogenesis and salinity tolerance—traits crucial for bioenergy crops produced on marginal soils. Expression profiling under salt stress using RNA-seq data (GSE102422) demonstrated tissue-specific expression patterns [51]. Genes like CsNF-YA01 and CsNF-YC15 were highly upregulated in roots during salt stress, showing their role in ion homeostasis and osmotic regulation. In contrast, the more muted expression in shoots supports organ-specific control, possibly suited for preserving energy efficiency under stress.

Gene Ontology (GO) enrichment supports functions in transcription control [52], photomorphogenesis [53], and somatic embryogenesis [54]—underscoring the significance of CsNF-Y genes in essential developmental and stress-responsive pathways. Notably, multiple genes are implicated in the positive regulation of transcription from RNA polymerase II promoters [55], a hallmark of active transcriptional

control during stress adaption. When compared to *A. thaliana* and polyploid relatives like *Brassica napus* or *Panicum virgatum*, the extensive NF-Y family in *C. sativa* stands out as a consequence of polyploid-driven regulatory complexity. Functional redundancy and specialization coexist, permitting robust transcriptional responses and the creation of distinct NF-Y complexes adapted to specific environmental or developmental circumstances.

In summary, the NF-Y gene family in *Camelina sativa* reveals how genome triplication can improve regulatory flexibility and stress resilience [14]. The detailed characterization of this gene family not only enriches our understanding of transcription factor evolution in polyploids but also provides a strong platform for genetic interventions aiming at boosting stress tolerance and agronomic performance in biofuel crops.

5. CONCLUSION

The comprehensive evaluation of the CsNF-Y gene family in *Camelina sativa* stresses its astonishing structural and functional complexity, partly owing to the species' hexaploid genome and the various regulatory mechanisms governing its 73 genes. This work reveals how polyploidy has promoted gene duplication and diversification, improving the family's potential to influence important processes such as stress tolerance and developmental plasticity. The findings suggest considerable potential for improving *C. sativa*'s agronomic traits, particularly its tolerance to salinity, utilizing targeted genetic approaches that exploit the identified NF-Y subunits and associated regulatory components. Such modifications could enhance the crop's viability for marginal soils, improving its value as a sustainable bioenergy source. However, fulfilling these potential demands further functional research to establish gene-specific roles, study expression dynamics under diverse settings, and develop genetic engineering methodologies, paving the way for improved crop improvement and greater ecological adaptation.

References

1. Boeva V (2016) Analysis of genomic sequence motifs for deciphering transcription factor binding and transcriptional regulation in eukaryotic cells. *Front Genet* 7:24
2. Latchman DS (2010) Eukaryotic transcription factors. Academic
3. Asensi-Fabado M-A, Amtmann A, Perrella G (2017) *Plant responses to abiotic stress: the chromatin context of transcriptional regulation*. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 1860(1):106–122
4. Gusmaroli G, Tonelli C, Mantovani R (2002) Regulation of novel members of the *Arabidopsis thaliana* CCAAT-binding nuclear factor Y subunits. *Gene* 283(1–2):41–48
5. Kim I-S et al (1996) Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. *Mol Cell Biol* 16(8):4003–4013

6. Kumimoto RW et al (2010) NF–YC3, NF–YC4 and NF–YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in *Arabidopsis thaliana*. *Plant J* 63(3):379–391
7. Mu J et al (2013) *Arabidopsis* transcription factor genes NF-YA1, 5, 6, and 9 play redundant roles in male gametogenesis, embryogenesis, and seed development. *Mol Plant* 6(1):188–201
8. Zhang H et al (2023) Crucial abiotic stress regulatory network of NF-Y transcription factor in plants. *Int J Mol Sci* 24(5):4426
9. Bhattacharjee B, Hallan V (2023) NF-YB family transcription factors in *Arabidopsis*: Structure, phylogeny, and expression analysis in biotic and abiotic stresses. *Front Microbiol* 13:1067427
10. Siefers N et al (2009) Tissue-specific expression patterns of *Arabidopsis* NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol* 149(2):625–641
11. Pereira SL et al (2018) 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* 13(6):e0199187
12. Song T et al (2022) Genome-Wide Identification and Analysis of the NF-Y Transcription Factor Family in *Medicago sativa* L. *Agronomy* 12(5):1237
13. Yan X et al (2024) A member of NF-Y family, OsNF-YC5 negatively regulates salt tolerance in rice. *Gene* 892:147869
14. Wang J et al (2020) 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* 148:112253
15. Quach TN et al (2015) Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. *Mol Genet Genomics* 290(3):1095–1115
16. Xu L et al (2014) Multiple NUCLEAR FACTOR Y transcription factors respond to abiotic stress in *Brassica napus* L. *PLoS ONE* 9(10):e111354
17. Li M et al (2023) Genome-wide identification and chilling stress analysis of the NF-Y gene family in melon. *Int J Mol Sci* 24(8):6934
18. Feng Z-J et al (2015) Foxtail millet NF-Y families: genome-wide survey and evolution analyses identified two functional genes important in abiotic stresses. *Front Plant Sci* 6:1142
19. Ren C et al (2016) Genome-wide identification and characterization of the NF-Y gene family in grape (*Vitis vinifera* L). *BMC Genomics* 17(1):605
20. Panahi B et al (2019) Genome-wide identification and co-expression network analysis of nuclear factor-Y in barley revealed potential functions in salt stress. *Physiol Mol Biology Plants* 25(2):485–495
21. Wang T et al (2023) Genome-wide identification, characterization, and expression analysis of NF-Y gene family in *Ginkgo biloba* seedlings and GbNF-YA6 involved in heat-stress response and tolerance. *Int J Mol Sci* 24(15):12284

22. Brock JR et al (2018) Phylogenetics of *Camelina Crantz*.(Brassicaceae) and insights on the origin of gold-of-pleasure (*Camelina sativa*). *Mol Phylogenet Evol* 127:834–842
23. Obour AK et al (2015) Oilseed camelina (*Camelina sativa* L Crantz): Production systems, prospects and challenges in the USA Great Plains. *Adv Plants Agric Res* 2(2):00043
24. Hossain Z et al (2019) Comparative analysis of oil and protein content and seed yield of five Brassicaceae oilseeds on the Canadian prairie. *Ind Crops Prod* 136:77–86
25. Na G et al (2019) Enhancing micro RNA 167A expression in seed decreases the α -linolenic acid content and increases seed size in *Camelina sativa*. *Plant J* 98(2):346–358
26. Na G et al (2018) Seed-specific suppression of ADP-glucose pyrophosphorylase in *Camelina sativa* increases seed size and weight. *Biotechnol Biofuels* 11(1):330
27. Goodstein DM et al (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40(D1):D1178–D1186
28. Gasteiger E et al (2005) *Protein identification and analysis tools on the ExPASy server*, in *The proteomics protocols handbook*. Springer, pp 571–607
29. Bailey TL et al (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37(suppl2):W202–W208
30. Thompson JD, Gibson TJ, Higgins DG *Multiple sequence alignment using ClustalW and ClustalX*. *Current protocols in bioinformatics*, 2003(1): p. 2.3. 1-2.3. 22
31. Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol* 38(7):3022–3027
32. Zhou T et al (2023) itol. toolkit accelerates working with iTOL (Interactive Tree of Life) by an automated generation of annotation files. *Bioinformatics* 39(6):btad339
33. Hu B et al (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31(8):1296–1297
34. Horton P et al (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35(suppl2):W585–W587
35. Chen C et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 13(8):1194–1202
36. BJ WM et al (2024) Design and Development of a Website-Based Agriplan Information System to Help Plant Care. *J Comput Sci Inf Eng (J-Cosine)*, 8(2)
37. Nekrutenko A, Makova KD, Li W-H (2002) The KA/KS ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. *Genome Res* 12(1):198–202
38. Zhao T, Schranz ME (2017) Network approaches for plant phylogenomic synteny analysis. *Curr Opin Plant Biol* 36:129–134
39. Wang Y et al (2024) Detection of colinear blocks and synteny and evolutionary analyses based on utilization of MCScanX. *Nat Protoc* 19(7):2206–2229

40. Ge SX, Jung D, Yao R (2020) ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36(8):2628–2629
41. Clough E, Barrett T (2016) *The gene expression omnibus database*, in *Statistical Genomics: Methods and Protocols*. Springer, pp 93–110
42. Chen C, Xia R (2022) *Interactive data analyses using TBtools*, in *Integrative bioinformatics: history and future*. Springer, pp 343–375
43. Wang P et al (2019) Identification, expression, and putative target gene analysis of nuclear factor-Y (NF-Y) transcription factors in tea plant (*Camellia sinensis*). *Planta* 250(5):1671–1686
44. Tonsor S, ALONSO-BLANCO C, Koornneef M (2005) Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*, vol 28. *Plant, Cell & Environment*, pp 2–20. 1
45. Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene* 239(1):15–27
46. Romier C et al (2003) The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem* 278(2):1336–1345
47. Hutcheon C et al (2010) Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. *BMC Plant Biol* 10(1):233
48. Brock JR et al (2022) Chloroplast phylogenomics in *Camelina* (Brassicaceae) reveals multiple origins of polyploid species and the maternal lineage of *C. sativa*. *Hortic Res* 9:uhab050
49. Guo YL (2013) Gene family evolution in green plants with emphasis on the origination and evolution of a *Arabidopsis thaliana* genes. *Plant J* 73(6):941–951
50. Sabagh AE et al (2019) Drought and salinity stress management for higher and sustainable canola ('*Brassica napus*' L.) production: A critical review. *Aust J Crop Sci* 13(1):88–96
51. Zhu J et al (2016) *Identification of tissue-specific protein-coding and noncoding transcripts across 14 human tissues using RNA-seq*. *Scientific reports*. 6(1):28400
52. Gaudet P et al (2021) *Gene Ontology representation for transcription factor functions*. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 1864(11–12):194752
53. Song L et al (2009) Genome-wide analysis revealed the complex regulatory network of brassinosteroid effects in photomorphogenesis. *Mol Plant* 2(4):755–772
54. Li H-Z et al (2023) Transcriptome analysis revealed enrichment pathways and regulation of gene expression associated with somatic embryogenesis in *Camellia sinensis*. *Sci Rep* 13(1):15946
55. Butler JE, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* 16(20):2583–2592

Supplementary Tables

Supplementary Table S1 and S2 are not available with this version.

Figures



Figure 1

(A) Graphical representation of Amino acid residues of 73 CsNF-Y **(B)** Isoelectric points of NF-Y proteins. X-axis represents the CsNF-Y protein IDs and Y-axis represents the isoelectric point **(C)** Molecular weight of 73 CsNF-Y proteins **(D)** Grand average of hydropathy values of NF-Y proteins. X-axis represents the CsNF-Y protein IDs and Y-axis represents the GRAVY values.

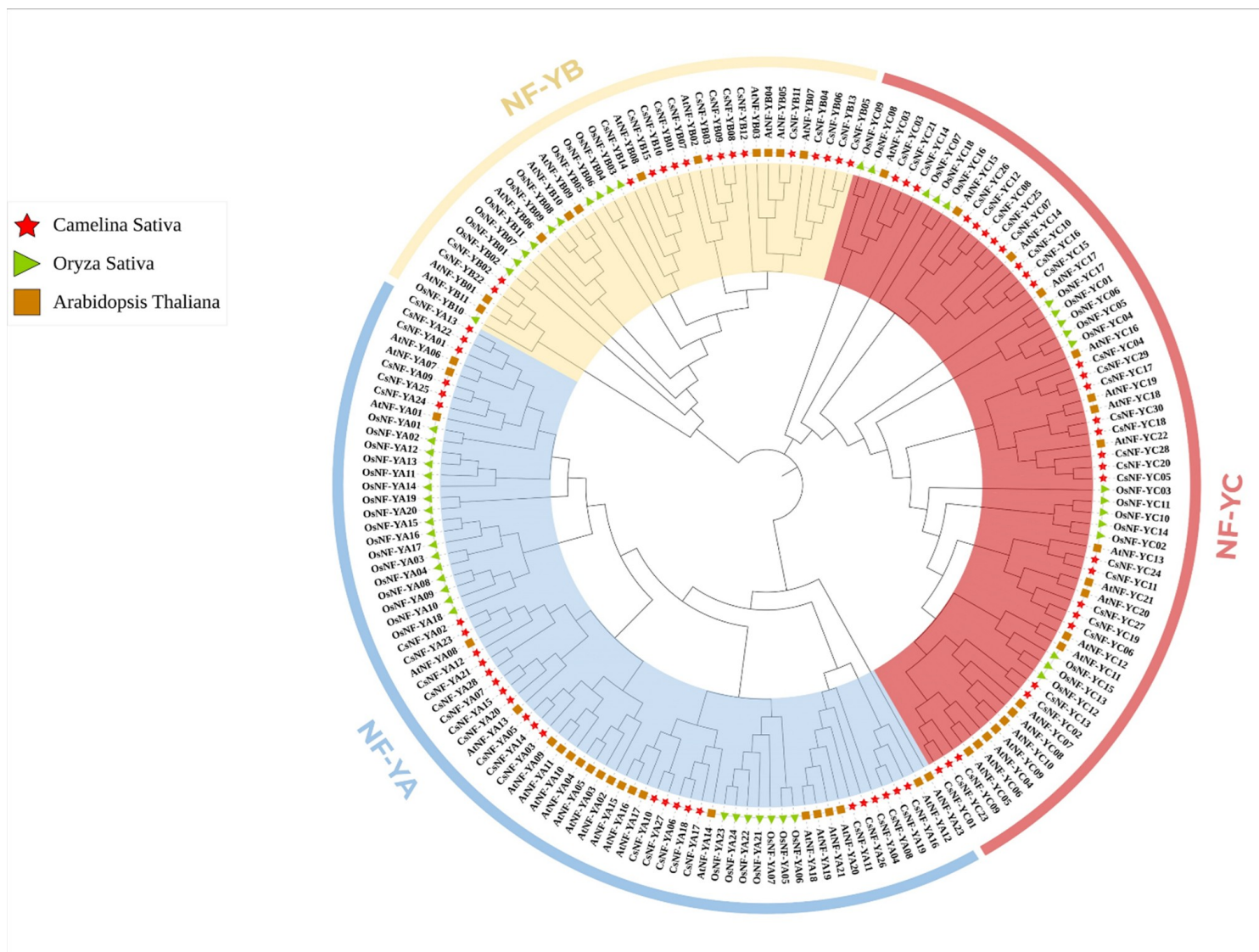


Figure 2

A phylogenetic analysis of NF-Y transcription factor proteins from *Camelina sativa*, *Oryza sativa*, and *Arabidopsis thaliana*. We utilized Clustal Omega with 1000 bootstrap runs to make sure that the topology of the Neighbor-Joining (NJ) phylogenetic tree was strong. There are three main groups of NF-Y proteins: NF-YA, NF-YB, and NF-YC. Each group is illustrated by a separate color segment: blue for NF-YA, yellow for NF-YB, and red for NF-YC. Red stars exhibit *Camelina sativa* proteins, green triangles show *Oryza sativa* proteins, and brown squares reveal *Arabidopsis thaliana* proteins.

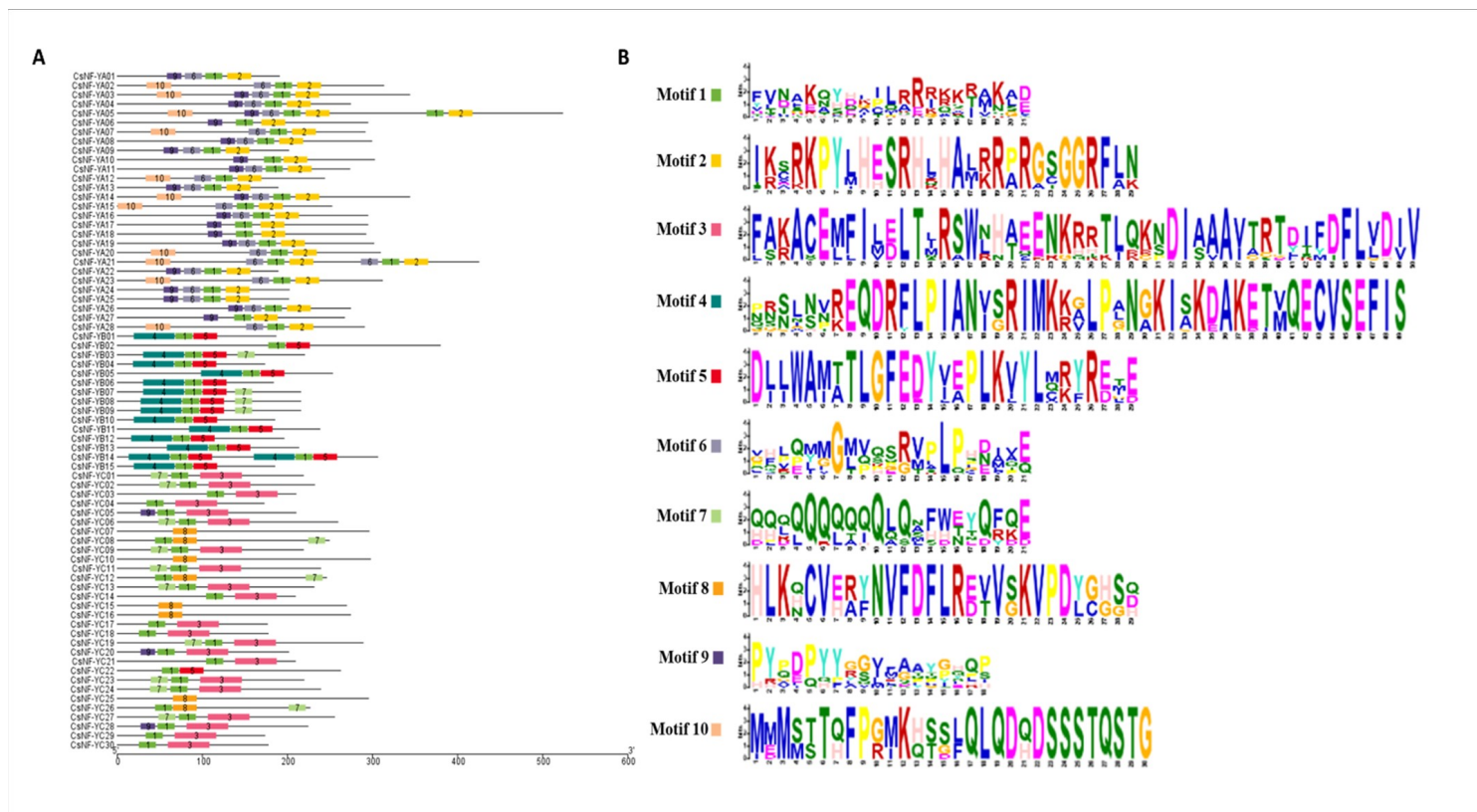


Figure 3

The distribution of conserved motifs in the CsNF-Ys genes. **(A)** Analysis of conserved motifs. Different color bars represented motif types in each subunit of PvNF-Ys proteins. The length of proteins can be estimated using the scale at the bottom. **(B)** Sequence information of 10 conserved motifs

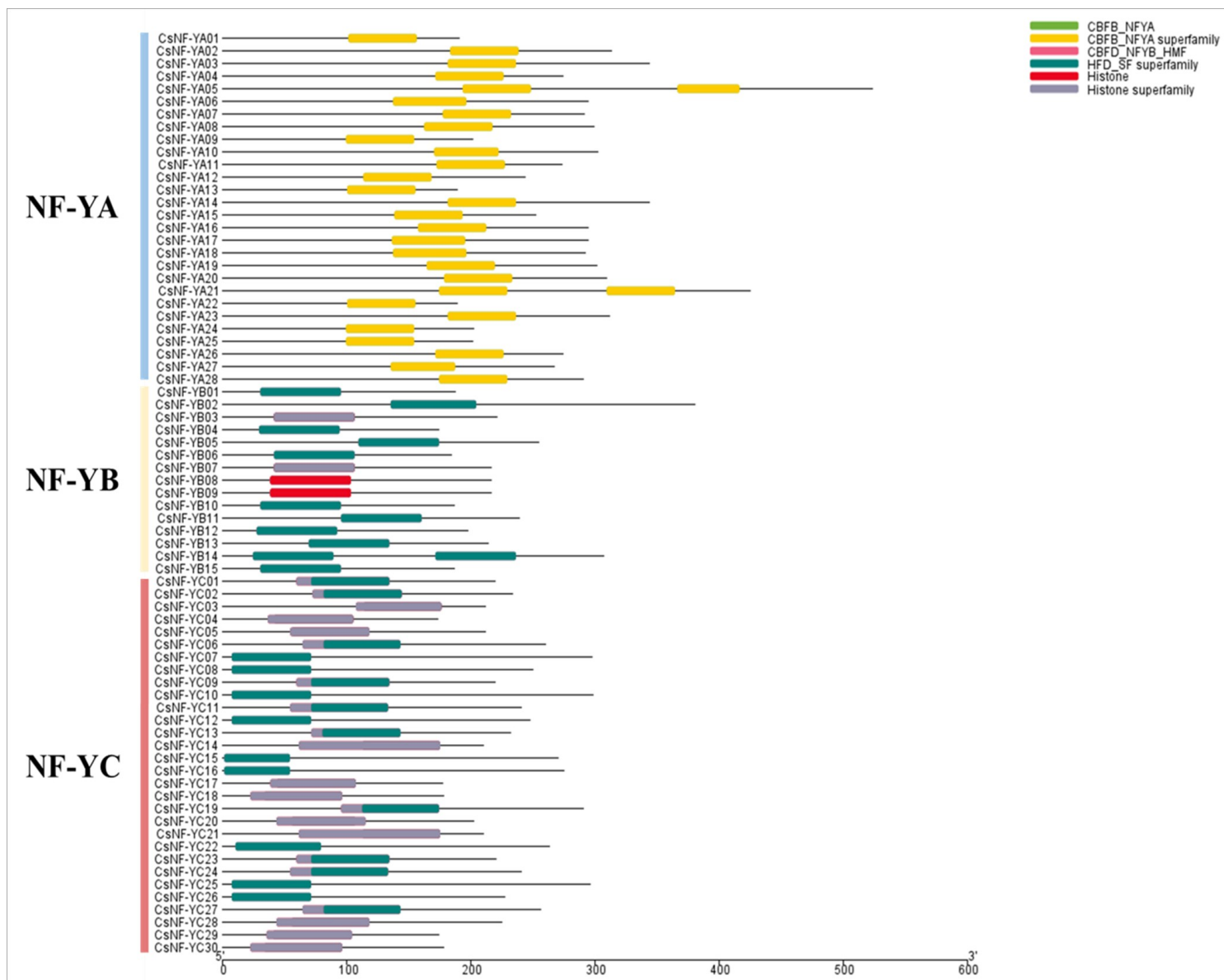


Figure 4

Conserved domains of CsNF-Ys genes. Different color boxes represent the conserved domains.

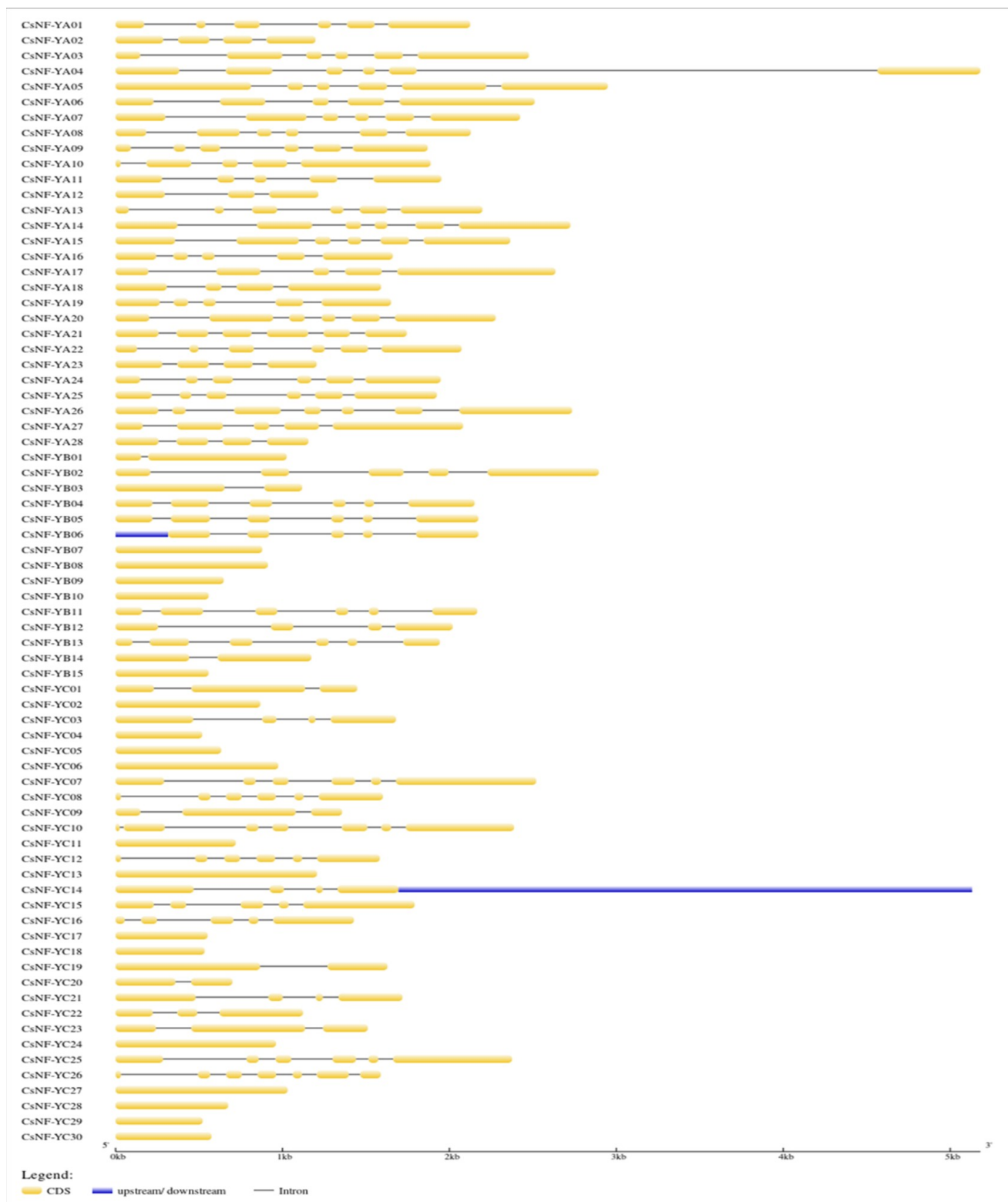


Figure 5

The number of introns, CDS and upstream/downstream were represented in black lines, yellow bars, and blue bars in each gene.

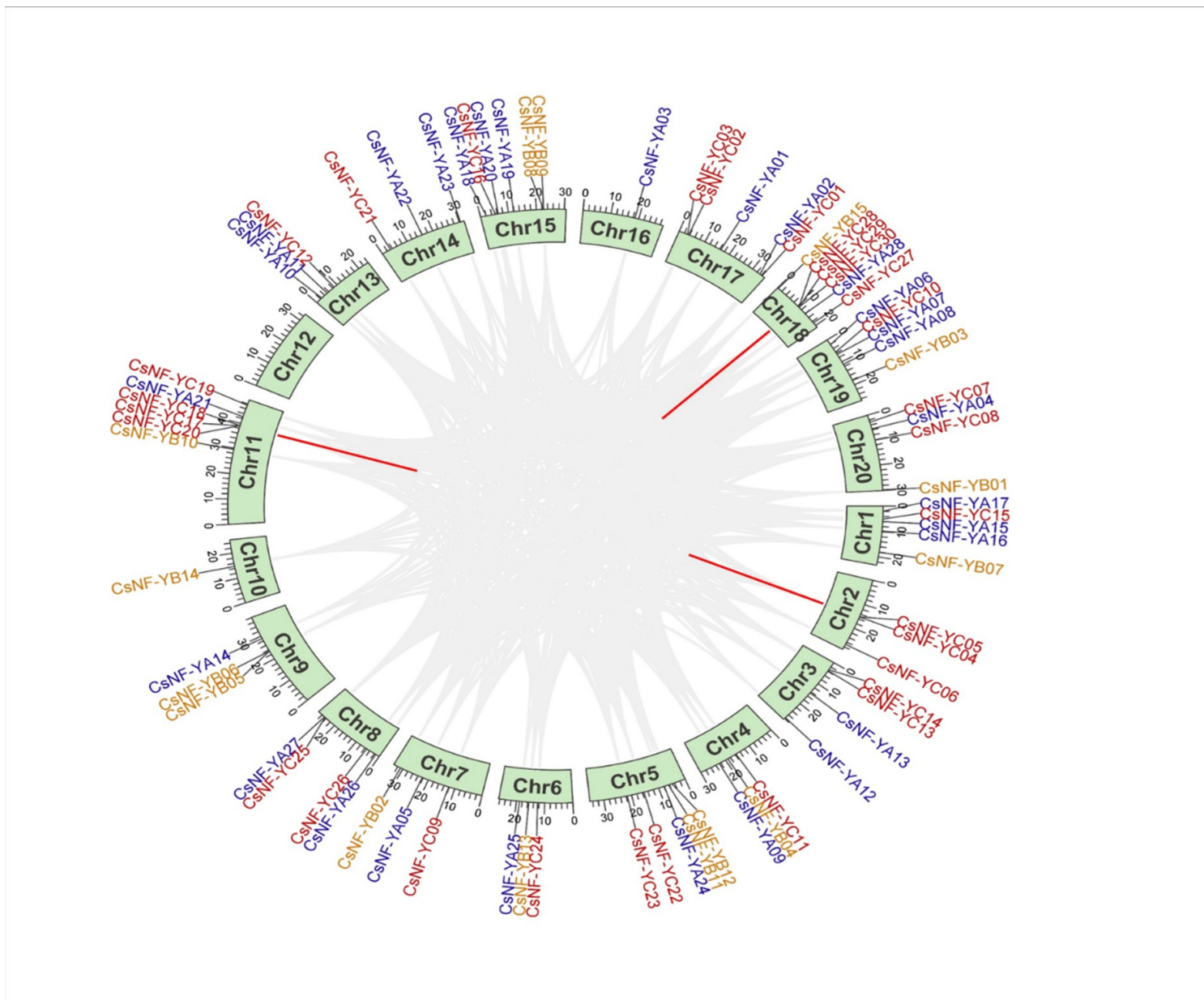


Figure 6

Chromosome distribution of PvNF-Ys genes is marked with different colors (NF-YA, blue; NF-YB, yellow; NF-YC, red)

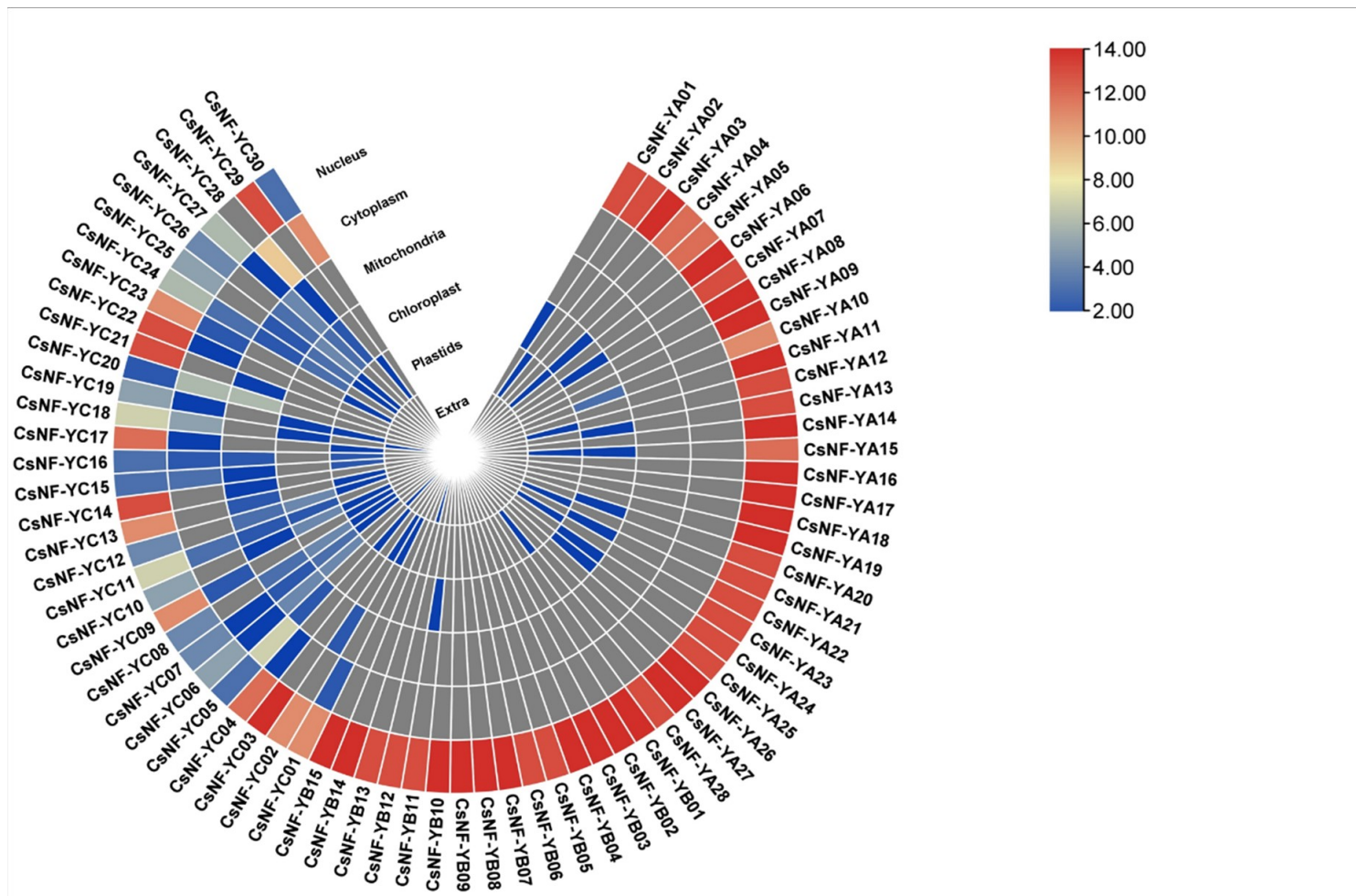


Figure 7

Subcellular localization analysis along with phylogenetic tree in the CsNF-Y proteins.

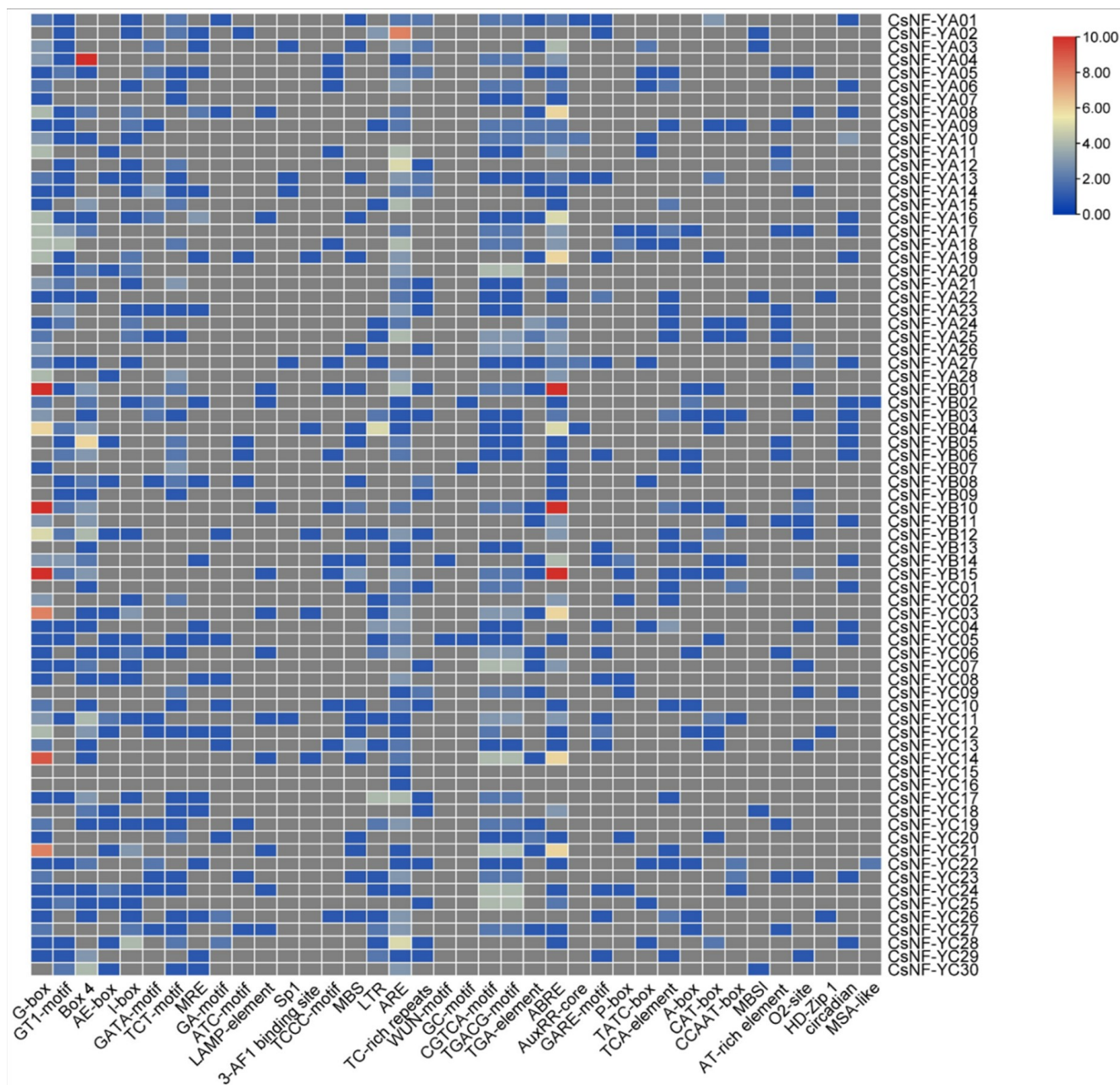


Figure 8

Different phytohormones, reactions to light and stress, and growth and developmental processes were linked to the analysis of cis-regulatory elements motifs in their respective groupings.

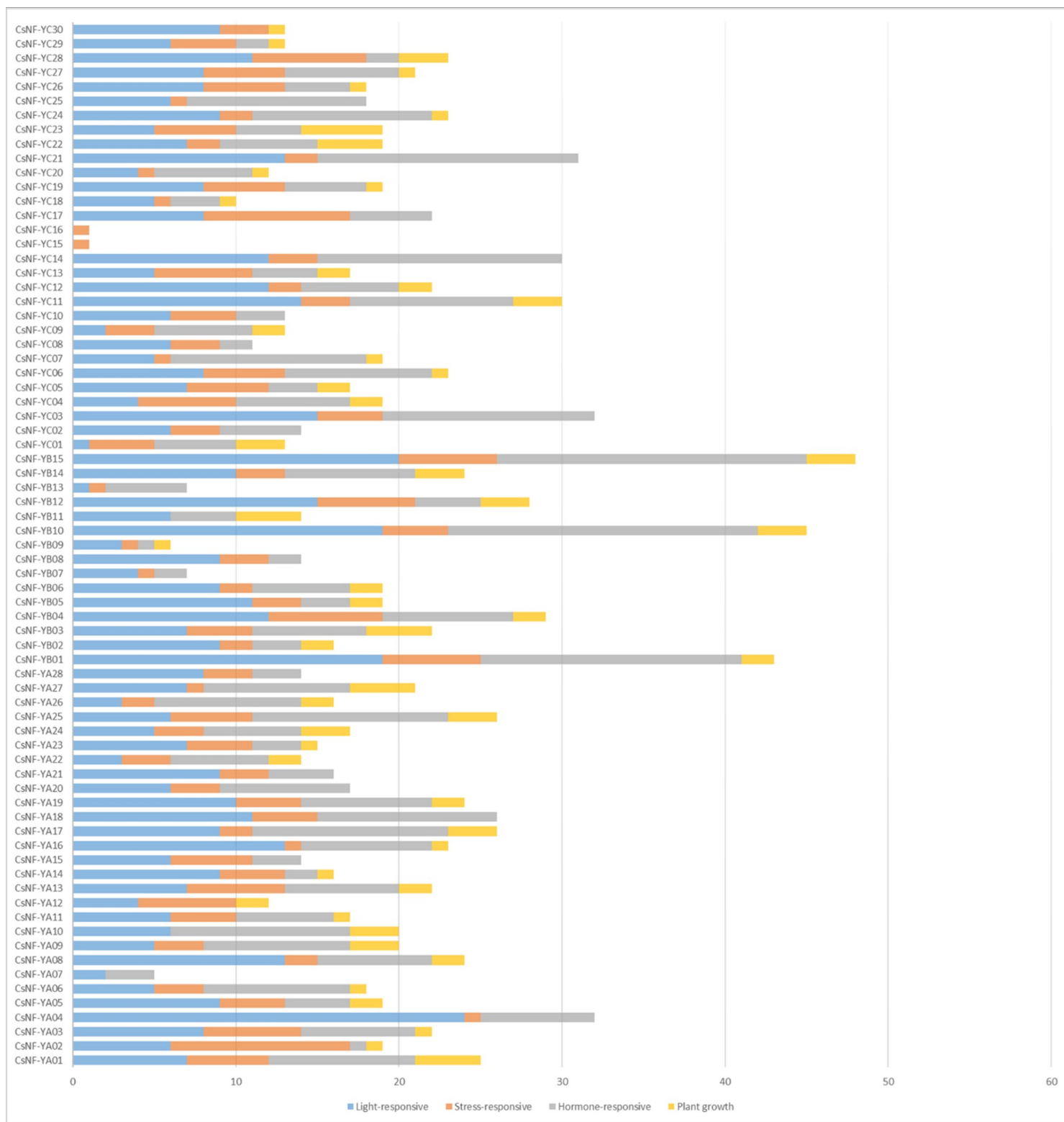


Figure 9

Depending on how frequently a specific cis-element appears in the promoter of each member of the CsNF-Y gene family, different color bars indicate the highest count.

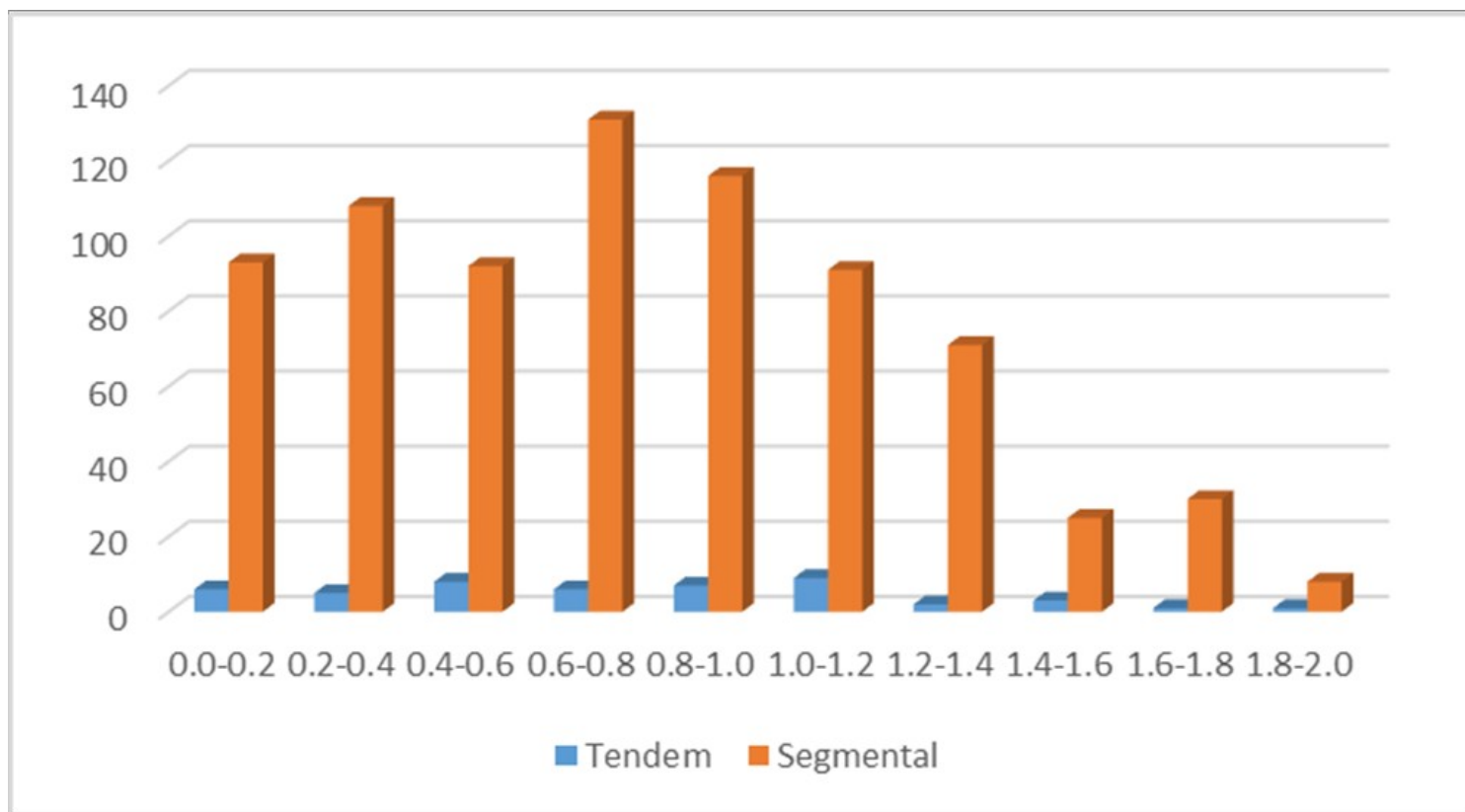


Figure 10

The distribution of Ka/Ks values in all tandem and segmental duplicated NF-Ys

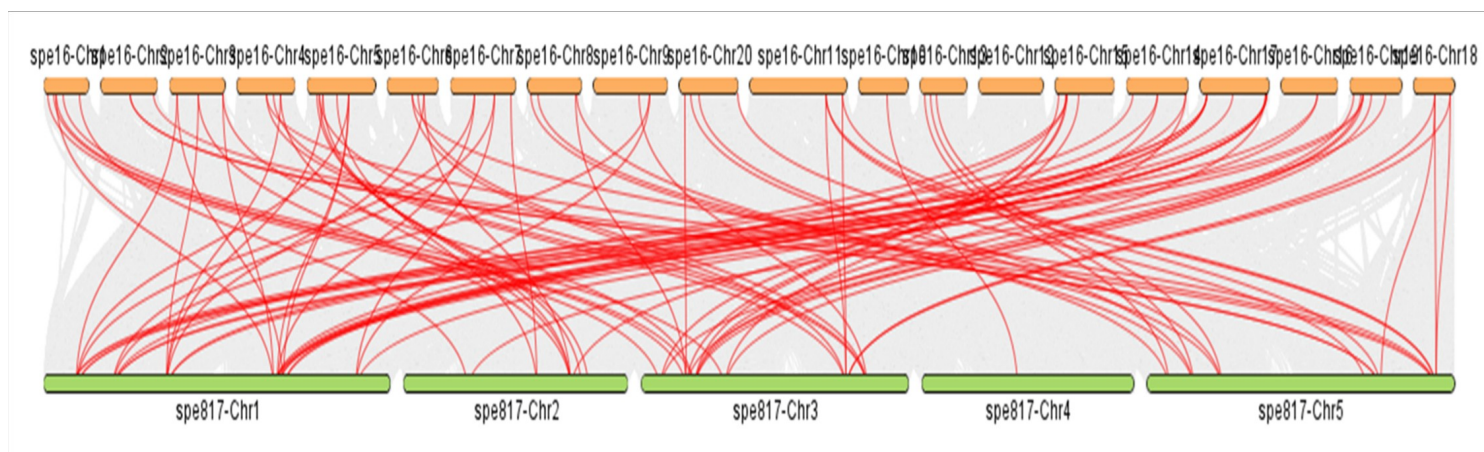


Figure 11

Dual Synteny analysis of *Camelina sativa* NF-Y genes with *Arabidopsis Thaliana*



Figure 12

Gene ontology analysis of the predicted CsNF-Ys genes significantly enriches < 1 Pvalue.

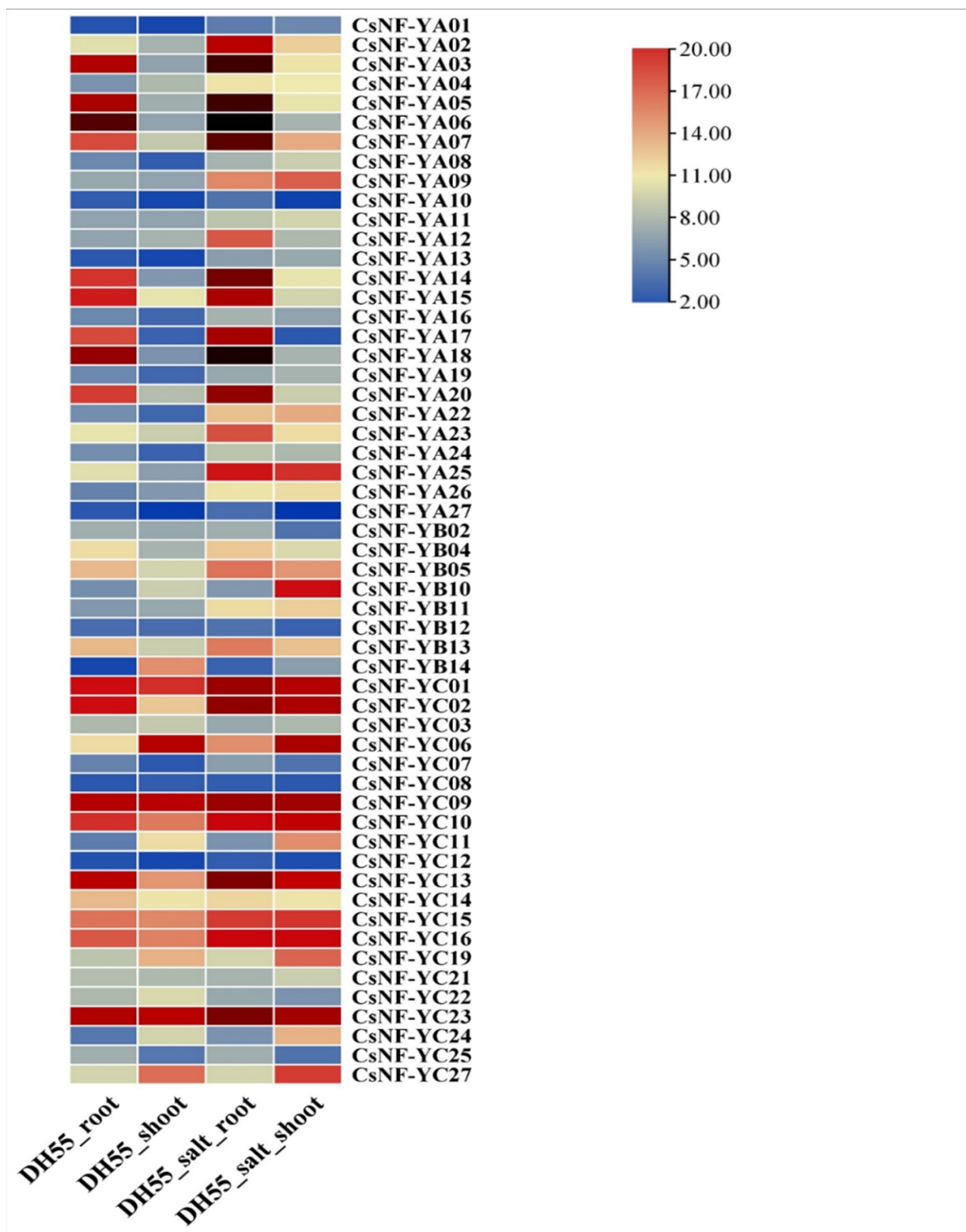


Figure 13

Expression heatmap of *Camelina sativa* NF-Y genes under normal and salt stress conditions based on transcriptomic data retrieved from the GEO dataset (GSE102422). The expression levels of CsNF-YA, CsNF-YB, and CsNF-YC genes were analyzed in four sample types: DH55_root, DH55_shoot, DH55_salt_root, and DH55_salt_shoot.