

Tissue-Specific Expression Patterns of Arabidopsis NF-Y Transcription Factors Suggest Potential for Extensive Combinatorial Complexity^{1[W][OA]}

Nicholas Siefers, Kristen K. Dang², Roderick W. Kumimoto, William Edwards Bynum IV³, Gregory Tayrose⁴, and Ben F. Holt III*

Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599 (N.S., K.K.D., W.E.B., G.T., B.F.H.); and Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma 73019 (N.S., R.W.K., B.F.H.)

All aspects of plant and animal development are controlled by complex networks of transcription factors. Transcription factors are essential for converting signaling inputs, such as changes in daylength, into complex gene regulatory outputs. While some transcription factors control gene expression by binding to cis-regulatory elements as individual subunits, others function in a combinatorial fashion. How individual subunits of combinatorial transcription factors are spatially and temporally deployed (e.g. expression-level, posttranslational modifications and subcellular localization) has profound effects on their control of gene expression. In the model plant *Arabidopsis thaliana*, we have identified 36 Nuclear Factor Y (NF-Y) transcription factor subunits (10 NF-YA, 13 NF-YB, and 13 NF-YC subunits) that can theoretically combine to form 1,690 unique complexes. Individual plant subunits have functions in flowering time, embryo maturation, and meristem development, but how they combine to control these processes is unknown. To assist in the process of defining unique NF-Y complexes, we have created promoter: β -glucuronidase fusion lines for all 36 *Arabidopsis* genes. Here, we show NF-Y expression patterns inferred from these promoter: β -glucuronidase lines for roots, light- versus dark-grown seedlings, rosettes, and flowers. Additionally, we review the phylogenetic relationships and examine protein alignments for each NF-Y subunit family. The results are discussed with a special emphasis on potential roles for NF-Y subunits in photoperiod-controlled flowering time.

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

MADS box transcription factors represent a well-studied example of this phenomenon (Messenguy and Dubois, 2003; Yamaguchi and Hirano, 2006). Through the formation of homodimers, heterodimers, and heteromultimers, MADS box proteins bind unique cis-regulatory elements and control, for example, most of the floral organ fates in plants (Kaufmann et al., 2005). While less well studied in the plant lineage, the numerous heterotrimeric NF-Y (for Nuclear Factor Y) transcription factors might provide similar levels of combinatorial diversity for transcriptional fine-tuning.

NF-Y transcription factors are likely found in all eukaryotes and have roles in the regulation of diverse genes (McNabb et al., 1995; Edwards et al., 1998; Maity and de Crombrughe, 1998; Mantovani, 1999). In mammals, where their biochemistry is well described, the NF-Y transcription factor complex is composed of three unique subunits: NF-YA, NF-YB, and NF-YC. Assembly of the NF-Y heterotrimer in mammals follows a strict, stepwise pattern (Sinha et al., 1995, 1996). Initially, a heterodimer is formed in the cytoplasm between the subunits NF-YB and NF-YC. This dimer then translocates to the nucleus, where the third subunit, NF-YA, is recruited to generate the mature, heterotrimeric NF-Y transcription factor (Frontini et al., 2004; Kahle et al., 2005). Mature NF-Y binds promoters with the core pentamer nucleotide sequence CCAAT, and this can result in either positive or negative transcriptional regulation (Peng and Jahroudi, 2002, 2003; Ceribelli et al., 2008). Bioinformatic analy-

¹ This work was supported by the University of Oklahoma's Department of Botany and Microbiology, Office of the Vice President of Research, and College of Arts and Sciences. Additional funding was provided by the Oklahoma Center for the Advancement of Science and Technology.

² Present address: Human Vaccine Institute, Duke University, Durham, NC 27710.

³ Present address: University of South Carolina School of Medicine, Columbia, SC 29201.

⁴ Present address: University of North Carolina School of Medicine, Chapel Hill, NC 27599.

* Corresponding author; e-mail benholt@ou.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Ben F. Holt III (benholt@ou.edu).

^[W] The online version of this article contains Web-only data.

^[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.130591

ses indicate that 25% to 30% of all mammalian promoters have predicted NF-Y binding sites (Bucher, 1990; Testa et al., 2005), and recent chromatin immunoprecipitation data demonstrate 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). Predictably, NF-Y function is essential for mammalian development (Hu and Maity, 2000; Bhattacharya et al., 2003).

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 et al., 2000). For example, humans and mice encode only one copy of each subunit. Thus, there is minimal combinatorial diversity in the subunit composition of the heterotrimeric mammalian NF-Y. In contrast, *Arabidopsis thaliana* has multiple genes encoding each subunit (10 *NF-YA*, 13 *NF-YB*, and 13 *NF-YC* homologs; Riechmann et al., 2000; this article). This *Arabidopsis* NF-Y expansion is a general feature of the plant lineage, including monocots and dicots. Because of the heterotrimeric composition, the 36 *Arabidopsis* NF-Y subunits can theoretically combine to generate 1,690 unique transcription factors.

NF-Y function in the plant lineage is poorly understood, yet many of the mechanistic details are likely conserved across plant, animal, and fungal lineages. This inference comes from strong, cross-kingdom conservation of amino acid residues with well-characterized importance in mammalian and yeast NF-Y functions (Maity and de Crombrughe, 1992; Maity et al., 1992; Mantovani et al., 1994; Sinha et al., 1995, 1996; Coustry et al., 1996; Kim et al., 1996; Romier et al., 2003). Additionally, several groups have demonstrated that each of the three plant NF-Y proteins can substitute for their yeast counterparts in gene expression assays (Masiero et al., 2002; Ben-Naim et al., 2006). Finally, our own preliminary data suggests that dominant negative *Arabidopsis* NF-YC subunits can be readily predicted from existing mammalian and yeast data (Kim et al., 1996; B.F. Holt and R.W. Kumimoto, unpublished data). Hence, animal and yeast models provide excellent starting points for further investigations of the numerous plant NF-Y proteins.

No complete plant NF-Y complex has been described, but individual subunits are increasingly known to be involved in a number of important processes. *LEAFY COTYLEDON1* (*LEC1* or *NF-YB9*; Table I) was the earliest cloned and described plant NF-Y. *LEC1* has strong expression in the developing embryo and is necessary for controlling the transition from embryo to adult status (West et al., 1994; Lotan et al., 1998; Lee et al., 2003). Evidence comes from the precocious development of trichomes on *lec1* cotyledons and the presence of embryo-like tissue on the adult leaves of *LEC1*-overexpressing plants. The

closely related *LEC1-LIKE* (*L1L* or *AtNF-YB6*) is also involved in embryogenesis (Kwong et al., 2003). While *LEC1* has been studied for well over a decade, the requirements for specific *NF-YA* and *NF-YC* subunits remain elusive. This may be partly due to the complexity of genetic and biochemical studies in the embryo. However, another likely reason for the general lack of complete plant NF-Y complexes is the likelihood of extensive redundancies between NF-Y subunits. Several new phenotypes associated with overexpression and mutant versions of NF-Y subunits promise to provide the anchor points for describing complete in planta complexes.

The *Medicago truncatula* NF-YA subunit, *MtHAP2-1*, is expressed in a narrow region of the root nodule meristem and is essential for postinitiation nodule development (Combiere et al., 2006). Interestingly, spatial and temporal maintenance of *MtHAP2-1* nodule expression is dually controlled by microRNA169 (Jones-Rhoades and Bartel, 2004) and a small RNA-binding peptide (uORF1p) encoded by the *MtHAP2-1* leader sequence (Combiere et al., 2008). Thus, as with the narrow expression of *LEC1* and *L1L* in embryos, *MtHAP2-1* expression is quite finely controlled. This is in contrast to animal systems, where NF-Y expression is largely ubiquitous and, unsurprisingly, suggests refinements and specialization of NF-Y function in the plant lineage.

Plant NF-Y function also appears to be important for responses to drought stress. Although a specific mechanism remains unclear, overexpression of *NF-YB1* and its ortholog in maize (*Zea mays*), *ZmNF-YB2*, leads to enhanced drought resistance (Nelson et al., 2007). Although the ability of *NF-YB1* to promote drought resistance is clear, no loss-of-function data were provided to demonstrate an actual biological role in *Arabidopsis*. In contrast, a recent publication shows both overexpression and loss-of-function data for *NF-YA5* (Li et al., 2008). Overexpression of *NF-YA5* reduced drought susceptibility, anthocyanin production, and stomatal aperture, while *nf-ya5* mutants had the expected opposite phenotype in each instance. As with the role of *MtHAP2-1* in nodule meristem development, miRNA169 also plays a regulatory role in drought resistance. Specifically, the miRNA169 precursors miRNA169a and miRNA169c are strongly down-regulated by drought treatments in an abscisic acid (ABA)-dependent manner. Similar to *NF-YA5* loss of function, 35S:miRNA169 plants were more drought susceptible (Li et al., 2008). Although still unknown, the involvement of both *NF-YA* and *NF-YB* in plant drought resistance suggests that complete NF-Y complexes with *NF-YC* subunits will eventually be discovered.

Perhaps the most interesting recent discovery is the involvement of NF-Y in the control of photoperiod-regulated flowering time. In *Arabidopsis*, the key regulator of photoperiod-induced flowering time is the zinc-finger-type transcriptional activator encoded by *CONSTANS* (*CO*; Redei, 1962; Koornneef et al.,

Table 1. Suggested *NF-Y* nomenclature

Nomenclature is the same as originally proposed (Gusmaroli et al., 2001, 2002), except the *At* designations are removed and seven new genes (bold) have been added (see note on nomenclature). AGI, Arabidopsis Genome Initiative number.

NF-YA Family			NF-YB Family			NF-YC Family		
Current Name	AGI	Other Name	Current Name	AGI	Other Name	Current Name	AGI	Other Name
NF-YA1	At5g12840	AtHAP2a	NF-YB1	At2g38880	AtHAP3a	NF-YC1	At3g48590	AtHAP5a
NF-YA2	At3g05690	AtHAP2b	NF-YB2	At5g47640	AtHAP3b	NF-YC2	At1g56170	AtHAP5b
NF-YA3	At1g72830	AtHAP2c	NF-YB3	At4g14540		NF-YC3	At1g54830	
NF-YA4	At2g34720		NF-YB4	At1g09030		NF-YC4	At5g63470	
NF-YA5	At1g54160		NF-YB5	At2g47810		NF-YC5	At5g50490	
NF-YA6	At3g14020		NF-YB6	At5g47670	LEC1-LIKE	NF-YC6	At5g50480	
NF-YA7	At1g30500		NF-YB7	At2g13570		NF-YC7	At5g50470	
NF-YA8	At1g17590		NF-YB8	At2g37060		NF-YC8	At5g27910	
NF-YA9	At3g20910		NF-YB9	At1g21970	LEC1	NF-YC9	At1g08970	AtHAP5c
NF-YA10	At5g06510		NF-YB10	At3g53340		NF-YC10	At1g07980	
			NF-YB11	At2g27470		NF-YC11	At3g12480	
			NF-YB12	At5g08190		NF-YC12	At5g38140	
			NF-YB13	At5g23090		NF-YC13	At5g43250	

1991; Putterill et al., 1995). *CO* mRNA levels are controlled by the circadian clock and oscillate on a daily basis: *CO* expression peaks during the day in long-day (LD) conditions and during the night in short-day conditions (Suarez-Lopez et al., 2001). Peak expression during the day is essential for *CO* activity because the protein is rapidly degraded in the dark (Valverde et al., 2004). Under LD conditions, *CO* protein accumulates and induces the expression of *FLOWERING LOCUS T (FT)*. Recent evidence strongly suggests that *FT* protein is the primary component of “florigen,” the long-sought mobile flowering signal (Izawa, 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007; Tamaki et al., 2007). Therefore, in LD-grown *co* mutants, *FT* does not accumulate to sufficient levels to induce flowering (Kardailsky et al., 1999; Samach et al., 2000). While the genetics of this biological process have been carefully examined, precisely how *CO* integrates with DNA to trigger *FT* expression remains unclear.

Several publications strongly suggest that *NF-Y* transcription factors are intimately involved in photo-period-regulated flowering. Research groups studying flowering time in both tomato (*Solanum lycopersicum*) and Arabidopsis identified *NF-YB* and *NF-YC* subunits as *CO*-interacting proteins via yeast two-hybrid assays (Ben-Naim et al., 2006; Wenkel et al., 2006). Additionally, two independent research groups described moderate flowering time delays for the same *nf-yb2* allele (Cai et al., 2007; Chen et al., 2007). Furthermore, overexpression of *NF-YB2* resulted in significantly more rapid flowering. Finally, Kumimoto et al. (2008) found that *nf-yb2 nf-yb3* double mutant plants flower as late as *co* mutants. Because of *NF-Y*'s well-characterized role as a transcription-activating, DNA-binding complex in mammals and yeast, these results immediately suggest a possible platform for *CO* interactions with DNA. Additionally, the region of interaction between *CO* and *NF-Y* proteins is highly conserved in many *CO-like* genes (Wenkel et al., 2006). This suggests that the *CO/NF-Y* regulatory module

might be paradigmatic for other *CO-like* DNA interactions and developmental processes.

The emerging picture has plant *NF-Y* complexes acting as essential regulatory hubs for many processes. While functions for individual *NF-Y* genes are beginning to emerge, overlapping functionality remains a persistent problem for further investigations. Additionally, demonstration of a complete *NF-Y* complex remains elusive. In this article, we present updated phylogenetic trees and alignments for all 36 Arabidopsis *NF-Y* proteins. As determined by functional analyses in yeast and mammals, we clearly identify the essential amino acids for each subunit type and specifically discuss these data in the context of recent findings on the *NF-Y* requirement in flowering. Furthermore, we test the utility of mammalian and yeast-derived positional weight matrices for defining CCAAT sites in plant promoters. Finally, we examine the tissue- and development-specific expression patterns for all 36 *NF-Y* genes using stable promoter:*GUS* fusions. The resulting plant lines will facilitate the discovery of complete *NF-Y* complexes and are freely available to academic researchers. Collectively, we hope the following data and review will serve as an entry point for other researchers interested in plant *NF-Y* function.

RESULTS

Notes on Nomenclature

Various nomenclatures are used for the three *NF-Y* families. The three most widely used names in various organisms are *CBF* (for CCAAT-binding factor), *HAP* (for histone or heme-associated protein), and *NF-Y*. Additionally, the unique subunits are alternatively assigned numerical or letter designations, and those designations often do not match across genera. For example, *NF-YA* in *Homo sapiens* is homologous to *CBF-B* in *Rattus norvegicus*, *HAP2* in *Saccharomyces*

cerevisiae, and *HAPB* in *Aspergillus nidulans*. Because there are still relatively few Arabidopsis NF-Y papers and we provide information on all 36 genes, we discussed nomenclature options with curators from The Arabidopsis Information Resource (TAIR). In Arabidopsis, these genes are alternatively called *AtHAP*, *AtNF-Y*, *NF-Y*, and *CBF* (Edwards et al., 1998; Kusnetsov et al., 1999; Gusmaroli et al., 2001; Kumimoto et al., 2008). Due to their usage for unrelated genes, *HAP* (for *HAPLESS*) and *CBF* (for *C-Repeat/DRE Binding Factor*) nomenclature options are both problematic and confusing. Additionally, TAIR curators are encouraging limited usage of the “At” designation. Thus, we jointly chose the *NF-Y* nomenclature following the number scheme initially suggested by Gusmaroli et al. (2001, 2002) and expanded here with seven new genes. To avoid unnecessary complexity, we respectfully request that future Arabidopsis researchers follow this updated naming scheme (Table I).

NF-Y Phylogenies and Alignments

Published phylogenies are available for Arabidopsis, rice (*Oryza sativa*), and wheat (*Triticum aestivum*) plant NF-Y families (Gusmaroli et al., 2001, 2002; Yang et al., 2005; Stephenson et al., 2007). A comprehensive examination of all Arabidopsis transcription factors concluded that there are 36 total NF-Y genes (10 NF-YA, 13 NF-YB, and 13 NF-YC homologs; Riechmann et al., 2000). Nevertheless, published phylogenies only list 29 to 30 genes (Gusmaroli et al., 2001, 2002; Yang et al., 2005). Therefore, we repeated the BLAST searches for each NF-Y family (Altschul et al., 1990). In each case, BLAST searches were performed with several members of each NF-Y family using amino acid sequences from highly conserved (across genera) regions. Although there is no absolute rule for inclusion/exclusion from a gene family, in each case we chose the last member of the family based on very obvious and large breaks in the BLAST-derived E value score (e.g. using NF-YB1 in BLAST analysis, we accepted NF-YB family members with E values ranging from 2e-50 to 2e-06; the next closest nonfamily member had an E value of 0.11). Our searches confirmed the previous finding of 36 total NF-Y genes (Riechmann et al., 2000) and are presented as three phylogenetic trees (Fig. 1) and three amino acid alignments (Figs. 2–4).

NF-YA Family

NF-YA proteins represent a unique transcription factor class lacking obvious homology to other described proteins (Maity and de Crombrughe, 1998). NF-YA proteins are characterized by the presence of Gln (Q)- and Ser/Thr (S/T)-rich NH₂ termini, a subunit interaction domain (NF-YB/NF-YC interaction), and a DNA-binding domain (Olesen and Guarente, 1990; Maity and de Crombrughe, 1992; Xing et al., 1993, 1994). The protein interaction and DNA binding

domains are well conserved between plant and other eukaryote lineages (Fig. 2). As with other eukaryotes, NF-YA NH₂-terminal regions are also characterized by an overall high composition of Q and S/T residues. The Q-rich regions of NF-YA and NF-YC are thought to act redundantly in transcriptional activation (di Silvio et al., 1999). However, it is notable that plant NF-YA proteins have low Q:S ratios (1.0:2.1, Q:S), while the yeast, human, and rat NF-YA proteins have the opposite relationship (2.9:1.0, Q:S). It is currently unknown how these changes might affect the transcriptional activation potential of plant NF-Y complexes.

The alignment in Figure 2 highlights the cross-kingdom conservation of NF-YA proteins in their interaction and DNA-binding domains. Functionally required amino acids were previously determined by mutagenesis in yeast and mammalian NF-YA (*HAP2* and *CBF-B*, respectively; Maity and de Crombrughe, 1992; Xing et al., 1993). Of the three plant NF-Y families, functionally required amino acids are most highly conserved in the NF-YA proteins (“required” throughout refers to data from yeast or mammals; almost no amino acid requirements have been experimentally determined for plant NF-Y proteins). There is one notable change that is apparently specific to the plant lineage: the required Arg (R) residue at position 9 is almost always Gly (G) or Ala (A) in the plant lineage. This is true for all Arabidopsis NF-YA proteins as well as the majority of other plant NF-YA sequences (rice, poplar [*Populus* spp.], grape [*Vitis vinifera*], moss, wheat, etc.). All other required amino acids from yeast are absolutely conserved in all 10 Arabidopsis NF-YA proteins.

The conservation of specific amino acid residues across highly divergent NF-YA lineages strongly suggests functional conservation. This is particularly relevant when considering the recent finding that CO, a master regulator of floral transitions, physically interacts with NF-YB and NF-YC proteins (Ben-Naim et al., 2006; Wenkel et al., 2006). In this sense, CO seems to act like an NF-YA protein. Accordingly, the CCT (for CO, CO-LIKE, TOC1) domain of CO and CO-LIKE (COL) proteins shares a region of apparent homology to the NF-YA proteins (see “CCT Cons.” line in Fig. 2; Wenkel et al., 2006). In a stretch of 41 amino acids, the CCT-containing proteins CO and COL1-5 share 32% identity with the NF-YA proteins (Fig. 2; Supplemental Figs. S1 and S2). Furthermore, previously identified *co* alleles, and mutant alleles in related CCT domain proteins, represent amino acid changes in residues that are conserved between the CCT domain and NF-YA proteins (Wenkel et al., 2006). The similarities between the conserved domains of NF-YA and COL suggest the possibility of another layer of combinatorial interactions with plant NF-Y complexes.

NF-YB and NF-YC Families

As with NF-YA, NF-YB and NF-YC families have well-described subunit interaction and DNA-binding

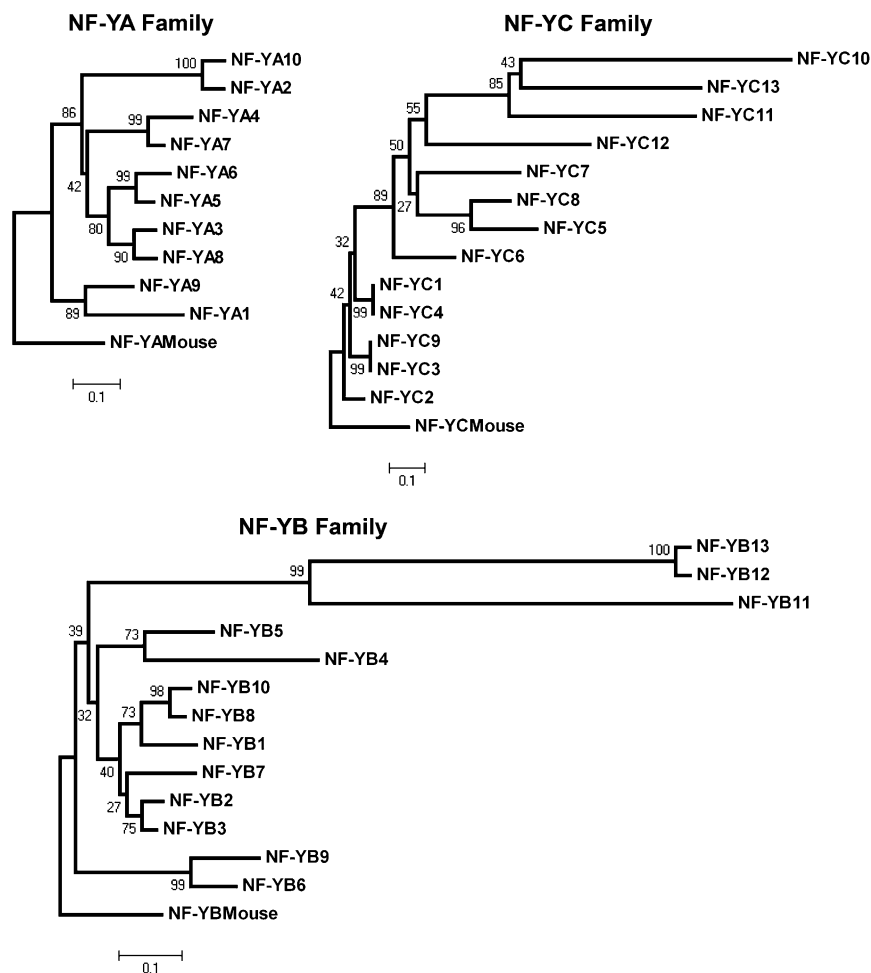


Figure 1. NF-Y family phylogenies. Phylogenetic trees for each family were constructed by neighbor joining using the conserved regions illustrated in Figures 2 to 4. Phylogenetic trees were also generated for full-length proteins and did not differ substantially from those shown (data not shown). Reliability values at each branch represent bootstrap samples (2,000 replicates). All trees were determined and constructed using MEGA 4 (Tamura et al., 2007). Note that NF-YB12 and NF-YB13 are sometimes included in a separate DR1-related, two-member protein family (see “Gene Families” at TAIR and Riechmann et al. [2000]). Nevertheless, NF-YB12 and -YB13 clearly show much stronger homology to the larger NF-YB family than to any other proteins, and there is currently no functional evidence to differentiate them as unique. Thus, we consider them to be divergent, but related, NF-YB family members.

domains (Figs. 3 and 4; Kim et al., 1996; Sinha et al., 1996; McNabb et al., 1997; Romier et al., 2003). The conserved regions of NF-YB and NF-YC have structural and amino acid homology to histone fold motifs. Specifically, NF-YB is related to the histone fold motifs of H2B histones, while NF-YC subunits are related to H2A histones (Mantovani, 1999). NF-Y can physically interact with H3/H4 tetramers and H3-/H4-like TATA-binding protein-associated factors (Caretta et al., 1999; Frontini et al., 2002). Mature NF-Y binds the minor groove of DNA, resulting in localized DNA bending and distortions of nucleosome-DNA interactions (Ronchi et al., 1995; Coustry et al., 2001). Collectively, these nucleosomal distortions allow recruitment of chromatin-remodeling enzymes and RNA polymerase II, leading to transcriptional initiation (Matuoka and Chen, 2002).

For the majority of NF-YB and NF-YC proteins, required amino acids are well conserved (Figs. 3 and 4; Sinha et al., 1995, 1996; Kim et al., 1996). Phylogenetically distant Arabidopsis family members, such as NF-YB11 to -YB13 and NF-YC10 to -YC13 (Fig. 1), are much more likely to have undergone changes in required amino acids (Fig. 3). Because the required amino acids are so highly conserved across evolutionary time and space, it is likely that nonconservative

changes will significantly alter protein function. In this regard, LEC1 (NF-YB9) and LEC1-like (NF-YB6) have Asp (D) residues where Lys (K) is found in yeast, mammals, and most plants and is required for mammalian NF-YB function (position 29 in Fig. 3; actual change is K55D). Although LEC1 has other differences from plant and animal NF-YB proteins, K55D is the only change in a required amino acid and reversion of this alteration in LEC1 eliminates rescue of the *lec1* embryonic desiccation-intolerance phenotype (Lee et al., 2003). Thus, LEC1 appears to have evolved a novel function.

NF-YB2 and NF-YB3, redundant players in photoperiod-related floral transitions (Kumimoto et al., 2008), have replaced the required Ser (S) at position 48 (Fig. 3) with Gly (S66G and S72G, respectively). Although Gly and Ser are both very small amino acids, Gly is more hydrophobic. In fact, many of the changes in the plant NF-YB and NF-YC proteins result in changes in hydrophathy (Yang et al., 2005). It would be interesting to see if reversions to the apparently ancestral Ser residue would eliminate the rescue of flowering time defects in either the *nf-yb2* or *nf-yb3* mutant. Furthermore, it is possible that this change is necessary for the CO/NF-Y interaction and might prevent formation of the normal NF-Y heterotrimer

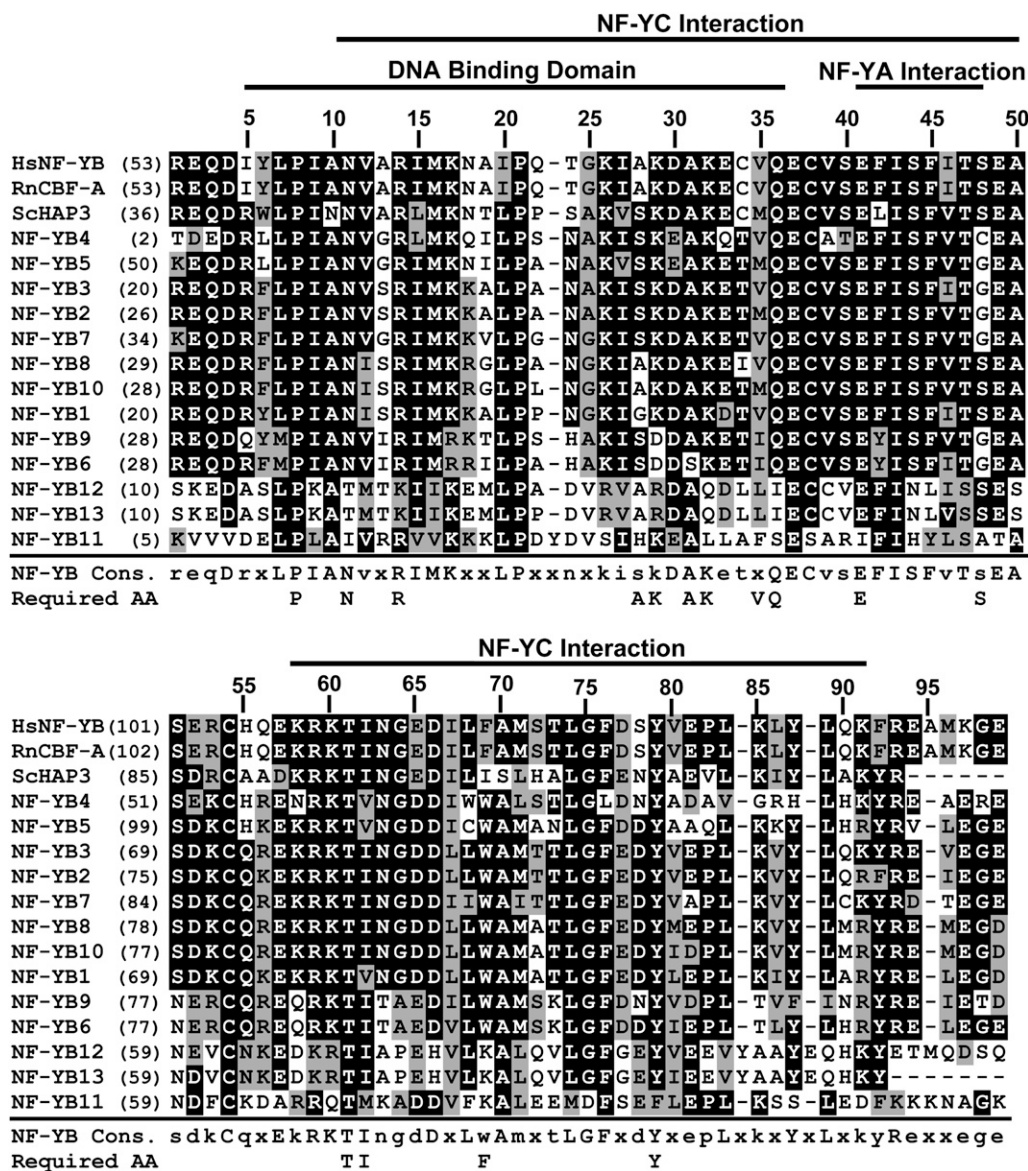


Figure 3. Arabidopsis NF-YB family alignment, constructed as in Figure 2. Note that the NF-YC interaction domain extends across two independent regions and partly overlaps with the DNA-binding and NF-YA interaction domains. To eliminate a gap of nonhomology, the amino acid sequence SECS was removed from NF-YB11 between the L and P residues at position 20/21. Required AA, Required amino acid.

previously described set of 13,010 sequences corresponding to the same region relative to the TSS (FitzGerald et al., 2004). Computer-based scanning of each human sequence for PWM matches ($\geq 90\%$ match threshold) were consistent with previous publications (i.e. there is a strong peak of putative CCAAT sites at approximately -100 from the TSS; Fig. 5A). In contrast, Arabidopsis promoters are largely devoid of this strong enrichment of putative CCAAT sites. Although the PWM does predict a small number of CCAAT sites in Arabidopsis (approximately one of 20 promoters examined), the human peak is almost eight times higher. These data suggest that NF-Y-binding sites

have significantly evolved from those of yeast and mammals.

Because there is currently only one confirmed Arabidopsis NF-Y binding site (Kusnetsov et al., 1999), we were not able to test if a specialized plant PWM could find more putative CCAAT sites. Furthermore, because of the many possible plant NF-Y complexes, it might be more likely that plant NF-Y binding sites have coevolved with specific complexes. Nevertheless, we could assess the frequency of CCAAT-only sequences in Arabidopsis and humans (i.e. no flanking sequences; Fig. 5B). Interestingly, when flanking sequences are removed from the search, Arabidopsis has

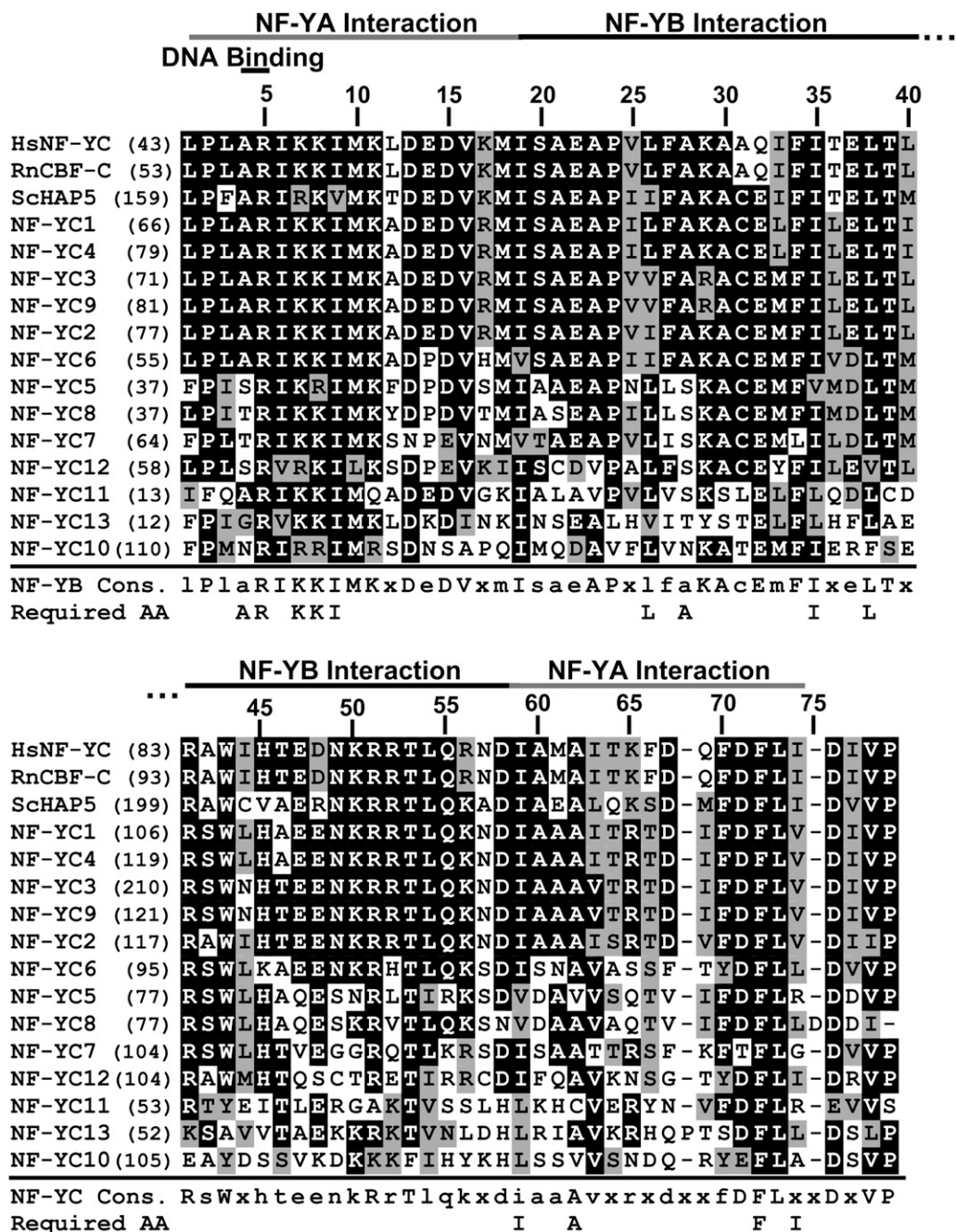


Figure 4. Arabidopsis NF-YC family alignment, constructed as in Figure 2. Note that the NF-YA interaction domain extends across two separate regions. The DNA-binding domain in NF-YC consists of the two amino acids AR (found in most NF-YC homologs). To eliminate two large gaps of nonhomology, the amino acid sequence DTLTRS was removed from NF-YC7 between the S and D residues at position 57/58 and YVNFQK was removed from NF-YC12 between the paired I residues at 18/19. Required AA, Required amino acid.

more CCAAT/ATTGG sequences than humans. Additionally, both Arabidopsis and humans have peaks in the expected -50 to -100 window. Thus, the core pentamer is quite common in Arabidopsis, and there is currently no reason to believe that Arabidopsis has fewer CCAAT sites than humans.

Promoter:GUS Fusions

As with other multigene families, researchers interested in Arabidopsis NF-Y genes must contend with overlapping functionality. Based on our own experience, loss-of-function mutations in NF-Y rarely have any obvious phenotypes. From the lack of complete

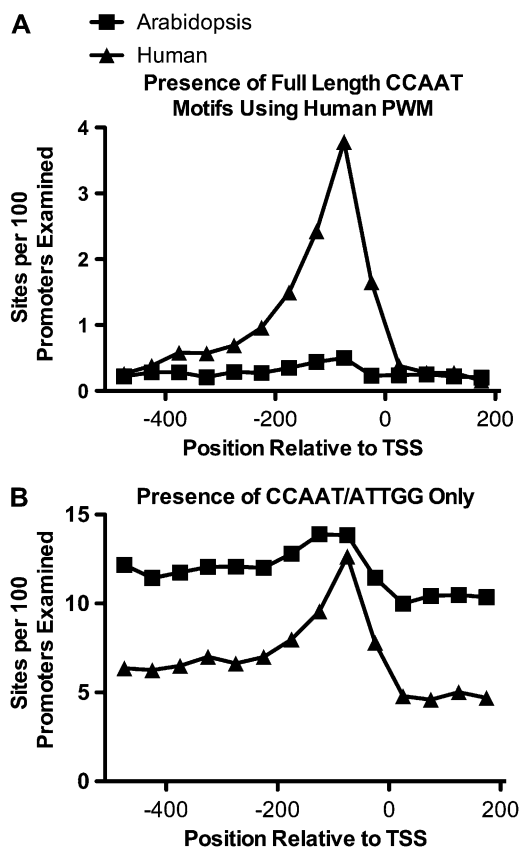


Figure 5. Promoter searches for CCAAT motifs in Arabidopsis and humans. A, Arabidopsis and human promoter data sets were searched for sequence matches to a PWM based on a large human and yeast data set of experimentally defined CCAAT sites. To control for differences in the size of each data set, numbers of positive matches (90% threshold for positive match) are presented as sites per 100 promoters examined. B, The same data sets were searched for the simple presence of the sequence CCAAT or its reverse complement ATTGG. Note that Arabidopsis is the top line in B.

plant NF-Y complexes described in the literature, we infer that other groups have also confronted this problem. We reasoned that assembling complete complexes would be greatly facilitated if reporter gene fusions were available for all 36 NF-Y genes. Accordingly, for each gene we cloned 1,000 bp of upstream sequence (5' untranslated region and promoter region) in front of an enhanced GFP (*eGFP*):GUS reporter gene fusion (Karimi et al., 2002). For each gene, we collected at least two independent, stable (third-generation) transformants and analyzed these in numerous plant tissues throughout normal development. We expect the data available from these lines to greatly facilitate the discovery of complete, in planta NF-Y complexes.

Root Expression Patterns

As a first qualitative measure of the reporter lines, we compared the GUS expression patterns in roots with a high-quality data set based on fluorescence-

activated cell sorting (FACS; Birnbaum et al., 2005) and numerous standardized microarray data sets available for online visualization at Genevestigator and The Botany Array Resource (Schmid et al., 2005; Toufighi et al., 2005; Grennan, 2006). Figure 6 shows the promoter:GUS expression patterns in the root tip for all 36 genes. Although precise quantitative comparisons are difficult, we generally found that GUS expression levels and patterns were consistent with expectations from FACS and online data. For example, GUS and FACS data sets agree that *NF-YA2* is primarily expressed in the maturation zone of the root (Fig. 6). It is important to note that all GUS pictures are static representations of dynamic expression patterns. For example, *NF-YB2* is very weakly expressed in the meristematic region of elongated roots (Fig. 6). Nevertheless, in young roots that are just emerging from the pericycle, *NF-YB2* expression is very strong throughout the root tip (data not shown).

We found only one obvious problem for the root GUS expression patterns: *NF-YB10* had essentially no GUS expression. Nevertheless, FACS, Genevestigator, and our own reverse transcription (RT)-PCR results show that it is strongly expressed in the roots. Additionally, *NF-YA5* did not express anywhere in our experiments, but another publication shows broad expression in the aboveground plant that is highly ABA inducible (Li et al., 2008). Our construct was not ABA inducible (data not shown). This suggests that ABA induction of *NF-YA5* relies on promoter elements beyond 1 kb upstream of the start codon. Overall, for 34 of 36 NF-Y genes, we find that our GUS patterns are largely consistent with published expression patterns. Our *NF-YA5* and *NF-YB10* results are included in all figures but are boxed in red wherever they clearly differ from the published record.

Dark- Versus Light-Grown Expression Patterns

To further demonstrate the usefulness of the promoter:GUS fusions, we compared our results with a previous publication examining NF-Y functions in blue light perception and ABA signaling (Warpeha et al., 2007). RT-PCR was used by the authors to identify NF-Y expression in 6-d-old dark-grown seedlings. For the NF-YC genes, the authors examined NF-YC1 through NF-YC9 and reported that NF-YC1, -YC4, and -YC9 were expressed. We confirmed this finding and added NF-YC3 to the list (Fig. 7). It is not surprising to find that NF-YC3 is expressed in a similar fashion to NF-YC1, -YC4, and -YC9, as all four genes are very closely related and likely arose from recent duplication events (Fig. 1). Additionally, we also tested NF-YC10 and the newly identified NF-YC11 to -YC13. Demonstrating that NF-YC expression patterns in dark-grown seedlings are more complicated than previously realized, we also found that NF-YC10, -YC11, and -YC12 are strongly expressed in the dark. There were no significant GUS expression differences between light- and dark-grown seedlings (Fig. 7).

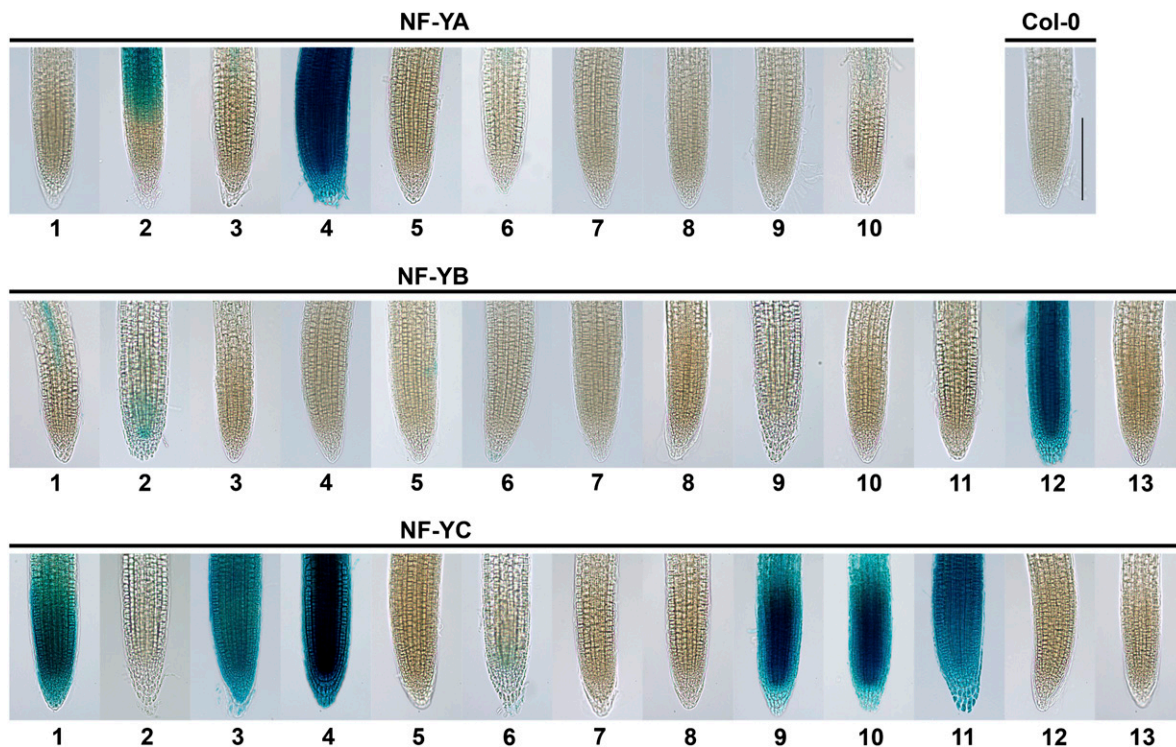


Figure 6. *NF-Y* expression patterns in 6-d-old *Arabidopsis* root tips. Note that none of the roots continue to have broad, whole root expression above the approximate level where these pictures were cut off (i.e. mature roots do not generally have *NF-Y* expression in the cortical and epidermal layers of the root). Many *NF-Y* genes have strong expression in the root stele; these are generally the same genes that are expressed in the leaf vasculature of rosettes (Fig. 8). All panels are shown at the same magnification. Bar in Col-0 panel = 200 μm .

In addition to the *NF-YC* data, it was previously reported that only *NF-YA5*, *NF-YB6*, and *NF-YB9* are expressed in 6-d-old dark-grown seedlings (Warpeha et al., 2007). To further validate our promoter:*GUS* fusions, we additionally examined the *NF-YA* and *NF-YB* expression patterns (data not shown). Our findings were inconsistent with the previous report in that we saw clear expression for all 10 *NF-YA* genes (*NF-YA5* was weak) as well as *NF-YB2*, *-YB3*, *-YB5*, *-YB6*, *-YB8*, *-YB9*, and *-YB12*. In our experiments, *NF-YB6* (*L1L*) and *NF-YB9* (*LEC1*) were among the weakest of the *GUS*-positive *NF-YB* lines: *GUS* staining for *NF-YB9* lines required extensive incubation (2 d) in 5-bromo-4-chloro-3-indoxyl- β -*D*-glucuronide solution. This finding is consistent with online microarray data and the known, narrow roles for *LEC1* and *L1L* in embryogenesis, but it does not rule out a role in blue light perception (Lotan et al., 1998; Kwong et al., 2003; Lee et al., 2003; Warpeha et al., 2007). We additionally confirmed our *GUS* results by examining the presence or absence of *NF-Y* gene expression by RT-PCR (data not shown). Our RT-PCR data were consistent with the *GUS* data presented in Figure 7. Additionally, we generated a comparative table for our *GUS* expression data versus publicly available microarray data for the dark-grown seedlings, root tips, rosettes, and flowers (Figs. 6–9; Supplemental Table S1).

Rosette Expression Patterns

We examined the rosette expression patterns of all 36 *NF-Y* genes (Fig. 8). In general, the *NF-Y* expression patterns are spatially complex with highly variable levels of intensity. For example, numerous genes have trichome expression. In some instances, this is part of a larger staining pattern (e.g. *NF-YA7*, *NF-YB2*, and *NF-YC3*), while in others, the staining is much more specific to the trichome (e.g. *NF-YA1* and *NF-YB12*). Several genes from each family have clear vascular expression patterns. These genes are particularly interesting because CO function in floral induction is a phloem-specific process (An et al., 2004; Ayre and Turgeon, 2004). Furthermore, it is now well established that FT accumulates in the phloem and then translocates to the shoot meristem (Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007; Tamaki et al., 2007). If, as current data suggest, *NF-Y* proteins form an interaction platform for CO at the *FT* promoter, this process would also be expected to take place in the leaf phloem.

Strongly supporting this theory, *NF-YB2* and *NF-YB3* are strongly expressed in the vasculature (Fig. 8), and they are known to redundantly control photoperiod-regulated flowering time (Kumimoto et al., 2008). *NF-YB2* and *NF-YB3* are very closely related proteins

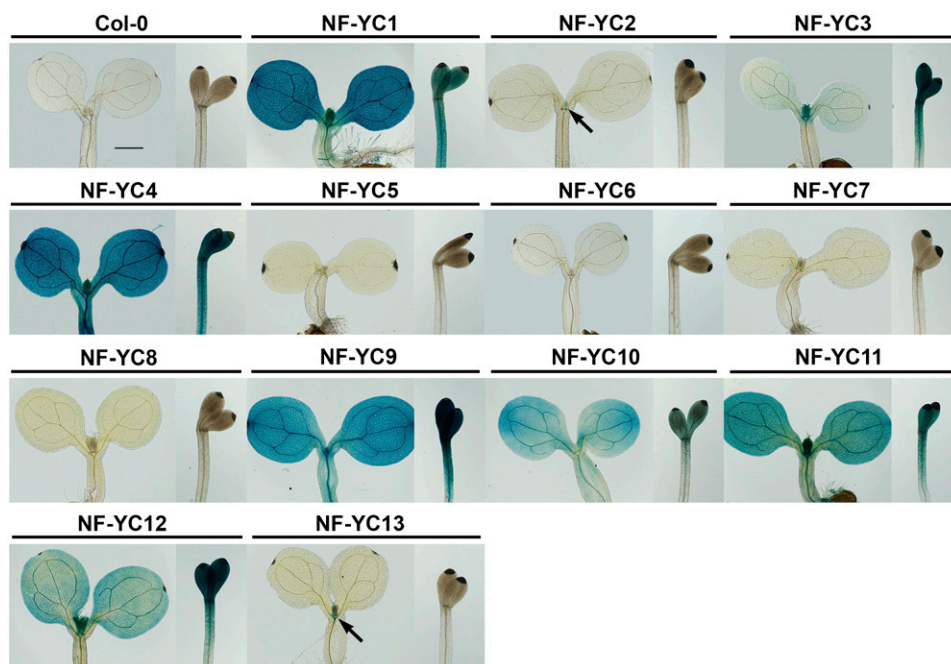


Figure 7. *NF-YC* expression patterns in light- versus dark-grown plants. Pairs of light- and dark-grown seedlings are shown. To facilitate direct comparison, plants were grown as described previously (Warpeha et al., 2007). Dark-grown plants were infiltrated and stained in the dark to prevent possible light activation of promoter:GUS fusions. Bar in Col-0 panel = 500 μ m.

sharing 94% amino acid identity in their conserved domains (Figs. 1 and 3). *NF-YB8* and *NF-YB12* also have vascular expression patterns and might be expected to have a role in flowering. Nevertheless, the vascular expression of *NF-YB8* is much weaker than that of either *NF-YB2* or *-YB3* and does not appear to extend into the second set of true leaves. *NF-YB12* is phylogenetically distant from both *NFYB2* and *-YB3*, and *nf-yb2 nf-yb12* double mutants show no additive delay in flowering time over the *nf-yb2* single mutant (B.F. Holt, R.W. Kumimoto, and N. Siefers, unpublished data). Therefore, especially considering their *co* mutant phenocopy (Kumimoto et al., 2008), *nf-yb2 nf-yb3* are likely the major *NF-YB* components of the photoperiod-regulated flowering response. The next important challenge is to describe the entire *NF-Y* complex involved in flowering.

NF-YC1, *-YC3*, *-YC4*, *-YC9*, *-YC11*, and *-YC12* are all vascular expressed and, therefore, are potential targets for flowering time control. Beyond the first set of true leaves, *NF-YC1* and *NF-YC11* expression does not generally extend into the vasculature. Of the remaining four genes, *NF-YC3*, *-YC4*, and *-YC9* are all very strongly expressed in the vasculature and are very closely related; in fact, *NF-YC3* and *YC9* are identical across their 78-amino acid conserved regions. Unlike single mutations in *NF-YB2* and *-YB3* (Cai et al., 2007; Chen et al., 2007; Kumimoto et al., 2008), we did not measure any flowering delays for *nf-yc1*, *nf-yc3*, *nf-yc4*, *nf-yc9*, or *nf-yc11* (B.F. Holt, R.W. Kumimoto, and N. Siefers, unpublished data). Although this does not formally exclude involvement for any of the other *NF-YC* genes, we suspect that overlapping functionality will be important here as well. Assuming that they are involved, the *NF-YA* genes may be the most complicated: most of them have some level

of vasculature expression, and the overall family conservation is much higher than for the *NF-YB* and *NF-YC* families (Figs. 1, 2, and 8). Interestingly, plants expressing 35S:miR169, which is predicted to target all *NF-YA* mRNAs except *NF-YA4*, *-YA6*, and *-YA7* (Jones-Rhoades and Bartel, 2004), do not show a flowering-time phenotype (Li et al., 2008; W.-X. Li and J.-K. Zhu, personal communication).

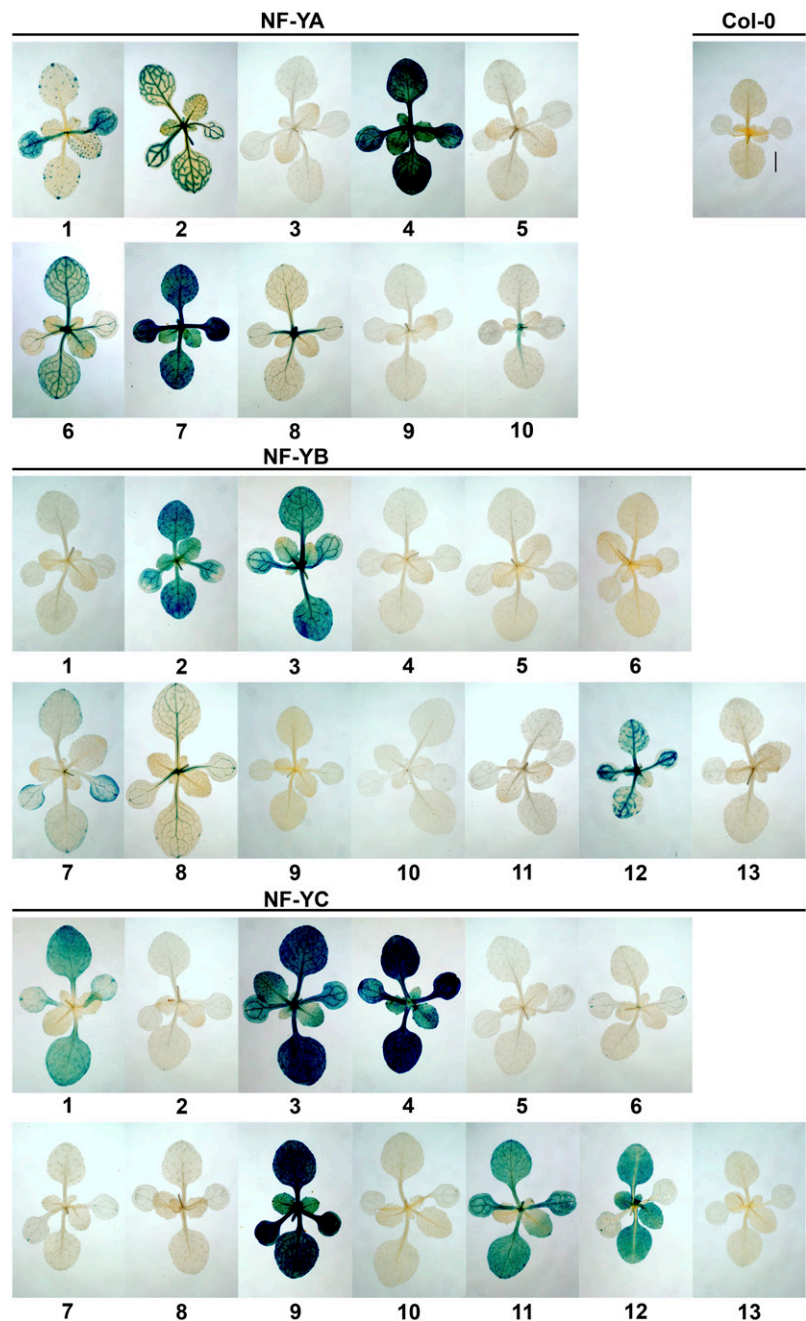
Flower Expression Patterns

As in the rosettes, floral *NF-Y* expression patterns are quite variable and complex. Some of the genes with ubiquitous expression in rosettes are much more restricted in the floral organs. For example, *NF-YB3* is widely expressed in the rosette but is restricted to the filaments in flowers. Alternatively, genes such as *NF-YB7* are minimally expressed in the rosettes and ubiquitously expressed in the flowers. The potential for using these tissue-specific patterns to infer likely *NF-Y* complexes is simply illustrated by the stigmas; in the pictured developmental stage, only *NF-YA7*, *NF-YB7*, *NF-YB12*, *NF-YC3*, and *NF-YC12* have strong stigma expression. To our knowledge, no phenotypes related to floral organs have been reported for the *NF-Y* genes. This small set may be a good place to start.

DISCUSSION

There have been numerous *NF-Y* duplications unique to the plant lineage. However, compared with yeast and mammals, much less functional information is currently available. This is likely due to overlapping functionality between *NF-Y* subunits re-

Figure 8. *NF-Y* expression patterns in 10-d-old rosettes. Plants were grown on soil in a standard LD light regime. Bar in Col-0 panel = 2 mm.



sulting in very few plant *NF-Y* genes isolated in forward genetic mutant screens. Additionally, when plant *NF-Y* proteins have been associated with specific functions, identifying their interacting partners has been complicated by the numbers of possible *NF-YA/B/C* combinations. Although their existence is virtually certain, no single complete *NF-Y* complex has ever been described in plants.

To simplify future analyses and aid in the discovery of complete complexes, we have created a set of publicly available expression lines. The gross morphological expression patterns appear to be accurate for 34 of 36 lines. Arabidopsis *NF-Y* genes are expressed in

almost all plant tissues we examined, and this was typically true for at least one of each subunit type per tissue examined (Figs. 6–9). The epidermis and cortex of mature roots were the only tissues largely devoid of *NF-Y* expression. The tissue- and development-specific expression patterns presented here will simplify the process of identifying complete *NF-Y* complexes.

It is interesting that plants utilize so many *NF-Y* genes while other complicated organisms need only single copies of each subunit. This expansion is also true for other Arabidopsis transcription factors, such as the Myb, Myc, and MADS proteins (Riechmann et al., 2000; Riechmann and Ratcliffe, 2000). Evolution-

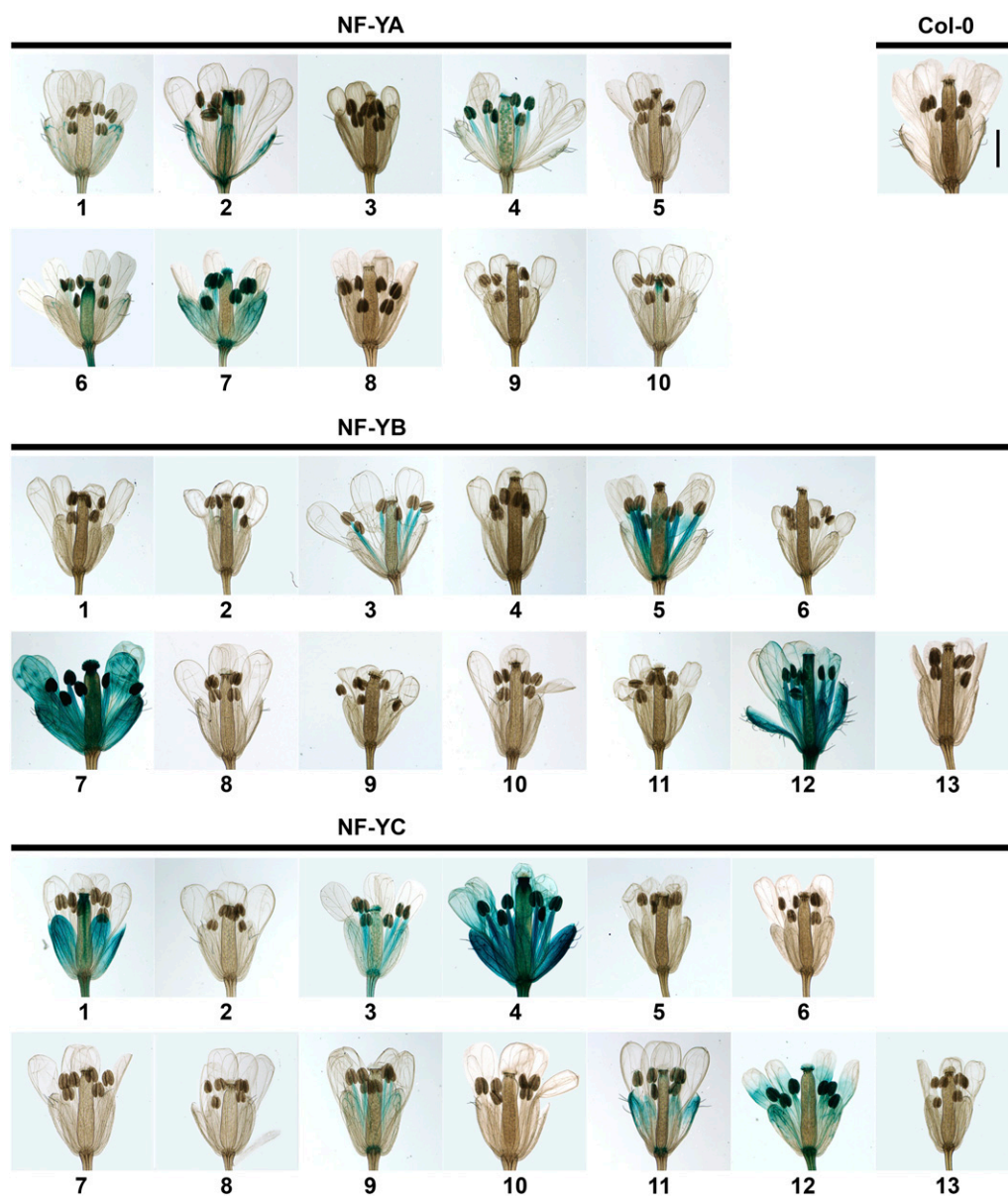


Figure 9. NF-Y expression patterns in flowers. Flowers were harvested from plants at approximate “principal growth stage” 6.50 (Boyes et al., 2001; Kjemtrup et al., 2003). Bar in Col-0 panel = 2 mm.

ary biologists have long postulated that infrequent gene duplication events typically result in the loss or inactivation of one of the duplicated copies (Lynch and Force, 2000; Lynch et al., 2001; Meyer, 2003; Ward and Durrett, 2004; Moore and Purugganan, 2005). However, whole genome sequencing has demonstrated that gene duplications take place more commonly than expected and are often maintained indefinitely in the apparent absence of functional diversification (Meyer, 2003). For many of the *NF-Y* genes, especially *NF-YA*, the high degree of conservation strongly suggests the maintenance of ancestral functions. Perhaps plant NF-Y proteins have maintained their core ancestral DNA-binding and complex-forming functions

while refining the ways in which they interact with specific cis- and trans-elements.

Our searches for CCAAT motifs using a yeast/mammalian PWM suggest divergence of the cis-elements bound by plant NF-Y (Fig. 5). Alternatively, there may simply be fewer NF-Y-regulated genes in plants. Because of the amino acid conservation across lineages, we favor the idea that plant NF-Y complexes still bind sequences with a central CCAAT motif but the surrounding bases of their cognate cis-elements have evolved from those of other lineages. Obviously, there is a great need for direct, *in vivo* measurements of NF-Y/DNA interactions if we hope to understand how these proteins have uniquely evolved in the plant lineage.

The addition of novel, plant-specific trans-interacting factors to NF-Y complexes may also greatly modify their functionality. For example, there is at least one known example of the NF-YB/YC dimer being coopted by another transcription factor (OsMADS18; Masiero et al., 2002). Additionally, there is evidence that CO and CO-Like proteins can physically associate with specific NF-Y subunits (Ben-Naim et al., 2006; Wenkel et al., 2006), but it remains to be seen if these complexes can bind CCAAT motifs *in vivo*.

One exciting prospect is that many of the CCT domain proteins might require NF-Y complexes to exert their regulatory effect on specific promoters. For example, CO might exert its regulatory effects on *FT* expression by competing for the position of NF-YA occupancy in NF-Y complexes (Wenkel et al., 2006). The coincidence of light (LD conditions) and peak CO expression might allow CO to outcompete NF-YA for utilization of specific NF-YB/C dimers. In this formulation, specific NF-YA proteins might act as negative regulators of CO-mediated flowering. As predicted by this replacement model, overexpression of several NF-YA subunits delayed flowering (Wenkel et al., 2006). Additionally, as a variation on the replacement model, there may be no need for NF-YA in a CO/NF-YB/YC complex. To date, no supporting loss-of-function data have been published regarding NF-YA subunits in flowering.

While there are clear similarities between NF-YA and CCT-containing proteins, there are important unanswered questions with a replacement model. First, the region of similarity between the two domains is almost exclusively within the NF-YA DNA-binding domain. There is essentially no similarity between CCT proteins and the NF-YA subunit interaction domain (Fig. 2). Thus, the alignments do not support a direct equivalency between NF-YA and CCT proteins for interactions with NF-YB/YC dimers. Additionally, while the DNA-binding domains of NF-YA proteins share clear similarities with CCT domains, none of the required His residues are shared. His residues are absolutely essential for NF-Y complex binding to CCAAT motifs (Xing et al., 1993). Therefore, assuming that the replacement model is correct, one would not expect the resulting CO/NF-YB/YC complexes to bind at CCAAT motifs. An alternative model for CO/NF-Y interactions places CO docking on preassembled, mature NF-Y/DNA complexes but not actually displacing the NF-YA subunit. If this docking model is correct, we would expect appropriate NF-YA loss-of-function alleles to result in flowering delays. Although this is just one example of NF-Y interactions, it may prove paradigmatic for how NF-Y complexes uniquely function and fine-tune plant gene expression.

After numerous rounds of duplication, plant NF-Y proteins have likely evolved numerous unique cis- and trans-interactions and have clearly become much more highly regulated in their tissue- and development-specific expression patterns. This in turn suggests a

refinement and narrowing in their gene targets. We expect that this refinement will be in marked contrast to the more broad and universal transcriptional activation potential expected in animal systems. As there are now several very interesting NF-Y-associated developmental and stress-responsive processes, we expect that our collective understanding of NF-Y complexes and their plant-specific functions will expand rapidly in the next few years.

MATERIALS AND METHODS

Plant Growth Conditions

All *Arabidopsis* (*Arabidopsis thaliana*) plants used are in the Columbia (Col-0) ecotype. Plants for rosette and flower expression patterns (Figs. 8 and 9) were grown at 23°C in a standard LD light regime (16 h of light/8 h of dark). Plants were grown in medium containing equal parts Farfard C2 Mix and Metromix 200 supplemented with 40 g of Marathon pesticide and dilute Peters fertilizer (NPK = 20:20:20). Plants were watered throughout with dilute fertilizer (approximately one-tenth the recommended regular feeding levels). Root and light/dark experimental plants (Figs. 6 and 7) were grown on sterile plates. To allow appropriate comparisons, root and light/dark plants were grown exactly as described by Birnbaum et al. (2003) and Warpeha et al. (2007), respectively.

Phylogenies and Alignments

Individual subunits were identified by standard BLAST searches at TAIR and the National Center for Biotechnology Information (Altschul et al., 1990). Full-length protein sequences were imported into MEGA 4, where phylogenies and alignments were created (Tamura et al., 2007). Multiple sequence alignments were done by ClustalX (Thompson et al., 2002). Alignments and phylogenies were created from both full-length protein sequences and truncated sequences identical to those shown in Figures 2 to 4. All phylogenetic trees presented in this article were derived from the truncated alignments of highly homologous regions. Neighbor-joining and bootstrap methods were employed as described previously (Hall, 2008). Figures 2 to 4 were created with BOXSHADE (created by K. Hofmann and M. Baron) within the Mobyle Web portal (<http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py>).

CCAAT Searches

Using the Perl TFBS module (Lenhard and Wasserman, 2002), we generated a 16-base PWM from the nucleotide frequencies in 178 confirmed yeast and mammalian NF-Y binding sites (Mantovani, 1998). We searched the human and *Arabidopsis* data sets with this PWM, counting only those sequences that achieved a match score of 90% or greater of the maximum possible score. Match scores were calculated as a sum across positions of the weights for the observed nucleotide at that position. *Arabidopsis* promoter sequences were assembled as described in the text by TAIR data curators Aleksey Kleitman and Leonore Reiser.

Promoter:GUS Fusions

To clone each NF-Y promoter region, primers were designed with partial B1 and B2 sites (according to standard Gateway protocols; Invitrogen) and gene-specific sequence. These PCRs were then used in a second PCR step containing full-length B1 and B2 primers. The resulting PCR products were then cloned into pDONR207 (Invitrogen). After confirming the correct sequence for each promoter, each was transferred by LR recombination reaction to the eGFP/GUS fusion-containing binary vector pGWFS7 (Karimi et al., 2002). For each promoter, the sequence represented 1,000 bp starting at 2 bp downstream of the ATG start codon and extending 995 bp upstream of the start codon. The 2 bp downstream of the ATG placed the gene-specific ATG in frame with the eGFP/GUS start codon. We chose this conservative strategy to avoid possible aberrant expression patterns arising from the gene-specific

start codon being out of frame with the reporter gene. After transferring each construct into plants by standard *Agrobacterium tumefaciens*-based methods (Bechtold et al., 1993), we confirmed the correct transgene in each line by PCR amplification from plant DNA and restriction digestion. All pictures show stable, third-generation plants with single T-DNA insertions. At least two lines were examined per promoter:GUS fusion.

Microscopy

Macrophotography was used to visualize the Arabidopsis rosettes in Figure 8. For this, an Olympus DP71 CCD camera was fitted with a Pentax KC 50-mm Adapter and mounted above the subject with a Nikon MKII optic light providing illumination from above. SPOT software (version 4.6) was used to record the pictures. For Figures 6, 7, and 9, we used an Olympus BX41 microscope with an Insight 2 Megapixel Color Mosaic CCD camera. We used SPOT software (version 4.6) to record the pictures.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers 831124, 819738, 843614, 818037, 841856, 820616, 839929, 838335, 821640, 830539, 818472, 834815, 827101, 837424, 819393, 834818, 815843, 818282, 838800, 824502, 817292, 830715, 832373, 824019, 842070, 841922, 836466, 835117, 835116, 835115, 832857, 837417, 837313, 820427, 833794, and 834343.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Phylogenetic tree for the CO-Like proteins.

Supplemental Figure S2. Full alignment of CCT domains from CO and COL1-5 with the Arabidopsis, human (Hs), rat (Rn), and yeast (Sc) NF-Y.

Supplemental Table S1. GUS staining versus microarray scores.

ACKNOWLEDGMENTS

This research originated in Jeff Dangl's laboratory at the University of North Carolina in Chapel Hill; we owe him a great debt of gratitude. We also thank Aleksey Kleitman and Leonore Reiser at TAIR for assembling the Arabidopsis promoter data set and Peter FitzGerald at the National Cancer Institute for supplying the human promoters data set. Microscopy equipment and expertise were provided by Phil Gibson, Scott Russell, and Greg Strout at the University of Oklahoma. We thank John Skvarla for critical reading of the manuscript.

Received September 30, 2008; accepted November 12, 2008; published November 19, 2008.

LITERATURE CITED

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410
- An H, Roussot C, Suarez-Lopez P, Corbesier L, Vincent C, Pineiro M, Hepworth S, Mouradov A, Justin S, Turnbull C, Coupland G (2004) CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. *Development* **131**: 3615–3626
- Ayre BG, Turgeon R (2004) Graft transmission of a floral stimulant derived from CONSTANS. *Plant Physiol* **135**: 2271–2278
- Bechtold N, Ellis J, Pelletier G (1993) In planta *Agrobacterium*-mediated gene transfer by infiltration of *Arabidopsis thaliana* plants. *C R Acad Sci Ser III Sci Vie* **316**: 1194–1199
- Ben-Naim O, Eshed R, Parnis A, Teper-Bamnolker P, Shalit A, Coupland G, Samach A, Lifschitz E (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J* **46**: 462–476
- Bhattacharya A, Deng JM, Zhang Z, Behringer R, de Crombrugge B, Maity SN (2003) The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. *Cancer Res* **63**: 8167–8172
- Birnbaum K, Jung JW, Wang JY, Lambert GM, Hirst JA, Galbraith DW, Benfey PN (2005) Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat Methods* **2**: 615–619
- Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN (2003) A gene expression map of the Arabidopsis root. *Science* **302**: 1956–1960
- Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorchach J (2001) Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* **13**: 1499–1510
- 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, 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 MC, 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 AM, Pavesi G, Mantovani R (2008) The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol* **28**: 2047–2058
- Chen NZ, Zhang XQ, Wei PC, Chen QJ, Ren F, Chen J, Wang XC (2007) AtHAP3b plays a crucial role in the regulation of flowering time in Arabidopsis during osmotic stress. *J Biochem Mol Biol* **40**: 1083–1089
- Combiér JP, de Billy F, Gamas P, Niebel A, Rivas S (2008) Trans-regulation of the expression of the transcription factor MtHAP2-1 by a uORF controls root nodule development. *Genes Dev* **22**: 1549–1559
- Combiér JP, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernie T, Ott T, Gamas P, Crespi M, 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
- Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle J, Giakountis A, Farrona S, Gissot L, Turnbull C, et al (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* **316**: 1030–1033
- Coustry F, Hu Q, de Crombrugge B, Maity SN (2001) CBF/NF-Y functions both in nucleosomal disruption and transcription activation of the chromatin-assembled topoisomerase IIalpha promoter: transcription activation by CBF/NF-Y in chromatin is dependent on the promoter structure. *J Biol Chem* **276**: 40621–40630
- Coustry F, Maity SN, Sinha S, de Crombrugge 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
- di Silvio A, Imbriano C, 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, Mathis D (1987) A multiplicity of CCAAT box-binding proteins. *Cell* **50**: 863–872
- Edwards D, Murray JA, Smith AG (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. *Plant Physiol* **117**: 1015–1022
- FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C (2004) Clustering of DNA sequences in human promoters. *Genome Res* **14**: 1562–1574
- Frontini M, Imbriano C, diSilvio A, Bell B, Bogni A, Romier C, Moras D, Tora L, Davidson I, Mantovani R (2002) NF-Y recruitment of TFIID, multiple interactions with histone fold TAF(II)s. *J Biol Chem* **277**: 5841–5848
- Frontini M, Imbriano C, Manni I, Mantovani R (2004) Cell cycle regulation of NF-YC nuclear localization. *Cell Cycle* **3**: 217–222
- Grennan AK (2006) Genevestigator: facilitating Web-based gene-expression analysis. *Plant Physiol* **141**: 1164–1166
- Gusmaroli G, Tonelli C, Mantovani R (2001) Regulation of the CCAAT-binding NF-Y subunits in Arabidopsis thaliana. *Gene* **264**: 173–185
- Gusmaroli G, Tonelli C, Mantovani R (2002) Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. *Gene* **283**: 41–48
- Hall BG (2008) *Phylogenetic Trees Made Easy: A How-to Manual*, Ed 3. Sinauer Associates, Sunderland, MA
- Hu Q, Maity SN (2000) Stable expression of a dominant negative mutant of

- CCAAT binding factor/NF-Y in mouse fibroblast cells resulting in retardation of cell growth and inhibition of transcription of various cellular genes. *J Biol Chem* **275**: 4435–4444
- Izawa T** (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *J Exp Bot* **58**: 3091–3097
- Jaeger KE, Wigge PA** (2007) FT protein acts as a long-range signal in *Arabidopsis*. *Curr Biol* **17**: 1050–1054
- Jones-Rhoades MW, Bartel DP** (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* **14**: 787–799
- Kahle J, Baake M, Doenecke D, Albig W** (2005) Subunits of the heterotrimeric transcription factor NF-Y are imported into the nucleus by distinct pathways involving importin beta and importin 13. *Mol Cell Biol* **25**: 5339–5354
- Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D** (1999) Activation tagging of the floral inducer FT. *Science* **286**: 1962–1965
- Karimi M, Inze D, Depicker A** (2002) Gateway vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci* **7**: 193–195
- Kaufmann K, Melzer R, Theissen G** (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347**: 183–198
- Kim IS, Sinha S, de Crombrughe B, Maity SN** (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
- Kjemtrup S, Boyes DC, Christensen C, McCaskill AJ, Hylton M, Davis K** (2003) Growth stage-based phenotypic profiling of plants. *Methods Mol Biol* **236**: 427–442
- Koornneef M, Hanhart CJ, van der Veen JH** (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol Gen Genet* **229**: 57–66
- Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, Marion CM, Hempel FD, Ratcliffe OJ** (2008) The nuclear factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in *Arabidopsis*. *Planta* **228**: 709–723
- Kusnetsov V, Landsberger M, Meurer J, Oelmuller R** (1999) The assembly of the CCAAT-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 RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ** (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* **15**: 5–18
- Lee H, Fischer RL, Goldberg RB, Harada JJ** (2003) *Arabidopsis* LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci USA* **100**: 2152–2156
- Lenhard B, Wasserman WW** (2002) TFBS: computational framework for transcription factor binding site analysis. *Bioinformatics* **18**: 1135–1136
- Li WX, Oono Y, Zhu J, He XJ, Wu JM, Iida K, Lu XY, Cui X, Jin H, et al** (2008) The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell* **20**: 2238–2251
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ** (1998) *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* **93**: 1195–1205
- Lynch M, Force A** (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473
- Lynch M, O'Hely M, Walsh B, Force A** (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**: 1789–1804
- Maity SN, 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 SN, de Crombrughe B** (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem Sci* **23**: 174–178
- Maity SN, Sinha S, Ruteshouser EC, 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
- 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
- Mantovani R, Li XY, Pessara U, Hoof van Huisduijnen R, Benoist C, Mathis D** (1994) Dominant negative analogs of NF-YA. *J Biol Chem* **269**: 20340–20346
- Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM** (2002) Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. *J Biol Chem* **277**: 26429–26435
- Mathieu J, Warthmann N, Kuttner F, Schmid M** (2007) Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*. *Curr Biol* **17**: 1055–1060
- Matuoka K, Chen KY** (2002) Transcriptional regulation of cellular ageing by the CCAAT box-binding factor CBF/NF-Y. *Ageing Res Rev* **1**: 639–651
- McNabb DS, Tseng KA, Guarente L** (1997) The *Saccharomyces cerevisiae* Hap5p homolog from fission yeast reveals two conserved domains that are essential for assembly of heterotetrameric CCAAT-binding factor. *Mol Cell Biol* **17**: 7008–7018
- McNabb DS, Xing Y, Guarente L** (1995) Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev* **9**: 47–58
- Messinguy E, Dubois E** (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* **316**: 1–21
- Meyer A** (2003) Molecular evolution: duplication, duplication. *Nature* **421**: 31–32
- Moore RC, Purugganan MD** (2005) The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol* **8**: 122–128
- Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, et al** (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 USA* **104**: 16450–16455
- Olesen JT, 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
- Peng Y, 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, 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
- Putterill J, Robson F, Lee K, Simon R, Coupland G** (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**: 847–857
- Redei GP** (1962) Supervital mutants of *Arabidopsis*. *Genetics* **47**: 443–460
- Remenyi A, Scholer HR, Wilmanns M** (2004) Combinatorial control of gene expression. *Nat Struct Mol Biol* **11**: 812–815
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al** (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**: 2105–2110
- Riechmann JL, Ratcliffe OJ** (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol* **3**: 423–434
- Romero C, Cocchiarella F, Mantovani R, 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
- Ronchi A, Bellorini M, Mongelli M, Mantovani R** (1995) CCAAT-box binding protein NF-Y (CBF, CP1) recognizes the minor groove and distorts DNA. *Nucleic Acids Res* **23**: 4565–4572
- Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G** (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* **288**: 1613–1616
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU** (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**: 501–506
- Singh KB** (1998) Transcriptional regulation in plants: the importance of combinatorial control. *Plant Physiol* **118**: 1111–1120
- Sinha S, Kim IS, Sohn KY, de Crombrughe B, Maity SN** (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
- Sinha S, Maity SN, Lu J, 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 USA* **92**: 1624–1628
- Stephenson TJ, McIntyre CL, Collet C, Xue GP** (2007) Genome-wide identification and expression analysis of the NF-Y family of transcription factors in *Triticum aestivum*. *Plant Mol Biol* **65**: 77–92
- Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G** (2001) CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**: 1116–1120
- Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K** (2007) Hd3a protein is a mobile flowering signal in rice. *Science* **316**: 1033–1036
- Tamura K, Dudley J, Nei M, Kumar S** (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599
- Testa A, Donati G, Yan P, Romani F, Huang TH, Vigano MA, 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
- Thompson JD, Gibson TJ, Higgins DG** (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **August**: Chapter 2, Unit 2.3
- Toufighi K, Brady SM, Austin R, Ly E, Provart NJ** (2005) The Botany Array Resource: e-northern, expression angling, and promoter analyses. *Plant J* **43**: 153–163
- Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G** (2004) Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science* **303**: 1003–1006
- Ward R, Durrett R** (2004) Subfunctionalization: How often does it occur? How long does it take? *Theor Popul Biol* **66**: 93–100
- Warpeha KM, Upadhyay S, Yeh J, Adamiak J, Hawkins SI, Lapik YR, Anderson MB, Kaufman LS** (2007) The GCRI1, 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, 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
- West M, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ** (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. *Plant Cell* **6**: 1731–1745
- Wolberger C** (1998) Combinatorial transcription factors. *Curr Opin Genet Dev* **8**: 552–559
- Xing Y, Fikes JD, Guarente L** (1993) Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding domain. *EMBO J* **12**: 4647–4655
- Xing Y, Zhang S, Olesen JT, Rich A, Guarente L** (1994) Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short alpha-helix in HAP2. *Proc Natl Acad Sci USA* **91**: 3009–3013
- Yamaguchi T, Hirano HY** (2006) Function and diversification of MADS-box genes in rice. *ScientificWorldJournal* **6**: 1923–1932
- Yang J, Xie Z, Glover BJ** (2005) Asymmetric evolution of duplicate genes encoding the CCAAT-binding factor NF-Y in plant genomes. *New Phytol* **165**: 623–631