

The Evolutionary History of CBF Transcription Factors: Gene Duplication of CCAAT – Binding Factors NF-Y in Plants

Alexandro Cagliari, Andreia Carina Turchetto-Zolet, Felipe dos Santos Maraschin, Guilherme Loss, Rogério Margis and Marcia Margis-Pinheiro
*Universidade Federal do Rio Grande do Sul/UFRGS
Brazil*

1. Introduction

Eukaryotic gene expression is often controlled by complex and refined combinatorial transcription factor networks composed of multiprotein complexes that derive their gene regulatory capacity from both intrinsic properties and from their *trans*-acting partners (Singh, 1998; Wolberger, 1998; Remenyi *et al.*, 2004). Participation in such higher complex order allows an organism to use single transcription factors to control multiple genes with different temporal and spatial expression patterns (Siefers *et al.*, 2009).

In this chapter, we provide a synopsis of the genetic and genomic mechanisms that might be responsible for the gene copy diversification observed in the eukaryotic NF-Y transcription factor family. We identify the genes coding for NF-Y transcription factors in eukaryotes with an emphasis on the duplication of the NF-Y family in the plant lineage and discuss the important consequences of its gene diversification.

2. The CCAAT *cis*-element promoter

Eukaryotic genes contain numerous *cis*-regulatory elements that mediate their induction, repression or basal transcription (Dyner and Tjian, 1985; Myers *et al.*, 1986; Maity and de Crombrughe, 1998). These regulatory elements can be found in the proximity of transcribed genes, such as the promoter region and/or in distant regions of the genes where they may act as enhancers (de Silvio *et al.*, 1999).

The transcriptional regulation of several eukaryotic genes is coordinated through sequence-specific binding of proteins to the promoter region located upstream of the gene. During evolution, many of these protein-binding sequences, which are found in a wide variety of organisms, have shown a high degree of conservation (Edwards *et al.*, 1998).

The CCAAT box is one of the most common upstream elements, found in approximately 25–30% of eukaryotic promoters (Bucher, 1990; Mantovani, 1998). It is typically located between 60–100 bp upstream of the transcription start site and it can function in direct or in inverted orientations (Dorn *et al.*, 1987b; Bucher, 1990; Edwards *et al.*, 1998; Mantovani, 1998; Stephenson *et al.*, 2007) with possible cooperative interactions between multiple boxes (Tasanen *et al.*, 1992) or other conserved motifs (Muro *et al.*, 1992; Rieping and Schoffl, 1992;

Edwards *et al.*, 1998). CCAAT boxes are highly conserved within homologous genes across species in terms of position, orientation, and flanking nucleotides (Mantovani, 1998). In addition, the spacing between the CCAAT box and other promoter-specific *cis*-elements is also conserved among species (Dorn *et al.*, 1987a; Chodosh *et al.*, 1988; Maity and de Crombrughe, 1998). The expression of genes under the control of promoters that contain CCAAT boxes may be ubiquitous or tissue/stage specific, suggesting that the gene expression pattern is also determined by other *cis* and *trans* elements (Stephenson *et al.*, 2007).

In *Sacharomyces cerevisiae*, CCAAT boxes are found in the promoters of cytochrome genes, in genes coding for proteins that are activated by non-fermentable carbon sources (McNabb *et al.*, 1995) and in genes involved in nitrogen metabolism (Dang *et al.*, 1996). In the filamentous fungus *Aspergillus nidulans*, CCAAT boxes are present in genes involved with penicillin biosynthesis (Steidl *et al.*, 1999). In higher eukaryotes, a multitude of promoters contain CCAAT boxes, including those of developmentally controlled and tissue-specific genes (Berry *et al.*, 1992), housekeeping and inducible genes (Roy and Lee, 1995) and cell-cycle regulated genes (Mantovani, 1998). In addition, many cell-cycle regulated promoters lack a recognizable TATA-box, but contain more than one CCAAT box in a position close to and sometimes overlapping with the start site of transcription (Zwicker and Muller, 1997).

3. The CBF/NF-Y transcription factor

Several CCAAT-binding proteins have been isolated and described, including CBF/NF-Y (CCAAT Binding Factor/Nuclear Factor of the Y box), CTF/NF1 (CCAAT Transcription Factor/Nuclear Factor 1), C/EBP (CCAAT/Enhancer Binding Protein) and CDP (CCAAT Displacement Protein) (Mantovani, 1999). Among them, NF-Y is the most ubiquitous and specific one acting as a key proximal promoter factor in the transcriptional regulation of an array of different eukaryotic genes. Unlike other CCAAT-binding proteins, NF-Y requires a high degree of conservation of the CCAAT pentanucleotide sequence and shows strong preference for specific flanking sequences (Dorn *et al.*, 1987a; Stephenson *et al.*, 2007). Therefore, the NF-YC transcription factor can be distinguished from the other CCAAT-binding proteins based on its DNA sequence requirements (Maity and de Crombrughe, 1998).

The CBF/NF-Y transcription factor, which will be referenced in this chapter as NF-Y, is a conserved oligomeric transcription factor found in all eukaryotes that is involved in the regulation of diverse genes (Maity *et al.*, 1992; McNabb *et al.*, 1995; Edwards *et al.*, 1998; Mantovani, 1998; Siefers *et al.*, 2009). NF-Y typically acts in concert with other regulatory factors to modulate gene expression in a highly controlled manner (Nelson *et al.*, 2007). In many eukaryotic promoters, the functional NF-Y-binding sites are relatively close to the TATA motif (Bucher, 1990) and are invariably flanked by at least one additional functionally important *cis*-element. Several reports have shown that various factors, including transcription factors, co-activators, and TATA-binding proteins, interact with NF-Y or its subunits in promoting transcriptional regulation (Mantovani, 1999; Yazawa and Kamada, 2007). NF-Y was originally identified as the protein that recognizes the MHC class II conserved Y box in *Ea* promoters (Dorn *et al.*, 1987a; Matuoka and Chen, 2002). It specifically recognizes the consensus sequence 5'-CTGATTGGYYRR-3' or 5'-YYRRCCAATCAG-3' (Y is 5 pyrimidines and R is 5 purines) present in the promoter region of eukaryotic genes. Bioinformatic analyses indicate that about 30% of mammalian promoters have predicted

NF-Y binding sites (Bucher, 1990; Testa *et al.*, 2005), and chromatin immunoprecipitation data have demonstrated additional widespread NF-Y binding in nonpromoter sites.

Suggesting the importance of binding context, NF-Y-regulated gene expression can be tissue specific, developmentally regulated, or constitutive (Maity and de Crombrughe, 1998; Siefers *et al.*, 2009). The transcriptional activity of NF-Y can be regulated by differential expression, alternative splicing, protein-protein interactions, and cellular redox potential (Matuoka and Yu Chen, 1999).

NF-Y has been shown to be involved in the regulation of some G1/S genes whose expressions are attenuated during the senescence process (Matuoka and Yu Chen, 1999). NF-Y plays a pivotal role in the cell cycle regulation of the mammalian cyclin A, *cdc25C*, and *cdc2* genes, in the S-phase of the cell cycle (Currie, 1998). Additionally, there are a number of genes involved in the cellular response to damage and stress, including the phospholipid hydroperoxide glutathione peroxidase genes (Huang *et al.*, 1999), which are regulated by NF-Y, indicating its pivotal role in the removal of damaging agents from cells (Matuoka and Chen, 2002). Although NF-Y functions basically as a transactivator of gene expression, it is also involved, directly or indirectly, in the downregulation of transcription. For instance, NF-Y binds to the mouse CCAAT box renin enhancer and blocks the binding of positive regulatory elements (Shi *et al.*, 2001). In this case, NF-Y dysfunction would lead to the damage of systems that control blood pressure (Matuoka and Chen, 2002).

NF-Y is composed of three different subunits named NF-YA (also known as HAP-2 or CBF-B), NF-YB (HAP3 or CBF-A), and NF-YC (HAP5 or CBF-C) that interact to form a complex that can bind CCAAT DNA motifs and control the expression of target genes (Figure 1). Each subunit is required for DNA binding, subunit association and transcriptional regulation in both vertebrates and plants (Sinha *et al.*, 1995; Stephenson *et al.*, 2007). Yeast possesses a fourth subunit, called HAP4, which provides a transcriptional activation domain to the complex (Forsburg and Guarente, 1989; Lee *et al.*, 2003). The yeast HAP4 protein is not needed for DNA-binding but contains an acidic domain that is essential to promote transactivation when associated with the HAP2/HAP3/HAP5 complex (Olesen and Guarente, 1990; Serra *et al.*, 1998). In vertebrates, the function of this fourth domain was incorporated into other subunits (Forsburg and Guarente, 1989; Yazawa and Kamada, 2007). Despite the wide cellular distribution and functional variability of NF-Y-regulated genes, most eukaryotic genomes have only one or two genes encoding each NF-Y subunit (Maity and de Crombrughe, 1998; Riechmann and Ratcliffe, 2000). Fungi and animals, for example, present single genes encoding each protein subunit. Thus, there is minimal combinatorial diversity in the subunit composition of the heterotrimeric NF-Y in these organisms (Siefers *et al.*, 2009). In contrast, the NF-Y complex in vascular plants is generally encoded by gene families (Riechmann and Ratcliffe, 2000).

3.1 NF-Y subunits

NF-Y is the only transcription factor thus far identified for which the interaction of three heterologous subunits creates the DNA binding domain (Maity and de Crombrughe, 1992; McNabb *et al.*, 1995; Sinha *et al.*, 1995). All three NF-Y subunits are essential for the DNA binding activity and one molecule of each subunit forms the NF-Y-DNA complex (Maity and De Crombrughe, 1996). Each NF-Y subunit contains a conserved domain with identities greater than 70% across species. This highly conserved domain is located at the C-terminus of NF-YA; in the central part of NF-YB; and at the N-terminus of NF-YC (Li *et al.*, 1992).

The NF-YA conserved domain can be divided in two functionally distinct regions: an N-terminal region that is required for NF-YB and NF-YC association and a C-terminal region required for DNA-binding (Maity and de Crombrughe, 1992). Additionally, NF-YA usually contains a glutamine (Q)-rich and a serine/threonine (S/T)-rich regions. There are numerous variants of NF-YA due to alternative splicing at the Q-S/T domains (Li *et al.*, 1992) and, although the expression of these isoforms is variable depending of tissue and cell types, they all seem intact in terms of transcriptional function (Matuoka and Chen, 2002).

Both NF-YB and NF-YC subunits possess the highly conserved histone-fold motif (HFM) and are structurally similar to core histone subunits H2B and H2A, respectively, and to the archaeobacterial histone-like protein Hmf-2 (Arents and Moudrianakis, 1995; Baxevanis *et al.*, 1995; Mantovani, 1998). In terms of identity, NF-YB is 30% identical to H2B, 14% to H2A, 17% to H4 and 18% to H3; NF-YC is 21% identical to H2A, 15% to H4 and H3 and 20% to H2B (Liberati *et al.*, 1999). Other proteins showing a remarkable identity (25-30%) to both NF-YB and NF-YC are present in *Archaea*. These proteins homodimerize and associate with DNA, forming nucleosome-like structures (Sandman *et al.*, 1990). The NF-YB and NF-YC subunits also contain residues that are important for their contact with DNA (Romier *et al.*, 2003; Stephenson *et al.*, 2007). In contrast, the conserved segment of NF-YA has no homology with the histone-fold motif, or with any of the known dimerization motifs present in other heteromeric DNA-binding proteins (Maity and de Crombrughe, 1992).

Some portions of NF-YA, NF-YB and NF-YC present a high degree of identity with yeast HAP3, HAP2 and HAP5, respectively. These HAP genes, which are components of the yeast CCAAT-binding protein, are necessary for the expression of genes encoding components of the electron transport chain. Yeast strains mutated for either of the three genes failed to grow on media containing a nonfermentable carbon source such as lactate or glycerol, a characteristic respiratory-defect phenotype (McNabb *et al.*, 1995).

Assembly of the NF-Y heterotrimer in mammals (where this complex is better studied) follows a strict, stepwise pattern (Sinha *et al.*, 1995; Sinha *et al.*, 1996) (Figure 1). Initially, the NF-YB and NF-YC subunits form a tight heterodimer (Figure 1a) similar to those of the HFM, a conserved protein-protein and DNA-binding interaction module (Luger *et al.*, 1997) composed by 65 amino acid stretch common to all histones that is required for nucleosome formation (Baxevanis *et al.*, 1995; Luger *et al.*, 1997; de Silvio *et al.*, 1999). This dimer then moves to the nucleus, where the third subunit (NF-YA, Figure 1b) is recruited to generate the complete, heterotrimeric NF-Y (Figure 1c). Interestingly, NF-YA is unable to interact with the NF-YB or NF-YC alone, interacting only with the NF-YB-NF-YC heterodimer (Serra *et al.*, 1998). The complete NF-Y is able to bind promoters containing the core pentamer nucleotide sequence CCAAT (Figure 1d) with high specificity and affinity resulting in either positive or negative transcriptional regulation (Figure 1e) (Peng and Jahroudi, 2002; 2003; Ceribelli *et al.*, 2008; Siefers *et al.*, 2009).

Because the NF-Y transcription factor contains H2B-like and H2A-like molecules (NF-YB and NF-YC, respectively), the complex presents all the core histone components and could mimic the interaction of the nucleosome core with genomic DNA (Struhl and Moqtaderi, 1998). In this scenario, it has been demonstrated that the NF-YA/NF-YB/NF-YC trimer or the NF-YB/NF-YC dimer can bind to H3/H4 tetramer during nucleosome assembly (Caretti *et al.*, 1999). In addition, the NF-Y complex also can bind to the chromatin even after nucleosome formation, indicating the ability of NF-Y to interact with genomic DNA assembled in the nucleosome. The interaction between the NF-Y transcription factor and the DNA molecule causes local disruption of the nucleosomal architecture (Coustry *et al.*, 2001).

This disruption results in a partial dissociation of DNA from the histone core, which might enable the access of the general transcription machinery to initiate the transcription process.

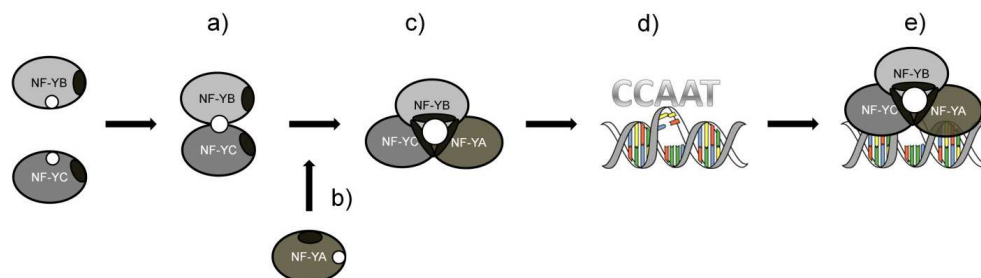


Fig. 1. Assembly of NF-Y subunits and its binding to DNA. Initially, the NF-YB and NF-YC subunits form a tight heterodimer via protein-protein interactions (a). The dimer then moves to the nucleus, where is recruited the third subunit (NF-YA) (b) to generate the complete, heterotrimeric NF-Y (c) that is able to bind promoters containing the core pentamer nucleotide sequence CCAAT (d) resulting in either positive or negative transcriptional regulation (e). Adapted from Mantovani (1999). White circles and oblong black circles into each NF-Y subunit represent the DNA-binding domain and the NF-Y interaction domain of NF-Y subunits, respectively.

4. Gene duplication and evolution

DNA duplication act as one of the main forces driving the evolution of organisms by creating the raw genetic material that natural selection can subsequently modify. Gene duplications arise in eukaryotes at a rate of 0.01 paralogs per gene per million years (Lynch and Conery, 2000), the same order of magnitude of the mutation rate per nucleotide per year (De Grassi *et al.*, 2008). Duplication of individual genes, chromosomal segments, or entire genomes represent the primary source for the origin of evolutionary novelties, including new gene functions and expression patterns (Holland *et al.*, 1994; Sidow, 1996; Lynch and Conery, 2000). However, how duplicated genes successfully evolve from an initial state of complete redundancy, wherein one copy is likely to be expendable, to a stable situation in which both copies are maintained by natural selection, is unclear (Sidow, 1996; Lynch and Conery, 2000; Ober, 2010).

In the evolutionary history of plants, genome duplications have been relatively common, leading to the hypothesis that most angiosperms are to some extent polyploid (Soltis, 2005). The genome of Arabidopsis, for example, possesses traces of at least three polyploidy events (Vision *et al.*, 2000; Simillion *et al.*, 2002), followed by subsequent gene loss (Bowers *et al.*, 2003; Ober, 2010).

Similar to a point mutation, a duplication that occurs in an individual can be fixed or lost in the population. Compared with pre-existing alleles, if a new allele of the duplicate gene is selectively neutral, it has a small probability ($1/2N$) to be fixed in a diploid population (where N is the effective population size). This suggests that the majority of duplicated genes will be lost. For those duplicated genes that do become fixed, the fixation time averages is $4N$ generations (Kimura, 1989; Zhang, 2003).

On an evolutionary scale, gene duplication may result in new functions via different scenarios. Although the most likely outcome is a loss of function in one of the two gene copies (nonfunctionalization, Figure 2a), in rare instances one copy may acquire a novel evolutionarily advantageous function and become preserved by natural selection (neofunctionalization, Figure 2b), while the other copy retains the original function. Alternatively, after duplication, mutations may occur in both genes leading to specialization to perform complementary functions (subfunctionalization, Figure 2c) (Lynch and Conery, 2000; Lynch and Force, 2000). This process produces novel genetic variants that drive genetic innovation (Lynch and Conery, 2000; Conrad and Antonarakis, 2007). Because gene duplication generates functional redundancy, it is often not advantageous to the organism to possess two identical genes. In nonfunctionalization (Figure 2a), the accumulation of deleterious mutations might lead to the loss of the original function of one paralogue. Alternatively, instead of being completely lost, many duplicated genes are silenced or become pseudogenes and are thus either unexpressed or functionless (Gallagher *et al.*, 2004; Nicole *et al.*, 2006; Yang *et al.*, 2006; Beisswanger and Stephan, 2008; Xiong *et al.*, 2009). Pseudogenization is the most frequent fate of duplicated genes. In *Caenorhabditis elegans*, for example, genomic analyses have identified 2168 pseudogenes or approximately one pseudogene for every eight functional genes (Harrison *et al.*, 2001). In humans, one pseudogene was identified for approximately every two functional genes (Harrison *et al.*, 2002). As pseudogenes generally do not confer a selective advantage, they have a low probability of being fixed in large populations (Ober, 2010).

Unless the presence of an extra amount of gene product is advantageous, it is unlikely that two genes with the same function will be stably maintained in the genome of the organism (Nowak *et al.*, 1997). In subfunctionalization (Figure 2c), both duplicated copies may become, by accumulation of mutations, partially compromised to the point at which their total capacity is reduced to the level of the single-copy ancestral gene (Force *et al.*, 1999; Stoltzfus, 1999; Lynch and Force, 2000). Subfunctionalization can occur through the modification of the regulatory elements by mutations (Force *et al.*, 1999; Hinman and Davidson, 2007) or by epigenetic silencing (Rodin and Riggs, 2003). In an evolutionary scale, one of the most important forms of subfunctionalization is the division of gene expression after duplication (Force *et al.*, 1999). For example, zebrafish ENGRAILED 1 and ENGRAILED 1-B, generated by a chromosomal segmental duplication, are a pair of transcription factors that occurred in the lineage of ray-finned fish. While ENGRAILED-1 is expressed in the pectoral appendage bud, ENGRAILED 1-B is expressed in a specific set of neurons in the hindbrain/spinal cord (Force *et al.*, 1999). In yeast, more than 40% of gene pairs exhibit significant expression divergence (Gu *et al.*, 2002). Also, the comparison of 17 fungal genomes revealed that duplicated genes rarely diverge with respect to biochemical function, but typically diverge with respect to regulatory control (Wapinski *et al.*, 2007). On the other hand, if two redundant gene copies were retained without significant functional divergence in the genome, the organism may acquire increased genetic robustness against harmful mutations (Figure 1d) (Conrad and Antonarakis, 2007).

In neofunctionalization (Figure 2b), the ancestral gene keeps its ancestral function, while the duplicated gene gains a new function under positive selection for advantageous mutations (De Grassi *et al.*, 2008). However, in many cases, rather than an entirely new function, a related function evolves after gene duplication. For example, the red and green-sensitive opsin genes of humans were the result of a gene duplication that occurred in hominoids and Old World monkeys (Yokoyama and Yokoyama, 1989). After the duplication process,

functional divergence of the two opsins resulted in a 30-nanometer difference in their maximum absorption wavelength. This difference conferred a sensitivity to a wide range of colors for humans and related primates (Zhang, 2003).

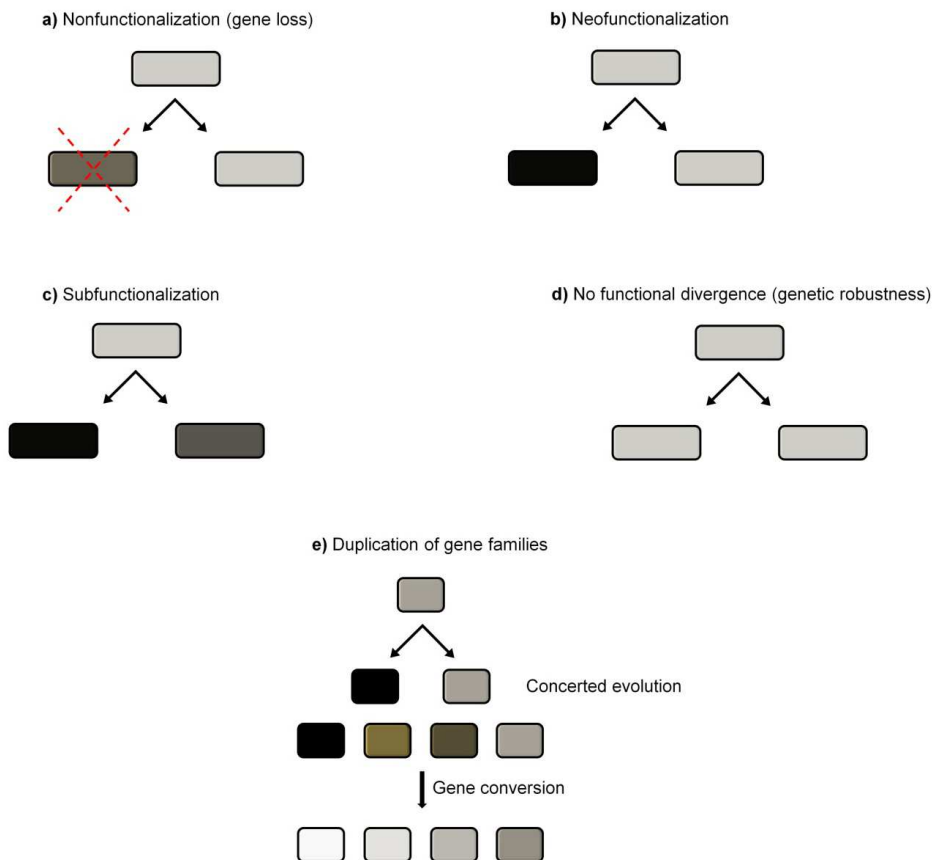


Fig. 2. Evolutionary fate of duplicated genes. Gene duplication may result in new functions via different scenarios. **(a)** nonfunctionalization; **(b)** neofunctionalization; **(c)** subfunctionalization; **(d)** genetic robustness and **(e)** gene conversion. Adapted from Conrad and Antonarakis (2007).

The fate of a gene that suffers duplication seems to be the result of diverse and, in some cases, interdependent factors (Taylor *et al.*, 2001). These variables include its functional category (Papp *et al.*, 2003; Kondrashov and Koonin, 2004; Marland *et al.*, 2004), degree of conservation (Conant and Wagner, 2002; Davis and Petrov, 2004; Jordan *et al.*, 2004; Braybrook and Harada, 2008), sensitivity to dosage effects (Kondrashov and Koonin, 2004), as well as its regulatory and architectural complexity (He and Zhang, 2005). Some observations indicate that natural selection created a preferential association of duplications with certain gene categories. For example, genes encoding proteins that interact with the environment are more frequently retained after the duplication process than genes which interact at intracellular compartments

(Li *et al.*, 2003; Marland *et al.*, 2004). In addition, genomes tend to retain duplicated genes involved in signal transduction and transcription, but to lose duplicated DNA repair genes (Blanc and Wolfe, 2004; Maere *et al.*, 2005; Paterson *et al.*, 2010).

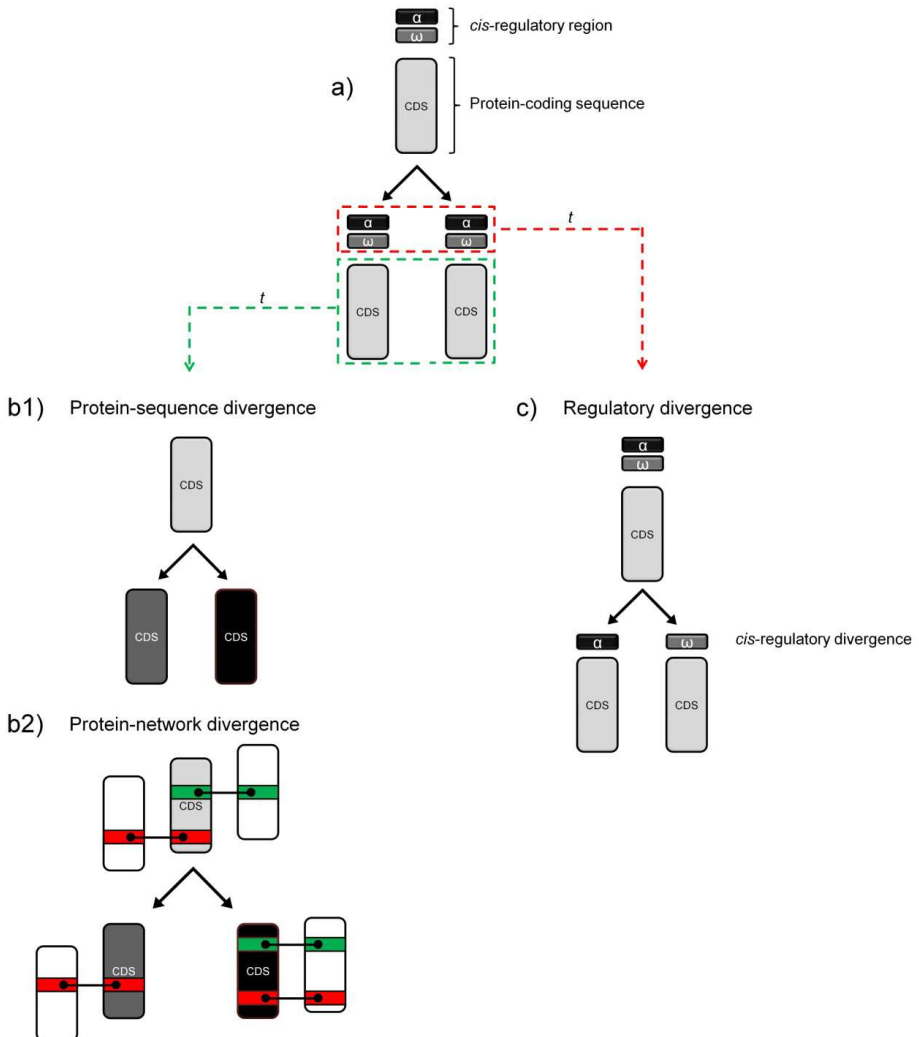


Fig. 3. Functional divergence of duplicated genes. **(a)** The independent evolution of *cis*-regulatory and the protein-coding regions. **(b1)** protein sequence divergence after duplication; **(b2)** protein network divergence with increase or loss of partners and **(c)** DNA sequence regulatory divergence after duplication (certain regulatory motifs are lost in one copy of the duplicated gene sequence). *t*; evolutionary time. Adapted from Conrad and Antonarakis (2007).

It has been shown that shortly after duplication the protein-coding sequence and *cis*-regulatory regions of some duplicated genes can evolve independently (Figure 3a) (Wagner, 2000). This independent evolution can generate protein sequence divergence of duplicated genes (Figure 3b1) or protein network divergence (Figure 3b2), where the protein interaction domains (*cis*-regulatory elements) of the original sequence evolve by maintenance, gain, or loss of interacting partners. Alternatively, the divergence of *cis*-regulatory motifs in the promoter-proximal region (Figure 3c) can generate expression divergence between the duplicated genes (Conrad and Antonarakis, 2007).

5. Gene duplication of NF-Y in plants

While duplication of NF-Y genes is poorly understood in the plant lineage, many of the functional mechanistic details are likely conserved across plant, animal and fungal lineages. This inference comes from strong cross-kingdom conservation of functional important amino acid residues in mammalian and yeast NF-Ys (Maity and de Crombrughe, 1992; Maity *et al.*, 1992; Sinha *et al.*, 1995; Coustry *et al.*, 1996; Kim *et al.*, 1996; Sinha *et al.*, 1996; Mantovani, 1998; Romier *et al.*, 2003). CCAAT-like motifs are found in several plant promoters, and binding activity to CCAAT sequences has been identified in plant nuclear extracts (Yazawa and Kamada, 2007). Besides, at least some plant NF-YA and NF-YB subunits have been shown to complement yeast mutant strains lacking the corresponding NF-Y subunit. Additionally, several groups have demonstrated that each of the three plant NF-Y proteins can substitute their yeast counterparts in gene expression assays (Edwards *et al.*, 1998; Masiero *et al.*, 2002; Ben-Naim *et al.*, 2006; Siefers *et al.*, 2009). These observations indicate that plant NF-Y subunits might act as general transcription factors, as in mammals (Yamamoto *et al.*, 2009).

Although a complete functional plant NF-Y complex has not yet been described, the individual subunits are known to be involved in a number of important physiological processes, such as specific developmental processes and response to environmental stimuli (Lotan *et al.*, 1998; Kusnetsov *et al.*, 1999; Miyoshi *et al.*, 2003; Ben-Naim *et al.*, 2006; Combiere *et al.*, 2006; Wenkel *et al.*, 2006; Cai *et al.*, 2007; Nelson *et al.*, 2007; Warpeha *et al.*, 2007; Siefers *et al.*, 2009). A well-established example is the NF-YB subunit gene called LEAFY COTYLEDON-1 (LEC1), which specifically controls embryo development, especially the maturation phase. LEC1 plays specialized roles not only because of its developmentally regulated expression but also due to its distinct molecular activity, as the *in vivo* function of LEC1 cannot be replaced by other NF-YB subunits, except for the most closely related Leafy Cotyledon 1 Like (L1L) (Kwong *et al.*, 2003; Lee *et al.*, 2003; Yamamoto *et al.*, 2009). In Arabidopsis, many NF-Y subunit genes are expressed ubiquitously, although some are differentially expressed. For example, while the AtNF-YC-4 transcript accumulates in seeds 7 days after germination, AtNF-YB-9 is only expressed in green siliques (Gusmaroli *et al.*, 2001).

Plant NF-Y function also appears to be important for responses to drought stress. Although a specific mechanism of action remains unclear, overexpression of the AtNF-YB1 subunit and its orthologue in maize (*Zea mays*), ZmNF-YB2, leads to enhanced drought resistance (Nelson *et al.*, 2007). Another study showed that overexpression of maize NF-YA5 reduced drought susceptibility, anthocyanin production and stomatal aperture, while *nf-ya5* mutants had the expected opposite phenotype in each situation (Li *et al.*, 2008). In addition, several

publications strongly suggest that NF-Y transcription factors are also involved in photoperiod-regulated flowering (Ben-Naim *et al.*, 2006; Wenkel *et al.*, 2006; Siefers *et al.*, 2009).

We adopted a high throughput comparative genomic approach to conduct a broad survey of fully sequenced genomes, including representatives of amoebozoa, yeasts, fungi, algae, mosses, plants, vertebrate and invertebrate species to identify the presence of homologous genes coding for each of the three subunits that form the NF-Y transcription factor (Table 1). NF-Y gene and protein sequences were obtained through blast searches (blastp, blastx and tblastx) against the Protein and Genome databases with the default parameters at the NCBI (National Center for Biotechnology Information - <http://www.ncbi.nlm.nih.gov>) and against completed genome projects database at the JGI (Joint Genome Institute - <http://www.jgi.doe.gov>).

The results point to a scenario where all fungi and the majority of metazoa possess single genes coding for each of the NF-Y subunits (Table 1). The metazoa exceptions include the amphioxus *Branchiostoma floridae*, the nematode *Caenorhabditis elegans* and the gastropod *Lottia gigantea*, all of each present a proportional duplication of the three subunits, possessing two genes for each subunit (Table 1).

In contrast, plants possess gene families coding for each NF-Y subunit (Table 1). For instance, in the model plant *Arabidopsis thaliana* 10 genes coding for NF-YA, 13 for NF-YB, and 13 for NF-YC were identified. Because of the heterotrimeric composition, the 36 *Arabidopsis* NF-Y subunits could theoretically combine to generate 1.690 unique transcription factors (Siefers *et al.*, 2009). This *Arabidopsis* NF-Y expansion is a general feature of the plant lineage, including monocots and eudicots. In rice (*Oryza sativa*), for example, 11 genes were identified coding for the NF-YA subunit, 10 for NF-YB and 8 for NF-YC. Four of the rice NF-YB subunits have been characterized and at least one of these genes is involved in chloroplast development (Miyoshi *et al.*, 2003; Yazawa and Kamada, 2007). Interestingly, the moss *Physcomitrella patens* and the lycophyte *Selaginella mollendorffii* possess single genes coding for NF-YA subunits whereas the other subunits are encoded by multiple genes (Table 1).

Since the evolutionary rates can be species dependent, the difference observed in the number of genes of NF-Y subunits in eukaryotic class (Table 1), especially in vascular plants, can be result of recent duplication process that contribute to the establishment of genes families coding each NF-Y subunit. However, some duplicated genes might have suffered high level of diversification what could be responsible to prevent their identification in our analyses.

Representative plants genes (monocot and eudicot) were selected to perform phylogenetic analyses of the NF-Y subunits. The phylogenetic analysis was reconstructed after protein sequence alignments using a Bayesian approach in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The mixed amino acid substitution model plus gamma and invariant sites was used in two independent runs of 5,000,000 generations each with two Metropolis-coupled Monte Carlo Markov chains (MCMCMC) that were run in parallel (starting each from a random tree). Markov chains were sampled every 100 generations, and the first 25% of the trees were discarded as burn-in. The remaining ones were used to compute the majority rule consensus tree, the posterior probability of clades and branch lengths (Figure 4 to 6).

Phylo	Specie	Code	Subunit A Genes	Subunit B Genes	Subunit C Genes
Metazoa	<i>Homo sapiens</i>	Hsa	1	1	1
	<i>Mus musculus</i>	Mmu	1	1	1
	<i>Rattus norvegicus</i>	Rno	1	1	1
	<i>Canis familiaris</i>	Cfa	1	1	1
	<i>Monodelphis domestica</i>	Mdo	1	1	1
	<i>Gallus gallus</i>	Gga	1	1	1
	<i>Xenopus tropicalis</i>	Xtr	1	1	1
	<i>Gasterosteus aculeatus</i>	Gac	1	1	1
	<i>Oryzias latipes</i>	Ola	1	1	1
	<i>Takifugu rubripes</i>	Tru	1	1	1
	<i>Danio rerio</i>	Dre	1	1	1
	<i>Ciona savignyi</i>	Csa	1	1	1
	<i>Branchiostoma floridae</i>	Bfl	2	2	2
	<i>Strongylocentrotus purpuratus</i>	Spu	1	1	1
	<i>Drosophila melanogaster</i>	Dme	1	1	1
	<i>Anopheles gambiae</i>	Aga	1	1	1
	<i>Tribolium castaneum</i>	Tca	1	1	1
	<i>Caenorhabditis elegans</i>	Cel	2	2	2
	<i>Lottia gigantea</i>	Lgi	2	2	2
	<i>Nematostella vectensis</i>	Nve	1	1	1
Fungi	<i>Neurospora crassa</i>	Ncr	1	1	1
	<i>Candida tropicalis</i>	Ctr	1	1	1
	<i>Tuber melanosporum</i>	Tme	1	1	1
	<i>Pyrenophora teres</i>	Pte	1	1	1
	<i>Aspergillus nidulans</i>	Ani	1	1	1
	<i>Chaetomium globosum</i>	Cgl	1	1	1
	<i>Penicillium marneffeii</i>	Pma	1	1	1
	<i>Talaromyces stipitatus</i>	Tst	1	1	1
	<i>Sordaria macrospora</i>	Sma	1	1	1
	<i>Naegleria gruberi</i>	Ngr	1	1	1
Heterolobosea	<i>Manihot esculenta</i>	Mes	12	15	9
	<i>Ricinus communis</i>	Rco	6	12	7
	<i>Populus trichocarpa</i>	Ptr	8	17	9
	<i>Medicago truncatula</i>	Mtr	5	10	5
	<i>Glycine max</i>	Gma	21	25	11
	<i>Cucumis sativus</i>	Csa	6	11	3
	<i>Prunus persica</i>	Ppe	6	13	6
	<i>Arabidopsis thaliana</i>	Ath	10	13	13
	<i>Carica papaya</i>	Cpa	5	9	3
	<i>Vitis vinifera</i>	Vvi	7	12	5
	<i>Sorghum bicolor</i>	Sbi	9	10	7
	<i>Zea mays</i>	Zma	10	20	14
	<i>Oryza sativa</i>	Osa	11	10	8
	<i>Brachipodyum</i>	Bdi	7	13	10

<i>distachyon</i>					
	<i>Selaginella mollendorffii</i>	Smo	1	5	3
	<i>Physcomitrella patens</i>	Ppa	1	6	6
Heterokonta	<i>Phaeodactylum tricornutum</i>	Ptri	1	1	1

Table 1. NF-Y genes identified in the fully eukaryotic sequenced genomes.

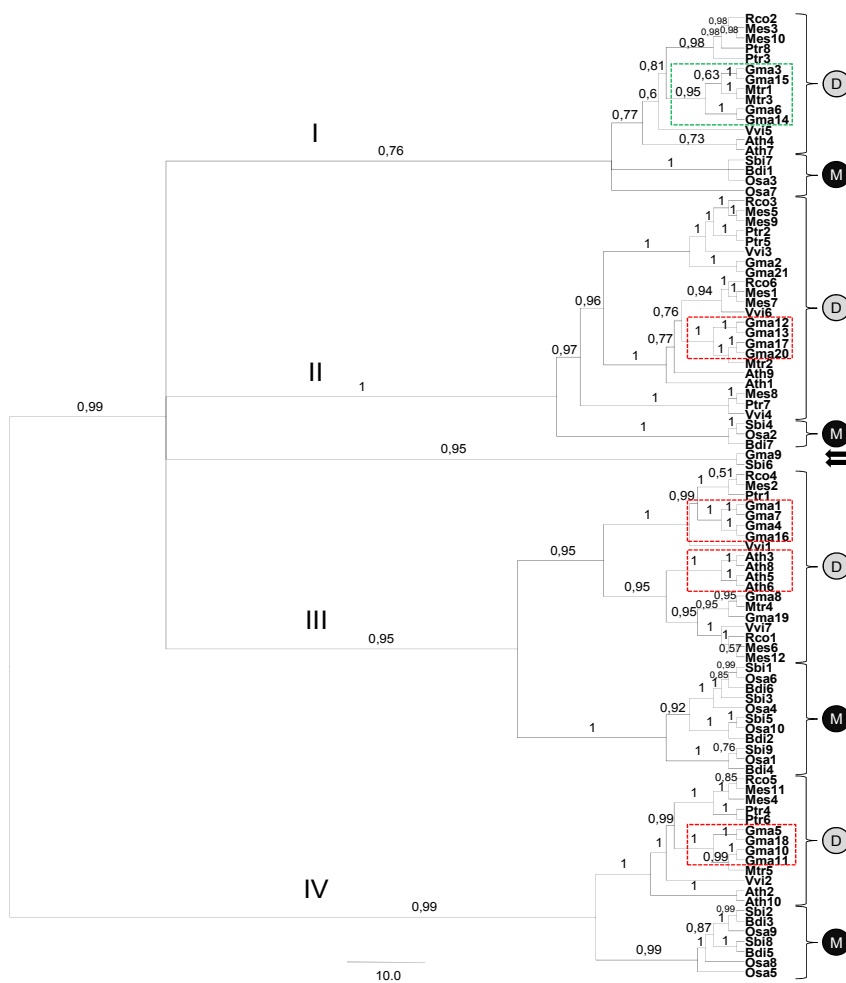


Fig. 4. Phylogenetic tree of monocot and eudicot representatives of NF-YA subunit. M: monocots; D: eudicots; Rco: *Ricinus communis*; Mes: *Manihot esculenta*; Ptr: *Populus trichocarpa*; Gma: *Glycine max*; Mtr: *Medicago truncatula*; Vvi: *Vitis vinifera*; Ath: *Arabidopsis thaliana*; Sbi: *Sorghum bicolor*; Bdi: *Brachipodyum distachyon*; Osa: *Oryza sativa*; red square: event of duplication inside the species; green square: event of duplication inside the same plant family; black arrows: genes that possess an unresolved position in the phylogenetic tree; I to IV: independent phylogenetic gene clusters.

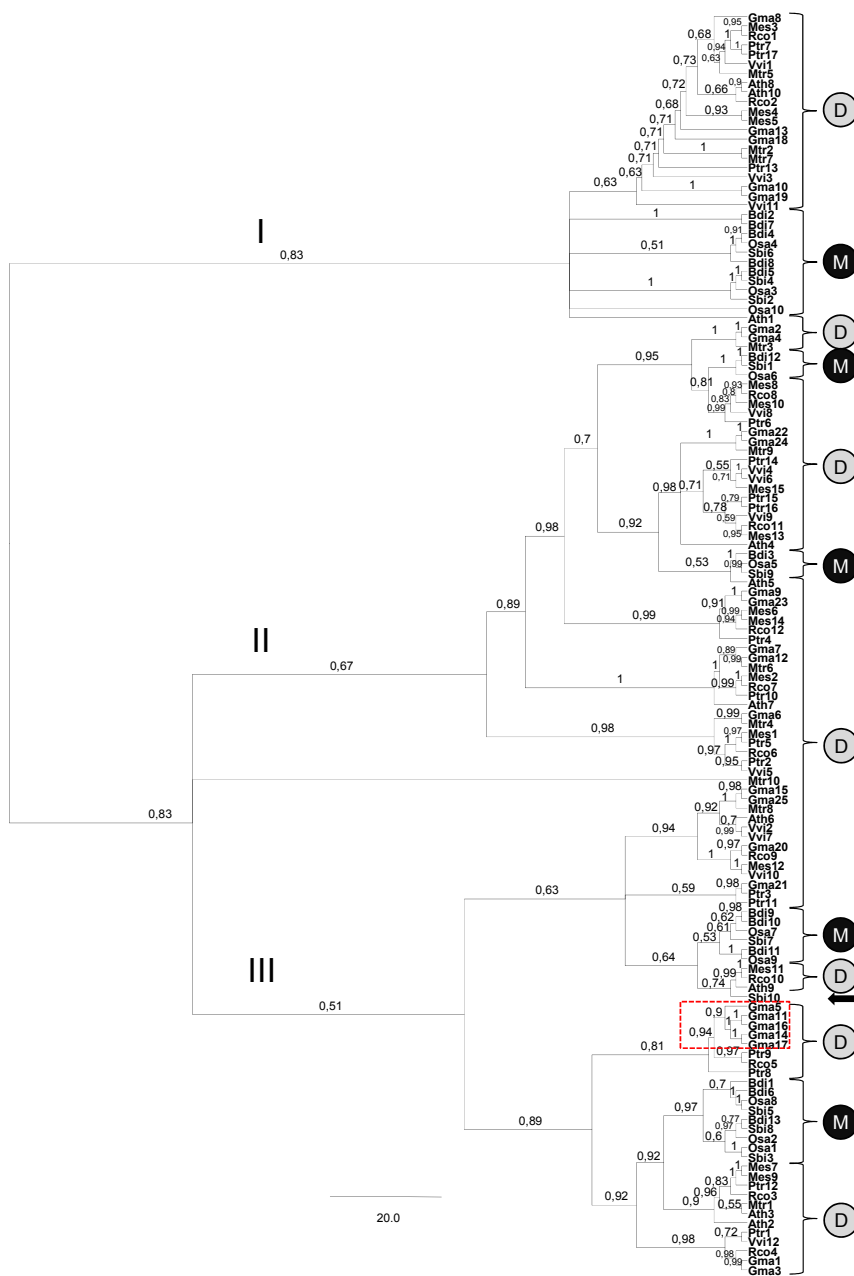


Fig. 5. Phylogenetic tree of monocot and eudicot representatives of NF-YB subunit.
For details see legend of Figure 4.

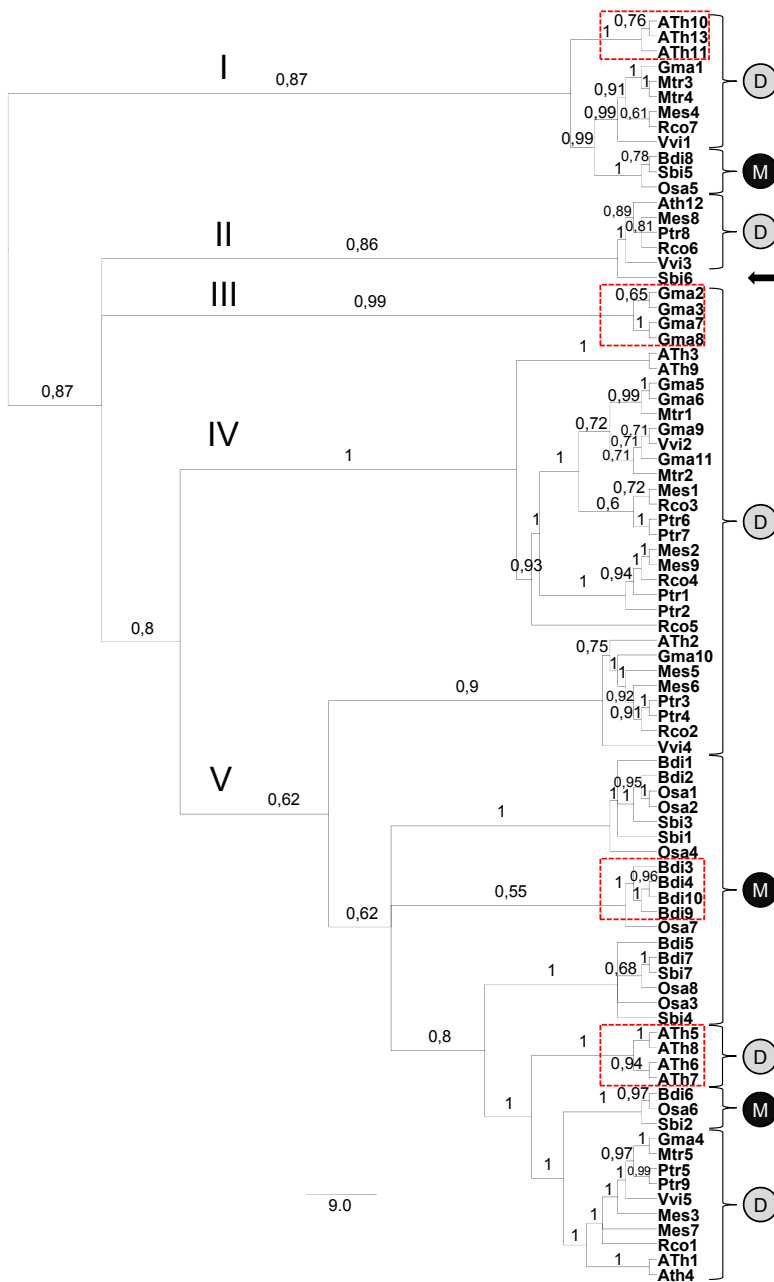


Fig. 6. Phylogenetic tree of monocot and eudicot representatives of NF-YC subunit.

For details see legend of Figure 4.

Phylogenetic analysis showed that the gene diversification of all NF-Y subunits likely resulted from several duplication events along evolution and diversification of plants (Figure 4 to 6). It was possible to observe the formation of four independent highly supported clusters for the NF-YA subunit (I to IV, Figure 4), three for NF-YB (I to III, Figure 5) and five for NF-YC (I to V, Figure 6). Based on these results, we suggest that each cluster might possess an independent ancestral subunit that the duplicated members of each group originated from. However, independent duplication events have occurred in many species after the divergence of monocots and eudicots. For example, the soybean and *Arabidopsis* genomes have experienced a series of recent duplication events (red squares in figures 4 to 6) that could be the result of chromosome duplication or could be derived from polyploidization events (soybean is a good example of a recent polyploidization). These duplications can help us to explain the differences observed in the number of genes coding for the NF-Y subunits in plants (Table 1). Additionally, these duplications seem to be relatively recent and can provide the raw material for neofunctionalization (Figure 2b) and functional divergence of duplicated genes (Figure 3). With few exceptions (genes that possess an unresolved position in the phylogenetic tree are plotted with black arrows, Figures 4 to 6), all clusters of a specific NF-Y subunit are formed by well-defined sub-clusters of monocot and eudicot representatives (Figure 4 to 6). Events of duplication inside a specific plant family were also observed between the two fabaceae species *Glycine max* and *Medicago truncatula* (green square, Figure 4), which could indicate concerted evolution of duplicated genes between these related species (Figure 2e). This is similar to the clade-specific shifts in selective constraint following concerted duplication events observed for MADS box transcription factors in angiosperms (Shan *et al.*, 2009).

The duplication process is a prominent feature of plant genomic architecture (Figure 1). This has led many researchers to speculate that gene duplication may have played an important role in the evolution of phenotypic novelty within the plant lineage (Flagel and Wendel, 2009). As a result of pervasive and recurring small-scale duplications, which may be followed by functional divergence, many nuclear genes in plants are members of gene families and may exhibit copy number variation lineages (Blanc and Wolfe, 2004; Schlueter *et al.*, 2004), as can be observed in table 1.

Evidence for frequent gene duplication has also been observed in the evolutionary history of numerous gene families that have expanded during the diversification of the angiosperms (De Bodt *et al.*, 2005; Zahn *et al.*, 2005; Duarte *et al.*, 2010). In multigene families descended from a common ancestor, individual genes in the group exert similar functions and have similar DNA sequences (Conrad and Antonarakis, 2007). One concept, concerted evolution, applies particularly to localized and typically tandem copies of a gene. The concept posits that all genes in a given group evolve coordinately, and that homogenization is the result of gene conversion (Figure 2e) (Conrad and Antonarakis, 2007).

The emerging picture points to plant NF-Y complexes acting as essential regulatory hubs for many processes. Multiple NF-Y subunits in vascular plants may associate with each other in various combinations that regulate the expression of specific gene sets and might provide similar levels of combinatorial diversity for transcriptional fine-tuning (Siefers *et al.*, 2009). The amplification observed in the plant lineage (Table 1) raises the possibility that new and divergent functions of heterotrimeric complexes have evolved in plants (Nelson *et al.*, 2007) indicating a more complex regulatory role for the various NF-Y proteins in plants than in other organisms (Stephenson *et al.*, 2007).

The existence of multiple genes for each subunit in the plant genome indicates that the specificity of subunit interaction may be determined by preferential protein-protein interaction, tissue or cell-specific expression of each gene or a combination of both (Yazawa and Kamada, 2007). The large number of possible combinations has hindered the analysis of plant NF-Y complexes and suggests that they might act in a more intricate system than in vertebrates and yeast, which have only one gene that encodes each HAP subunit (Yazawa and Kamada, 2007). Additionally, the multiple copies for each NF-Y subunit raises a question if a specific NF-Y subunit interacts with any other two NF-Y subunits or if the NF-Y subunit interacts with only specific member(s) of the other two subunits (Thirumurugan *et al.*, 2008).

Although the presence of many genes encoding NF-Y subunits suggests a high degree of genetic redundancy in plants, the analysis of mutants in single NF-Y genes in *Arabidopsis* has been associated with defects in development and enhanced stress sensitivity, suggesting a specialized function for each member (Lotan *et al.*, 1998; Kwong *et al.*, 2003; Lee *et al.*, 2003; Zanetti *et al.*, 2010). This could indicate that duplicated genes have passed through a neofunctionalization process (Figure 2b).

Some proteins may require several key substitutions before acquiring a new function, while others may be more mutationally labile. An example includes the terpene synthase gene family in Norway spruce (*Picea abies*). These genes appear to have undergone repeated rounds of neofunctionalization (Figure 2b) (Keeling *et al.*, 2008) and a small number of key amino acid substitutions among paralogs was sufficient to alter the substrate specificity and terpenoid product profiles (Flagel and Wendel, 2009). Another example of neofunctionalization (Figure 2b) in plants is observed in *Arabidopsis*, where a specific amino acid residue identified in LEC1 and LEC1-LIKE (L1L) is responsible for differentiating their functions (seed development) from those of other NF-YB members (Kwong *et al.*, 2003; Lee *et al.*, 2003; Yamamoto *et al.*, 2009). In addition, the analysis of amino acid substitution rates in plants has been appointed for the asymmetric evolution of certain duplicates of NF-YB and NF-YC subunits, which appears to be coupled with the asymmetric divergence in gene function (Yang *et al.*, 2005; Yamamoto *et al.*, 2009).

With respect to expression patterns, the *Arabidopsis* NF-Y gene family presents some members that are ubiquitously expressed and others that are tissue specific or induced only after the switch to reproductive growth in flowers and siliques (Gusmaroli *et al.*, 2001; 2002; Yazawa and Kamada, 2007). The difference observed in the expression pattern of these genes could represent an example of *cis*-regulatory divergence (Figure 3c), where the *cis*-element of gene evolves independently from the other members of gene family, and becomes regulated by different stimuli and/or *trans*-activators.

Because genes that harbor NF-Y binding domains include genes that are constitutive, inducible, and cell-cycle-dependent, the regulation of the expression of these genes cannot be exclusively due to NF-Y binding to DNA. In this scenario, the interaction with other transcription factors, either functionally or physically, will contribute to the NF-Y action (Matuoka and Chen, 2002). In addition, the independent evolution of protein-binding domains present in duplicated gene architecture can contribute to protein network divergence (Figure 3b2), increasing the numbers of possible interacting partners of NF-Y genes.

When compared with other forms of mutation, a notable feature of duplication is that it creates genetic redundancy. This redundancy fosters evolutionary innovation, creating the opportunity for duplicates to explore new evolutionary terrain (Flagel and Wendel, 2009).

The most important contribution of gene duplication to evolution is to supply new genetic material for mutation, drift and selection to act upon. This leads to the creation of new genes and new gene functions (Hurley *et al.*, 2005; Woollard, 2005; Schmidt and Davies, 2007), two important factors in the origin of genomic and organismal complexity (Gu *et al.*, 2002; Taylor and Raes, 2004; Sterck *et al.*, 2007).

The plasticity of a genome or species in adapting to environmental changes would be severely limited without gene duplication, because no more than two variants (alleles) exist at any locus within a diploid individual. A good example is the dozens of duplicated immunoglobulin genes that constitute the vertebrate adaptive immune system. It seems difficult to imagine how this system could have acquired this high complexity level without gene duplication (Zhang, 2003).

Plant gene families are largely conserved even over evolutionary time scales that encompass the diversification of all angiosperms and nonflowering plants (Rensing *et al.*, 2008). This property of plant genomes indicates that plants have not created new gene families, but have been endowed with a basic genetic toolkit of ancient origin. Despite the evolutionary conservation of gene families, lineage-specific fluctuations in gene family size are frequently observed among taxa (Velasco *et al.*, 2007; Ming *et al.*, 2008; Rensing *et al.*, 2008), which suggests that the diversity and lineage-specific phenotypic variation observed in land plants may not be explained by an equally diverse set of entirely novel genes. Indeed, much of plant diversity may have arisen from the duplication and adaptive specialization processes of pre-existing genes (neofunctionalization and subfunctionalization, Figure 2b and c, respectively). This perspective assigns gene duplication a central role in plant diversification, being a key process that generates the raw material necessary for adaptive evolution (Flagel and Wendel, 2009).

6. Conclusions

Whereas various classes of structural and metabolic genes preferentially return to a single copy state following whole-genome duplication (Paterson *et al.*, 2010), transcription factors tend to be preferentially retained among the duplicated genes in *A. thaliana* (Flagel and Wendel, 2009). Our findings support the hypotheses that this preference seems to be true for all plant species, based on the number of genes identified for each NF-Y subunit. Certainly, further studies encompassing functional assays are required to ascertain the role of these genes in plant metabolism.

The number of interacting partners in a molecular network (connectivity) of a particular gene also influences the probability of duplication gene retention (Flagel and Wendel, 2009). In this scenario, the high number of genes coding for the three subunits of NF-Y transcription factor in higher plants leads to numerous interaction possibilities among different genes of each subunit and among these genes and other transcription factors what could contribute to gene retention of the NF-Y transcription factor family in plants.

7. Acknowledgement

This work was supported by a CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPERGS (Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul), FINEP (Financiadora de Projetos) and MCT (Ministério de Ciência e Tecnologia). A. Cagliari

received a Ph.D. fellowship from CNPq. and G. Loss a Ph.D. fellowship from CAPES. A. Turchetto-Zolet and F. Maraschin have a PNPd-CAPES fellowship and M. Margis-Pinheiro and R. Margis are recipients of CNPq research fellowships number 308708/2006-7 and 303967/2008-0, respectively.

8. References

- Arents, G. and Moudrianakis, E.N. (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A*, 92, 11170-11174.
- Baxevanis, A.D., Arents, G., Moudrianakis, E.N. and Landsman, D. (1995) A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res*, 23, 2685-2691.
- Beisswanger, S. and Stephan, W. (2008) Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in *Drosophila*. *Proc Natl Acad Sci U S A*, 105, 5447-5452.
- Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., Samach, A. and Lifschitz, E. (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J*, 46, 462-476.
- Berry, M., Grosveld, F. and Dillon, N. (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature*, 358, 499-502.
- Blanc, G. and Wolfe, K.H. (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, 16, 1667-1678.
- Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, 422, 433-438.
- Braybrook, S.A. and Harada, J.J. (2008) LECs go crazy in embryo development. *Trends Plant Sci*, 13, 624-630.
- Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol*, 212, 563-578.
- Cai, X., Ballif, J., Endo, S., Davis, E., Liang, M., Chen, D., DeWald, D., Kreps, J., Zhu, T. and Wu, Y. (2007) A putative CCAAT-binding transcription factor is a regulator of flowering timing in *Arabidopsis*. *Plant Physiol*, 145, 98-105.
- Caretti, G., Motta, M.C. and Mantovani, R. (1999) NF-Y associates with H3-H4 tetramers and octamers by multiple mechanisms. *Mol Cell Biol*, 19, 8591-8603.
- Ceribelli, M., Dolfini, D., Merico, D., Gatta, R., Viganò, A.M., Pavesi, G. and Mantovani, R. (2008) The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol*, 28, 2047-2058.
- Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Human CCAAT-binding proteins have heterologous subunits. *Cell*, 53, 11-24.
- Combiér, J.P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernie, T., Ott, T., Gamas, P., Crespi, M. and Niebel, A. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes Dev*, 20, 3084-3088.

- Conant, G.C. and Wagner, A. (2002) GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Res*, 30, 3378-3386.
- Conrad, B. and Antonarakis, S.E. (2007) Gene duplication: a drive for phenotypic diversity and cause of human disease. *Annu Rev Genomics Hum Genet*, 8, 17-35.
- Coustry, F., Maity, S.N., Sinha, S. and de Crombrughe, B. (1996) The transcriptional activity of the CCAAT-binding factor CBF is mediated by two distinct activation domains, one in the CBF-B subunit and the other in the CBF-C subunit. *J Biol Chem*, 271, 14485-14491.
- Coustry, F., Hu, Q., de Crombrughe, B. and Maity, S.N. (2001) CBF/NF-Y functions both in nucleosomal disruption and transcription activation of the chromatin-assembled topoisomerase II α promoter. Transcription activation by CBF/NF-Y in chromatin is dependent on the promoter structure. *J Biol Chem*, 276, 40621-40630.
- Currie, R.A. (1998) NF-Y is associated with the histone acetyltransferases GCN5 and P/CAF. *J Biol Chem*, 273, 1430-1434.
- Dang, V.D., Bohn, C., Bolotin-Fukuhara, M. and Daignan-Fornier, B. (1996) The CCAAT box-binding factor stimulates ammonium assimilation in *Saccharomyces cerevisiae*, defining a new cross-pathway regulation between nitrogen and carbon metabolisms. *J Bacteriol*, 178, 1842-1849.
- Davis, J.C. and Petrov, D.A. (2004) Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol*, 2, E55.
- De Bodt, S., Maere, S. and Van de Peer, Y. (2005) Genome duplication and the origin of angiosperms. *Trends Ecol Evol*, 20, 591-597.
- De Grassi, A., Lanave, C. and Saccone, C. (2008) Genome duplication and gene-family evolution: the case of three OXPHOS gene families. *Gene*, 421, 1-6.
- de Silvio, A., Imbriano, C. and Mantovani, R. (1999) Dissection of the NF-Y transcriptional activation potential. *Nucleic Acids Res*, 27, 2578-2584.
- Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987a) A multiplicity of CCAAT box-binding proteins. *Cell*, 50, 863-872.
- Dorn, A., Durand, B., Marfing, C., Le Meur, M., Benoist, C. and Mathis, D. (1987b) Conserved major histocompatibility complex class II boxes--X and Y--are transcriptional control elements and specifically bind nuclear proteins. *Proc Natl Acad Sci U S A*, 84, 6249-6253.
- Duarte, J.M., Wall, P.K., Edger, P.P., Landherr, L.L., Ma, H., Pires, J.C., Leebens-Mack, J. and dePamphilis, C.W. (2010) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol Biol*, 10, 61.
- Dynan, W.S. and Tjian, R. (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature*, 316, 774-778.
- Edwards, D., Murray, J.A. and Smith, A.G. (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in *Arabidopsis*. *Plant Physiol*, 117, 1015-1022.
- Flagel, L.E. and Wendel, J.F. (2009) Gene duplication and evolutionary novelty in plants. *New Phytol*, 183, 557-564.

- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L. and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151, 1531-1545.
- Forsburg, S.L. and Guarente, L. (1989) Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. *Genes Dev*, 3, 1166-1178.
- Gallagher, C.E., Matthews, P.D., Li, F. and Wurtzel, E.T. (2004) Gene duplication in the carotenoid biosynthetic pathway preceded evolution of the grasses. *Plant Physiol*, 135, 1776-1783.
- Gu, Z., Nicolae, D., Lu, H.H. and Li, W.H. (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet*, 18, 609-613.
- Gusmaroli, G., Tonelli, C. and Mantovani, R. (2001) Regulation of the CCAAT-Binding NF-Y subunits in *Arabidopsis thaliana*. *Gene*, 264, 173-185.
- Gusmaroli, G., Tonelli, C. and Mantovani, R. (2002) Regulation of novel members of the *Arabidopsis thaliana* CCAAT-binding nuclear factor Y subunits. *Gene*, 283, 41-48.
- Harrison, P.M., Echols, N. and Gerstein, M.B. (2001) Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome. *Nucleic Acids Res*, 29, 818-830.
- Harrison, P.M., Hegyi, H., Balasubramanian, S., Luscombe, N.M., Bertone, P., Echols, N., Johnson, T. and Gerstein, M. (2002) Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res*, 12, 272-280.
- He, X.L. and Zhang, J.Z. (2005) Gene complexity and gene duplicability. *Current Biology*, 15, 1016-1021.
- Hinman, V.F. and Davidson, E.H. (2007) Evolutionary plasticity of developmental gene regulatory network architecture. *Proc Natl Acad Sci U S A*, 104, 19404-19409.
- Holland, P.W., Garcia-Fernandez, J., Williams, N.A. and Sidow, A. (1994) Gene duplications and the origins of vertebrate development. *Dev Suppl*, 125-133.
- Huang, H.S., Chen, C.J. and Chang, W.C. (1999) The CCAAT-box binding factor NF-Y is required for the expression of phospholipid hydroperoxide glutathione peroxidase in human epidermoid carcinoma A431 cells. *FEBS Lett*, 455, 111-116.
- Hurley, I., Hale, M.E. and Prince, V.E. (2005) Duplication events and the evolution of segmental identity. *Evol Dev*, 7, 556-567.
- Jordan, I.K., Wolf, Y.I. and Koonin, E.V. (2004) Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol*, 4, 22.
- Keeling, C.I., Weisshaar, S., Lin, R.P. and Bohlmann, J. (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc Natl Acad Sci U S A*, 105, 1085-1090.
- Kim, I.S., Sinha, S., de Crombrughe, B. and Maity, S.N. (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, 4003-4013.
- Kimura, M. (1989) The neutral theory of molecular evolution and the world view of the neutralists. *Genome*, 31, 24-31.

- Kondrashov, F.A. and Koonin, E.V. (2004) A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet*, 20, 287-290.
- Kusnetsov, V., Landsberger, M., Meurer, J. and Oelmuller, R. (1999) The assembly of the CAAT-box binding complex at a photosynthesis gene promoter is regulated by light, cytokinin, and the stage of the plastids. *J Biol Chem*, 274, 36009-36014.
- Kwong, R.W., Bui, A.Q., Lee, H., Kwong, L.W., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell*, 15, 5-18.
- Lee, H., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A*, 100, 2152-2156.
- Li, W.H., Gu, Z., Cavalcanti, A.R. and Nekrutenko, A. (2003) Detection of gene duplications and block duplications in eukaryotic genomes. *J Struct Funct Genomics*, 3, 27-34.
- Li, W.X., Oono, Y., Zhu, J., He, X.J., Wu, J.M., Iida, K., Lu, X.Y., Cui, X., Jin, H. and Zhu, J.K. (2008) The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell*, 20, 2238-2251.
- Li, X.Y., Mantovani, R., Hooft van Huijsduijnen, R., Andre, I., Benoist, C. and Mathis, D. (1992) Evolutionary variation of the CCAAT-binding transcription factor NF-Y. *Nucleic Acids Res*, 20, 1087-1091.
- Liberati, C., di Silvio, A., Ottolenghi, S. and Mantovani, R. (1999) NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices. *J Mol Biol*, 285, 1441-1455.
- Lotan, T., Ohto, M., Yee, K.M., West, M.A., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell*, 93, 1195-1205.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389, 251-260.
- Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, 290, 1151-1155.
- Lynch, M. and Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154, 459-473.
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y. (2005) Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A*, 102, 5454-5459.
- Maity, S.N. and de Crombrughe, B. (1992) Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments. *J Biol Chem*, 267, 8286-8292.
- Maity, S.N., Sinha, S., Ruteshouser, E.C. and de Crombrughe, B. (1992) Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. *J Biol Chem*, 267, 16574-16580.
- Maity, S.N. and De Crombrughe, B. (1996) Purification, characterization, and role of CCAAT-binding factor in transcription. *Methods Enzymol*, 273, 217-232.

- Maity, S.N. and de Crombrughe, B. (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem Sci*, 23, 174-178.
- Mantovani, R. (1998) A survey of 178 NF-Y binding CCAAT boxes. *Nucleic Acids Res*, 26, 1135-1143.
- Mantovani, R. (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene*, 239, 15-27.
- Marland, E., Prachumwat, A., Maltsev, N., Gu, Z. and Li, W.H. (2004) Higher gene duplicabilities for metabolic proteins than for nonmetabolic proteins in yeast and *E. coli*. *J Mol Evol*, 59, 806-814.
- Masiero, S., Imbriano, C., Ravasio, F., Favaro, R., Pelucchi, N., Gorla, M.S., Mantovani, R., Colombo, L. and Kater, M.M. (2002) Ternary complex formation between MAD5-box transcription factors and the histone fold protein NF-YB. *J Biol Chem*, 277, 26429-26435.
- Matuoka, K. and Yu Chen, K. (1999) Nuclear factor Y (NF-Y) and cellular senescence. *Exp Cell Res*, 253, 365-371.
- Matuoka, K. and Chen, K.Y. (2002) Transcriptional regulation of cellular ageing by the CCAAT box-binding factor CBF/NF-Y. *Ageing Res Rev*, 1, 639-651.
- McNabb, D.S., Xing, Y. and Guarente, L. (1995) Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev*, 9, 47-58.
- Miyoshi, K., Ito, Y., Serizawa, A. and Kurata, N. (2003) OsHAP3 genes regulate chloroplast biogenesis in rice. *Plant J*, 36, 532-540.
- Muro, A.F., Bernath, V.A. and Kornblihtt, A.R. (1992) Interaction of the -170 cyclic AMP response element with the adjacent CCAAT box in the human fibronectin gene promoter. *J Biol Chem*, 267, 12767-12774.
- Myers, R.M., Tilly, K. and Maniatis, T. (1986) Fine structure genetic analysis of a beta-globin promoter. *Science*, 232, 613-618.
- Nelson, D.E., Repetti, P.P., Adams, T.R., Creelman, R.A., Wu, J., Warner, D.C., Anstrom, D.C., Bensen, R.J., Castiglioni, P.P., Donnarummo, M.G., Hinchey, B.S., Kumimoto, R.W., Maszle, D.R., Canales, R.D., Krolikowski, K.A., Dotson, S.B., Gutterson, N., Ratcliffe, O.J. and Heard, J.E. (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. *Proc Natl Acad Sci U S A*, 104, 16450-16455.
- Nicole, M.C., Hamel, L.P., Morency, M.J., Beaudoin, N., Ellis, B.E. and Seguin, A. (2006) MAP-ping genomic organization and organ-specific expression profiles of poplar MAP kinases and MAP kinase kinases. *BMC Genomics*, 7, 223.
- Nowak, M.A., Boerlijst, M.C., Cooke, J. and Smith, J.M. (1997) Evolution of genetic redundancy. *Nature*, 388, 167-171.
- Ober, D. (2010) Gene duplications and the time thereafter - examples from plant secondary metabolism. *Plant Biol (Stuttg)*, 12, 570-577.
- Olesen, J.T. and Guarente, L. (1990) The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex. *Genes Dev*, 4, 1714-1729.
- Papp, B., Pal, C. and Hurst, L.D. (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature*, 424, 194-197.

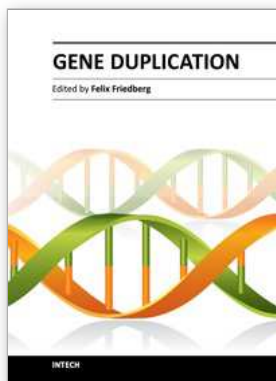
- Paterson, A.H., Freeling, M., Tang, H. and Wang, X. (2010) Insights from the comparison of plant genome sequences. *Annu Rev Plant Biol*, 61, 349-372.
- Peng, Y. and Jahroudi, N. (2002) The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter. *Blood*, 99, 2408-2417.
- Peng, Y. and Jahroudi, N. (2003) The NFY transcription factor inhibits von Willebrand factor promoter activation in non-endothelial cells through recruitment of histone deacetylases. *J Biol Chem*, 278, 8385-8394.
- Remenyi, A., Scholer, H.R. and Wilmanns, M. (2004) Combinatorial control of gene expression. *Nat Struct Mol Biol*, 11, 812-815.
- Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol*, 3, 423-434.
- Rieping, M. and Schoffl, F. (1992) Synergistic effect of upstream sequences, CCAAT box elements, and HSE sequences for enhanced expression of chimaeric heat shock genes in transgenic tobacco. *Mol Gen Genet*, 231, 226-232.
- Rodin, S.N. and Riggs, A.D. (2003) Epigenetic silencing may aid evolution by gene duplication. *J Mol Evol*, 56, 718-729.
- Romier, C., Cocchiarella, F., Mantovani, R. and Moras, D. (2003) The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J Biol Chem*, 278, 1336-1345.
- Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572-1574.
- Roy, B. and Lee, A.S. (1995) Transduction of calcium stress through interaction of the human transcription factor CBF with the proximal CCAAT regulatory element of the *grp78/BiP* promoter. *Mol Cell Biol*, 15, 2263-2274.
- Sandman, K., Krzycki, J.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermobacter fervidus*, is most closely related to histones. *Proc Natl Acad Sci U S A*, 87, 5788-5791.
- Schlueter, J.A., Dixon, P., Granger, C., Grant, D., Clark, L., Doyle, J.J. and Shoemaker, R.C. (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome*, 47, 868-876.
- Schmidt, E.E. and Davies, C.J. (2007) The origins of polypeptide domains. *Bioessays*, 29, 262-270.
- Serra, E., Zemzoumi, K., di Silvio, A., Mantovani, R., Lardans, V. and Dissous, C. (1998) Conservation and divergence of NF-Y transcriptional activation function. *Nucleic Acids Res*, 26, 3800-3805.
- Shan, H., Zahn, L., Guindon, S., Wall, P.K., Kong, H., Ma, H., DePamphilis, C.W. and Leebens-Mack, J. (2009) Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol Biol Evol*, 26, 2229-2244.
- Shi, Q., Gross, K.W. and Sigmund, C.D. (2001) NF-Y antagonizes renin enhancer function by blocking stimulatory transcription factors. *Hypertension*, 38, 332-336.
- Sidow, A. (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev*, 6, 715-722.

- Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum, W.E.t., Tayrose, G. and Holt, B.F., 3rd (2009) Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol*, 149, 625-641.
- Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M. and Van de Peer, Y. (2002) The hidden duplication past of Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, 99, 13627-13632.
- Singh, K.B. (1998) Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol*, 118, 1111-1120.
- Sinha, S., Maity, S.N., Lu, J. and de Crombrughe, B. (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. *Proc Natl Acad Sci U S A*, 92, 1624-1628.
- Sinha, S., Kim, I.S., Sohn, K.Y., de Crombrughe, B. and Maity, S.N. (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol*, 16, 328-337.
- Soltis, P.S. (2005) Ancient and recent polyploidy in angiosperms. *New Phytol*, 166, 5-8.
- Steidl, S., Papagiannopoulos, P., Litzka, O., Andrianopoulos, A., Davis, M.A., Brakhage, A.A. and Hynes, M.J. (1999) AnCF, the CCAAT binding complex of *Aspergillus nidulans*, contains products of the hapB, hapC, and hapE genes and is required for activation by the pathway-specific regulatory gene amdR. *Mol Cell Biol*, 19, 99-106.
- Stephenson, T.J., McIntyre, C.L., Collet, C. and Xue, G.P. (2007) Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*. *Plant Mol Biol*, 65, 77-92.
- Sterck, L., Rombauts, S., Vandepoele, K., Rouze, P. and Van de Peer, Y. (2007) How many genes are there in plants (... and why are they there)? *Curr Opin Plant Biol*, 10, 199-203.
- Stoltzfus, A. (1999) On the possibility of constructive neutral evolution. *J Mol Evol*, 49, 169-181.
- Struhl, K. and Moqtaderi, Z. (1998) The TAFs in the HAT. *Cell*, 94, 1-4.
- Tasanen, K., Oikarinen, J., Kivirikko, K.I. and Pihlajaniemi, T. (1992) Promoter of the gene for the multifunctional protein disulfide isomerase polypeptide. Functional significance of the six CCAAT boxes and other promoter elements. *J Biol Chem*, 267, 11513-11519.
- Taylor, J.S., Van de Peer, Y. and Meyer, A. (2001) Genome duplication, divergent resolution and speciation. *Trends Genet*, 17, 299-301.
- Taylor, J.S. and Raes, J. (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet*, 38, 615-643.
- Testa, A., Donati, G., Yan, P., Romani, F., Huang, T.H., Vigano, M.A. and Mantovani, R. (2005) Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF-Y binding CCAAT sites outside of core promoters. *J Biol Chem*, 280, 13606-13615.

- Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A. and Kurata, N. (2008) Identification, characterization and interaction of HAP family genes in rice. *Mol Genet Genomics*, 279, 279-289.
- Vision, T.J., Brown, D.G. and Tanksley, S.D. (2000) The origins of genomic duplications in Arabidopsis. *Science*, 290, 2114-2117.
- Wagner, A. (2000) Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc Natl Acad Sci U S A*, 97, 6579-6584.
- Wapinski, I., Pfeffer, A., Friedman, N. and Regev, A. (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature*, 449, 54-61.
- Warpeha, K.M., Upadhyay, S., Yeh, J., Adamiak, J., Hawkins, S.I., Lapik, Y.R., Anderson, M.B. and Kaufman, L.S. (2007) The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in Arabidopsis. *Plant Physiol*, 143, 1590-1600.
- Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierc, J., Samach, A. and Coupland, G. (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *Plant Cell*, 18, 2971-2984.
- Wolberger, C. (1998) Combinatorial transcription factors. *Curr Opin Genet Dev*, 8, 552-559.
- Woollard, A. (2005) Gene duplications and genetic redundancy in *C. elegans*. *WormBook*, 1-6.
- Xiong, A.S., Peng, R.H., Zhuang, J., Gao, F., Zhu, B., Fu, X.Y., Xue, Y., Jin, X.F., Tian, Y.S., Zhao, W. and Yao, Q.H. (2009) Gene duplication, transfer, and evolution in the chloroplast genome. *Biotechnol Adv*, 27, 340-347.
- Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S. and Hattori, T. (2009) Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J*, 58, 843-856.
- Yang, J., Xie, Z. and Glover, B.J. (2005) Asymmetric evolution of duplicate genes encoding the CCAAT-binding factor NF-Y in plant genomes. *New Phytol*, 165, 623-631.
- Yang, X., Tuskan, G.A. and Cheng, M.Z. (2006) Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol*, 142, 820-830.
- Yazawa, K. and Kamada, H. (2007) Identification and characterization of carrot HAP factors that form a complex with the embryo-specific transcription factor C-LEC1. *J Exp Bot*, 58, 3819-3828.
- Yokoyama, S. and Yokoyama, R. (1989) Molecular evolution of human visual pigment genes. *Mol Biol Evol*, 6, 186-197.
- Zahn, L.M., Leebens-Mack, J., DePamphilis, C.W., Ma, H. and Theissen, G. (2005) To B or Not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *J Hered*, 96, 225-240.
- Zanetti, M.E., Blanco, F.A., Beker, M.P., Battaglia, M. and Aguilar, O.M. (2010) A C subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean-Rhizobium etli symbiosis. *Plant Cell*, 22, 4142-4157.

Zhang, J.Z. (2003) Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18, 292-298.

Zwicker, J. and Muller, R. (1997) Cell-cycle regulation of gene expression by transcriptional repression. *Trends Genet*, 13, 3-6.



Gene Duplication

Edited by Prof. Felix Friedberg

ISBN 978-953-307-387-3

Hard cover, 400 pages

Publisher InTech

Published online 17, October, 2011

Published in print edition October, 2011

The book Gene Duplication consists of 21 chapters divided in 3 parts: General Aspects, A Look at Some Gene Families and Examining Bundles of Genes. The importance of the study of Gene Duplication stems from the realization that the dynamic process of duplication is the "sine qua non" underlying the evolution of all living matter. Genes may be altered before or after the duplication process thereby undergoing neofunctionalization, thus creating in time new organisms which populate the Earth.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Alexandro Cagliari, Andreia Carina Turchetto-Zolet, Felipe dos Santos Maraschin, Guilherme Loss, Rogério Margis and Marcia Margis-Pinheiro (2011). The Evolutionary History of CBF Transcription Factors: Gene Duplication of CCAAT – Binding Factors NF-Y in Plants, Gene Duplication, Prof. Felix Friedberg (Ed.), ISBN: 978-953-307-387-3, InTech, Available from: <http://www.intechopen.com/books/gene-duplication/the-evolutionary-history-of-cbf-transcription-factors-gene-duplication-of-ccaat-binding-factors-nf-y>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.