UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FUNCTIONAL ANALYSIS OF DUPLICATED NF-YB GENES IN PHOTOPERIOD-DEPENDENT FLOWERING

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

EMILY MILLS CARPENTER
Norman, Oklahoma
2018

FUNCTIONAL ANALYSIS OF DUPLICATED NF-YB GENES IN PHOTOPERIOD-DEPENDENT FLOWERING

A THESIS APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

_____
Dr. Ben F. Holt III, Chair

_____
Dr. Scott Russell

_____
Dr. John P. Masly

Acknowledgements

This research was supported by a National Science Foundation Grant (IOS-1149822) to Ben Holt.

Table of Contents

List of Figures

Abstract

The heterotrimeric transcription factor NUCLEAR FACTOR Y (NF-Y) regulates photoperiod-dependent flowering and is composed of NF-YA, NF-YB, and NF-YC subunits. In higher plants, such as Arabidopsis thaliana, whole genome duplication (WGD) leads to an evolutionary expansion of NF-Y subunits. NF-YB and NF-YC proteins have a highly conserved histone-like fold domain (HFD) and flanking variable amino (N) and carboxy (C) terminals. NF-YB proteins have known roles in the control of floral induction, with the *nf-yb2 nf-yb3* mutant displaying a late-flowering phenotype. To better understand their ability to regulate photoperiod-dependent flowering from a structure/function perspective, 10 NF-YB HFD were overexpressed in *nf-yb2 nf-yb3*. The HFD alone could rescue the late-flowering phenotype of *nf-yb2 nf-yb3*, except NF-YB4. While full-length *NF-YB1* cannot rescue late flowering, while Full-length *NF-YB2* rescues the *nf-yb2 nf-yb3* phenotype, similar to wild-type. Chimeric experiments between NF-YB1 and NF-YB2 show that *NF-YB2* N and C termini decrease the success of *NF-YB1* HFD to rescue *nf-yb2 nf-yb3*. Meaning the N and C termini have evolved to modulate the functionality of the NF-YB HFD, and cannot enhance *NF-YB1* HFD ability to promote flowering. My structure and function analysis of the NF-YB family displays a unique case of subfunctionalization called Escape from Adaptive Conflict (EAC), the ability of a duplicated gene to obtain a novel function while still maintaining its ancestral role. The ancestral role is designated in the HFD where most of the action for floral initiation

resides, however, the variable N and C termini can only modulate the primary functionality of their native NF-YB HFD.

Introduction

Transcription factors (TF) are proteins that regulate developmental processes by binding to specific DNA sequences and controlling the rate of converting DNA into messenger RNA (Latchman 1997). NUCLEAR FACTOR Y (NF-Y) is a transcriptional complex that regulates many different developmental processes. The NF-Y complex consists of three subunits (NF-YA, NF-YB, and NF-YC) that trimerize in a stepwise pattern into a mature transcriptional complex. NF-YB and NF-YC evolved from the core histones H2B and H2A, respectively, with conserved histone-like fold domains (HFDs), allowing protein-protein interactions and protein-DNA interactions (Baxevanis et al., 1995; Nardone et al., 2017). The HFD facilitates dimerization of NF-YB and NF-YC in the cytoplasm where they translocate to the nucleus and trimerize with NF-YA, and NF-YA binds to DNA with high affinity for *CCAAT* sequences (Sinha et al., 1995, 1996). Positive or negative transcriptional regulation of genes bound by NF-Y is a result of the mature complex binding to DNA at the highly conserved *CCAAT* box region (Ceribelli et al., 2008). The *CCAAT* box is highly conserved in eukaryotic promoters, typically located 80-100 base pairs upstream of the transcriptional start site (Bucher, 1990; Maity and De Crombrugghe, 1998; Mantovani, 1999).

Mammals and fungi usually have a single gene encoding each NF-Y subunit. However, angiosperms contain numerous genes encoding each subunit. Our

research model, *Arabidopsis thaliana*, contains 10 NF-YA, 10 NF-YB, and 10 NF-YC genes (Gusmaroli et al., 2001, 2002; Petroni et al., 2012). The large expansion of NF-Y family members in Arabidopsis is likely principally from Whole Genome Duplication events (WGD). WGD is a process in which plants can duplicate their entire genome, creating different copies of each gene. WGD is a fundamental evolutionary process in plants for adaptation, leading to the creation of paralogs, with each paralog having different selection restraints (Ohno 1970). Different selection restraints allow synonymous and non-synonymous mutations to affect duplicated genes a. Genes can become deleterious, change its function, change its expression patterns, or be maintained as a functionally redundant copy (Li et al., 2005; Ha et al., 2009; Siefers et al., 2009). The hypothesis is that WGD led to the expansion of the NF-Y families and allowed the evolution of roles in several known developmental- and stress-related processes. Including the endoplasmic reticulum (ER) stress response (Liu and Howell, 2010), embryogenesis (West et al., 1994; Lotan et al., 1998; Lee et al., 2003; Kwong et al., 2003), nodule development in legumes (Combier et al., 2006), drought responses (Nelson et al., 2007; Li et al., 2008), and the control of photoperiodic dependent flowering time (Ben-Naim et al., 2006; Wenkel et al., 2006; Cai et al., 2007; Chen et al., 2007; Kumimoto et al., 2008, 2010, Cao et al., 2011).

Photoperiod-dependent flowering is a response in plants that links the ability to initiate flowering under different day/night lengths. In *Arabidopsis* flowering activation is induced in long photoperiods, making it a long-day (LD) plant.

CONSTANS (CO) is a transcriptional regulator of photoperiod-dependent flowering time according to day length (Samach et at., 2000). CO day length perception is regulated by CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) which physically interacts and causes degradation of CO during the dark (Jang et al. 2008). LD conditions restrict COP1 degradation inducing the expression of *FLOWERING LOCUS T (FT)* resulting in activation of flowering (Valverde et al., 2004).

The evolution of many plant species has led to the initiation of flowering under specific photoperiods (Imaizumi and Kay, 2006; Kobayashi and Weigel, 2007). It has been shown by several groups that for photoperiod-dependent flowering to occur the NF-YB subunit is required. *The nf-yb2* and *nf-yb3* single mutant plants display a moderate delay in flowering (Cai et al., 2007; Chen et al., 2007), while the *nf-yb2 nf-yb3* double mutant plants cause a severe delay in flowering (Kumimoto et al., 2008). The *nf-yb2 nf-yb3* double mutant pheno-copies *co* mutants, inhibiting *FT* expression, causing flowering time to be delayed. Genetic analyses affirmed that the *nf-yc3 nf-yc4 nf-yc9* triple mutants produced a late flowering phenotype, and NF-YC3, 4, and 9 can physically interact with CO. Indicating genetic interaction of the NF-Y proteins and CO are required for expression of *FT* and to initiate photoperiod-dependent flowering (Kumimoto et al., 2010). Interestingly, these specific NF-YC proteins bind NF-YB2 or NF-YB3 and CO to activate *FT*. Additionally, mutations in NF-YA and NF-YC to form specific combinations of NF-Y subunits leads to late flowering (Siriwardana et al., 2016; Gnesutta et al., 2017). Although genetic analyses

show that CO interacts with NF-YB and NF-YC (NF-CO), the nature of the interactions between the NF-Y and NF-CO complexes remain unclear in the regulation of *FT* and activation photoperiod-dependent flowering (Gnesutta et al., 2017).

The NF-YB proteins highly-conserved HFD is approximately 96 amino acids in length and is flanked by highly variable flanking amino (N) and carboxyl (C) terminal sequences. The NF-YB's share a resemblance to the plant MADS-box transcription factors, as they both have conserved motifs and contain a C terminal, which has variability in sequence and length. Studies on the MADS-box C-termini reveal they are involved in transcriptional activation, the formation of transcription factors multimeric complexes, protein-DNA stability, sub-cellular localization, and post-translational modifications (Cho et al.,1999; Egea-Cortines et al., 1999; Alvarez-Buylla et al., 2000; Becker et al., 2003). Suggesting the NF-YB proteins to have a similar function in developmental processes, and potential differences of the NF-YB C-terminals can lead to specialization in the ability to form multimeric complexes, interaction with DNA, and other operations. The late-flowering phenotype of the *nf-yb2 nf-yb3* double mutant becomes a prime model for understanding the specific roles of each NF-YB subunit. By observing flowering phenotypes of overexpression of NF-YB's in the *nf-yb2 nf-yb3* mutant will hopefully help elucidate the function of NF-YB HFD, N, and C terminal.

Jan Risinger in the lab previously examined the ability of each overexpressed NF-YB full-length protein to rescue the late-flowering phenotype of the *nf-yb2 nf-yb3* double mutant. Specifically, Clade I and NF-YB5 of Clade III are positive regulators, NF-YB1 8 of Clade II are weakly positive regulators, and NF-YB4 of Clade III is a negative regulator, prolonging flowering activation. Clade IV members NF-YB6 and NF-YB9 are known to have roles in embryogenesis and seed maturation and were unable to produce transgenic plants when overexpressed (Risinger 2012). The functional ability of each NF-YB clade to alter flowering suggests that they are functionally diversifying, and is particularly true for Clade III where NF-YB5 is a positive regulator of flowering and NF-YB4 is a negative regulator.

Following these previous investigations, I examined several new questions surrounding the HFD, and N and C termini, as follows:

1) The highly conserved NF-YB HFD shares regulatory roles shown to be necessary for dimerizing with the NF-YC HFD, and binding with NF-YA to form the NF-Y complex; Knowing this data, is the HFD alone sufficient in regulating flowering, and to what degree?
2) What are the functions of the amino (N) and carboxy (C) termini?
3) Can you rescue or prolong flowering by switching the terminals of a known positive regulator with a weak flowering regulator, and how efficiently will they rescue *nf-yb2 nf-yb3* double mutant phenotype?

4) Finally, does a native promoter's spatial and temporal function effect the ability of functionally redundant proteins to positively regulate flowering?

I examined the functional redundancy of the HFD of the NF-YB family by overexpressing each member in the *nf-yb2 nf-yb3* double mutant background and observing the ability to rescue the late-flowering phenotype. Next, I determined if swapping the N and C termini affected the function of the proteins ability to rescue the *nf-yb2 nf-yb3* late-flowering phenotype. Finally, I investigated native promoter swaps of NF-YB2 and NF-YB7 and whether their native spatial and temporal regulatory patterns affected their ability to promote early flowering.

Results

Identification of Four Distinct Clades of NF-YB HFD Proteins

The role of the histone-like fold domain (HFD) makes it a prime choice in understanding the importance of its function in photoperiodic flowering because of its moderate evolutionary divergence they all seem to share a common ancestor. Previously the NF-YBs were aligned as full-length proteins. I wish to extend this analysis by focusing on the HFD alone and the differences in the Amino (N) and Carboxy (C) flanking regions (Figure 1A-D). The HFD alignment showed amino acid identities are highest in Clade I- NF-YB2 HFD and NF-YB3 HFD are 94% identical, NF-YB2 HFD and NF-YB7 HFD are 84% identical, and NF-YB3 HFD and NF-YB7 HFD are 88% identical. Clade II amino acid identities are the highest between NF-YB8 HFD and NF-YB10 HFD with 94%, NF-YB1 HFD is 86% identical with NF-YB8, and 84% identical with NF-YB10. Clade III is the most divergence, with 69% amino acid identity between NF-YB4 HFD and NF-YB5 HFD. Clade IV consists of NF-YB9 and NF-YB6, also known as LEC1 and LEC1- LIKE, respectively, with 83% identity (Figure 1B and 1E).

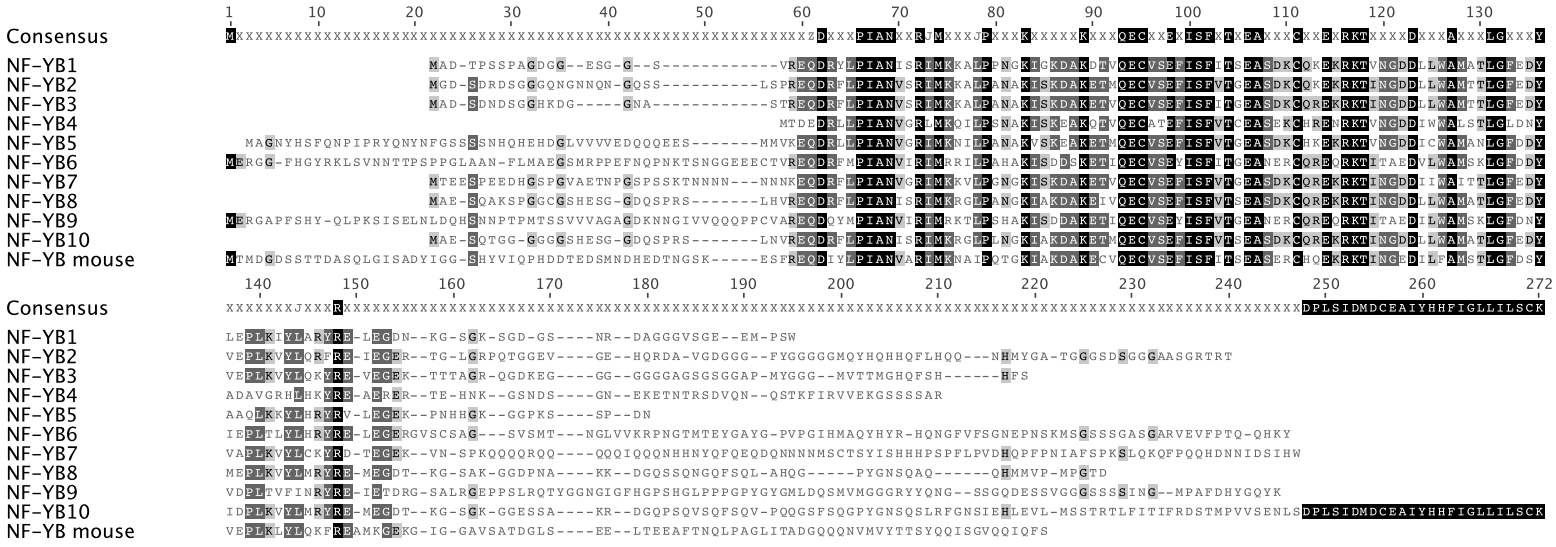


Figure 1A. NF-YB full-length protein alignment constructed using Geneious Pro 7.1.9, alignment was performed with ClustalW. Black boxes (Consensus) represent 100% amino acid (AA) identity. Dark grey boxes represent synonymous AA differences (1). Light grey boxes represent synonymous AA differences (2). Non-shaded AA residues represent <50%

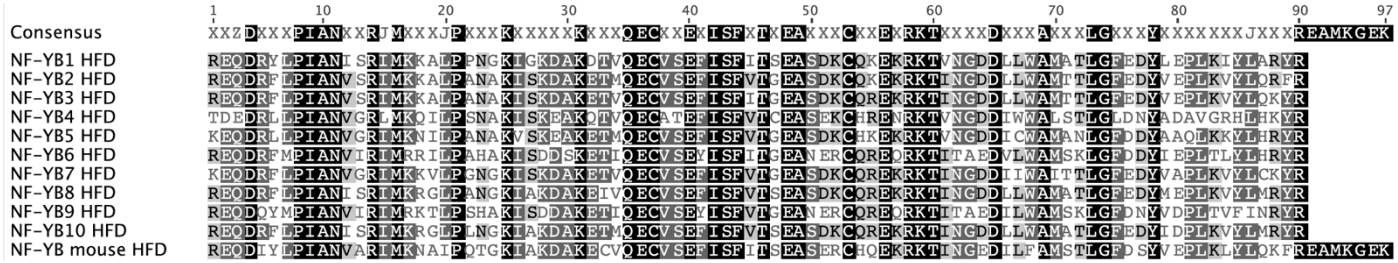


Figure 1B. NF-YB HFD protein alignment constructed using Geneious Pro 7.1.9, f alignment was performed with ClustalW. Black boxes (Consensus) represent 100% amino acid (AA) identity. Dark grey boxes represent synonymous AA differences (1). Light grey boxes represent synonymous AA differences (2). Non-shaded AA residues represent <50% AA identity.

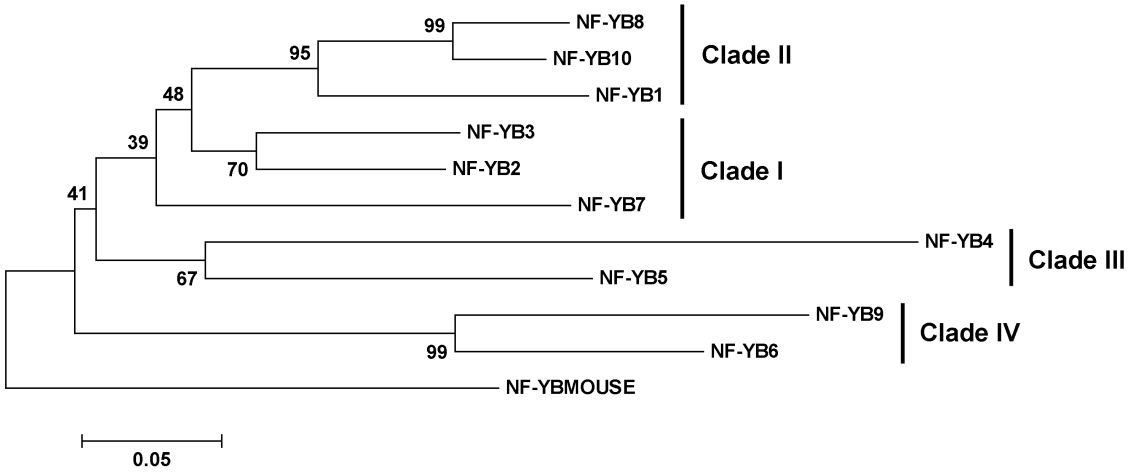


Figure 1C. NF-YB phylogenetic tree constructed by neighbor joining using the full-length proteins illustrated in Figure 1A. Reliability values at each branch represent bootstrap samples (2,000 replicates). All trees were determined and constructed using MEGA 4 (Tamura et al., 2007).

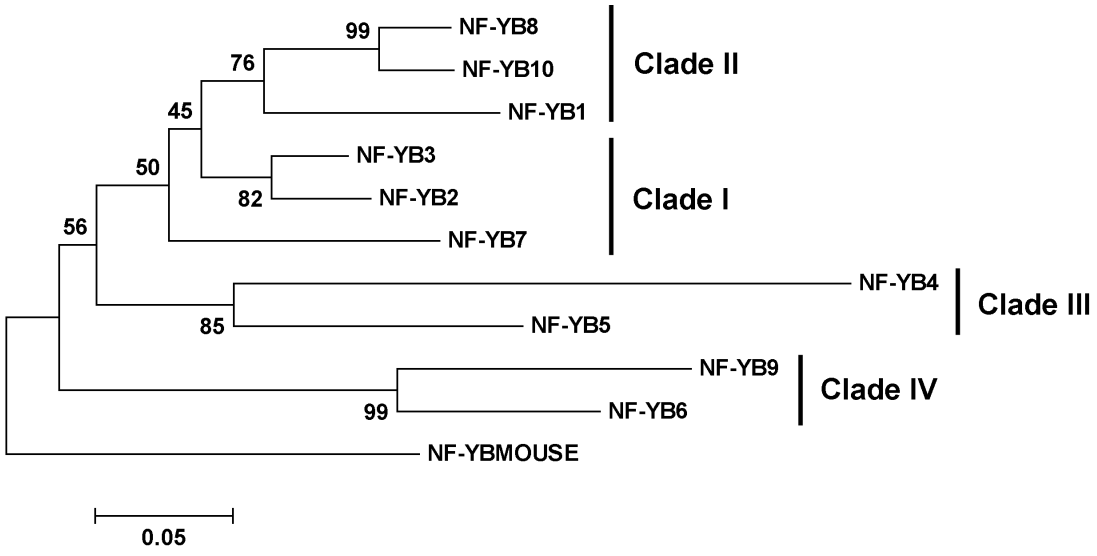


Figure 1D. NF-YB phylogenetic tree constructed by neighbor joining using the HFD regions illustrated in Figure 1B. Reliability values at each branch represent bootstrap samples (2,000 replicates). All trees were determined and constructed using MEGA 4 (Tamura et al., 2007).

	NF-YB1	NF-YB2	NF-YB3	NF-YB4	NF-YB5	NF-YB6	NF-YB7	NF-YB8	NF-YB9
NF-YB1									
NF-YB2	57 / 83								
NF-YB3	61 / 84	69 / 94							
NF-YB4	46 / 58	43 / 59	44 / 62						
NF-YB5	53 / 72	55 / 78	54 / 74	61 / 69					
NF-YB6	45 / 62	43 / 69	44 / 71	37 / 52	40 / 63				
NF-YB7	50 / 77	43 / 84	48 / 88	44 / 64	52 / 77	33 / 63			
NF-YB8	65 / 86	52 / 86	55 / 87	41 / 58	54 / 72	39 / 67	47 / 81		
NF-YB9	40 / 59	36 / 67	37 / 67	40 / 54	38 / 59	38 / 83	31 / 64	37 / 61	
NF-YB10	65 / 84	48 / 86	51 / 84	44 / 59	54 / 73	36 / 67	43 / 81	77 / 94	35 / 63

Figure 1E. NF-YB percent identity values for full length (first value) and HFD-only (second value). Values calculated from figures 1C and 1D using Geneious Pro 7.1.9.

Photoperiodic Flowering of NF-YB Histone Fold Domain Experiments

Based on the conserved sequences of the 10 NF-YB HFD, I wanted to test the ability of each HFD to activate flowering independently. To form the NF-Y complex the HFD is necessary for DNA and protein interaction. I hypothesize that NF-YB1, 2, 3, 5, 7, 8 and 10 HFD alone will promote positive floral regulation. However, overexpression of *NF-YB4* HFD and *NF-Y5* HFD, as a result of their significant divergences, will act as negative regulators of photoperiodic flowering. The lack an N terminal region of the native full-length *NF-B4* proposes that absence of the C terminal flanking region will not aid in its ability to promote flowering.

To examine the function of each NF-YB, I cloned each of the 10 NF-YB HFDs (i.e., the expressed proteins lacked the amino and carboxy terminal regions that flanked the HFD) and transferred them to a plant overexpression vector, and transformed them into the *nf-yb2 nf-yb3* background (Figure 3A). The double mutant is the late flowering control and the non-mutant, parental Arabidopsis ecotype Columbia (Col-0) was used as the wild-type flowering time control. The HFD overexpression experiments were compared to the overexpressed full-length lines (Figure 2A) to determine if the N and C termini played a role in the regulation of flowering time.

Compared to overexpression of the full-length *NF-YB1*, *NF-YB1 HFD* was able to activate flowering more efficiently (Figure 2A). Suggesting that the N and C

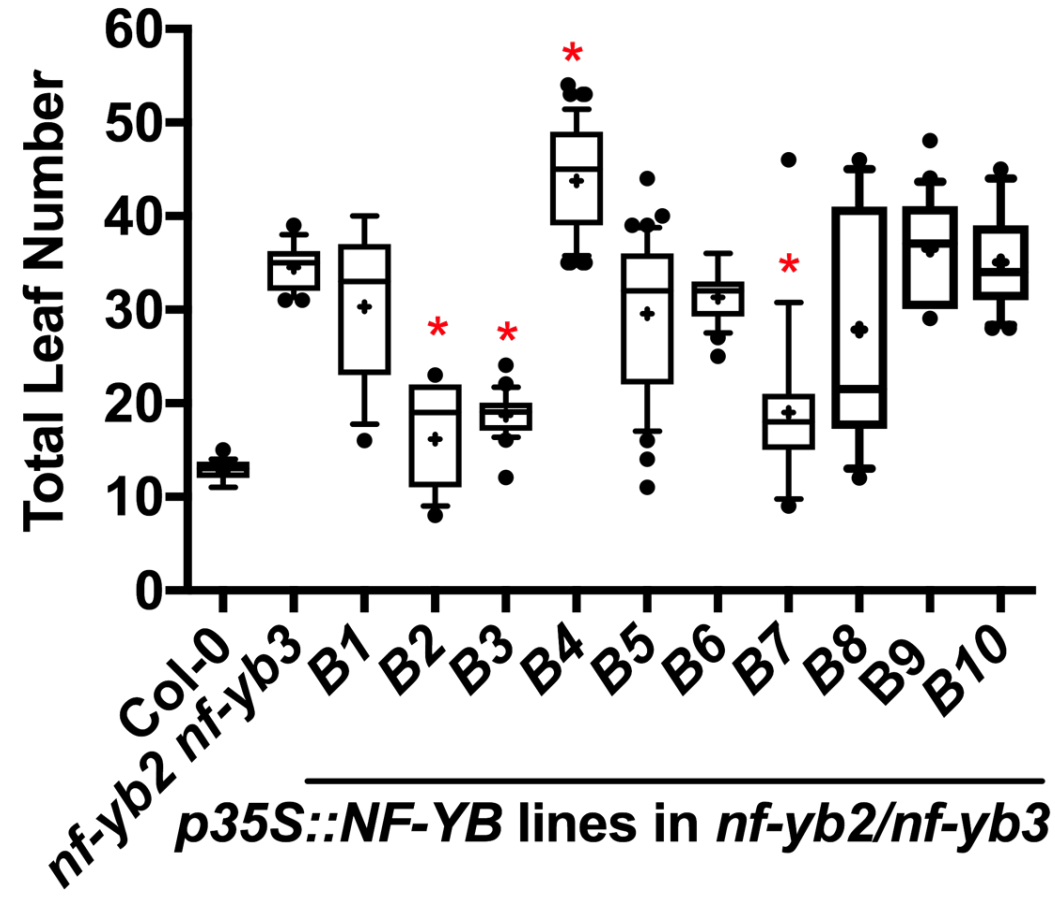
terminal portions of the *NF-YB1* protein repress its ability to act as a positive regulator of flowering. Overexpression of *NF-YB10 HFD* also resulted in significantly earlier flowering time. As with *NF-YB1*, full-length *NF-YB10* was incapable of significantly rescuing, while *NF-YB10 HFD* could partially rescue the late-flowering phenotype. The NF-YB10 HFD constructs expressed protein in the earliest flowering transgenic plants, unlike. Full-Length *NF-YB10* protein which showed little to no protein (Figure 3 A-B). Suggesting overexpression of full-length NF-YB10 causes lethality, in association with the N or C terminal flanking regions.*NF-YB7 HFD* and *NF-YB8 HFD* had a no significant difference in flowering time compared to their full-length counterparts.

As expected, overexpression of the full-length *NF-YB2* can rescue the *nf-yb2 nf-yb3* late-flowering phenotype, similar to the Col-0 ecotype, making it a capable positive regulator of flowering. *NF-YB2, 3, and 7 HFD* is better in the ability to rescue the *nf-yb2 nf*-yb3 double mutant phenotype than the rest of the *NF-YB HFDs*, indicating that NF-B2, 3, and 7 HFD is a positive regulator of flowering. The high amino acid identity between NF-YB2, 3, and 7 suggests that the N and C termini cause superior ability to initiate flowering. A correlation between protein accumulation and flowering phenotype shows that overexpression of each *NF-YB2 HFD*, *NF-YB3 HFD,* and *NF-YB7* the highest accumulation of protein in the earliest flowering plants (Figure 3B), the full-length proteins also observe the same phenotype (Figure 2B).

Overexpression of *NF-YB4 HFD* consistently produced significantly later flowering than the *nf-yb2 nf-yb3* background. In fact, the highest levels of

proteins accumulate in the plants flowering the latest (Figure 2B). Remarkably, *NF-YB5 HFD*, which is 67% identical to NF-YB4 HFD, was able to drive significantly earlier flowering in the *nf-yb2 nf-yb3* late-flowering mutant (Figure 3A).

A



B

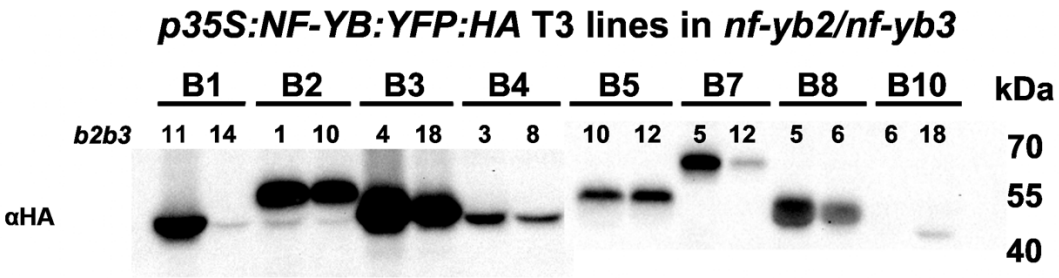
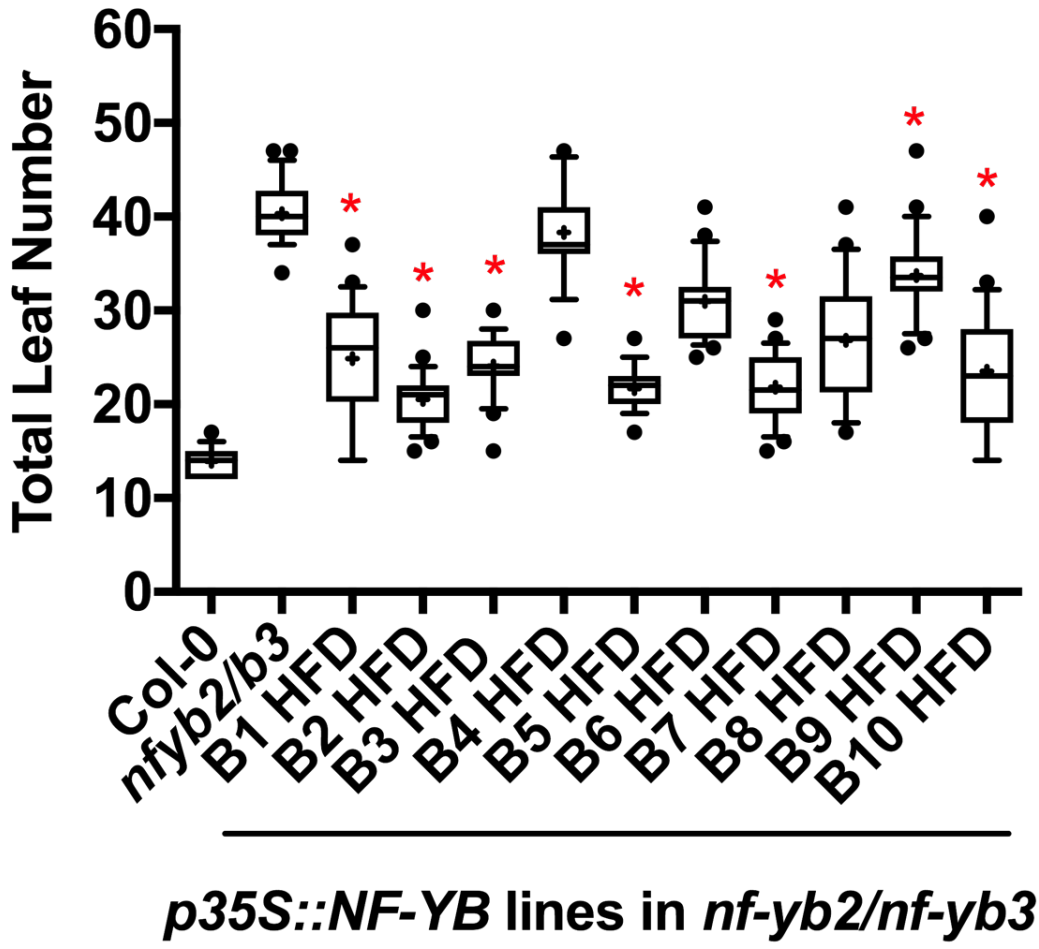


Figure 2. **A)** T1 flowering time quantification of p*35S::NF-YB* constructs in the *nf-yb2 nf-yb3* background. In the box and whisker plots throughout, a cross represents the mean, a horizontal line represents the upper and lower 5%, the

box represents 50%, the whiskers represent 25%, and dots represent outliers (<10^{th} and >90^{th} percentiles). Sample sizes were ≥40 independent first generation transformants/transgenic line. Asterisks denote significant differences derived from Kruskal- Wallis test ($P < 0.05$) followed by Dunnett's multiple comparison post hoc test against *nf-yb2 nf-yb3*. **B)** Protein expression levels of each of T3 lines assayed, with the first line representing an earlier flowering and the second line representing a later flowering plant.

A



B

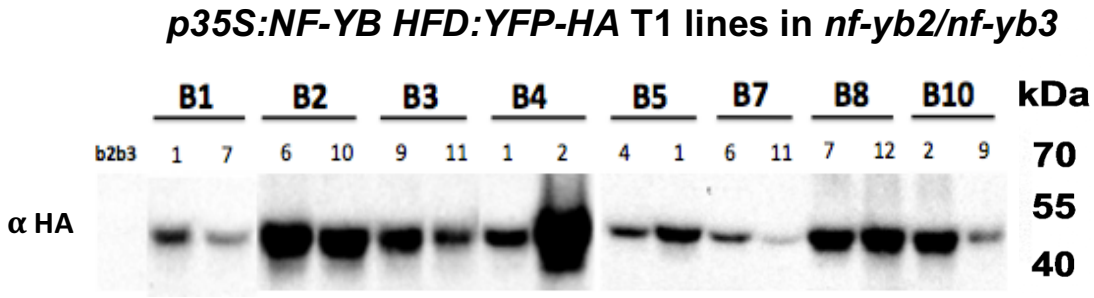


Figure 3. **A)** T1 flowering time quantification of p35S::NF-YB full-length HFD constructs in the *nf-yb2 nf-yb3* background. In the box and whisker plots

throughout, a cross represents the mean, a horizontal line represents the upper and lower 5%, the box represents 50%, the whiskers represent 25%, and dots represent outliers (<10th and >90th percentiles), respectively. Sample size ≥24 independent first generation transformants/transgenic line. Asterisks denote significant differences derived from Kruskal- Wallis test ($P < 0.05$) followed by Dunnett's multiple comparison post hoc test against *nf-yb2 nf-yb3*. **B)** Western blot analysis of NF-YB HFD flowering time quantification from part **A**, with the first line representing an earlier flowering and the second line representing a later flowering plant**.**

NF-YB HFD Localization to the Nucleus

Nuclear localization is an essential NF-YB function, and it is possible that the NF-YB rescue experiments phenotypes are affected by the ability or inability of the protein to localize to the nucleus. All the NF-YB constructs contain a yellow fluorescent protein (YFP), allowing determination of localization by epifluorescent microscopy. I examined the localization of NF-YB 2, 3, 4, and 7 HFD, because of their ability to rescue or delay the *nf-yb2 nf-yb3* late-flowering phenotype. Overlaying fluorescent light microscopy and differential interference contrast (DIC) microscopy revealed that all NF-YB HFD tested can translocate to the nucleus. While, NF-YB2, 3, and 7 HFD show strong fluorescence in the nucleus, NF-YB4 does not. NF-YB4 inability to translocate to the nucleus efficiently might show correlation to its extreme late-flowering phenotype (Figure 4).

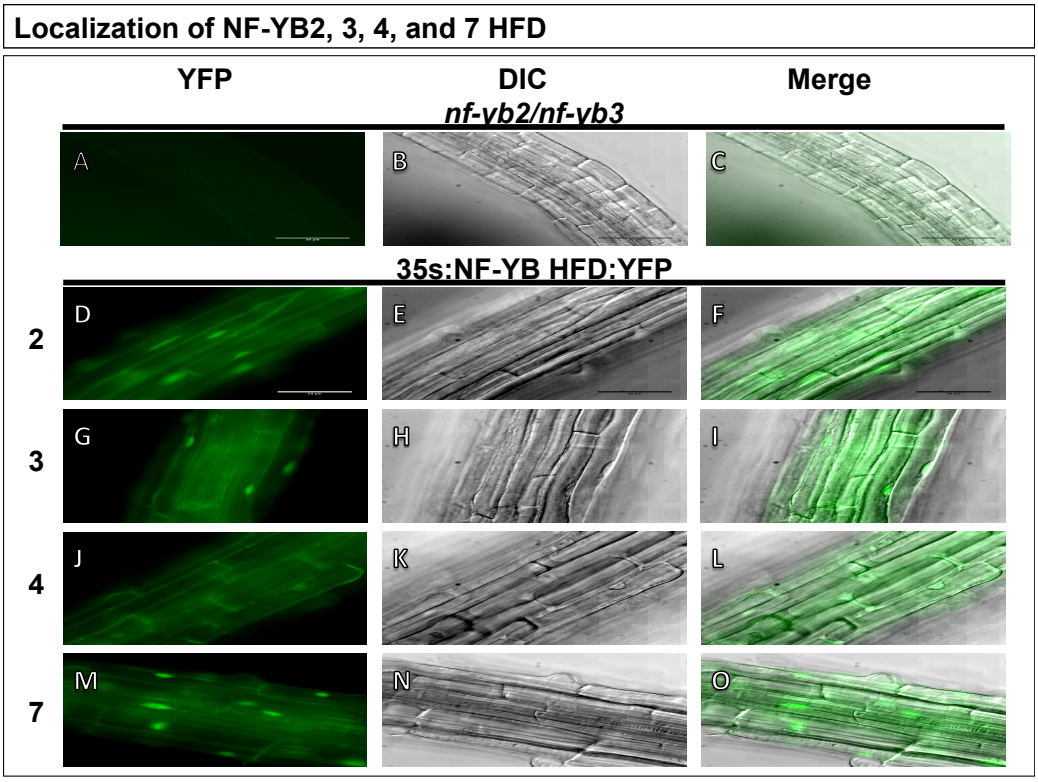


Figure 4. Localization of 35s::NF-YB HFD:YFP gene fusions. GFP fluorescence was detected by a compound epifluorescent microscopy using a 40x dry objective (column1). Location of the nucleus and cell walls was identified using DIC and bright field microscopy, respectively (middle row). NF-YB HFD 2 **(D)**, 3 **(G)**, and 7 **(M)** show higher florescence in the nucleus than NF-YB4 HFD **(J).** The right column displays the merge of green fluorescence and the DCI. Data has not been quantified.

Promoter Swap Reveals NF-YB7 has a Possible Role in Photoperiodic Flowering

Rescue experiments using the full-length NF-YB and HFD revealed that NF-B7 could positively regulate flowering when overexpressed. In most cases NF-YB7 is more effective at rescuing the late-flowering phenotype of *the nf-yb2 nf-yb3* double mutant than *NF-YB2,* having known floral promotion. *NF-YB7* also shows overlapping expression patterns to *NF-YB2* and *NF-YB3* further indicating that native NF-YB7 regulation may play a role in floral induction. I hypothesize that expression of the *NF-YB7* coding sequence (CDS) under *NF-YB2* and *NF-YB7* native promoter will show that NF-YB7 rescues the moderate late-flowering phenotype of the *nf-yb2* single mutant.

A native promoter swap experiment was performed to explore the role of NF-YB7 in photoperiodic-dependent flowering. The *nf-yb2* mutant is the late flowering control, with parental *Arabidopsis* ecotype Columbia (Col-0) establishing a wild-type flowering time control. To examine the function of each promoter, I cloned each of the *NF-YB2* and *NF-YB7* promoter (pB2/7) with each *NF-YB2 and NF-YB7* coding sequence (CDS), transferred them to a plant overexpression vector, and were transformed into the *nf-yb2* mutant background (Figure 5). pB2::*NF-YB2*, p*B2*::*NF-YB7,* and pB7::*NF-YB7* were significantly able to rescue the *nf-yb2* mutant with a mean of18, 17, and 15 rosette leaves, respectively, However, the *pB7::NF-YB2* construct significantly drove the earliest flowering, with a mean of 13 rosette leaves, showing superior ability to promote flowering compared with the other constructs. The full-length NF-YB and HFD overexpression experiments showed significant ability to

rescue late flowering. However, this promoter analysis indicates that the native *NF-YB7* promoter and the coding region more than likely have a role in photoperiodic flowering. Additional research, such as an *NF-YB7* knock out mutant, would help elucidate a more definitive function of this protein.

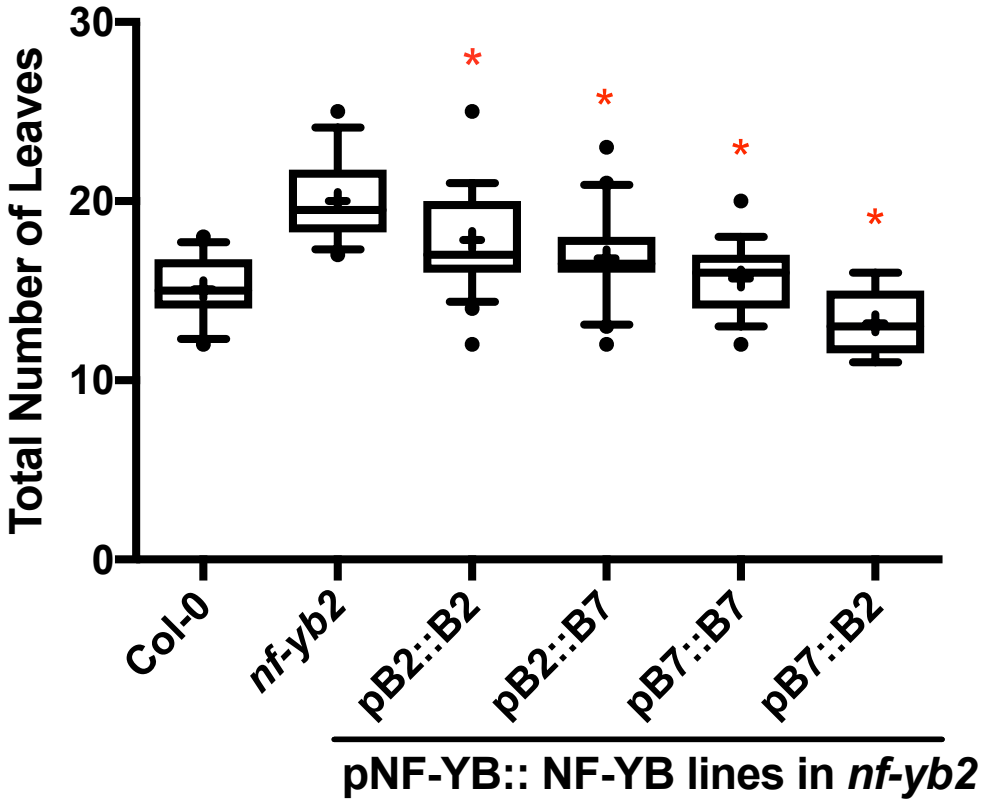


Figure 5. T1 flowering time quantification of pB2/B7::NF-YB2/7 promoter swap constructs in the *nf-yb2* background. In the box and whisker plots throughout, a cross represents the mean, a horizontal line represents the upper and lower 5%, the box represents 50%, the whiskers represent 25%, and dots represent outliers (<10th and >90th percentiles), respectively. Sample size ≥24 independent first generation transformants/transgenic line. Asterisks denote significant differences derived from one-way ANOVA ($P < 0.05$) followed by Dunnett's multiple comparison post hoc test against *nf-yb2.*

Chimeric Experiments

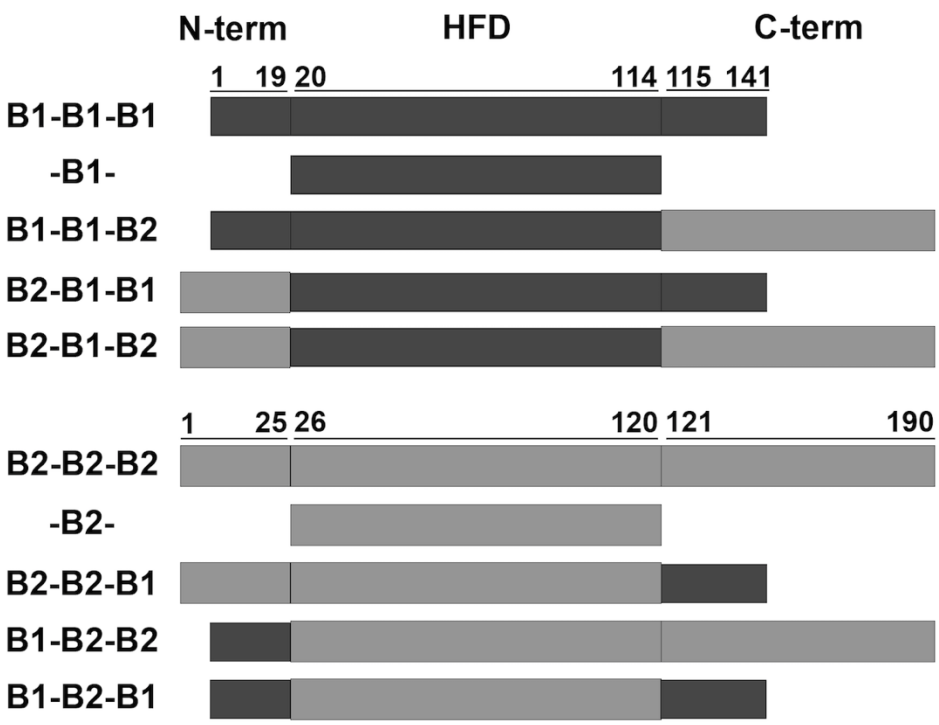
Based on the differences of NF-*YB1* and *NF-YB2* in full-length and HFD rescue experiments, domain swaps between *NF-YB2* and *NF-YB1* could distinguish potential novel roles of the amino (N) and carboxyl (C) termini. Overexpression of full-length *NF-YB1* shows variation in its ability to promote flowering, however, the plants with the highest protein accumulation flower the earliest. In conjunction, NF-YB1 HFD has an 83% identity with *NF-YB2*, I would conclude that swapping *NF-YB2* N and C termini on the *NF-YB1* HFD will promote early flowering, finding that *NF-YB1* N and C termini have diverged from regulation of photoperiod-dependent flowering.

NF-YB1 and NF-YB2 chimeric protein were constructed to determine the effects of each terminus on the HFD. The *nf-yb2 nf-yb3* double mutant is the late flowering control and the non-mutant, parental Arabidopsis ecotype Columbia (Col-0) was used as the wild-type flowering time control. To examine the function of each NF-YB N and C flanking terminals. I cloned six NF-YB chimeric proteins (i.e., expressed the NF-YB1 HFD and NF-YB2 HFD proteins with different N and C terminal regions from each full-length NF-YB1 and NF-YB2). Also, *NF-YB1* and *NF-YB2* full-length and HFD were also examined(Figure 7A). Each construct was transferred to a plant overexpression vector and transformed into the *nf-yb2 nf-yb3* background (Figure 7B). Each construct allows observation of the effects of the N terminal, C terminal, and both termini

on each HFD. *NF-YB1* and *NF-YB2* N terminal length doesn't differ greatly in length but has amino acid sequence divergence. However C termini display a significant number of amino acid differences, with *NF-YB2 having* ~40 amino acids differences.

Full-length *NF-YB1* shows no significant difference in *nf-yb2 nf-yb3* plants, while *NF-YB1* HFD alone is more effective in its ability to rescue the late flowering mutant. Adding *NF-YB2* terminals either makes a slight difference or no difference in its ability to rescue *nf-yb2 nf-yb3.* As shown above, statistically full-length *NF-YB2* and HFD show the opposite phenotype of *NF-YB1,* whereas the full-length *NF-YB2* promotes flowering comparative to Col-0, *NF-YB2* HFD slight delay in its ability to rescue *nf-yb2 nf-yb3.* Adding both *NF-YB1* N and C terminals, or the C terminal, does not show significant differences to the *nf-yb2 nf-yb3 plants;* however, adding the *NF-YB1* N terminal to the NF-YB2 HFD and C terminal, there is a significant degree of rescue of the late flowering double mutant phenotype (Figure 6B).

A


N-term
HFD
C-term
1
19
20
114
115
141
B1-B1-B1
-B1-
B1-B1-B2
B2-B1-B1
B2-B1-B2
1
25
26
120
121
190
B2-B2-B2
-B2-
B2-B2-B1
B1-B2-B2
B1-B2-B1

B

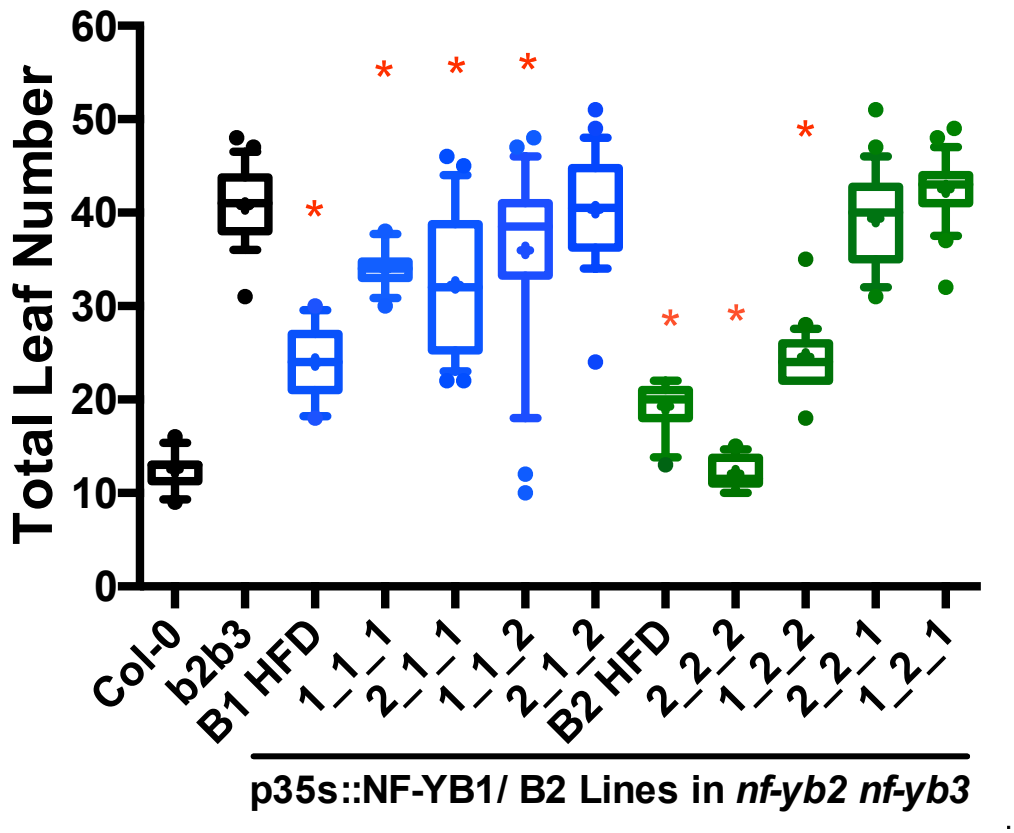


Figure 6. A) Comparison of NF-YB1/ NF-YB2 chimeric protein lengths drawn to scale. **B)** T1 flowering time quantification of *p35s::NF-YB1/ NF-YB2* domain swap constructs in the *nf-yb2 nf-yb3* background. In the box and whisker plots throughout, a cross represents the mean, a horizontal line represents the upper and lower 5%, the box represents 50%, the whiskers represent 25%, and dots represent outliers (<10th and >90th percentiles), respectively. Sample size ≥24 independent first generation transformants/transgenic line. Asterisks denote significant differences derived from one-way ANOVA ($P < 0.05$) followed by Dunnett's multiple comparison post hoc test against *nf-yb2 nf-yb3.*

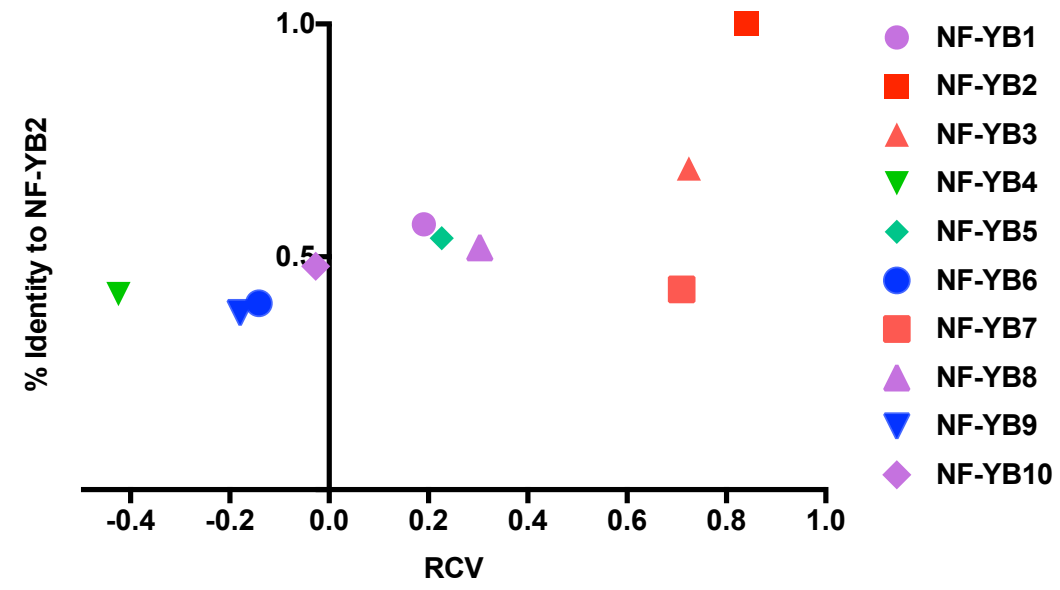
Relative Complementation Value

To compare independent experiments a quantitative value was created called the Relative Complementation Value (RCV). The RCV of a given analysis allows one to compare the flowering time of any given transgenic NF-YB (constitutively expressed: full-length and HFD) to Col-0 and *nf-yb2 nf-yb3* double mutant, generates a quantitative value that determines the ability of any transgenic NF-YB relative complementation of Col-0 flowering time in the *nf-yb2 nf-yb3* double mutant background, allowing a balance point between the two controls. For example, the RCV of Col-0 always equals 1, as the non-mutant parental line, and the RCV of *nf-yb2* or *nf-yb2 nf-yb3* always equals 0.

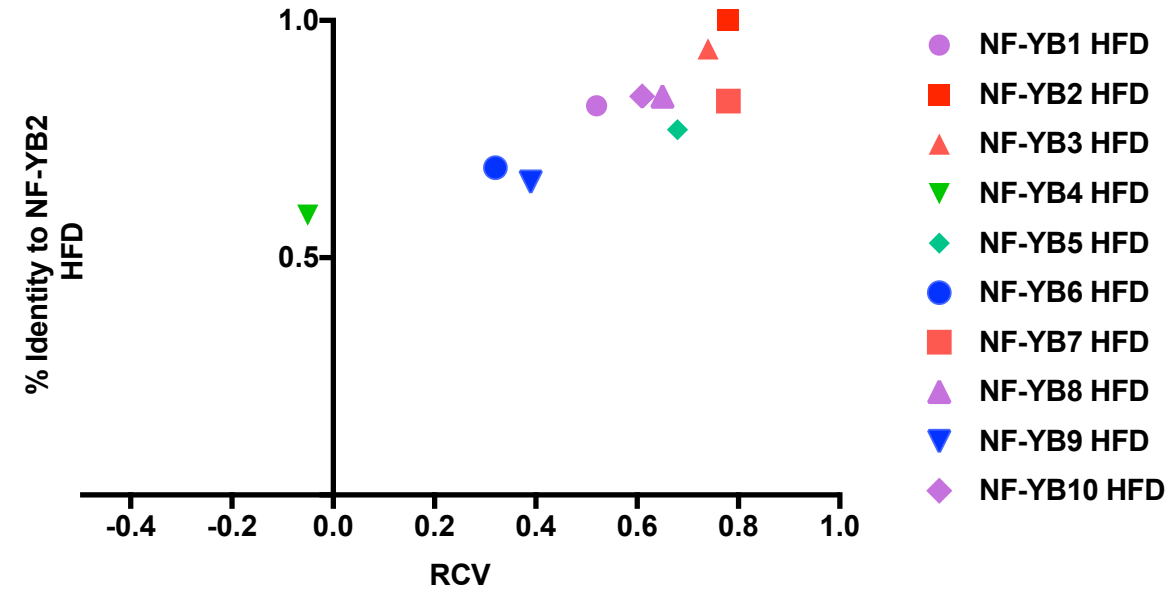
The RCV figures also allow you to visualize the data which might not be apparent boxplot figures. For example, for the full-length RCV (Figure 7A) a widespread distribution of the NF-YB's ability to complement Col-0, compared to the variation in the percent identity to *NF-YB2* is seen. Also, one can observe the clustering of each clade, except clade III: NF-YB4 and NF-YB5. On the other hand, NF-YB HFD RCV (Figure 7B) shows that of all the HFD's cluster and shift together corresponding to the RCV and their percent identity to *NF-YB2* HFD representing their sufficient ability to rescue *nf-yb2 nf-yb3* double mutant, except *NF-YB4* HFD. Additionally, the domain swap between *NF-YB1* and *NF-YB2* experiment's RCV (Figure 7C) further proposes that *NF-YB1* N

and C termini have diverged from regulation, while *NF-YB2* N and C termini fine-tune the regulation of floral promotion.

A


1.0
0.5
% Identity to NF-YB2
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
RCV
NF-YB1
NF-YB2
NF-YB3
NF-YB4
NF-YB5
NF-YB6
NF-YB7
NF-YB8
NF-YB9
NF-YB10

B


1.0
0.5
% Identity to NF-YB2 HFD
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
RCV
NF-YB1 HFD
NF-YB2 HFD
NF-YB3 HFD
NF-YB4 HFD
NF-YB5 HFD
NF-YB6 HFD
NF-YB7 HFD
NF-YB8 HFD
NF-YB9 HFD
NF-YB10 HFD

C

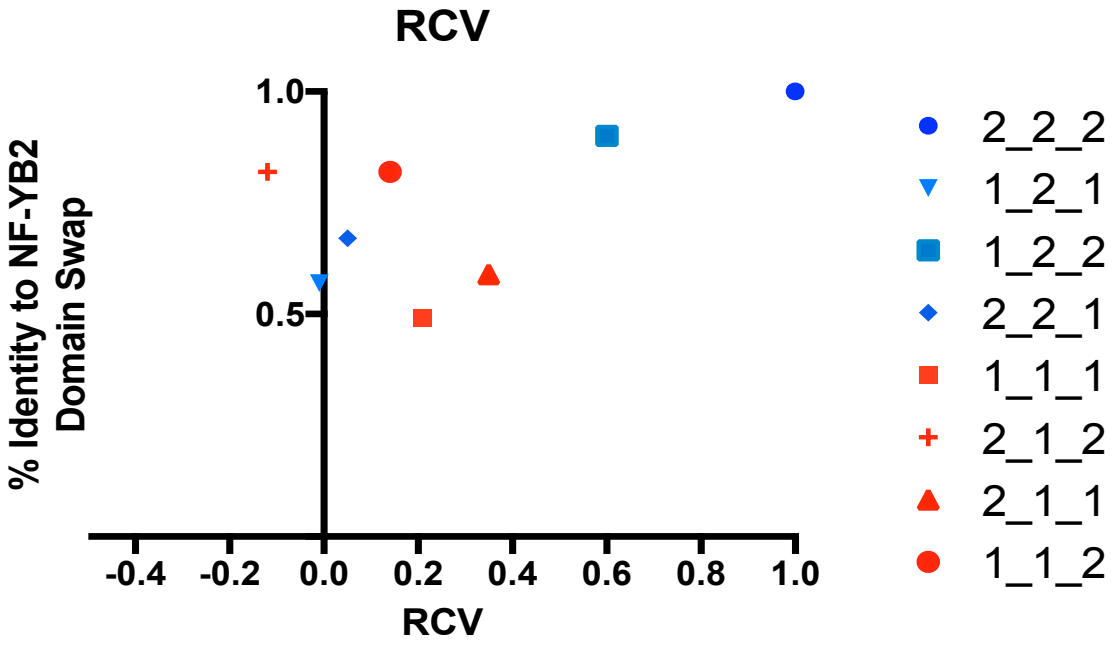


Figure 7. A) Relative Complementation Value (RCV) compared to the percent identity to NF-YB2 full length protein. Each symbol represents over expression of a NF-YB. **B)** RCV compared to the percent identity to NF-YB2 full-length protein. Each symbol represents over expression of a NF-YB HFD. **C)** RCV compared to the percent identity to NF-YB2 full length protein. Each symbol represents over expression of each of the domain swap constructs described in figure 6A

Distinct Overexpression Phenotypes

Interestingly, Overexpression of NF-YB2, 4, 5, 7, and 10 HFD shared characteristics of a fascinating phenotype, with rosettes that are dark green and ruffled in appearance, and is much smaller in size than the Col-0 and nf-yb2 nf-yb3 mutants. High protein accumulation drives the phenotype for each of these genes. NF-YB4 HFD displayed this phenotype more consistently than any of the other NF-YB HFDs. In fact, the NF-YB4 HFD with rosettes that are dark green and ruffled in appearance were usually the latest flowering, with >35 leaves. Sometimes flowering time was unable to be calculated because the meristem did not elongate and from the center of the rosette, instead, the steam would grow from underneath the base of the plant. Due to the compression of the meristem and the successive inability to flower correctly, these NF-YB4 gave the plant a bush like phenotype.

Discussion

Histone-like fold domain alone can activate photoperiodic flowering

All the NF-YBs histone-like Fold Domains (HFD) tested in the rescue assays all displayed a relative complementation value (RCV) higher than 0, except NF-YB4 HFD, having an RCV of -0.5. Most HFDs were better at rescuing than the full-length versions, except NF-YB2, 3, *4, and 7*, with their HFDs having slightly worse ability to rescue the late-flowering phenotype. Indicating the HFD is sufficient and necessary in photoperiodic-dependent flowering time. The small deviation between the efficiency of HFDs RCV reveals that NF-YB HFD have critical amino acids for *NF-YB*'s function, and ability to bind with *NF-YC* HFD. The tight "handshake" heterodimerization of the HFD of NF-YB and NF-YC mediates the NF-Y complex by providing the exterior for NF-YA interaction. Moreover, the homologous sequence and structure to histone H2B, is known to be involved in transcription and chromatin remodeling, further indicating the essential function of the HFD (Romier et al., 2003, Hackenberg et al., 2012, Nardini et al., 2013, Gnesutta et al., 2013).

The interaction between *NF-YB* and *NF-YC* HFD, which is conserved among all family members of both subunits, is the only portion of the protein necessary for not only floral promotion but in other development processes, such as embryogenesis. Embryonic regulators *NF-YB9* (LEAFY COTYLEDON-1, Lec1) and *NF-YB6* (LEAFY COTYLEDON-1-LIKE) HFD are necessary and sufficient

for seed viability in the null *lec1-1* mutant phenotype. Specifically the 16 amino acid sequence that NF-YB6 and NF-YB9 share are essential for embryogenesis (Lee et al., 2003, Cagilari et al., 2014). To further understand the role of the NF-YB HFD, experiments pinpointing essential amino acids will be necessary.

NF-YB4 Histone-like Fold Domain translocation to the Nucleus

The NF-Y complex in mammalians forms in a step-wise manner. First, the HFD of NF-YB and NF-YC heterodimerize in the cytoplasm, translocate to the nucleus where it forms its stable heterotrimeric complex with NF-YA (Kim et al., 1996). It was necessary to determine that the *NF-YB4* HFD alone was sufficient to translocate to the nucleus and form the NF-Y complex, as both the full-length and HFD is unable to complement flowering in *the nf-yb2 nf-yb3 mutant,* with a -0.5 RCV value, and having higher protein accumulation in the later flowering lines. The ability for the NF-Y complex to form is a major mechanism required for downstream gene transcription, and the structure of the NF-YC and NF-YB HFD have critical structural qualities that can account for the ability for the NF-Y complex to form (Romier et al., 2003, Hackenberg et al., 2012). NF-YB4 possible inadequate ability to create functional NF-Y complexes could explain its extreme delay in flowering.

Risinger 2012 showed through Yeast 2 Hybrid (Y2H) experiments that *NF-YB4* interacts with the NF-YC floral promoting proteins, and *NF-YB4's* ability to

cause later flowering (Figure 2A) has been hypothesized to have a dominant negative response when overexpressed. Figure 4 shows that *NF-YB4* appears to have less accumulation in the nucleus compared to flowering regulators HFD *NF-YB2, 3,* 7 HFD (although data has not been quantified).

It is hypothesized that *NF-YB4* binds the floral promoting NF-YC3, 4, and 9 (Kumimoto et al., 2010) and forms non-beneficial protein complexes but is unable to bind the appropriate CCAAT-box, -5.3kb sequence upstream of the transcriptional start site halting the promotion of FLOWERING LOCUS T (FT) (Cao et al., 2014). *NF-YB4* is the most divergent of all the full-length and HFD NF-YB proteins, missing an N terminal, meaning that it may be lacking the necessary amino acids to dimerize strongly with NF-YC, preventing movement into the nucleus. To further characterize this functional analysis of the NF-YB4 quantifiable microscopy data is needed and a*n NF-YB4* knock out mutant would help define if it functions as a negative regulator of flowering.

NF-YB7 has a Novel Positive Role in Photoperiodic Flowering

Bioinformatic analysis has suggested that NF-YB7 would have a positive role in flowering, with its closest relatives being NF-YB2 and NF-YB3*,* which are known positive regulators of flowering. NF-YB7 has also been shown to have similar expression patterns with NF-YB2 and NF-YB3, with expression in the leaf vascular tissue (Siefers et al., 2009). The NF-Y complex regulates transcription of FT in the leaves, where the subsequent protein travels to the shoot apical

meristem turning on downstream flowering genes (Corbesier et al., 2007). Also, Y2H experiments showed that NF-YB7 could interact with floral promoting NF-YC proteins, has the highest *FT* levels of transcription, and its full-length and HFD can complement the *nf-yb2 nf-yb3* mutant phenotype, pointing to a functional role in flowering (this study and Risinger 2012).

By testing the promoters of NF-YB2 and NF-YB7 transcriptional regulation, in conjunction, with each protein's floral-inducing CDS, NF-YB7 proves to have a functional role in floral induction. The NF-YB2 and NF-YB7 promoter swap displayed that the NF-YB7 CDS under the regulation of native NF-YB2 promoter complement the *nf-yb2* mutant. Also, the NF-YB7's native promoter regulating NF-YB2 CDS flowers early than all other constructs. This study indicates that the promoter of NF-YB7 has overlapping function with NF-YB2 and NF-YB3 and is hypothesized to have a spatiotemporal expression that is advantageous in the activation of *FT.* Additional research, such as an *NF-YB7* knock out mutant along with an *nf-yb2 nf-yb3 nf-yb7* triple mutant would further validate a biological role to activation of photoperiodic-dependent flowering time.

Chimeric protein rescue experiments determine C and N termini modulate HFD

The NF-YB functional analysis from Risinger 2012 concluded that *NF-YB2* was a positive regulator of flowering, with an RCV of 1, and *NF-YB1* showed little ability to complement the *nf-yb2 nf-yb3* double mutant, with an RCV of 0.2, however, protein accumulation for both *NF-YB1* and *NF-YB2* were highest in the earliest flowering plants. Also, the HFD rescue experiments concluded that *NF-YB1 HFD* and *NF-YB2 HFD* alone were sufficient in complementing the *nf-yb2 nf-yb3* late-flowering phenotype, with an RCV of 0.5 and 0.75, respectively. Conversely, as *NF-YB2* HFD diminishes the *NF-YB1* HFD increases capability in rescuing *nf-yb2 nf-yb3*, suggesting the N and C termini have diverged regarding photoperiod-dependent flowering. Creation of chimeric proteins gives insights to the hypothesis that fusing *NF-YB2* N and/or C terminal to the *NF-YB1* HFD will increase the RCV compared to the *NF-YB1* native protein. On the other hand, fusing *NF-YB1* N and/or C terminal to *NF-YB2* HFD will decrease the RCV compared to the *NF-YB2* full-length protein. Proposing the question: do the N and C termini evolve independently making new advantageous genetic variants, while the corresponding HFD is maintaining an ancestral function?

Chimeric proteins swapping the *NF-YB2* N and/or C terminus with *NF-YB1* HFD, *NF-YB1* N and/or C terminus with *NF-YB2* HFD and were not able to rescue, and in most cases caused a significant delay in flowering, compared to

full-length proteins. However, *NF-YB1* N terminus fused to the native *NF-YB2* HFD and C terminus, produced a significant ability to flower compared to *nf-yb2 nf-yb3,* but was delayed compared to full-length NF-YB2. The *NF-YB1* and *NF-YB2* HFD alone was effective as a positive regulator, indicating that most of the action resides in the HFD. However, *NF-YB2* HFD rescues better, suggesting that this particular HFD has evolved towards an optimal function in flowering. *NF-YB2* N and C termini decrease the success of *NF-YB1* HFD to rescue the *nf-yb2 nf-yb3* double mutant, meaning the N and C termini have evolved to modulate the core functionality of the NF-YB HFD, and cannot enhance *NF-YB1* HFD ability to promote flowering positively.

Evolution of NF-YB family

These function analyses are starting to elucidate the evolutionary force, such as subfunctionalization and neofunctionalization, acting upon this family. One possibility is a unique case of subfunctionalization called Escape from Adaptive Conflict (EAC). EAC is the ability of a duplicated gene to obtain a novel function while still maintaining its ancestral role(Piatigorsky and Wistow 1991; Hittinger and Carroll 2007; Des Marais and Rausher 2008; Deng et al. 2010; Abascal et al. 2013). EAC would explain the NF-YB HFD's ability to promote early flowering is related to its ancestral function and could be tied to the highly conserved sequence and structure of the H2B core histone. Subfunctionalization has also been hypothesized as a transitional state for

neofunctionalization (Rastogi and Liberles 2005). There is evidence of subfunctionalization and neofunctionalization in the NF-YB family, with differences in expression patterns (Siefers et al., 2009), conversion of the HFD sequence, distinct sequence variability in the N and C terminals, and the inability of N and C termini to promote floral initiation in combination with other NF-YB HFDs. Transcription factor C terminals have been known for controlling protein-protein interactions, DNA stabilization, and transcriptional modification (Alvarez-Buyella et al., 2000). The high variability in the N and C termini might attribute to the ability of each NF-YB to bind/interact with proteins in different processes, causing the NF-Y complex to undertake novel functions (neofunctionalization), such as *NF-YB4, NF-YB9* (LEC1), and *NF-YB6* (L1L) (West et al., 1994; Lotan et al., 1998; Lee et al., 2003; Kwong et al., 2003).

To create a better understanding of the evolution of duplicated genes more experimentation would need to be conducted. First, I hypothesize that the C terminal of NF-YB2 enhances the ability to promote flowering compared to the N terminal. Individual domain experiments might elucidate the evolutionary effects of each terminal in modulating the HFD. Finally, to further understand the function of NF-YB2, 3, and point mutations within the HFD would determine vital amino acids in promoting floral initiation.

Materials and Methods

Plant growth and flowering-time experiments

All plants were of the *Arabidopsis thaliana* Col-0 ecotype and grown in a custom walk-in growth chamber at 23C in standard long day (LD, 16 hours light, 8 hours darkness) conditions. Soil was composed of ProMix Flex with the addition of 40 g Marathon pesticide and dilute Peter's fertilizer solution ($1/10^{th}$ recommended feeding level; NPK 20:20:20). All plants were watered with dilute Peter's at $1/10^{th}$ of the recommended level, as needed. Flowering time was determined by counting all primary rosette and cauline leaves at bolting (Onouchi et al., 2000). All constructs were transformed into the Arabidopsis Col-0 ecotype, *nf-yb2,* or *nf-b2 nf-yb3* double mutant, using the Agrobacterium-mediated floral dip method (Clough and Bent, 1998), and then sprayed with glufosinate ammonium (BASTA) to identify the transformants.

DNA cloning and transgenic plants

All 10 of the full-length NF-YB coding regions were previously generated from Col-0 cDNA populations by PCR using Pfu II Ultra (Agilent Technologies Cat#600670-51) and cloned into the pENTR/D-TOPO Gateway™ entry vector (Invitrogen, Cat#45-0218; Risinger et al. 2012). All 10 NF-YB HFD DNA fragments were generated from plasmid DNA using the same method, PCR reactions were performed using PHUSION High-Fidelity Polymerase (NEB, M0535S). The Gateway BP Clonase™ Enzyme recombined these chimeric

fragments into the Gateway vector pDONR 207™ (Invitrogen, Cat#12213-013). Sequencing found all clones to be identical to the expected sequences found in The Arabidopsis Information Resource (TAIR) database (Swarbreck et al., 2008). The Gateway™ LR Clonase II reaction kit (Invitrogen cat#56485) was used to subclone the above pENTR/D-TOPO and pDONR 207 constructs into pEarleyGate 101 (ABRC, Stock#CD3-683, Earley et al. 2006). Cloning into pEarleyGate 101 creates clones with expression driven from the constitutive 35S promoter (Benfey et al., 1990) and creates a C terminal YFP/HA fusion. All plants were transformed into Arabidopsis using the Agrobacterium-mediated floral dip method.

Phylogenies and Alignments

Full length and conserved domain sequences were imported from TAIR into Geneious Pro 7.1.9 (Drummond et al., 2012), where phylogenies and alignments were created. Full-length and conserved domain multiple sequence alignments were created identically using the ClustalW alignment format within Geneious Pro using the BLOSUM90 scoring matrix. Neighbor-joining was employed by majority greedy clustering and 2,000 bootstrap replicates were generated to give reliability estimates on tree branches.

NF-YB2 and NF-YB7 Promoter Fusions

For isolation of the NF-YB2 and NF-YB7 promoter regions, approximately 1,300 bp of sequence upstream of the start site was amplified using PHUSION High-Fidelity Polymerase and Col-0 genomic DNA. Coding regions were amplified from previously cloned NF-YB2 and NF-YB7 plasmids with sequence-confirmed (post-PCR) cDNA clones (Risinger 2012). The promoters, coding regions, and a 3XHA tag were then assembled in the pBAR1 vector (Holt et al., 2002) by standard Gibson protocols (NEB Cat#E2611S). All plants were transformed by the Agrobacterium-mediated floral dip method (Clough and Bent, 1998) and positive transformants were chosen by spraying T1 individuals with BASTA.

Western Blotting

Total protein was extracted using improved lysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA pH 8.0, add fresh 5 mM DTT and 1X homemade Protease Inhibitor Cocktail (0.2mM AEBSF, 0.7μM Bestatin, 0.7μM Pepstatin A, 10μM Leupeptin, 1.4μM E-64, 1.4μM Phenanthroline; all reagents were dissolved separately in DMSO.). For each sample, four standard hole-punch sized tissue samples were suspended in 400 ul of cold improved lysis buffer and were pulverized by two ball-bearings in a GenoGrinder for 3 minutes at 1200 RPM. After grinding, SDS was added to each sample (1% final concentration) and placed on a rotator for 10 minutes at 4C, then centrifuged at 20XG at 4C for 20 minutes. The supernatant was then transferred to a new tube

with 6X Loading Buffer (375 mM Tris. HCl pH 6.8, 9% SDS, 50% Glycerol, 03% Bromophenol blue). Total protein samples were normalized by performing a Bradford protein assay and loaded on 12% polyacrylamide gels in standard SDS-PAGE conditions. All protein blots were observed by the Bio-Rad ChemiDoc XRS imaging system (www.bio-rad.com) using the horseradish peroxidase-based ECL Plus reagent (GE Healthcare, Cat#RPN2132).

Microscopy

After two days of cold stratification (4C), seeds were germinated on B5 plates in a custom walk-in growth chamber at 23° C in standard LD conditions. After 10 days of growth, lines over expressing NF-YB HFD with a GFP tag were removed from the plate and placed in water on a clean microscope slide. The plants were imaged using a 40× dry objective (NA = 0.75) on a Nikon Eclipse Ni-U compound epifluorescent microscope equipped with an X-Cite 120 LED fluorescent lamp and narrow band eGFP filter cube. Brightfield images were taken using DIC optics and overlaid with fluorescence to determine NF-YB-GFP localization.

RNA isolation and qRT-PCR

Total RNA was collected from 10-day-old plants grown on soil in LD conditions. Each construct was collected in 4-hour increments, for 24 hours. Total RNA

was isolated using the E.Z.N.A. Plant RNA Kit (Omega Biotek, inc., Cat#R6827), following manufacturer's instructions. RNA isolation columns were DNAse (Qiagen, Cat. # 79254) treated. Spectrophotometry (Thermo Scientific, NanoDropTM 1000) determined quality and quantity of RNA samples. First-strand cDNA synthesis was performed using the Superscript II Reverse Transcriptase Synthesis System (Invitrogen, Cat# 18064-022) with supplied oligo dT primers. qRT-PCR was performed using the Bio-Rad (http://www.bio-rad.com) CFX96 Real-Time PCR Detection System and the Thermo Fisher Scientific Maxima SYBR Green/ROX qPCR Master Mix (Cat#K022; https://www.thermofisher.com). Two independent, biological replicates were analyzed for each genotype, and repeated twice with the same results. All samples were normalized to the constitutively expressed genes At3g18780, At5g09810, and At1g49240 (Czechowski et al., 2005). Samples were analyzed using CFX Manager ™ (Instruction Manual - http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_10010424.pdf).

Relative Complementation Values (RCV)

$$\text{RCV} = \boxed{1 - \left(\frac{(\mu\text{BLN} - \mu\text{CLN})}{(\mu b2b3\text{LN} - \mu\text{CLN})}\right)}$$

To compare many independent experiments, a quantitative value (Relative Complementation Value, RCV) was calculated for each NF-YB transgene. Where μBLN is the mean leaf number of each NF-YB transgene population,

μCLN is mean leaf number of each Col-0 population, and $\mu b_2 b_3 LN$ *is* the mean leaf number of the *nf-yb2 nf-yb3* double mutant population.

First, a value is generated by comparing the difference between each independent NF-YB transgene population to the wild-type (Col-0) population. Next, a value is generated from comparing the difference of the parental late flowering background (*nf-yb2 nf-yb3 double mutant)* from wild-type (Col-0). These values are then combined by dividing the first by the second, creating a value for each independent NF-YB transgene that has been normalized to both wild type and the parental double mutant background. Finally, these values are subtracted from 1 to create a loss ratio, a balance point between the reference genotypes Col-0 and *nf-yb2 nf-yb3*, with Col-0 RCV always equaling 1 and *nf-yb2 nf-yb3* RCV always equaling 0.

References

Abascal, F., Corpet, A., Gurard-Levin, Z. A., Juan, D., Ochsenbein, F., Rico, D., . . . Almouzni, G. (2013). Subfunctionalization via adaptive evolution influenced by genomic context: the case of histone chaperones ASF1a and ASF1b. *Mol Biol Evol, 30*(8), 1853-1866. doi:10.1093/molbev/mst086

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., . . . Yanofsky, M. F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences, 97*(10), 5328-5333.

Baxevanis, A. D., Arents, G., Moudrianakis, E. N., & Landsman, D. (1995). A variety of DNA-binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Res, 23*(14), 2685-2691.

Becker, A., & Theißen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular phylogenetics and evolution, 29*(3), 464-489.

Ben-Naim, O., Eshed, R., Parnis, A., Teper-Bamnolker, P., Shalit, A., Coupland, G., . . . Lifschitz, E. (2006). The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *The Plant Journal, 46*(3), 462-476.

Benfey, P. N., & Chua, N. H. (1990). The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. *Science, 250*(4983), 959-966. doi:10.1126/science.250.4983.959

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol, 212*(4), 563-578. doi:10.1016/0022-2836(90)90223-9

Cagliari, A., Turchetto-Zolet, A. C., Korbes, A. P., Maraschin Fdos, S., Margis, R., & Margis-Pinheiro, M. (2014). New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants. *Genomics, 103*(5-6), 380-387. doi:10.1016/j.ygeno.2014.03.005

Cai, X., Ballif, J., Endo, S., Davis, E., Liang, M., Chen, D., . . . Wu, Y. (2007). A putative CCAAT-binding transcription factor is a regulator of flowering timing in Arabidopsis. *Plant Physiology, 145*(1), 98-105.

Cao, S., Kumimoto, R. W., Gnesutta, N., Calogero, A. M., Mantovani, R., & Holt, B. F., 3rd. (2014). A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in Arabidopsis. *Plant Cell, 26*(3), 1009-1017. doi:10.1105/tpc.113.120352

Cao, S., Kumimoto, R. W., Siriwardana, C. L., Risinger, J. R., & Holt III, B. F. (2011). Identification and characterization of NF-Y transcription factor families in the monocot model plant Brachypodium distachyon. *PloS one, 6*(6), e21805.

Ceribelli, M., Dolfini, D., Merico, D., Gatta, R., Viganò, A. M., Pavesi, G., & Mantovani, R. (2008). The histone-like NF-Y is a bifunctional transcription factor. *Molecular and cellular biology, 28*(6), 2047-2058.

Cho, S., Jang, S., Chae, S., Chung, K. M., Moon, Y.-H., An, G., & Jang, S. K. (1999). Analysis of the C-terminal region of Arabidopsis thaliana APETALA1 as a transcription activation domain. *Plant molecular biology, 40*(3), 419-429.

Combier, J.-P., Frugier, F., De Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., . . . Crespi, M. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. *Genes & development, 20*(22), 3084-3088.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., . . . Turnbull, C. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science, 316*(5827), 1030-1033.

Deng, C., Cheng, C. H., Ye, H., He, X., & Chen, L. (2010). Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc Natl Acad Sci U S A, 107*(50), 21593-21598. doi:10.1073/pnas.1007883107

Des Marais, D. L., & Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature, 454*(7205), 762-765. doi:10.1038/nature07092

Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., & Pikaard, C. S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J, 45*(4), 616-629. doi:10.1111/j.1365-313X.2005.02617.x

Frontini, M., Imbriano, C., Manni, I., & Mantovani, R. (2004). Cell cycle regulation of NF-YC nuclear localization. *Cell Cycle, 3*(2), 217-222.

Gnesutta, N., Kumimoto, R. W., Swain, S., Chiara, M., Siriwardana, C., Horner, D. S., . . . Mantovani, R. (2017). CONSTANS Imparts DNA Sequence Specificity to the Histone Fold NF-YB/NF-YC Dimer. *Plant Cell, 29*(6), 1516-1532. doi:10.1105/tpc.16.00864

Gnesutta, N., Nardini, M., & Mantovani, R. (2013). The H2A/H2B-like histone-fold domain proteins at the crossroad between chromatin and different DNA metabolisms. *Transcription, 4*(3), 114-119.

Goda, H., Nagase, T., Tanoue, S., Sugiyama, J., Steidl, S., Tüncher, A., . . . Kato, M. (2005). Nuclear translocation of the heterotrimeric CCAAT binding factor of Aspergillus oryzae is dependent on two redundant localising signals in a single subunit. *Archives of microbiology, 184*(2), 93-100.

Gusmaroli, G., Tonelli, C., & Mantovani, R. (2001). Regulation of the CCAAT-Binding NF-Y subunits in Arabidopsisthaliana. *Gene, 264*(2), 173-185.

Hackenberg, D., Wu, Y., Voigt, A., Adams, R., Schramm, P., & Grimm, B. (2012). Studies on differential nuclear translocation mechanism and assembly of the three subunits of the Arabidopsis thaliana transcription factor NF-Y. *Mol Plant, 5*(4), 876-888. doi:10.1093/mp/ssr107

Hilioti, Z., Ganopoulos, I., Bossis, I., & Tsaftaris, A. (2014). LEC1-LIKE paralog transcription factor: how to survive extinction and fit in NF-Y protein complex. *Gene, 543*(2), 220-233. doi:10.1016/j.gene.2014.04.019

Hittinger, C. T., & Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. *Nature, 449*(7163), 677-681. doi:10.1038/nature06151

Hoekema, A., Hirsch, P., Hooykaas, P., & Schilperoort, R. (1983). A binary plant vector strategy based on separation of vir-and T-region of the Agrobacterium tumefaciens Ti-plasmid.

Holt, B. F., 3rd, Boyes, D. C., Ellerstrom, M., Siefers, N., Wiig, A., Kauffman, S., . . . Dangl, J. L. (2002). An evolutionarily conserved mediator of plant disease resistance gene function is required for normal Arabidopsis development. *Dev Cell, 2*(6), 807-817.

Imaizumi, T., & Kay, S. A. (2006). Photoperiodic control of flowering: not only by coincidence. *Trends Plant Sci, 11*(11), 550-558.

Imaizumi, T., Tran, H. G., Swartz, T. E., Briggs, W. R., & Kay, S. A. (2003). FKF1 is essential for photoperiodic-specific light signalling in Arabidopsis. *Nature, 426*(6964), 302-306.

Jang, S., Marchal, V., Panigrahi, K. C., Wenkel, S., Soppe, W., Deng, X. W., . . . Coupland, G. (2008). Arabidopsis COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. *EMBO J, 27*(8), 1277-1288. doi:10.1038/emboj.2008.68

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*(13), 5339-5354. doi:10.1128/MCB.25.13.5339-5354.2005

Kim, I. S., Sinha, S., de Crombrugghe, B., & 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*(8), 4003-4013.

Kumimoto, R. W., Adam, L., Hymus, G. J., Repetti, P. P., Reuber, T. L., Marion, C. M., . . . Ratcliffe, O. J. (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*(5), 709-723. doi:10.1007/s00425-008-0773-6

Kumimoto, R. W., Zhang, Y., Siefers, N., & Holt, B. F., 3rd. (2010). NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. *Plant J, 63*(3), 379-391. doi:10.1111/j.1365-313X.2010.04247.x

Kwong, R. W., Bui, A. Q., Lee, H., Kwong, L. W., Fischer, R. L., Goldberg, R. B., & Harada, J. J. (2003). LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell, 15*(1), 5-18.

Latchman, D. S. (1997). Transcription factors: an overview. *The international journal of biochemistry & cell biology, 29*(12), 1305-1312.

Lee, H., Fischer, R. L., Goldberg, R. B., & Harada, J. J. (2003). Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proceedings of the National Academy of Sciences, 100*(4), 2152-2156.

Li, W.-X., Oono, Y., Zhu, J., He, X.-J., Wu, J.-M., Iida, K., . . . Zhu, J.-K. (2008). The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell, 20*(8), 2238-2251.

Liu, J.-X., & Howell, S. H. (2010). bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis. *Plant Cell, 22*(3), 782-796.

Lotan, T., Ohto, M.-a., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., . . . Harada, J. J. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell, 93*(7), 1195-1205.

Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics, 154*(1), 459-473.

Maity, S., Sinha, S., Ruteshouser, E., & De Crombrugghe, B. (1992). Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. *Journal of Biological Chemistry, 267*(23), 16574-16580.

Mantovani, R. (1999). The molecular biology of the CCAAT-binding factor NF-Y. *Gene, 239*(1), 15-27.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., . . . Mantovani, R. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell, 152*(1-2), 132-143. doi:10.1016/j.cell.2012.11.047

Nardone, V., Chaves-Sanjuan, A., & Nardini, M. (2017). Structural determinants for NF-Y/DNA interaction at the CCAAT box. *Biochim Biophys Acta, 1860*(5), 571-580. doi:10.1016/j.bbagrm.2016.09.006

Nelson, D. E., Repetti, P. P., Adams, T. R., Creelman, R. A., Wu, J., Warner, D. C., . . . Donnarummo, M. G. (2007). Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. *Proceedings of the National Academy of Sciences, 104*(42), 16450-16455.

Piatigorsky, J., & Wistow, G. (1991). The recruitment of crystallins: new functions precede gene duplication. *Science, 252*(5009), 1078-1079.

Samach, A., Onouchi, H., Gold, S. E., Ditta, G. S., Schwarz-Sommer, Z., Yanofsky, M. F., & Coupland, G. (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science, 288*(5471), 1613-1616.

Siefers, N., Dang, K. K., Kumimoto, R. W., Bynum, W. E. t., Tayrose, G., & Holt, B. F., 3rd. (2009). Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiol, 149*(2), 625-641. doi:10.1104/pp.108.130591

Sinha, S., Kim, I.-S., Sohn, K.-Y., De Crombrugghe, B., & 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. *Molecular and cellular biology, 16*(1), 328-337.

Siriwardana, C. L., Gnesutta, N., Kumimoto, R. W., Jones, D. S., Myers, Z. A., Mantovani, R., & Holt III, B. F. (2016). NUCLEAR FACTOR Y, Subunit A (NF-YA) Proteins Positively Regulate Flowering and Act Through FLOWERING LOCUS T. *PLoS Genetics, 12*(12), e1006496.

Steidl, S., Tuncher, A., Goda, H., Guder, C., Papadopoulou, N., Kobayashi, T., . . . Brakhage, A. A. (2004). A single subunit of a heterotrimeric CCAAT-binding complex carries a nuclear localization signal: piggy back transport of the pre-assembled complex to the nucleus. *J Mol Biol, 342*(2), 515-524. doi:10.1016/j.jmb.2004.07.011

Suárez-López, 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*(6832), 1116-1120.

Tüncher, A., Spröte, P., Gehrke, A., & Brakhage, A. A. (2005). The CCAAT-binding complex of eukaryotes: evolution of a second NLS in the HapB subunit of the filamentous fungus Aspergillus nidulans despite functional conservation at the molecular level between yeast, A. nidulans and human. *Journal of molecular biology, 352*(3), 517-533.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., & Coupland, G. (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. *Science, 303*(5660), 1003-1006.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, 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*(11), 2971-2984.

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., & Hattori, T. (2009). Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate

transcription by interacting with seed-specific ABRE-binding factors. *The Plant Journal, 58*(5), 843-856.