RESEARCH ARTICLE

CONSTANS imparts DNA sequence-specificity to the histone-fold NF-YB/NF-YC dimer

Nerina Gnesutta1, Roderick W. Kumimoto2, Swadhin Swain1, Matteo Chiara1, Chamindika Siriwardana2, David S. Horner1, Ben F. Holt III2, Roberto Mantovani§

1) Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133, Milano, Italy.
2) Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA.
§ Corresponding Author: mantor@unimi.it

Short title: The CO trimeric DNA-binding complex

One-sentence summary: CONSTANS forms a trimeric complex with At-NF-YB2/NF-YC3 to efficiently bind the CORE element: the CCT provides trimerization and DNA-binding, akin to NF-YA but divergent in sequence selectivity.

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.plantcell.org) is: Roberto Mantovani (mantor@unimi.it).

ABSTRACT

Nuclear Factor Y (NF-Y) is a heterotrimeric transcription factor that binds CCAAT elements. The NF-Y trimer is composed of a Histone Fold Domain (HFD) dimer (NF-YB/NF-YC) and NF-YA, which confers DNA sequence-specificity. NF-YA shares a conserved domain with the CONSTANS, CONSTANS-LIKE, TOC1 (CCT) proteins. We show that CONSTANS (CO/B-BOX PROTEIN 1 BBX1), a master flowering regulator, forms a trimer with Arabidopsis thaliana NF-YB2/NF-YC3 to efficiently bind the CORE element of the FLOWERING LOCUS T (FT) promoter. We term this complex NF-CO. Using saturation mutagenesis, electrophoretic mobility shift assays, and RNA-sequencing profiling of co, nf-yb, and nf-yc mutants, we identify CCACA elements as the core NF-CO binding site. CO physically interacts with the same HFD surface required for NF-YA association, as determined by mutations in NF-YB2 and NF-YC9, and tested in vitro and in vivo.

The co-7 mutation in the CCT domain, corresponding to an NF-YA arginine directly involved in CCAAT recognition, abolishes NF-CO binding to DNA. In summary, a unifying molecular mechanism of CO function relates it to the NF-YA paradigm, as part of a trimeric complex imparting sequence-specificity to HFD/DNA interactions. It is likely that members of the large CCT family participate in similar complexes with At-NF-YB and At-NF-YC, broadening HFD combinatorial possibilities in terms of trimerization, DNA-binding specificities, and transcriptional regulation.

INTRODUCTION

In all eukaryotes, the precise regulation of transcription of any given gene is ultimately determined by the combinatorial binding of sequence-specific transcription factors (TFs) to their target sequences within promoters, enhancers, and other genomic regulatory regions. These DNA-protein complexes serve as platforms for recruitment of coactivators,
most of which contain enzymatic activities that impact local chromatin organization. Specifically, they either modify DNA directly or the tails of DNA-bound, nucleosomal core histones. Histones have a central, globular Histone Fold Domain (HFD) that is required for heterodimerization, tetramerization, and non sequence-specific binding to DNA as an octameric structure (Luger et al., 1997). HFDs are not specific to core histones, but are shared by other proteins, such as the Nuclear Factor-Y (NF-Y) TF. Canonically, NF-Y is a trimeric complex composed of an NF-YB/NF-YC dimer, homologous to the H2B/H2A type histone heterodimer (Romier et al., 2003), and NF-YA, the subunit conferring sequence specificity (Huber et al., 2012; Nardini et al., 2013). The targeted DNA sequence is the CCAAT box, originally discovered in human promoters, and later found in all eukaryotes. Mechanistically, different experiments in the mammalian system suggested that the CCAAT box and NF-Y serve a “pioneering” role in gene activation, namely that this TF is able to penetrate “hostile” chromatin territories and set the stage for binding of other TFs required for full gene activation (Fleming et al., 2013; Sherwood et al., 2014; Oldfield et al., 2014). This hypothesis is further supported by recent experiments with mouse zygotes, where NF-Y appears to be the major TF opening chromatin as early as the 2- and 4-cell stages (Lu et al., 2016).

The genes coding for the three NF-Y subunits have been identified in essentially all eukaryotes and are among the most evolutionarily conserved proteins described to date. Notably, conserved domains include the HFDs, required for heterodimerization and non sequence-specific DNA-binding, and a stretch of 56 amino acids of NF-YA, required for HFD association and specific CCAAT-binding. In mammals, invertebrates, and fungi, there are one or two genes coding for each subunit. Instead, plants have dramatically expanded the number of NF-Y genes: typically, there are 8-14 gene family members for each subunit, conferring an enormous combinatorial capacity on the trimer; some are expressed in a tissue-restricted manner, many are relatively ubiquitous (Gusmaroli et al., 2001, 2002; Stephenson et al., 2007; Siefers et al., 2009; Cao et al., 2011b; Hilioti et al., 2014; Liang et al., 2014; Quach et al., 2014; Ripodas et al., 2015; Zhang et al., 2015; Qu et al., 2015; Feng et al., 2015; Ren et al., 2016; Malviya et al., 2016; Li et al., 2016; Yang et al., 2016; Zhang et al., 2016). Typical features of other plant TFs, such as the presence of
duplicate members with similar functions and neofunctionalization of specific genes, were determined by genetic experiments, mostly performed in *Arabidopsis thaliana* (reviewed in Laloum et al., 2012; Petroni et al., 2012). A growing body of evidence indicates that specific At-NF-Y subunits are involved in disparate physiological events in plant development, growth, and reproduction, as well as in adaptation to physiological and adverse environmental conditions.

One aspect of plant development in which specific NF-Y genes were shown to be important is the regulation of photoperiod-dependent flowering: at least two At-NF-YB (At-NF-YB2, At-NF-YB3) and three At-NF-YC subunits (At-NF-YC3, At-NF-YC4, At-NF-YC9) are involved in floral timing (Wenkel et al., 2006; Kumimoto et al., 2008; 2010; Cao et al., 2014). At-NF-YB and At-NF-YC subunits can physically interact with CONSTANS (CO), which is also an essential regulator of photoperiod-dependent flowering (Putterill et al., 1995; reviewed in Song et al., 2015). CO contains a conserved CCT (CONSTANS, CONSTANS-LIKE, TOC1) domain, which is shared by >30 proteins in Arabidopsis and similar numbers in other angiosperms. Interestingly, the CCT domain is homologous to the HFD-interaction and *CCAAT*-binding domain of NF-YA; CO (as well as CO-like and TOC1 proteins) has been shown to bind several NF-Y HFD subunits (Wenkel et al., 2006; Ben-Naim et al., 2006; Kumimoto et al., 2010; Cao et al., 2011a; Distelfeld et al., 2011; Li et al., 2011; Chen et al., 2014; Hou et al., 2014) in a CCT domain-dependent manner (Wenkel et al., 2006). The sequence identity/similarity between CO and NF-YA is particularly evident in the sub-domain required in NF-YA for *CCAAT* recognition (Wenkel et al., 2006; Petroni et al., 2012). Furthermore, CO and the HFDs mentioned above participate in the same genetic pathway controlling flowering (Cao et al., 2014), as both CO and NF-Y bind and regulate the expression of *FLOWERING LOCUS T* (*FT*- aka, the principal florigen), through a *CCAAT* box in its enhancer and CO-Responsive Elements (*COREs*) in its promoter, respectively (Adrian et al., 2010, Song et al., 2012; Cao et al., 2014; Bu et al., 2014).

*COREs* have been identified through analysis of evolutionary conservation of the *FT* promoter in different plant species and through mutagenesis and functional analysis of the
promoter in vivo (Adrian et al., 2010; Tiwari et al., 2010; Cao et al., 2014). Despite the wealth of genetic knowledge, the molecular mechanisms of the activity of CO and CO-like proteins are not completely understood (Blackman and Michaels, 2010). In some experiments, CO, and the related TIMING OF CAB EXPRESSION 1 (TOC1), were shown to be stand-alone sequence-specific TFs capable of directly binding COREs (Tiwari et al., 2011; Gendron et al., 2013; Abelenda et al., 2016). In others, CO-like proteins compete with NF-YA for association with specific HFDs, thus influencing NF-Y transcriptional activity (Li et al., 2012). Finally, it is also conceivable that CO might form a quaternary complex with NF-Y (through CO/HFD interactions): in this model, CO would essentially act as a coactivator (Cao et al., 2014).

The lack of a general consensus as to the molecular mechanism of CO function, and the observation of the fundamental similarity between NF-YA and CCT conserved domains (Wenkel et al., 2006) drove the experiments reported here. We reasoned that CO, and by inference all CCT proteins, are “NF-YA-likes”, associating with HFD heterodimers, binding to DNA with robust and specific affinity only in the trimeric configuration. We set out biochemical and genetic experiments to test this hypothesis.

RESULTS

The CO/At-NF-YB2/NF-YC3 trimer binds to DNA in vitro

CO was previously shown to interact directly with several NF-YB/NF-YCs (Summarized in Supplemental Fig. 1). To test whether it forms a DNA-binding trimer, we expressed in E. coli and independently purified the CCT domain of CO, previously shown to be sufficient for HFD interactions (Wenkel et al., 2006); in parallel, the HFD dimer At-NF-YB2/NF-YC3 was also produced by co-expressing the subunits (Supplemental Fig. 2). We used this specific heterodimer because of the genetic evidence implicating the two genes in the regulation of the timing of flowering, of a known in vivo interaction, and of biochemical data suggesting CO interactions (Wenkel et al., 2006; Kumimoto et al., 2008; 2010). We used the purified proteins in electrophoretic mobility shift assays (EMSAs) with a Cy5-labeled 31-mer oligonucleotide containing the functionally important FT CO
Responsive Element 2 (CORE2) (Adrian et al., 2010; Tiwari et al., 2010; Cao et al., 2014). The results demonstrated that a CO/HFD dimer complex, but not CO alone, efficiently bound FT CORE2 (Figure 1 A). Note that at high CO concentrations a very faint, faster-migrating DNA complex could be observed in the absence of HFDs upon long exposures (Supplemental Fig. 3). Incubation of CO and HFD subunits with a labeled, functional CCAAT from the FT enhancer (Cao et al., 2014) yielded residual binding only at high CO concentrations. Instead, addition of At-NF-YA2 or At-NF-YA6 generated the NF-Y complex on CCAAT, as expected, but not on CORE2 (Fig. 1A). The specificity of the CO/HFD complex was then assayed by competition analysis with different unlabeled oligos containing wild-type (wt) or mutant CORE2, CCAAT or an unrelated sequence. Unlabeled, wt CORE2 competitors interfered with binding in a concentration- and position-specific manner; CORE2 oligos with mutations known to reduce FT expression in vivo (Tiwari et al., 2010; Cao et al., 2014), FT CCAAT oligos, or an unrelated sequence did not reduce binding of the labeled probe (Fig. 1B). We conclude that CO forms a complex with NF-Y HFD subunits, which binds to the CORE sequence with high affinity and specificity. By analogy to the NF-Y acronym, we refer to the CORE-binding trimer as NF-CO.

**NF-CO binds the core pentamer CCACA, with preferred flanking sequences**

To pinpoint precisely the DNA-binding requirements of NF-CO, we initially challenged the complex with ten unlabeled 30-mers containing 3 bp scanning mutations (mI oligos, Figure 2 A). Four oligos (mI-3 to 6) showed loss of competition, indicating reduced or absent NF-CO interactions (Fig. 2A, Supplemental Fig. 4), whereas mutations in the flanking areas had a negligible effect on complex formation. We then dissected this 12 bp central region with six oligos harboring 2 bp mutations (mII oligos). Again, four oligonucleotides (mII-3 to 6) competed poorly, trimming down the targeted element to 8 bp (Fig. 2A). Finally, we used 1 bp mutations, changing each of the 8 bases to all other 3 nucleotides (mIII series, Fig. 2 and Supplemental Fig. 4). This led to the definition of a central TGTGG pentanucleotide, or CCACA on the reverse strand, with preferred flanking sequences, as the optimal in vitro binding site of NF-CO (Fig. 2C).
In vivo RNA-Seq analysis identifies the CCACA pentamer in promoters of genes regulated by NF-CO subunits

To identify genes regulated by NF-CO subunits in vivo, we performed RNA-Seq analysis on previously described, late flowering co-sail, nf-yb2 nf-yb3, and nf-yc3 nf-yc4 nf-yc9 mutants. The requirement for the double and triple HFD subunit mutants is due to negligible phenotypic effects on flowering timing of single At-NF-YB and single or double At-NF-YC mutants (Kumimoto et al., 2008, 2010). The complete list of genes whose expression was affected in the mutants is in Supplemental Data Set 1. Of the 1,690 genes significantly down-regulated (FDR<0.05 in both Limma and DESeq2 analyses, see Materials and Methods) in co-sail, 955 were shared in nf-yb2 nf-yb3 ($P_{\text{hypergeometric}} = <10e^{-127}$), 624 in nf-yc3 nf-yc4 nf-yc9 ($P_{\text{hypergeometric}} = <10e^{-127}$), and 398 were common in all three mutants ($P_{\text{hypergeometric}} = <10e^{-127}$) (Figure 3). The overlap among the HFD subunit mutants was somewhat lower, but still striking: 473 of the 1,615 down-regulated genes in nf-yb2 nf-yb3 are shared with nf-yc3 nf-yc4 nf-yc9 ($P_{\text{hypergeometric}} = <10e^{-127}$). Up-regulated gene sets also showed robust overlaps (Fig. 3). We did find genes previously known to be dependent upon CO and HFD activity, including FT (Kumimoto et al., 2008; 2010). We validated 16 genes of the cohort down-regulated in both nf-yb2 nf-yb3 and co-sail by reverse transcription quantitative PCR (RT-qPCR) analysis, and all showed the expected changes, thus confirming the robustness of the RNA-Seq data (Supplemental Fig. 5). We then retrieved the promoter sequences (-1 kb to TSS/ATG) of affected genes and analyzed them with Weeder, an algorithm for de novo DNA motif discovery (Zambelli et al., 2013). For genes up-regulated in mutants, a matrix resembling a GATA box (Reyes et al., 2004) and a second motif, unrelated to CORE or CCAAT, emerged, suggesting indirect effects on other TFs. On the other hand, three similar, but not identical, matrices emerged in down-regulated genes for each dataset (Figs. 3 and S6). CCACACA was found in the co-sail and co-sail by nf-yb2 nf-yb3 intersection, which is similar to the “morning element” (Harmer and Kay, 2005; Michael et al., 2008) that is important for circadian clock regulation (Liu et al., 2016). The CCACATA4 sequence, differing by one bp, was found in the nf-yb2 nf-yb3 cohort. Note that in the NF-CO mutagenesis experiments of Fig. 2, a C or a T at this position are essentially equivalent. Finally, the CCACGTG motif, resembling a G-Box, and previously described in TOC1 and PRR chromatin immunoprecipitation (ChIP)-seq
experiments and in promoters of genes up-regulated after TOC1 overexpression (Gendron et al., 2012; Liu et al., 2016), was recovered from the nf-yc3 nf-yc4 nf-yc9 cohort and was the most enriched element in intersections involving this cohort. Collectively, these elements all contain or closely resemble the CCACA core motif identified by in vitro EMSAs as optimal for NF-CO. In summary, in vivo RNA-Seq analysis was consistent with in vitro biochemical data, both identifying CCACA as the NF-CO matrix.

**Analysis of NF-CO regulated genes**

Next, we further analyzed the NF-CO regulated genes for circadian expression with the Phaser tool, which uses a database derived from microarray analyses of circadian and diurnal gene expression patterns (Mockler et al., 2007). For a long-day light regime, genes upregulated in co-sail and nf-yb2 nf-yb3 were highly significantly enriched for pre-dawn expression (hours 21-22 of the long-day cycle). Alternatively, down-regulation of expression in these mutants was correlated with morning expressed genes (hours 2-5), as well as distinct peaks for genes normally expressed in the 10th and 15-16th hours (Supplemental Fig. 7, Supplemental Data Set 2a). Indeed, a number of genes believed to be involved in circadian regulation (including GIGANTEA, CIRCADIAN CLOCK ASSOCIATED 1, LATE ELONGATED HYPOCOTYL, ELONGATED HYPOCOTYL 5, CONSTANS-LIKE 2, and several pseudo-response regulator –PRR– genes), as well as diurnal markers such as early light-inducible protein genes ELIP1 and ELIP2, showed significant differential expression, often with a relatively large fold-change of expression levels between mutants and wild type (Supplemental Fig. 8A). For all mutants tested, and for all intersections between their down-regulated genes, multiple GO terms related to plastid locations and functions were highly significantly enriched (Supplemental Data Set 2b). For up-regulated genes, terms related to the plasma membrane and cell wall, as well as response to carbohydrate stimulus, were consistently over-represented. No particular enrichment of motifs corresponding to the previously identified CCACA matrices was noted in promoters of plastid genes. Taken together, these observations are consistent with fundamental alterations to the regulation of circadian processes in co-sail and the nf-yb mutants tested, at least under the light conditions employed in the current experiments.
We also specifically looked at the expression of paralogs of At-NF-Y subunits and CCT genes. Among the At-NF-Y paralogs, expression of At-NF-YA4/2/5/6/7/9 was detected. CO mRNA itself is expressed at low levels in the wild type, unlike many of the CO-like (COL) and PRR genes, which are more abundantly expressed than At-NF-YA genes (Supplemental Fig. 8B and Supplemental Data Set 2c). Few significant changes in expression were observed, with NF-YA4 showing down-regulation in co-sail and the nf-yb mutants, At-NF-YB8 and At-NF-YB10 up-regulated in nf-yb2 nf-yb3 and nf-yc3 nf-yc4 nf-yc9 mutants, while At-NF-YB7 was down-regulated in all mutants tested (Supplemental Fig. 8B, Supplemental Data Set 2c). Among the down-regulated genes in co-sail, we find many members of the B-box protein (BBX) family (Khanna et al., 2009), mostly those that do not contain a CCT domain. These include BBX19 (At4g38960), whose protein product interacts with CO and whose reduced expression accelerates flowering (Wang et al., 2014), and BBX30 (At4g15248) and BBX31 (At3g21890), which also interact with the CO protein and whose overexpression delays flowering (Graeff et al., 2016), as well as BBX32, which interacts with COL3 to regulate FT (Tripathi et al., 2017).

**At-NF-YB2-NF-YB3 are essential for CO recruitment onto the FT promoter**

To verify the importance of HFD subunits for CO association to DNA *in vivo*, we performed ChIP analysis on Col-0 and nf-yb2 nf-yb3 plants transgenic for CO-YFP/HA under the control of the CaMV 35S promoter (p35S:CO-YFP/HA). We used an HA antibody to assess the overexpression of the transgenes in the two genetic backgrounds (Supplemental Fig. 9), and then for ChIP analysis. The immunoprecipitated DNAs were checked by qPCR with three amplicons within the FT locus: at -5.3 kb, a region of the distal promoter which contains the CCAAT box, but no known CORE elements (negative control); at -0.3 kb, a region of the proximal promoter where CORE2 is located and CO is known to bind (Song et al., 2012; Cao et al., 2014); and at +2.0 kb, corresponding to Exon 4 (negative control). Figure 4 shows that CO binding was detected only in the core promoter in wild-type plants, as expected, but not in the nf-yb2 nf-yb3 mutants. Note that the -0.3 kb region contains CORE2, CORE1, and two additional essential CCACA motifs, termed P1/P2 (Adrian et al., 2010), all within 100 bp. With the current precision of the ChIP procedure it is not possible to discriminate the exact binding site(s) bound by CO, but
previous evidence suggests that CO interacts with several \textit{CCACA}-containing sites in the proximal promoter (Adrian et al., 2010; Cao et al., 2014). Regardless of the specific \textit{CCACA}(s) bound, these data show that At-NF-YB2 and At-NF-YB3 are required for CO binding at the \textit{FT} proximal promoter.

\textbf{Similar HFD structural elements provide association with NF-YA and CO, and are important for the timing of flowering}

In canonical NF-Y complexes, NF-YA cannot bind to single NF-YB or NF-YC subunits (Kim et al., 1996; Sinha et al., 1996). This result was later rationalized by knowledge of the quaternary NF-Y/\textit{CCAAT} 3D structure, showing instead that NF-YA binds to a composite surface formed by the \(\alpha2\) helix of NF-YB and the \(\alpha1/\alphaC\) helices of NF-YC; (Nardini et al., 2013; Huber et al., 2012). In keeping with these data, in a systematic study of subunit interactions using yeast two-hybrid (Y2H) assays, direct interactions between At-NF-YAs and the single HFD subunits were generally not observed (Hackenberg et al., 2012). The original Y2H screenings with CO identified interactions with At-NF-YB or At-NF-YC subunits (Wenkel et al., 2006; Ben-Naim et al., 2006), a result further confirmed by other Y2H studies (Supplemental Fig. 1). Because the A1 trimerization domain of NF-YA does not superimpose perfectly with the corresponding area in the CCT domain (Petroni et al., 2012), we wondered whether the CO interaction regions of the HFD heterodimers were equivalent to the ones contacted by At-NF-YA. To evaluate this, we designed specific mutations in At-NF-YC9 and At-NF-YB2 known to affect NF-Y trimer formation and assayed them \textit{in vivo} and \textit{in vitro}.

For \textit{in vivo} testing of At-NF-YC9, we focused on amino acids Phe151 and Val153 in the \(\alphaC\) helix, mutating these residues to Arg and Lys. These same mutations were originally described for mammalian NF-Y where they eliminated formation of the heterotrimer and DNA binding, but had no impact on formation of the HFD heterodimer (Kim et al., 1996). Note that within the HFD, At-NF-YC9 is identical to At-NF-YC3 and nearly identical to At-NF-YC4 (Figure 5 A), and all three proteins have known functional overlap in several processes (Kumimoto et al., 2010; Myers et al., 2016), hence any data obtained with mutants of the At-NF-YC9 HFD are likely valid for all three proteins. We transformed the
nf-yc3 nf-yc4 nf-yc9 triple mutant with At-NF-YC9 under the control of its own promoter, either in the wild-type (wt) or F151R/V153K mutant configuration. As shown in Fig. 5B, compared to the wild type, triple mutant plants had significantly delayed flowering, as previously reported (Kumimoto et al., 2010). This delay was almost completely reverted by wt At-NF-YC9, but not by the At-NF-YC9-F151R/V153K mutant, despite generally robust expression patterns of the mutant transgene (Fig. 5C). To check whether the double mutation impaired formation of subunits, we performed Y2H with wt and mutant At-NF-YC9. Fig. 5D shows that the wt At-NF-YC9 interacts with all tested partners (At-NF-YB2, At-NF-YA1, At-NF-YA2, and CO), whereas the mutant only interacted with the HFD partner, and was unable to contact At-NF-YAs or CO. As a further control, we used a mutant in a conserved position in the α2 helix (Ile89) previously shown to affect HFD hetero-dimerization in mammalian NF-Y subunits (Kim et al., 1996); indeed, the I89D mutant lost At-NF-YB2 interactions in Y2H assays, which agreed with predictions, and supported the specificity of the Y2H data.

Next, we switched to in vitro EMSAs with recombinant proteins. In this case, we generated the single F151R mutation in At-NF-YC9, since mutation of the (non conserved) Ile residue corresponding to Val153 (Fig. 5A) was previously shown not to impair NF-YA interaction (Romier et al., 2003). This result was later rationalized by the central role played by the perfectly conserved phenylalanine in the nucleation of the hydrophobic core driving the correct positioning of the NF-YC αC helix. The acidic NF-YC αC provides crucial A1-contacts and is further stabilized by NF-YA interactions with main chain atoms of the same Phe residue (Nardini et al., 2013). Coexpression and purification of both wt At-NF-YC9 and the F151R mutant with At-NF-YB2 was equally efficient, as expected, indicating similar heterodimerization capacities (Supplemental Fig. 2). The purified HFDs were then incubated with either CO or At-NF-YA2, and tested for interaction with the CORE2 and CCAAT probes. As expected, wt At-NF-YC9 was able to form efficient DNA-binding complexes with either CO or At-NF-YA2 on their respective DNA targets, while the At-NF-YC9F151R mutation led to very inefficient binding to either probe (Fig. 5E). Altogether, these data indicate that an At-NF-YC mutation that interfered with heterotrimer
formation and CCAAT binding for the canonical NF-Y complex had essentially the same effects on the NF-CO complex at its CORE site.

On the At-NF-YB side, we employed a similar strategy, by targeting the conserved Glu65 in the α2 helix of At-NF-YB2, corresponding to Glu90 of mammalian NF-YB. In mammalian NF-Y, this acidic residue provides contacts with two conserved arginines (Arg249 and Arg253) of NF-YA helix A1 (Nardini et al., 2013). Similarly to the NF-YC mutagenesis described above, detailed biochemical analyses have previously shown that the mammalian NF-YB E90R mutation does not alter HFD dimerization, but impairs trimerization and DNA binding (Sinha et al., 1996). As expected, both At-NF-YB2 and the At-NF-YB2E65R mutant efficiently heterodimerized with At-NF-YC3 (Supplemental Fig. 2), but the E65R mutant lost functional NF-Y binding to CCAAT in EMSAs (Figure 6). Replacing At-NF-YA2 with CO and testing binding to the CORE2 probe gave a similar result: NF-CO binding was reduced, albeit not completely eliminated (Fig. 6). Note that this mutant, unlike wt At-NF-YB2, could not rescue a late flowering nf-yb2 nf-yb3 mutant (Siriwardana et al., 2016), paralleling the At-NF-YC mutation shown above. In summary, the same conserved residues of At-NF-YC9 αC and At-NF-YB2 α2 are important for trimerization with At-NF-YA and CO, for binding of the NF-Y and NF-CO trimers to their respective DNA sites, and for function in vivo.

**Mutation of a single amino acid in CO that is highly conserved in both CCT and NF-YA families eliminates NF-CO DNA binding**

Having established that the docking sites on the HFD dimers are similar, we switched to analysis of the properties of the CCT. Several single-residue alterations, mutations, or natural variations in the CCT domain of CO and CO-like proteins were previously reported; importantly, these alterations of CCT family members were often pinpointed in genetic screenings as having functional consequences in flowering timing (Distelfeld et al., 2004). In particular, the Arabidopsis co-7 allele is one of the later flowering co alleles (Robson et al., 2001). This mutant allele results in an arginine to glutamine change at position 340 in CO, corresponding to mammalian NF-YA Arg283, a residue known to be important for DNA binding (Xing et al., 1993; Mantovani et al., 1994) and specifically for CCAAT
recognition (Nardini et al., 2013; Huber et al., 2012). We produced and purified the co-7 CCT (Supplemental Fig. 2) and assayed it with the At-NF-YB2/NF-YC3 dimer for binding to CORE2 in EMSAs. Once again mimicking the canonical NF-Y/CCAAT interaction, NF-CO bound CORE2 with wt CO, but not with the R340Q (co-7) mutant protein, even at high concentrations (Figure 7 A). We conclude that Arg340 of the CO CCT domain, corresponding to the perfectly conserved arginine in mammalian, yeast, and plant NF-YA proteins, is equally important in DNA-binding. This strongly suggests that the DNA-binding subdomains of CO and NF-YA are structurally and mechanistically analogous and likely explains the molecular mechanism of the late flowering phenotypes of co-7 mutant plants.

DISCUSSION

Here we have demonstrated that the master flowering regulator CO interacts physically with NF-Y histone-like subunit dimers to form a novel DNA binding sequence-specific trimeric complex, NF-CO (Figure 8). Mutations within the HFDs and CCT domain indicate that the overall modalities of trimerization and DNA-binding are similar to the canonical NF-Y. In vitro experiments have defined a DNA sequence matrix that is specifically recognized by NF-CO and which is independently recovered through analyses of promoters of genes downregulated in vivo in a co mutant. Indeed, highly significant overlaps of genes dysregulated in co-sail, nf-yb, and nf-yC mutants, as well as ChIP analysis showing lack of CO binding to the FT promoter in the nf-yb2 nf-yb3 mutant, corroborate conclusions from in vitro experiments. Additionally, the recovery of core motifs related to, but distinct from, the NF-CO binding consensus in subsets of the genes differentially expressed in NF-Y HFD subunit mutants is consistent with the possibility that other CCT proteins also form sequence specific DNA binding complexes with NF-Y components.

NF-CO and NF-Y in flowering

NF-Y subunits have been implicated in a plethora of physiological plant processes. Specifically, there is well established evidence that different NF-Y HFD subunits are involved in the regulation of timing of flowering in Arabidopsis (Wenkel et al., 2006; Chen
et al., 2007; Cai et al., 2007; Kumimoto et al., 2008; 2010; Hackenberg et al., 2012; Hou et al., 2014; Cao et al., 2014), rice (Dai et al., 2008; Wei et al., 2010; Yan et al., 2011; Feng et al., 2011; Chen et al., 2014; Kim et al., 2016; Hwang et al., 2016; Goretti et al., 2017), wheat (Li et al., 2012) and tomato (Ben-Naim et al., 2006). Wenkel et al. originally observed homology between the CCT domain and the conserved domain of NF-YA, suggesting that CO and NF-YA might both bind DNA with NF-YB/NF-YC HFD dimers (Wenkel et al., 2006). The CCT domain N-terminal portion, however, is not perfectly superimposable with the NF-YA A1 helix involved in the interactions with the HFD dimer, yet it does share its highly basic nature. Our data (Figs. 5 and 6) suggest that the CCT does contact the same acidic surface patch of the HFD dimer recognized by NF-YA A1. A variant in this basic CCT region of both Heading date 1 (Hd1, the rice CO homolog) and OsPRR37 (another rice CCT protein involved in flowering) also impairs HFD interactions (Goretti et al., 2017), further solidifying the idea that multiple CCTs can form NF-CO complexes.

The notion that CO is an NF-YA equivalent, contacting similar HFD surfaces, implies that it could compete for HFD occupancy, and vice versa, that NF-YA could compete with CO. This is consistent with the finding that overexpressing some NF-YAs can cause late flowering (Wenkel et al., 2006; Li et al., 2011; Leyva-Gonzalez et al., 2012). In turn, this implies that the interpretation of the phenotypes of plants in which NF-YAs or CO, or other CCT proteins, were ubiquitously overexpressed is likely complex, since a change in the levels of either subunit could alter the stoichiometry and function of NF-Y or NF-CO complexes. While interference with HFD dimers remains distinctly possible even at physiological protein concentrations, as a general mechanism of CCT/NF-YA interplay, we note that the FT gene has two canonical sites of regulation for the two trimers: at least one NF-CO site in the core promoter and a canonical CCAAT box in its enhancer, both functionally essential for photoperiod-dependent flowering (Adrian et al., 2010; Tiwari et al., 2010; Cao et al., 2014). Thus, FT appears to be regulated by both NF-Y and NF-CO, sharing common HFD subunits.

CCAAT vs CORE
The sequence targeted by NF-Y (RRCCAAT(C/G)(A/G)) has long been known, thanks to numerous biochemical and genomic studies performed in mammals (Dolfini et al., 2009; Dolfini and Mantovani, 2013). By examining the NF-CO matrices (Fig. 2C and Fig. 3), one can notice a clear similarity to the NF-Y CCAAT matrix, RRCCAAT(C/G)(A/G), with a deviation of two nucleotides (underlined). Similarity is present at the 5’ end, where the five nucleotides RRCCA are identical, and at the 3’ ends. Given the stunning evolutionary conservation between animal and plant NF-YA proteins, specifically in the sequence-specific DNA-binding subdomain, we take for granted that plant NF-Y also binds to sequences centered on the CCAAT pentanucleotide. This was suggested by a previous study showing that the pentanucleotide, but not the mammalian matrix with the flanking nucleotides, was enriched in plant promoters (Siefers et al., 2009). The fundamental differences of the NF-Y and NF-CO matrices are indeed within the central pentanucleotide (CCAAT vs CCACCA, respectively). At position 4, a C is crucial for NF-CO, and never an A, as required by NF-Y; the A dominant for NF-CO at position 5 is detrimental in vitro and essentially never found in sites in vivo for NF-Y, at least in mammals. Thus, these two residues are clearly discriminative and divergent for the two complexes, and the respective sequence-specificity is expected to drive binding to distinct target sites. In vivo, this was shown on FT (Fig. 4; Cao et al., 2014), and in more general terms, it is documented here by the RNA-Seq analysis, since mutants of the NF-CO subunits have clearly altered the expression of genes enriched for CORE elements in their promoters. Schmid et al. (Schmid et al., 2003) previously showed that, in the apical meristem, co and ft mutants cause almost identical sets of genes to be dysregulated in LD conditions, while Wigge et al. concluded that the major contribution of CO to floral transition is mediated by its activation of FT (Wigge et al., 2005). However, this later study, employing leaf tissue, identified over 400 genes that showed CO-dependent dysregulation on transition from short to long day. Interestingly, reanalysis of this expression data (ArrayExpress experiment E-TABM-21) using Limma yielded a significant overlap between these genes and those showing dysregulation in the co-sail mutant in continuous light conditions (Supplemental Data Set 3).

The At-NF-YBs and At-NF-YCs studied in this experiment seem to have a shared impact on CCACGTG (Fig. 3), possibly the PRR site (Liu et al., 2016), and CCACA(C/T),
possibly the CO and CO-like site. For the time being, it is hard to work out a coherent model of direct and indirect targets, because of the overlap between motifs, the combinatorial complexity of the complexes, and the fact that we do not know if a specific combination of HFDs in a CO complex can subtly alter the preferred binding-site. We see a strong over-representation of plastid genes within the group of genes misregulated in all three mutants analyzed and the best motifs strongly resemble elements known to be involved in diurnal regulation. Genes that are up- or down-regulated in co-sail, as well as in HFD mutants after seven days continuous light, are strongly enriched for distinct expression times under physiological conditions. Taken together with the similarities in the gene sets observed here and previously (Wigge et al., 2005), these observations suggest that NF-CO has physiological roles that go well beyond induction of flowering and might tie in with fine tuning of plastid gene expression timing, potentially via PRRs and/or CO-like containing complexes. Further, the extensive dysregulation of B-Box genes might be consistent with a more widespread control on FT- and NF-CO-regulated genes, given the significant level of mutual regulation (Shim et al., 2017).

We were surprised by the overall lack of CCAAT boxes in profiles of the HFD subunit mutants. Note that the presence of the CCAAT pentanucleotide in a partial set of Arabidopsis promoters was detailed before (Siefers et al., 2009). We can offer several non-mutually exclusive explanations. (i) To observe the emergence of CCAAT-dependent promoters, one should analyse NF-YA mutants, as these are the subunits imparting sequence-specificity to the trimer. (ii) CORE outnumbers CCAAT because CCT encoding genes outnumber NF-YAs by a factor of 4 to 1. (iii) CCT genes are overall more abundantly expressed compared to At-NF-YAs, thus, NF-CO complexes could simply be more abundant than NF-Ys, at least in the tissue and under the experimental conditions employed here.

**Lessons about NF-CO from the NF-Y/CCAAT structure**

Knowledge of the molecular details of the quaternary 3D NF-Y/CCAAT structures (Nardini et al., 2013; Huber et al., 2012) helps rationalize the NF-CO/DNA interactions. The HFDs are clearly crucial for stable complex formation, making >25 non sequence-specific contacts with DNA, spanning at least 25 and most likely 30 nucleotides. Indeed, the
stabilization of NF-CO DNA-binding by HFDs is quite dramatic (Figures 1, 4, 5, 6 and Supplemental Fig. 3). DNA-binding of recombinant CO and TOC1 was previously reported, however with very high protein concentrations (Tiwari et al., 2010; Gendron et al., 2012), and indeed we do see this effect in our assays. Intriguingly, this low affinity binding might be consistent with very high overexpression of CO being able to partially rescue late flowering in the nf-yb2 nf-yb3 mutant (Tiwari et al., 2010). However, we provide biochemical evidence that the trimer is a much more efficient DNA-binding entity in vitro, and a physiological one in vivo, as the matrices of the HFD mutants resemble those of co. The HFD importance could be linked to their histone-like nature, as they might play the known “pioneer” role of NF-Y in penetration of chromatin territories devoid of positive histone marks (Fleming et al., 2013; Oldfield et al., 2014; Sherwood et al., 2014; Lu et al., 2016).

Structurally, there are two features differentiating CO from NF-YA: (i) the shorter CO A1-A2 linker between the HFD association and DNA-binding subdomains, which might severely constrain the flexibility of the complex; (ii) the absence in CO of two glycines within the crucial (R)GxGGRF loop of NF-YA (aa 283-289), which is (R)VNGRF in CO (aa 340-345). Overall, fourteen NF-YA amino acids are involved in DNA-binding; seven make non sequence-specific contacts, of which only NF-YA Ser273 and Arg288 are conserved (Nardini et al., 2013). Interestingly, of the seven sequence-selective residues, five are conserved in CO (Fig. 7C), implying selective pressure to maintain similar, but not identical specificities. Notably, of the shared bp in CCAATC and CCACAC (underlined), C2, A3 and C6 are selectively bound by three NF-YA arginines conserved in CO (and in all other CCTs): Arg283 (CO Arg340), Arg281 (CO Arg338), Arg274 (CO Arg331), respectively, with C2 also being contacted by Gly287 (CO Gly343). The lack of DNA-binding in the CO R340Q mutant protein (co-7) is thus consistent with this arginine providing the same specificity as in NF-YA (Fig. 7A, B, C). We are thus tempted to conclude that the two non-conserved residues –corresponding to NF-YA Gly286 and His277- dictate the divergence in sequence specificity of NF-CO with respect to NF-YA. Gly286 is an asparagine in CO, or arginine/lysine in other CCT proteins, which are all bulkier residues in an area where the small glycines of NF-YA allow main chain insertion

16
in the minor groove space (Fig. 7B). The CCAAT A₄, a C in CORE, is contacted by NF-YA His277: this is Tyr334 in CO, a tyrosine, asparagine or arginine in CO-like, and a leucine in all TOC1/PRRs (Petroni et al., 2012). We hypothesize that these changes might command yet different selectivity, focused on this specific nucleotide of the pentamer.

**Interpretation of genetic experiments**

The inclusion of NF-CO among sequence-specific complexes containing HFD subunits changes the interpretation of genetic experiments performed on HFD subunits. Since the discovery of the expansion of NF-YB and NF-YC genes in plants, numerous HFD mutants or overexpressors, particularly for NF-YB (also termed Hap3, DTH8, Ghd8 in rice), were identified and characterized in different species. The obvious molecular interpretation relied on the notion that NF-Y would be crippled, or changed in its trimeric assembly, and activity of targeted CCAAT promoters (largely unknown at the moment) altered. The new data indicate that NF-CO would similarly be altered in these HFD mutants or overexpressors. The second important consequence stemming from our data concerns the phenotype of CO mutants and natural variants. Significant evidence already exists that mutations in the CCT domain are functionally important, primarily in the A2 helix and (R)VNGRF motif.

Based on our data, all these variants are now predicted to be loss-of-function (or hypomorphic) DNA-binding mutants. For example, VRN2 from wheat (ZCCT1 and 2) blocks flowering in long days until after vernalization (Yan et al., 2004), potentially regulating FT (VRN3 in wheat). Further, in vitro and in planta co-immunoprecipitations have shown that ZCCT proteins interact with selected HFDs and compete with NF-YA for binding (Li et al., 2011). The functionally disruptive, natural variants of VRN2 target either the same arginine as the Arabidopsis co-7 allele (ZCCT1) or the arginine equivalent to CO Arg344 (ZCCT2). In both cases, loss of the arginine eliminates the active roles of ZCCT proteins as repressors of flowering time (Distelfeld et al., 2009). Likewise, the barley PRR7 mutant allele ppd-H1 involves the glycine equivalent to Gly343 in CO, with a substitution to (a bulky) tryptophan residue (Turner et al., 2005), and the co-9 allele of Arabidopsis (a valine substitution of the perfectly conserved Ala335; Wenkel et al., 2006) is homologous to the toe1-1 allele of TIMING OF CAB1. Finally, one of the PRR37 natural rice variants,
contributing to adaptation of cultivation at different latitudes, harbours a missense L710P that corresponds to the above-mentioned NF-YA crucial residue His277. A second natural variant is a frameshift mutation at Gln705 (Koo et al., 2013), which would lose the C-term residues of the CCT: both of these mutations are predicted to be loss-of-function. Perhaps most intriguing, a third identified variant is on Tyr704, which becomes a histidine (equivalent to His271). A histidine at this position is never observed in CCTs, but is absolutely conserved in all NF-YAs and required for CCAAT binding (Xing et al., 1993; Mantovani et al., 1994); might this alter DNA specificity of this variant in a CCAAT-directed way? Alteration of DNA-binding, however, is not the only consequence of mutations in CCT proteins. A rice Hd1 variant with genetic adaptation to flowering in long days conditions (Mediterranean cultivar) shows deletion of a lysine in the subunit interaction (A1 helix) portion. This protein is unable to associate with OsNF-YB/NF-YC dimers and to bind a conserved CORE element in the Hd3a (FT) promoter (Goretti et al., 2017).

In summary, NF-CO represents a DNA-binding complex that includes CO, that may or may not require NF-Y function in genomic contexts – and there is every reason to believe the paradigm is generalizable to other CCT proteins. Our findings indicate a broad change of perspectives in CCT associations. Even more than with NF-Ys, the potential combinatorial diversity of NF-CO complexes is enormous. Searches at The Arabidopsis Information Resource suggest that there are 40 CCT proteins (17 CO-Like with BBX domains) (Khanna et al., 2009). Our data thus represent a considerable broadening of our understanding of combinatorial possibilities of NF-Y and NF-CO complexes in plants. Considering the diverse (and still mostly unknown) roles for CCT proteins, the potential for fine tuning of motif binding depending on specific HFD pairings and trimerization is considerable and sorting through this complexity represents an important challenge. The biochemical assays shown here open the possibility to molecularly characterize all NF-Y/NF-CO trimeric complexes generated by combinatorial associations. They also set the stage for structural studies to understand the fascinating details of CO/CORE recognition.

MATERIALS AND METHODS
**Protein production and purification**

The cDNA encoding the CCT domain of CONSTANS (aa 290-352), with the addition of a 5′ *ATG*, was obtained by PCR amplification; the cDNA encoding CO CCT aa 290-352 with the R340Q mutation (Robson et al., 2001), At-NF-YA2 (aa 134-207) and At-NF-YA6 (aa 170-237) were obtained by gene synthesis (Eurofins Genomics) and cloned into pmcnEA/tH (Diebold et al., 2011) by restriction-end ligation to obtain C-terminal 6His-tag fusions (Siriwardana et al., 2016). At-NF-YB2 mutant cDNA, encoding for aa 24-116 with residue E65 mutated to R was obtained by gene synthesis and subcloned in pET15b to obtain the N-terminal 6His-tag fusion (Siriwardana et al., 2016). At-NF-YC9 cDNA, encoding aa 62-158 with a 5′ *ATG*, a 3′ stop codon, and mutant At-NF-YC9 with residue F151 mutated to R (NF-YC9F151R) were obtained by gene synthesis and cloned in pmcnYC (Diebold et al., 2011). All constructs were verified by sequencing. 6His-NF-YB2 or 6His-NF-YB2E65R/NF-YC3 soluble HFD dimers were produced by co-expression in *E. coli* BL21(DE3) and purified by ion metal affinity chromatography (IMAC) as described (Calvenzani et al., 2012). CO-6His and co7-6His were expressed in BL21(DE3)Rosetta by IPTG induction (0.4 mM IPTG for 4h at 25°C) and purified by IMAC (HisSelect, SIGMA-Aldrich) in buffer A (10 mM Tris-HCl pH 8.0, 400 mM NaCl, 2 mM MgCl₂, 5 mM Imidazole). NF-YA2-6His, and NF-YA6-6His were produced in BL21(DE3). Purified proteins were eluted in buffer A containing 100 mM imidazole, and dialysed against buffer B (10 mM Tris-HCl pH 8.0, 400 mM NaCl, 2 mM DTT, 10 % glycerol).

**Electrophoretic Mobility Shift Assays (EMSA)**

EMSAs were performed essentially as previously described (Calvenzani et al., 2012; Cao et al., 2014). Heterotrimer formation and DNA binding of wt or mutant CO (or NF-YAs) was assessed in the presence of wt or mutant NF-YB2/NF-YC3 dimers using Cy5-labeled FT CORE2 (Cy5-AAGAAAAAGATTGTGGTTATGATTTCACCGA) or CCAAT probes (Cao et al., 2014) (Eurofins Genomics). DNA binding reactions (20 nM probe, 12 mM Tris-HCl pH 8, 50 mM KCl, 62.5 mM NaCl, 0.5 mM EDTA, 5 mM MgCl₂, 2.5 mM DTT, 0.2 mg/ml BSA, 5% glycerol, 6.25 ng/μl poly dA-dT), were added with wt or mutant NF-YB2/NF-YC3 HFD dimers (60 nM), in the presence of increasing amounts of the indicated CO or NF-YA proteins. Proteins were pre-mixed in buffer B containing 0.1 mg/ml BSA,
then added to DNA binding mixes. After 30 min incubation at 30°C, binding reactions were loaded on 6% polyacrylamide, and separated by electrophoresis in 0.25X TBE at 4 °C. For competition assays, after 10 min incubation at 30°C, binding reactions (containing 60 nM At-NF-YB2/NF-YC3 and 120 nM CO) were supplemented with increasing amounts of indicated unlabeled oligonucleotide competitors or TE buffer, and incubated for an additional 45 min at 30°C, then loaded on 6% polyacrylamide, or 2.3% agarose gels in 0.25X TBE for electrophoresis. Fluorescence gel images were obtained and analysed with a ChemiDoc MP System and bound DNA complexes were quantified with ImageLab software (Bio-Rad).

Quantification of competition efficiency by mutant CORE2 oligos was performed as follows: percent bound probe was quantified in each lane and plotted vs the competitor concentration (expressed as ratio of the unlabeled vs total oligo concentration). For each oligo, the competitor efficiency represents the slope of the regression line through the competition data point vs the slope of the wt oligo competition performed in the same experiment.

**RNA Sequencing and bioinformatics analysis**

Seedlings were grown for seven days on B5 media in continuous white light with standard, 32 Watt linear florescent tubes (GE product number 26668) producing a light intensity of ~150 µE.. Total RNA was isolated using the E.Z.N.A. Plant RNA Kit from (Omega Biotek). To ensure low levels of contaminating ribosomal RNA, two rounds of poly-A mRNA purification were performed using the µMACS mRNA Isolation Kit (Miltenyi Biotech). Indexed RNASeq libraries were prepared from 100 ng of poly-A RNA starting material using the NEXTflex Illumina qRNA-Seq Library Prep Kit (Bioo Scientific, cat # 5130). Sequencing of 150 bp paired end reads was performed on an Illumina HiSeq 2500 in rapid output mode at the Texas A&M Agrilife Research Facility (College Station, TX, USA). Sample de-multiplexing was performed using CASAVA software v1.8.2 and bcl2fastq was performed using conversion software v1.8.4. Reads were mapped on the reference Arabidopsis transcriptome (TAIR, version 10) using the bowtie2 program. Estimation of gene expression levels was performed by using the RSEM. Differential expression analysis was performed applying the latest versions of DESeq2 and Limma to
RSEM estimated reads counts. Only genes showing an FDR lower than 0.05 according to both tools were considered differentially expressed.

*De-novo* motif discovery was performed with Weeder 2.0 using the default parameters. PScan was used to generate P-values for the enrichment of motif PSSMs generated by Weeder, scanning the same 1-kb intervals upstream of TAIR v10 TSSs or translation start sites. Analyses of phased gene expression were performed using the Phaser tool associated with the DIURNAL database (Mockler et al., 2007), heatmaps were prepared using the heatmap.2 function from the gplots package for R. Gene Ontology enrichments were estimated using DAVID (Dennis et al., 2003).

**Chromatin Immunoprecipitation**

ChIP experiments were performed according to previous publications (Haring et al., 2007; Cao et al., 2014; Yamaguchi et al., 2014; Pchelintsev et al., 2016) with minor modifications. Briefly, we initially harvested 1.5-2 g of 10-day-old, LD-grown, transgenic (*p35S:CO-YFP/HA*) seedlings at 14 h after lights on. The *p35S:CO-YFP/HA* in *nf-yb2 nf-yb3* line was generated by crossing *nf-yb2 nf-yb3* to a stable, single-insertion *p35S:CO-YFP/HA* line in Col-0 and selecting F3 individuals of the appropriate genotype. Both lines showed accumulation of transgenic, epitope tagged CO (Supplemental Fig. 9). These whole seedlings were then ground to a fine powder in liquid nitrogen. The powder was immediately transferred into 23.5 mL of nuclear isolation buffer (10 mM HEPES, pH 7.6, 400 mM sucrose, 5 mM KCl, 5 mM MgCl₂, 5 mM EDTA, 5 mM 2-mercaptoethanol, 1% Triton X-100, 0.4 mM PMSF, 1X protease inhibitor cocktail, and 50 μM MG132) and incubated for 10 min at 4°C. To initiate cross-linking of chromatin complexes, fresh, methanol-free formaldehyde (1.56 ml, cat # 28906, Pierce) was added to the above solution and incubated at room temperature (~22 °C) for 10 min. Next, this solution was incubated with 2 M glycine (2 ml) for 5 min to stop the cross-linking reaction. The lysate was then filtered through two layers of Miracloth (cat # 475855-1R, Calbiochem) and nuclei were pelleted at 2800 g (4000 rpm with 15-cm-diameter rotor) for 10 min at 4°C. Chromatin shearing was then performed using a Bioruptor UCD300 (low power, 12 cycle of 24 s on, 24 s off, Diagenode). Immunoprecipitations were performed using μMACS anti-HA and anti-GFP microbeads, in combination, to improve immunoprecipitation efficiency. For non-
immune controls, the exact same procedure was followed, minus the addition of Miltenyi beads. The immunoprecipitation procedure follows the ChIP protocol described by Miltenyi Biotec. Quantitative PCR (qPCR) was performed on a Bio-Rad CFX Connect real-time system with Maxima SYBR Green/ROX qPCR Master Mix (cat # K0221, Thermo Fisher Scientific). The qPCR profile was 10 min at 95°C, 45 cycles of 10 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C, followed by the default dissociation step to generate a melting curve. Primers are listed in Supplemental Table 1. ChIP efficiency was calculated as percent of input. Statistical analysis and comparisons between samples was performed in the Bio-Rad CFX Manager Software (http://www.bio-rad.com/) through use of the $2^{(-\Delta\Delta CT)}$ method.

**Immunoblot analysis**

Soluble and nuclear protein fractions were isolated from 10-day-old, LD-grown seedlings by grinding in sucrose buffer (20 mM Tris, pH 8.0, 330 mM sucrose, 1 mM EDTA, pH 8.0, 5 mM DTT, 1X Protease inhibitor cocktail, and 50 µM MG132), followed by two rounds of low speed centrifugation (1000 g for 5 min each) to discard large plant debris. The cleared solution was separated into soluble and nucleus-containing fractions by high-speed centrifugation (20,000 g for 30 min). A standard 8% SDS-PAGE gel was loaded with 30 µg total protein for each soluble fraction resuspended, lysed nuclei from the equivalent of 50 mg starting material (approximately 3.3X concentrated relative to the soluble fraction in cell equivalents). Proteins were transferred to standard PVDF membranes and the presence of CO-YFP/HA was probed with high affinity anti-HA primary antibodies (cat # 11 867 423 001; Roche) and goat, anti-rat, HRP-conjugated secondary antibodies (cat # SC-2032; Santa Cruz Biotechnology). A Bio-Rad ChemiDoc XRS imaging system was used for visualizing the protein blot after incubations with ECL plus reagent (cat # RPN2132; GE Healthcare).

**Cloning**

The mutations were made by PCR using appropriate mutagenic primer sequences. Each construct was amplified from Pfu Ultra II (Invitrogen #600670) and cloned into the GATEWAY vector pENTR/D-TOPO (Invitrogen #45-0218). All resulting clones were
sequenced and, with the exception of the introduced mutations, were identical to the sequences at TAIR (http://www.arabidopsis.org; Huala et al., 2001). Inserts were then cloned into the Y2H expression vectors pDEST\textsuperscript{TM}22 or pDEST\textsuperscript{TM}32 (Invitrogen). The $pNF-YC9$:NF-$YC9$ construct was previously described (Kumimoto et al., 2010). The entry clone $pNF-YC9$:NF-$YC9^{F151R\,V153K}$ was cloned into the plant expression vector pEarlyGate301 (Earley et al., 2006).

**Plant transformation, cultivation and flowering time experiments**

*Arabidopsis thaliana* ecotype Col-0 was the wild type for all experiments. The triple mutant was previously described (Kumimoto et al., 2010). Agrobacterium mediated floral dipping was used to transform the triple mutant with $pNF-YC9$:NF-$YC9$ and $pNF-YC9$:NF-$YC9^{F151R\,V153K}$ (Clough and Bent, 1998). All experiments were carried out on plants grown in a custom-built walk-in chamber under standard long day conditions (16 h light/8 h dark, 22°C). Plant growth conditions were as described (Myers et al., 2016). Leaf number at flowering was measured as the total number of rosette and cauline leaves on the primary axis at flowering.

**Y2H analysis**

The activation domain (AD) or DNA binding domain (DBD) constructs were introduced into the yeast strain MaV203 (Invitrogen). Y2H assays were performed according to the instructions in the ProQuest manual (Invitrogen). For the X-Gal assay, nitrocellulose membranes were frozen in liquid nitrogen and placed on a filter paper saturated with Z-buffer containing X-Gal (5-bromo-4-chloro-3-indoxyl-beta-D-galactopyranoside, Gold Biotechnology, #Z4281L).

**Accession Numbers**

Sequence data from this article can be found in the EMBL/GenBank data libraries under accession numbers: AT5G12840; AT3G05690; AT3G14020; AT5G47640; AT4G14540; AT1G54830; AT5G63470; AT1G08970; AT5G15840; AT1G65480.
SUPPLEMENTAL DATA

Supplemental Figure 1. List of available data on CCT/NF-Y subunits interactions.

Supplemental Figure 2. Coomassie stained gels of purified proteins samples.

Supplemental Figure 3. CO efficiently binds DNA as a trimer with At-NF-YB2/NF-YC3.

Supplemental Figure 4. EMSA competition analysis of NF-CO sequence specificity.

Supplemental Figure 5. Validation of RNA-Seq data.

Supplemental Figure 6. Statistical enrichment of promoter motifs in DE gene sets and subsets.

Supplemental Figure 7. Enrichment of circadian expression phases among genes differentially expressed in co-sail, nf-yb2 nf-yb3, and nf-yc3 nf-yc4 nf-yc9.

Supplemental Figure 8. Fold expression changes of genes in co-sail, nf-yb2 nf-yb3, and nf-yc3 nf-yc4 nf-yc9 lines.

Supplemental Figure 9. Immunoblot analysis of CO overexpressing lines.

Supplemental Table 1. Oligonucleotides used to amplify regions of the FT gene.

Supplemental Data Set 1. List of Differentially Expressed genes in RNA-Seq analysis.

Supplemental Data Set 2a. Enrichment of peak expression phase (long-day growth) of genes differentially expressed in co-sail, nf-yb2 nf-yb3, and nf-yc3 nf-yc4 nf-yc9 lines.

Supplemental Data Set 2b. List of GO terms showing significant enrichment in sets of genes showing differential expression in co-sail, nf-yb2 nf-yb3, and nf-yc3 nf-yc4 nf-yc9 lines.

Supplemental Data Set 2c. Expression levels by RNA-Seq (TPM) for members of the CCT, B-box, NF-YA, NF-YB, and NF-YC gene families.

Supplemental Data Set 3. Intersection of differentially expressed (DE) genes in NF-CO subunits mutants with Wigge et al. 2005 expression data.
ACKNOWLEDGEMENTS
This work was supported by Institutional grant from the U. of Milano to NG and by NSF Grant IOS-1149822 to BFH.

AUTHOR CONTRIBUTIONS
NG performed the biochemical experiments. RWK and CS performed the RNA-Seq and validation experiments. SS performed the ChIP experiments. MC and DSH analysed the RNA-Seq experiments. NG, BFH and RM planned the experiments; BFH and RM wrote the manuscript.
Figure 1. CO binds DNA as a trimer with At-NF-YB2/NF-YC3 and recognizes the CORE element.

A. CO forms a trimer with At-NF-YB2/NF-YC3 HFD binding to the FT CORE2 element. EMSAs were performed using fluorescently labeled FT CORE2 (lanes 1-14) or FT CCAAT (lanes 15-28) 31-mer oligonucleotide DNA probes (20 nM) by addition of the indicated proteins. CO-CCT (CO) was incubated at increasing concentrations (90, 180, 270, 360 nM) with the CORE2 probe in the absence (lanes 2-5) or presence (lanes 9-12) of the At-NF-YB2/NF-YC3 HFD dimer (At-NF-YB2/YC3, 60 nM). As controls, At-NF-YA2 or -YA6.
(YA2, YA6) were incubated with the CORE2 probe at the highest concentration of the dose curve (360 nM), with or without (-) the HFD dimer (60 nM) (YA2: lanes 13, 6; YA6: lanes 14, 7, respectively). Lane 1: CORE2 probe alone, without protein additions. DNA binding of CO or At-NF-YAs, as indicated (lanes 16-27), was assayed on the FT CCAAT probe in the presence of At-NF-YB2/NF-YC3 (60 nM), with the same protein concentration dose curve (90, 180, 270, 360 nM). As controls, the FT CCAAT probe was incubated with the HFD dimer alone (60 nM, lane 15), or with At-NF-YA2 protein (360 nM, lane 28). NF-CO and NF-Y/DNA complexes are indicated by closed or open arrowheads. fp: free probe. B. EMSA competition analysis of the At-NF-YB2/NF-YC3/CO complex specificity on the labeled CORE2 probe. Top panel: sequences of the 31-mer CORE2 probe and unlabeled competitor derived from the FT promoter (-172/-141 from ATG). Oligos 1-6: CORE2 30mers and 25-mer, wt (WT) or mutant used as unlabeled competitors. The 31-mer derived from the FT enhancer CCAAT sequence, and the FT mutant competitor (Cao et al., 2014) are listed below, together with the Hsp70 CCAAT competitor. Sequence identity with the probe is indicated by dots, and 5' or 3' sequence extensions, or mutated nucleotides are indicated in capital letters. The previously described TTGTGGTT CORE element (Tiwari et al., 2010) and the CCAAT pentamer are highlighted in bold letters. Bottom panel: EMSA competition analysis was performed by incubation of the CORE2 probe with the trimer composed of indicated subunits (At-NF-YB3/NF-YC3, 60 nM; CO, 180 nM -At-NFY-B2/YC3/CO-: lanes 2-27) in the presence of TE buffer alone (lanes 2, 27), or with the addition of increasing concentrations of the indicated unlabeled competitors (5X or 25X molar excess; lanes 3-26). Lanes 1, 28: CORE2 probe alone, without proteins addition. The NF-CO/DNA complex is indicated by an arrowhead. fp: free probe.
Figure 2. Determination of NF-CO sequence specificity in vitro.

**A.** CORE2 competitors and mutagenesis strategy. The unlabeled wt CORE2 (WT) and 30-mer oligo sequences with mutated nucleotides are shown, as in Fig. 1, for the three sets of CORE2 mutant oligos (mI, mII, and mIII). In the bar graphs, mI and mII mutant oligo
competitor efficiency (competition) is expressed as ratio of the dose response curve slope of the mutant versus the wt oligo (see Materials and Methods). Competition of the wt oligo is set as 1. Indicated values represent the mean of three (ml oligos, top panel), or two (mII oligos, bottom panel) series of competition assay experiments (see also Supplemental Fig. 4). Error bars indicate +/- s.d. for ml oligos, or value ranges for mII oligos. Sequences in red boxes highlight mutations with reduced competition (< 0.67 of wt oligo efficiency). B. CORE2 mIII mutant oligo EMSA competition results. Competition efficiencies are shown as mean value of three independent series of experiments for each of the mIII single nucleotide mutant oligo, as indicated for A. For each mutated position the wt oligonucleotide competition value, set as 1, is also shown. Values are also displayed in the bar graph in (C) (average of three independent sets of experiments +/- s.d.). For each nucleotide position, dark and light shaded bars denote mutant and wt (*) competitors, respectively (see also Supplemental Fig. 4). In (C), the sequence matrix obtained with the mIII competitions (information content) is shown on the right, for the sense (+) and reverse (-) strands of the FT promoter.
Figure 3. Identification of CO and HFD matrices by RNA-Seq analysis.
RNA-Seq analysis of Differentially Expressed (DE) genes in the co-sail, nf-yb2 nf-yb3 and nf-yc3 nf-yc4 nf-yc9 lines compared to wild-type Arabidopsis. The motifs enriched in the associated promoters are shown for each intersection of coregulated genes. A and B. Venn diagram showing numbers of DE genes identified in comparisons between tested lines and wild-type plants, and overlaps between DE gene sets. For each gene set, an alphanumeric code signifies the most highly enriched motif identified. C. Sequence logos describing the motifs identified from analyses of promoters (from -1000 to TSS/ATG) of DE gene sets (see Supplemental Fig. 6).
**Figure 4.** CO binds the *FT* promoter in an At-NF-YB2/NF-YB3-dependent manner.

Chromatin immunoprecipitation was performed on Col-0 (parental) and *nf-yb2 nf-yb3* plants transgenic for p35S:CO-YFP/HA. Enrichment of the selected segments -CCAAT/-5.3kb, core promoter, Exon 4/+2.0kb- were evaluated by qPCR with appropriate amplicons. Error bars indicate standard error with five biological replicates. In each replicate, three technical replicates were performed. Statistical significance was obtained using BioRad CFX Manager Version 3.0 – in each case the comparison is between the non-immune control (NIC) and the immunoprecipitation (IP). *** is $p<0.001$. 
Figure 5
Figure 5. Properties of At-NF-YC9 trimerization mutants. 
A. Alignment of At-NF-YCs. Multiple sequence alignment of NF-YC protein core domains. MSA was computed using ClustalW within Geneious version 7.0. Amino acid residue positions of the HFD are indicated for the At-NF-YC9 and human proteins. Amino acids making physical contact with NF-YA are annotated by an asterisk (Nardini et al., 2013). Arrows mark the position of mutated residues, with the closed arrow indicating the conserved phenylalanine required for interaction with NF-YA in mammals, which was mutated in NF-YC9\textsuperscript{F151R\,V153K} and in the recombinant At-NF-YC9F151R HFD protein. At: Arabidopsis thaliana; Hs: Homo sapiens. B. In vivo analysis of timing of flowering. T1 generation flowering time analysis of pNF-YC9:NF-YC9\textsuperscript{F151R\,V153K} in the nf-yc triple (nf-yc3 nf-yc4 nf-yc9) mutant background. Asterisks represent significant differences derived from one-way ANOVA (p < 0.05) followed by Dunnett’s multiple comparison post hoc tests against the nf-yc triple mutant. C. Expression of At-NF-YC9 transgenes in transgenic plants. Protein expression in the plant lines used for the flowering time analysis was analyzed by immunoblot with antibodies directed to a translationally fused HA-epitope (top panel). Protein loading and transfer was assessed by Ponceau staining (bottom panel). D. Y2H assays of At-NF-YC9. Full length NF-YC9, and mutant variants, tested using Y2H against empty vector (EV) control, NF-YB2, NF-YA1, NF-YA2 and CO. Note that NF-YC9 has slight auto activation. E. EMSAs on CORE2 and CCAAT of wt and mutant At-NF-YC9. Trimerization and DNA-binding of the At-NF-YC9F151R mutant (YC9F151R; lanes 8-11) or wt (YC9; lanes 2-6) containing HFD dimer (60 nM) was assessed the with the CORE2 probe (lanes 1-14), by addition of the CO subunit at increasing concentrations (90, 180, 270, 360 nM; lanes 3-6; 8-11). At-NF-YA2 (YA2) trimerization with the wt or mutant dimers (lanes 17-20 and 22-25, respectively) was assessed with the CCAAT probe (lanes 15-28). As negative controls, CO or At-NF-YA2 were added alone to the reaction with the respective probes (lanes 12, 26). At-NF-YC9 trimer specificity was also assessed by addition of the CO or At-NF-YA2 containing trimers to the CCAAT or CORE2 probe, respectively (lanes 28, 14). DNA binding of At-NF-YC3 (YC3) containing trimers was also used as internal control (lanes 13, 27). In lanes 2 and 16, 7 and 21, wt or mutant HFD dimers were incubated alone with the probe. NF-CO and NF-Y/DNA complexes are indicated by labeled arrowheads. Lanes 1, 15: CORE2 and CCAAT probes without protein additions. fp: free probe.
Figure 6. Properties of At-NF-YB2 trimerization mutant.
Trimerization and DNA-binding of the E65R mutant (YB2E65R; lanes 7-11) or wt (WT; lanes 2-6) At-NF-YB2 containing HFD dimer (60 nM) was assessed with the CORE2 probe (lanes 1-14), by addition of the CO subunit at increasing concentrations (90, 180, 270, 360 nM; lanes 3-6; 8-11). At-NF-YA2 (YA2) trimerization with the wt or mutant dimers (lanes 17-20 and 22-25, respectively) was assessed with the CCAAT probe (lanes 15-28). As negative controls, CO or At-NF-YA2 was added alone to the reaction with the respective probes (lanes 12, 26). Trimer specificity was assessed by addition of the CO or At-NF-YA2 containing trimers to the CCAAT or CORE2 probe, respectively (lanes 27, 28; 13, 14, as indicated). In lanes 2 and 16, 7 and 21, wt or mutant HFD dimers were incubated alone with the probe. NF-CO and NF-Y/DNA complexes are indicated by labeled arrowheads. Lanes 1, 15: CORE2 and CCAAT probes without protein additions. fp: free probe.
Figure 7
Figure 7. The CO CCT drives sequence-specificity of NF-CO.

A. CO mutation R340Q of co-7 abolishes NF-CO DNA binding. Wild-type or R340Q CO were incubated at increasing concentrations (90, 180, 270, 360 nM) with the CORE2 probe in the absence (-) (lanes 2-5; 6-9), or in the presence (lanes 11-14; 15-18) of the At-NF-YB2/NF-YC3 HFD dimer (60 nM). In lane 10, the At-NF-YB2/NF-YC3 HFD dimer was incubated alone with the probe. Lane 1: probe alone, without protein additions. fp: free probe.

B. Schematic representation of selected NF-YA interactions with the C2A3 bp of CCAAT. Highlight of NF-YA A2 helix within the NF-Y/DNA 3D structure (PDB: 4AWL), with interactions of Arg281, Arg283 (corresponding to CO Arg340) and Gly287 (CO Gly343) amino acid residues (indicated in single letter code: R281, R283 and G287 respectively) with the G2 and A3 nucleotides. NF-YA protein (cyan) and the sugar-phosphate DNA strands backbones are represented as colored strings, with the CCAAT (I) and complementary (J) strands in red and green color, respectively. Orientation of DNA strands is indicated. Selected NF-YA residues and nitrogen bases of the C2A3;G2T3 nucleotides are labeled and displayed in ball and stick model in color matching the main chain color code. G287 main chain and R283 side chain contacts with G2 atoms, and R281 side chain with A3, are indicated by grey lines (Nardini et al., 2013). The NF-YB/NF-YC subunits within the 4AWL structure were omitted for clarity. The image was obtained with Protein Workshop (Moreland et al., 2005).

C. Amino acid sequence alignment of the C-terminal portion of CO CCT domain with mammalian NF-YA homology region is shown, with the proposed sequence-specific interactions, based on the NF-Y/DNA complex crystal structure (PDB: 4AWL). DNA sequence of the NF-CO and NF-Y respective element is shown at the top and bottom of the alignment, with the indicated orientation of the DNA strands, and base-pair positions in the bound elements numbered (see text). Side chain interactions of NF-YA with the CCAAT bases are indicated by full lines (bottom). On top of the alignment, dashed lines represent potential CO residues interactions with the CORE matrix. Conserved and non-conserved residues are highlighted in green and blue, respectively. R340Q in co-7 is indicated on top of the alignment. The closed circle represents hydrophobic base stacking interactions of phenylalanine residue side chains with the CA:GT nucleotides. Bold nucleotides highlight the divergence in sequence specificity of the two complexes.
Figure 8. Scheme of NF-Y vs NF-CO specificity. Association of CO or NF-YA with NF-YB/NF-YC dimers provides robust and specific recognition of the respective DNA element by the trimeric NF-CO and NF-Y complexes.


CONSTANS imparts DNA sequence-specificity to the histone-fold NF-YB/NF-YC dimer
Nerina Gnesutta, Roderick W Kumimoto, Swadhin Swain, Matteo Chiara, Chamindika Siriwardana, David
Stephen Horner, Ben F Holt and Roberto Mantovani

Plant Cell; originally published online May 19, 2017;
DOI 10.1105/tpc.16.00864

This information is current as of October 20, 2017

Supplemental Data  /content/suppl/2017/05/19/tpc.16.00864.DC1.html
/content/suppl/2017/07/11/tpc.16.00864.DC2.html
Permissions

eTOCs
Sign up for eTOCs at:
http://www.plantcell.org/cgi/alerts/ctmain

CiteTrack Alerts
Sign up for CiteTrack Alerts at:
http://www.plantcell.org/cgi/alerts/ctmain

Subscription Information
Subscription Information for The Plant Cell and Plant Physiology is available at:
http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY