Soybean (*Glycine max*) NF-Y (Nuclear Factor of Y Box) gene family and its potential role under stress conditions and nodulation

*Corresponding Author(s): Alexandro Cagliari*
State University of Rio Grande do Sul - UERGS, CEP 96816-501, Santa Cruz do Sul, RS, Brazil
Tel: +55-51-3715-6926;
Email: alexandro-cagliari@uergs.edu.br

Received: Jun 25, 2018
Accepted: Sep 14, 2018
Published Online: Sep 20, 2018
Journal: Journal of Plant Biology and Crop Research
Publisher: MedDocs Publishers LLC
Online edition: http://meddocsonline.org/
Copyright: © Cagliari A (2018). This Article is distributed under the terms of Creative Commons Attribution 4.0 International License

**Keywords:** NF-Y transcription factor; Gene evolution; Soybean nodulation; Biotic and abiotic stresses

**Abstract**

The Nuclear Factor of Y box (NF-Y) is a transcription factor composed of three subunits (NF-YA, NF-YB, and NF-YC) involved in the regulation of genes related to physiological processes. In plants, NF-Y genes act in development, in addition to biotic and abiotic stresses response. Regarding *Glycine max* (Gm), the NF-Y genes are poorly investigated. In this important crop, the characterization of genes related to stresses represents an important material for improvement of cultivars. In this study, *GmNF-Y* genes were investigated through an extensive data mining using several bioinformatics approaches. We identify a total of 21 genes of *GmNF-YA*, 23 to *GmNF-YB*, and 24 to *GmNF-YC* subunits. Results showed that during evolution, *GmNF-Y* genes arose by block duplication and subfunctionalization processes could be occurring in this family. In *silico* expression analyses indicate that *GmNF-Y* genes are not ubiquitously expressed in organs; their expression are instead organ-dependent. In several biotic and abiotic stresses, the majority of *GmNF-Y* genes are modulated, especially during *Phakopsora pachyrhizi* infection, representing a new characterization for these genes. In addition, in root hair inoculated with *Bradyrhizobium japonicum* (nitrogen-fixing bacteria) some genes are modulated. Our study provides an overview of the subtle genetic diversification of the NF-Y family in soybean, through the characterization of genes considering their evolution and structure, and analysis of expression pattern. We highlight a putative involvement of *GmNF-Y* genes under multivariate environmental conditions related to plant defense responses, providing data to be explored further for the generation of stress tolerant/resistant plants.
Introduction

The Nuclear Factor of Y box (NF-Y) is a eukaryotic conserved oligotrimeric transcription factor involved in regulation of genes related to different physiological processes. NF-Y is composed of three subunits: NF-YA, NF-YB, and NF-YC. Each subunit is required for DNA binding, subunit association, and transcriptional regulation [1]. The three NF-Y subunits interact, forming a heterocomplex in plants [2] that can bind CCAAT DNA motifs, controlling the expression of target genes. NF-YA performs specific recognition and binding to the CCAAT-box sequence in the promoter of target genes. NF-YB and NF-YC promote chromatin accessibility to NF-YA and to additional transcription factors involved with the transcriptional regulation process [2-4].

It has been previously showed that NF-Y genes are involved with the transcriptional control of gametogenesis [5], embryogenesis [5,6], seed development and rhizobia symbiosis [2,7], seed germination [8], hormonal signaling and flowering time [9-11], root architecture and elongation [12], fruit ripening [8], photosynthesis [13,14], endoplasmic reticulum stress response [15], drought response [16-18], grain yield [19], salinity and ABA treatments [18] and cell proliferation in endosperm development [20]. These results clarify that NF-Y genes are involved in several important networks. Their characterization in several species could contribute to the understanding of their wide role in diverse plant mechanisms.

While most eukaryotic genomes possess only one or two genes encoding each NF-Y subunit, in vascular plants each subunit is encoded by gene families [21]. In Physcomitrella patens, 2 NF-YA, 9 NF-B, and 12 NF-YC proteins were identified [22]. In the emerging monocot model plant Brachypodium distachyon, there are 7 NF-YA, 17 NF-YB, and 12 NF-YC proteins [23]. In rice, there are at least 10 NF-YA, 11 NF-YB, and 7 NF-YC genes [24], while in Arabidopsis thaliana there are 10 genes coding for each NF-Y subunit [25].

Soybean (Glycine max) is one of the most important crops in the world. Different types of stresses severely restrict soybean productivity. The identification of genes related to stresses regulation is a fundamental issue, representing the raw material for improvement of soybean cultivars. Regarding the NF-Y family, it is poorly investigate in soybean. Here, we carried out an extensive data mining to identify NF-Y genes (GmNF-Y) in soybean genome. We characterized gene structures, duplication and evolution. In silico expression analyses of GmNF-Y in organs and stresses conditions were investigated. Our results contribute to the knowledge about the GmNF-Y and lay a foundation for deep characterization of this gene family in soybean.

Material and methods

NF-Y gene annotation and domains characterization

The nucleotide and protein sequences of the A. thaliana genes available on Phytozome [26] were used as queries to BLAST searches against soybean genome in this database, NCBI and Gramene [27]. The exon-intron of each GmNF-Y gene was investigated using GSDS [28].

Duplication events, synteny, and Ka/Ks substitution rate analysis

We have analyzed the mechanisms involved with the evolution of GmNF-Y genes using the WGmapping tool in the PLAZA v 3.0 database [29]. The physical co-localization of NF-Y genes in soybean genome was analyzed in the Genome Duplication Database (PGDD) [30], considering 100 kb syntenic region between paralogous GmNF-Y. This analysis uses the BLASTP tool to search for potential anchors (E<1e−5, top 5 matches) among every possible chromosome pair in multiple genomes. The homolog pairs identified are used as the input in the multiple collinearity scan (MCscan) program, and an E value <1e−10 is considered a significant cut-off. The putative classification of the genes identified in the syntenic regions was accessed using Phytozome and the PLAZA databases.

The ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks) in paralogous GmNF-Y genes was analyzed using the PGDD database.

NF-Y domains characterization

Proteins with at least 25% of homology with the queries were examined to the presence of the NF-Y domains [1] using SMART [31] and InterProScan Signature [32]. GmNF-Y domain conservation was investigated using MEME [33].

Expression analysis of GmNF-Y genes

The expression pattern of GmNF-Y genes was analyzed using BAR (The Botany Array Resource, [34,35], considering leaves, flowers, green pods, SAM (Shoot Apical Meristem), roots, root tips, nodules and roots inoculated with Bradyrhizobium japonicum in different hours after inoculation (HAI). Additionally, Genevestigator database [26] was used to investigate the expression of GmNF-Y genes in soybean tissues (inflorescence, roots, seedling and shoots), under stress conditions and perturbations (biotic, chemical, temperature and stress).

Results

GmNF-Y gene annotation and protein in silico characterization

Using BLAST approaches, a total of 21 genes encoding GmNF-Y, 23 GmNF-YB, and 24 GmNF-YC subunits were identified (Figure 1, and ESM Table 1 - Electronic Supplementary Material 1). These genes were distributed in several loci, among 2 and 20 chromosomes. A wide difference in the length of GmNF-Y genes was observed, ranging to 500 bp up to 6000 bp. Analysis of structure showed that the number of exons was ranged between 4 to 6 exons in GmNF-YA genes, while in GmNF-YB and 1 to 6 exons in GmNF-YC genes. Larges 5’UTR compared to the total gene length were observed in some sequences, as GmNF-YA9, GmNF-YA10, GmNF-YA11, GmNF-YA14, GmNF-YB15, GmNF-YC8, GmNF-YC9, GmNF-YC11, GmNF-YC12, GmNF-YC13 and GmNF-YC15. In addition, wide 3’UTR were also observed in GmNF-YB14, GmNF-YB19 and GmNF-YC21. The consensuses analyses of the GmNF-Y domains showed a high degree of amino acid conservation, especially for GmNF-YA and GmNF-YC subunits (Figure 1).

NF-Y genes are located in a duplicated region

Figure 2 exemplify the results found considering the analysis of syntenic regions containing the duplicated NF-Y genes (ESM Table 2 shows the complete results). Block duplication processes generated all GmNF-Y genes encoding subunits A and B (ESM Table 1). Regarding subunit C, GmNF-YC13, GmNF-YC15, and GmNF-YC24 were originated by tandem duplication, and GmNF-YC18 by tandem and block duplication.

To analyze the processes that have driven the changes at the
molecular level, the Ka/Ks values between GmNF-Y syntenic genes were evaluated. The results showed that all Ka/Ks values were <1, indicating that a subfunctionalization process could be acting on these paralogous sequences (ESM Table 2).

In silico expression of GmNF-Y

All GmNF-Y genes with expression data available in BAR were analyzed. Figure 3 exemplifies the results and the complete results was showed in ESM (Figures 1-3). Regarding subunit A, we observed that in green pods, GmNF-YA2 was upregulated, while GmNF-YA8 and GmNF-YA12 were downregulated. In nodules, GmNF-YA4, 5, 6, 9, and 11 were highly upregulated, while GmNF-YA1 was downregulated. In roots, GmNF-YA2 was downregulated. In root tips, GmNF-YA3, 7, 8, 9, 10, 12, and 18 were downregulated. GmNF-YA5 was downregulated in flowers. Regarding the NF-B subunits, GmNF-YB4 was up- and GmNF-YB19 downregulated in nodules. In roots, GmNF-YB4 was up- and GmNF-YB16 was downregulated. In root tips, GmNF-YB1, 2, 3, 4, 5, and 17 were downregulated, while GmNF-YB21 was upregulated. In leaves, GmNF-YB16 was up, while GmNF-YB17 was downregulated. The expression of genes corresponding to subunit C demonstrated that GmNF-YC3 was upregulated, while GmNF-YC9 and GmNF-YC10 were downregulated in nodules. In roots, GmNF-YC4 and GmNF-YC11 were upregulated. In root tips, GmNF-YC3 and GmNF-YC7 were downregulated, while GmNF-YC20 was upregulated. GmNF-YC3, 20 and 21 were upregulated in SAM.

Figure 3 also presented examples GmNF-Y genes expression profile in plants growing in soil colonized with Bradyrhizobium japonicum (see all results in ESM Figures 1-3). GmNF-YA2 (12 HAI) and GmNF-YA17 were upregulated (12 and 24 HAI). GmNF-YA6 was downregulated in stripped roots (48 HAI). Regarding subunit B genes, GmNF-YB17 was highly upregulated in root hairs (12 HAI), and downregulated in stripped roots (48 HAI) and root hairs (48 HAI) (Figure 3). Considering subunit C, GmNF-YC4, 7, 9, 19, and 20 were downregulated, while GmNF-YC11 was highly upregulated in stripped roots (48 HAI). GmNF-YC11 (48 HAI), GmNF-YC19 (12 HAI), and GmNF-YC20 (24 HAI) were downregulated in root hairs.

The expression profile of GmNF-Y genes under several perturbations in diverse organs is shown in Figure 4. For elicitor, nutrient, chemical, and temperature, the majority of GmNF-Y genes were typically downregulated in all tissues. Under biotic stress, especially under P. pachyrhizi fungi infection. In roots, seedlings, and shoots, the majority of GmNF-YA genes were modulated. Regarding the subunit B, GmNF-YB4 was highly modulated under biotic stress conditions. Considering the the subunit C, GmNF-YC10 was the gene whose expression was modulated in more conditions.

Discussion

The gene family size reflects the number of duplicated genes. As observed to several gene families, soybean presents a high number of NF-Y genes in comparison with other species, as Arabidopsis (30 AtNF-Y genes) [25] and rice (28 OsNF-Y genes) [24]. In a previous work, Quach and colleagues (2015) have found the same number of GmNF-Y genes that we have presented in this study (68 genes). However, the distribution in NF-Y subunits is slightly different from our analysis, since we have used a more recent version of soybean annotated genome.

Even with variation in length, the structural analysis of GmNF-Y genes by exon-intron composition showed that some genes present a very similar structure, as observed in all GmNF-YA genes, in which the majority presents five exons. The composition of GmNF-YA18 suggests an exon gain, since it is the only one with six exons, while the composition of GmNF-YA6, 9, 10 and 11 suggests the loss of one exon. In relation to GmNF-YB and GmNF-YC, the majority genes possess only one exon, maybe representing the ancestral structure in these subunits. As the GmNF-Y genes are located in duplicated regions, appointed a common origin of paralogous genes, the acquisition of exons may have occurred during the evolution.

Block duplication events happen frequently in plants, because most of them are polyploids and retain duplicated chromosomal blocks. We found the majority of the GmNF-Y genes are located in duplicated blocks (ESM Tables 1 and 2, Figure 2), suggesting that segmental duplication contributed significantly to the family expansion. Similar results were observed in annex in [36] and LSD (Lesion Simulating disease) family in Viridiplan- tae [37]. Interestingly, genes corresponding to subunit C were not duplicated or the duplicated copy was lost, as observed for GmNF-YC14, GmNF-YC17, and GmNF-YC23 (ESM Table 2). In fact, several genes from other families present in the syntenic regions had no correspondent paralogous (Figure 2, white arrows), suggesting that the duplicated genes were lost. Multiple episodes of unequal crossovers might lead to increases/decreases in gene copy. Tandem duplication often results from unequal crossing-over. The occurrence of tandem duplication was observed in some genes encoding subunit C, including GmNF-YC13, GmNF-YC15, and GmNF-YC24. An unequal crossover mechanism should have driven the evolution in this subunit.

Soybean experienced whole genome duplication during their evolution. The high Ka/Ks values observed in the majority of the syntenic regions indicate that the duplicated blocks are a result of ancient polyploidy events [38]. During evolution, duplicated genes can undergo pseudogenization, neofunctionalization or subfunctionalization [39]. A Ka/Ks ratio >1 indicates a neofunctionalization process (through positive selection), and a Ka/Ks ratio <1 suggests a subfunctionalization process (through negative selection) [40]. In this study, the Ka/Ks ratio <1 suggests a subfunctionalization process can be acting in GmNF-Y sequences. In soybean, it has been shown that several paralogous genes are undergoing subfunctionalization [40]. The different expression patterns of GmNF-Y genes in different tissues and stresses, associated with the Ka/Ks ratio, reinforce the occurrence of subfunctionalization.

It is known that members of plant NF-Y family are involved with several processes, as development and biotic stress [25]. NF-YB2 confers drought tolerance and leads to improved corn yields in transgenic Z. mays [16,41,42]. In A. thaliana, the overexpression of NF-YA5 and NF-YB1 reduces water loss in leaves and improves drought tolerance. In addition, ectopic expression of an NF-YC from Amaranthus hypochondriacus and NF-YB from Picea wilsonii confers resistance to water deficiency in Arabidopsis [43,44]. In foxtail millet, NF-YA1 and NF-YB8 were highly activated in leaves and/or roots by drought and salt stresses and were induced under abscisic acid and H2O2 treatments [45]. In rice, the overexpression of OsNF-YA7 improved drought tolerance [46]. In addition, the overexpression of an NF-YA gene (HAP2E/Heme-associated protein 2E) confers resistance to pathogens, salinity, and drought [47]. Therefore, the involvement of NF-Y genes in several networks has been reported, although the precise mechanism of its action is not clear.

It was previously described that GmNF-YA3 of soybean is in
duced by various stress treatments, and their overexpression in Arabidopsis reduced leaf water loss and enhanced drought tolerance [48]. Otherwise, the GmNF-Y family is poorly explored. Our study characterized for the first time that GmNF-Y genes are not ubiquitously expressed in soybean organs; their expression is instead organ-dependent. Our results demonstrate for the first time that GmNF-Y genes were modulated under *P. pachyrhizi* fungi, representing a new characterization for these genes (Figure 4). *P. pachyrhizi* is one of the most important pathogens that affect soybean production, responsible for the development of Asian Soybean Rust (ASR) disease. The pathogen attacks leaves, stems, and pods and may defoliate soybean plants in a few days, leading to drastic crop losses. Identifying genes involved in susceptible or resistant response and characterizing their individual roles are key steps for engineering soybean resistant soybean plants. Thus, additional experiments are necessary to better understand the function of these genes in response to ASR.

An interaction between soybean plant and symbiotic soil bacteria, started by the infection of the plant root hair cells by symbiont, results in the nodulation process. In this study, especially after soil bacteria inoculation, the expression of some genes was induced (Figure 3), indicating these genes could play an important role in soybean nodule association besides plant/pathogen interaction. Recently, a phylogenetically conserved group of NF-Y, which interacts to control nodulation, was described [2]. *Medicago truncatula* genes MtNF-YB16 and MtNF-YC1/MtNF-YC2 interact with MtNF-YA1 and MtNF-YA2 to form NF-Y trimers in yeast and in planta. Moreover, a similar trimer was formed in common bean [2]. A number of independent NF-Y subunits had been reported to participate in nitrogen-fixing rhizobia symbiosis. This includes the MtNF-YA1 and MtNF-YA2 [49-51], NF-YA1 and NF-YB1 from Lotus japonicus [12], and NF-YC1 from common bean [7]. In soybean, nodulation and nodule activity are strongly reduced under stress conditions [52]. Thus, it is important to note that the modulation of NF-Y genes in both conditions (stresses and nodulation) is an interestingly characteristic to be deep explored, clarifying the importance of this family to soybean improvement.

**Conclusion**

Altogether, the results presented here have demonstrated that segmental duplication contributed significantly to the expansion of this gene family, and the subfunctionalization process can drive the evolution of duplicated genes. Several GmNF-Y members are modulated in soybean nodules and under different stresses. The modulation of GmNF-Y expression, especially in response to ASR, suggested their involved in the plant stress response. Future experiments might be important to clarify the relationship between GmNF-Y and the transcription regulation of physiological processes that culminate with the resistance/tolerance of soybean plants to biotic/abiotic stress conditions.

**Figures**

![Figure 1: Gene structure and domain conservation of A) GmNF-YA, B) GmNF-YB and C) GmNF-YC. The blue lines indicate the 5’ and 3’ UTR, the yellow (1A), red (1B) and pink (1C) lines indicate exons and the black lines show intron region. Considering the domain conservation, the total height of each cell indicates the conserved sequence at each position. The height of each letter is proportional to the corresponding relative frequency. The amino acids are colored according to their chemical properties.](image-url)
Figure 2: Expression profile of GmNF-Y genes under different stress conditions. Red color indicates up- and green indicates downregulated genes (Log 2 ratio).

Figure 3: Expression profile of GmNF-YA, GmNF-YB and GmNF-YC genes in leaves, flowers, green pods, SAM (Shoot Apical Meristem), roots, root tips and nodules, and roots inoculated with Bradyrhizobium japonicum. Red color indicates up- and blue indicates downregulated genes (Log 2 ratio). HAI: Hours After Inoculation. Data were generated using BAR database (Libault et al. 2010a; Libault et al. 2010b).
Figure 4: Syntenic regions between GmNF-Y genes analyzed using the PGDD database. The green horizontal line represents the chromosome. The blue and red vertical lines and arrows represent the duplicated orthologous; the red line represents the sequence that was used as the search query. The white arrow indicates genes without correspondent paralogous.
ESM Figures 1-3: Expression profile of GmNF-YA, GmNF-YB and GmNF-YC genes in leaves, flowers, green pods, SAM (Shoot Apical Meristem), roots, root tips, nodules, and roots inoculated with Bradyrhizobium japonicum. In order to generate more understandable figures, the expression data were adapted from BAR images. Red color indicates up- and blue indicates down-regulated genes. A direct comparison of each specific gene in each organs/tissues is applicable. However, comparison among different genes is not adequate. Data were generated using BAR database (Libault et al. 2010 (a); Libault et al. 2010 (b)).

References


