A distal enhancer controls transcription of *FLOWERING LOCUS T* in Arabidopsis

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Liangyu Liu

aus Weifang, China

Köln, März 2013

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor Prof. Dr. G. Coupland) angefertigt.

Berichterstatter: Prof. Dr. George Coupland

Prof. Dr. Ute Höcker

Prüfungsvorsitzender: Prof. Dr. Martin Hülskamp

Tag der Disputation: 09. April, 2013

Abstract

Control of flowering time is critical for reproductive success. In *Arabidopsis thaliana*, the gene *FLOWERING LOCUS T* (*FT*) encodes part of florigen, which moves from leaves to the shoot apex to induce the floral transition. Under inductive long days (LD), transcription activation is mainly mediated by CONSTANS (CO). The epigenetic repressor LHP1, which associates with H3K27me3, antagonizes CO activity in vascular tissues at *FT*.

The 5.7 kb sequence upstream of the *FT* translation start codon contains sufficient regulatory elements to mediate spatial and temporal expression of *FT*. On this full length *FT* promoter, phylogenetic analysis identified three conserved blocks, each around 300 bp, called *Block A, B* and *C*. The aim of my study was to identify and characterize functional *cis*-elements within the conserved blocks and elucidate the molecular mechanisms of *FT* trancription regulation.

My results indicate that *Block C* (5.2 to 5.6 kb upstream of the transcriptional start site) is necessary for *FT* expression in LD conditions, and that *CCAAT* boxes in *Block C* and close to *Block A* at the proximal promoter impair the efficiency of a full length *FT* promoter. Mutagenesis of either one of four conserved sub-blocks in *Block C* dramatically reduces promoter activity. Although *Block C* acts as distal enhancer, it requires specific sequences contained in *Block A* to convey phloem-specific expression of *FT* in leaves. Through Chromosome Conformation Capture (3C) assays, a weak long-distance physical interaction between *Block C* and *A* could be detected.

The distance between *Block C* and *A* displays natural variation among different *A. thaliana* accessions. Shortening the distance between the blocks by deleting 3.5 kb intersecting region or directly connecting *Block C* and *A*, generated shorter versions of the *FT* promoter that mimic the full-length promoter funtion in response to LDs.

How the repressive histone mark H3K27me3 and LHP1 contribute to *FT* regulation is still unlcear. To study differential enrichment of H3K27me3 at the *FT* promoter, nuclei were isolated by INTACT (isolation of nuclei tagged in specific cell types) from the leaf phloem of plants grown in LD or SD conditions. Chromatin immunoprecipitation (ChIP) revealed that the abudance of H3K27me3 at *Block C* and *A* regions negatively correlated with *FT* expression levels in the phloem.

Taken together, the distal enhancer *Block C* and *Block A* in the proximal promoter are accessible in leaf phloem cells and required and suffcent for *FT* expression in LD conditions.

Zusammenfassung

Für eine erfolgreiche pflanzliche Vermehrung ist es unumgänglich den Blühzeitpunkt zu kontrollieren. In *Arabidopsis thaliana* kodiert das Gen *FLOWERING LOCUS T* (*FT*) einen Teil des „florigen", welches sich von den Blättern zum Apex der Pflanze bewegt, um dort das Blühen einzuleiten. In induzierenden Langtagbedingungen (LD), wird die Aktivierung der *FT* Transkription hauptsächlich über CONSTANS (CO) gesteuert. Das epigenetische Repressionsprotein LHP1, welches mit H3K27me3 assoziiert, fungiert antagonistisch zu der CO Aktivität in Bezug auf *FT* im vaskulären Gewebe.

Die 5,7 kb Promotersequenz vor dem Startcodon von *FT* enthält regulatorische Elemente, die ausreichen, um eine örtlich und zeitlich begrenzte Expression von *FT* zu gewährleisten. Auf diesem funktionalen *FT* Promoter wurden drei konservierte Blöcke durch phylogenetische Analysen identifiziert. Jeder der Blöcke (*Block A, B* und *C*) ist etwa 300 bp lang. Meine Arbeit hatte die Identifikation und Charakterisierung funktionaler *cis*-Elemente innerhalb der konservierten Promoterblocks zum Ziel. Somit sollte zum Verständnis der molekularen Mechanismen beigetragen werden, welche die *FT* Transkription regulieren.

Meine Ergebnisse deuten an, dass *Block C* (5.2 – 5.6 kb vor dem *FT* Transkriptionsstart) essentiell für die *FT* Expression in LD Bedingungen ist. Zusätzlich scheinen die *CCAAT* Elemente in *Block C* und nahe *Block A* im proximalen Promoter die Effizienz des Gesamtpromoters von *FT* direkt zu beeinflussen. Wenn man vier konservierte Unterblöcke in *Block C* durch Mutagenese verändert, führt dies zu stark verminderter Aktivität des Promoters. Obwohl *Block C* die Aktivität von *FT* aus der Distanz beeinflusst, sind zusätzlich spezifische Sequenzen in *Block A* nötig, um eine phloemspezifische Expression von *FT* in Blättern zu gewährleisten. Durch Chromosome Confirmation Capture (3C) konnte eine schwache physikalische Interaktion zwischen *Block C* und *Block A* nachgewiesen werden.

Die Distanz zwischen *Block C* und *A* weist natürliche Variation zwischen verschiedenen *A. thaliana* Akzessionen auf. Die Verkürzung der Distanz zwischen den Blöcken, durch die Beseitigung des 3,5 kb umfassenden Zwischenregion oder die direkte Verbindung zwischen *Block C* und *A*, generierten kürzere Versionen des *FT* Promoters, die die Funktion des kompletten *FT* Promoters in LD Bedingungen imitierten.

Es bleibt weiterhin unklar, wie die repressive Histonmarkierung H3K27me3 und LHP1 zur *FT* Regulation beitragen. Um differenzierte Anreicherung von H3K27me3 am *FT* Promoter zu untersuchen, wurden Zellkerne durch INTACT (Isolation von in bestimmten Geweben

markierten Zellkernen) aus dem Blattphloem von Pflanzen, die in LD oder SD Bedingungen gewachsen waren, isoliert. Mit Hilfe von Chromatin Immunopräzipitation (ChIP) konnte gezeigt werden, dass die Häufigkeitsverteilung von H3K27me3 an *Block C* und *A* im Phloem negativ mit der der Expression von *FT* korrelierte.

Zusammen genommen sind der entfernte Expressionsverstärker *Block C* und *Block A* im proximalen Promoter in Blattphloemzellen verfügbar und sowohl nötig als auch ausreichend für die Expression von *FT* in LD Berdingungen.

Table of Contents

1. Introduction

1.1 Photoperiodic flowering pathway

In higher plants, two basic growth phases are known. They result in vegetative and reproductive developmental stages of plant growth. In *Arabidopsis thaliana* (Arabidopsis or *At*), during the process of floral transition changing the vegetative growth to the reproductive phase, the shoot apical meristem (SAM) is phenotypically modified into an inflorescence meristem (Huijser and Schmid, 2011). In the annual Arabidopsis genus, floral transition cannot be reverted under normal growth conditions. Flowering time is stringently controlled by an internal gene network and external environment cues such as plant age, day length, light quality and temperature (Jack, 2004). Genetic studies have been identified several flowering time pathways, for examples photoperiod (day length), GA (gibberellin acid), vernalization (a period of cold), thermo-sensory, autonomous and aging pathways (Andres and Coupland, 2012; Kumar et al., 2012; Putterill et al., 2004; Wang et al., 2009). The photoperiod pathway was introduced in this part.

Photoperiodism is the physiological reaction of plants to the length of day and night periods, and the controlling factor is the length of darkness (Hamner, 1940). Depending on the flowering stimulus from day-length, plants are categorized as long-day, short-day and day-neutral plants (Allard and Garner, 1940). In 1936, physiologist Mikhail Chailakhyan reported that only leaves were receiving a photoperiodic stimulus whereas the stem buds were acceptors of this signal in Chrysanthemum (reviewed by Chailakhyan, 1975). One year later, in 1937, combining grafting experiments from sunflower and other species, the "florigen" concept was mentioned for the first time. In this concept, a hormone-like molecule which is produced in leaves and moves up to the SAM to trigger flowering was described (reviewed by Chailakhyan, 1975). However, it took more than seven decades to identify the molecular nature behind the florigen concept.

1.1.1 FT: part of florigen

Arabidopsis as a facultative long-day plant is early flowering under inductive LD, but late flowering in SD conditions (Hisamatsu and King, 2008; King et al., 2008). Late flowering mutants such as *gi* (*gigantea)*, *co* (*constans*) and *ft* (*flowering locus t*) mutants were isolated by Maarten Koornneef and colleagues (Koornneef et al., 1991). *FT* is an early downstream

target of CO (Samach et al., 2000). Transgenic plants carrying *35S::CO* (the promoter of *35S* is a strong and constitutive *35S* RNA promoter from Cauliflower Mosaic Virus (CaMV)) flower early independent of day length (Simon et al., 1996). Mis-expression of *CO* by phloem-specific expression promoter *SUC2* rescues late flowering of *co* mutants, but *CO* expressing in the meristem driven by *KNAT1* promoter fails to complement *co* mutants. Overexpression of *FT* in the phloem or meristem rescues both *co* and *coft*. It suggests that *CO* controls the synthesis of systemic flowering signal probably through *FT* (An et al., 2004). A bZIP transcription factor FD is preferentially expressed in the meristem and as a partner required for FT action of triggering flowering (Abe et al., 2005).

In 2005, *FT* mRNA as florigen was reported to move from leaves to apex. These results, however, were later retracted (Bohlenius et al., 2007; Huang et al., 2005). In tomato, *SFT* (*SINGLE FLOWER TRUSS*) is the ortholog of Arabidopsis *FT*. *SFT* mRNA was not detected in the SAM of *sft* shoots that had been grafted onto *35S::SFT* donors (Lifschitz et al., 2006).

In 2007, in Arabidopsis, GFP signal was detected in the SAM of *ft-7* plants carrying *SUC2::FT:GFP*, and the GFP signal was also observed in the apex of plants with *ft-7* mutant shoots grafted onto *SUC2::FT:GFP;ft-7* roots (Corbesier et al., 2007). In rice, an ortholog of *FT*, *Hd3a* tagged GFP driven by phloem-specific-promoters such as *rolC* or *RPP16* showed florescence signals in the apexes. As control, the photoactivatable fluorescent protein Kaede driven by the same promoters only showed florescence signals that were restricted to phloem tissue beneath the SAM (Tamaki et al., 2007).

Evidence provided by different research groups supported the hypothesis that FT protein movement was necessary for the floral induction. Under the control of the *SUC2* promoter, FT tagged with 5x Myc at the N- terminal moved out of companion cells into the SAM, but Myc:FT with nuclear localization signal (NLS) was trapped in the nucleus, which resulted in plants in which FT could not travel nor trigger flowering. As control, the fusion protein Myc:FT with NLS expressed in the apex promoted flowering (Jaeger and Wigge, 2007). Furthermore, FT tagged by TEV-3×YFP carrying a cleavage site by TEV (tobacco etch virus) protease was too big to move out of phloem cells. While FT released from the big fusion protein was sufficient to induce floral transition (Mathieu et al., 2007). From hetero-grafting experiments in Cucurbits, mass spectrometry analysis of phloem sap proteins revealed the presence of FT protein, but no *FT* transcripts were detected according to RT-PCR measurements (Lin et al., 2007). In sum, FT protein can be viewed as a long-distance mobile signal, which is translocated from leaves to the SAM to promote floral transition. Thus, FT

protein fits well with the definition of florigen.

1.1.2 GI-CO-FT module in photoperiod pathway

Based on two decades of genetic studies on flowering time, components GI-CO-FT became established as central module in photoperiod pathway (Turck et al., 2008).

GI encoding a nuclear protein is identified as a component of the circadian clock network. There, it regulates the central loop factors *LHY* (*LATE ELONGATED HYPOCOTYL*) and *CCA1* (*CIRCADIAN CLOCK-ASSOCIATED 1*) (Park et al., 1999). Additionally, GI could also repress *TOC1* (*TIMING OF CAB EXPRESSION 1*) expression in a feedback loop in the evening (Pruneda-Paz and Kay, 2010). Genetic experiments demonstrates that through circadian clock GI positively upregulates expression levels of *CO* and *FT* (Mizoguchi et al., 2005).

CO encodes a nuclear protein containing a CCT (CONSTANS, CO-like, and TOC1) domain and two zinc binding B-boxes. *CO* is found to act downstream of *GI* as well as the circadian clock. The signals of GUS reporter gene driven by *CO* promoter are restricted to phloem cells in LD conditions (An et al., 2004; Takada and Goto, 2003). *CO* mRNA accumulates with a diurnal rhythm and displays a biphasic pattern in one cycle. On the transcriptional level, the increment of *CO* transcripts in the late afternoon of LD is mainly controlled by GI (Putterill et al., 1995; Turck et al., 2008). DOF transcription factor CYCLING DOF FACTOR 1 (CDF1) is redundancy with CDF2, CDF3 and CDF5. When expressed each CDF in phloem under *SUC2* promoter, plants delay flowering coupling with downregulating *CO* mRNA. CDFs inhibit the increment of *CO* transcripts in the late afternoon. The *cdf1;2;3;5* quintuple mutants flower early independent on the photoperiod, and they restore late flowering of *gi* plants (Fornara et al., 2009). In addition, CDF1 has been suggested to directly associate with *CO* promoter. The abundance of CDFs is dependent on GI as well as FKF1 (FLAVIN-BINDING, KELCH REPEAT, F-BOX 1). GI and FKF1 can form a complex which removes the CDFs through the 26S proteome degradation pathway (Fornara et al., 2009; Imaizumi et al., 2005; Sawa et al., 2007).

Even though *CO* mRNA has a high expression level in the night when plants grow in SD conditions, C protein does not accumulate in the dark. That the CO protein is unstable was first observed when GPF signal could not be detected in the nucleus of guard cells in darkness although a *GFP:CO* fusion gene was expressed under the control *35S* promoter. In

addition, CO protein is ubiquitinated and stabilized by a proteasome inhibitor indicating that posttranslational regulation of CO participates in determining photoperiodic flowering (Valverde et al., 2004). Light quality also regulates CO protein stability. In LD, the red light photoreceptor Phytochrome B (PHYB) participates in the degradation of CO in the early morning. However, PHYA, another photoreceptor for far-red light, as well as the blue light receptors Cryptochrome 1 and 2 (CRY1 and CRY2) stabilize CO at the end of long days (Valverde et al., 2004). SPA1 (SUPPRESSOR OF PHYA-105), redundantly with SPA2, SPA3 and SPA4, belongs to a family of WD-repeat proteins which suppress photomorphogenesis in darkness (Laubinger et al., 2004). SPA1-CO interaction requires the CCT domain of CO, and CO abundance increases in *spa1;3;4* mutants. SPAs mediate the destabilization of CO protein in the evening (Laubinger et al., 2006).

FT was cloned in 1999. Genetically *FT* is the main downstream target of CO, and CO is the main activator of *FT* in photoperiod pathway (Kardailsky et al., 1999; Kobayashi et al., 1999). *CO* likely activates *FT* transcription requiring its proximal promoter, especially some conserved sequences (Adrian et al., 2010).

Unlike *CO* and *FT* which are specifically expressed in the phloem, *GI* also displays expression in other tissue types (Winter et al., 2007). Independent of CO, *GI* indirectly regulates *FT* through the *miRNA* (*miR*) *172* pathway (see below) (Mathieu et al., 2009). Surprisingly, GI itself seems to activate *FT* directly in mesophyll cells by binding to *FT* repressors and associating with the *FT* proximal promoter (Sawa and Kay, 2011) and also other upstream regulators of *CO* such as FKF1 and CDFs have been reported to directly regulate *FT* based on ChIP results (Song et al., 2012). More evidence is needed to understand how the transcriptional components involved in CO regulation are also involved in *FT* transcription control.

In sum, the GI-CO-FT module contains the central factors in the flowering response to day length in Arabidopsis. Interestingly, in the SD plant rice, OsGI-Heading date 1(Hd1)-Hd3a are homologs of GI-CO-FT in Arabidopsis. The rice module activates flowering in SDs. Under LD conditions, Hd1 represses *Hd3a*. However, which type of Hd1 protein participates in upregulating *Hd3a* in SD and which one repressed *Hd3a* in LD remained unclear (Hayama et al., 2003; Tsuji et al., 2011).

1.1.3 FT action in the SAM

In inductive conditions, *FT* is transcribed in the companion cells of the minor veins in leaves, and FT is translocated with the assistance of FTIP1 (FT-INTERACTING PROTEIN 1) to sieve elements (Adrian et al., 2010; Liu et al., 2012; Takada and Goto, 2003). After long distance movement along the stem phloem, FT is unloaded to the SAM (Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007; Tamaki et al., 2007).

A recent study in rice demonstrated that once entering the meristem cell cytoplasm, Hd3a interacts with a 14-3-3 protein to form a complex that moves into the nucleus. After interacting with FD in the nucleus, a ternary complex Hd3a/14-3-3/OsFD, also called "florigen activation complex" (FAC) is formed. There is no direct interaction between Hd3a and OsFD based on the results of crystal structure. The FAC directly induces transcription of *OsMADS15*, an Ortholog of *AtAP1* in the SAM (Taoka et al., 2011). In Arabidopsis, FT protein could also migrate into the nuclei to form a complex with the bZIP transcription factor FD. Even though interaction between FT and FD happened in yeast-one-hybrid assay that yeast 14-3-3 proteins may provide a scaffold (Abe et al., 2005). FT-FD protein complex is suggested to directly activate *AP1* in Arabidopsis (Wigge et al., 2005). This process may also require a 14-3-3 protein to form the *At*FAC.

In the SAM, *SOC1* (*SUPPRESSOR OF OVEREXPRESSOR OF CONSTANS* 1), a floral identity gene operates downstream of FT and FD (Abe et al., 2005; Kardailsky et al., 1999; Yoo et al., 2005). Upregulation of *SOC1* detected by *in situ* hybridization is supposed to be the earliest event during the LD-induced floral transition (Borner et al., 2000; Samach et al., 2000; Searle et al., 2006).

1.1.4 Homologs of *FT*

FT belongs to the phosphatidyl ethanolamine-binding protein (PEBP) gene family which is present in all eukaryote kingdoms. PEBPs function on controlling growth and differentiation through various signaling pathways (Chautard et al., 2004). In Arabidopsis, all six *PEBP* genes have been shown to affect flowering time, and they are classified into three clades. The FT-like clade includes *FT* and *TWIN SISTER OF FT* (*TSF*), the TFL1-like clade includes *TERMINAL FLOWER1* (*TFL1*), *BROTHER OF FT AND TFL1* (*BFT*) and *ARABIDOPSIS THALIANA CENTRORADIALIS* (*ATC*) and the MFT-like clade includes *MOTHER OF FT AND TFL1* (*MFT*) (Karlgren et al., 2011).

TSF, which encodes the most closely related protein to FT based on the amino acid level comparison, plays a role as flowering pathway integrator gene and has largely overlapping function with *FT*. In seedlings, *TSF* is mainly expressed in the vascular tissue of leaves, petioles and hypocotyl. Although *tsf* single mutants do not show obvious later flowering phenotypes in LD or SD, a loss of *tsf* function enhances the *ft* flowering phenotype under both conditions (Jang et al., 2009; Yamaguchi et al., 2005).

In the TFL1-like clade*, TFL1* is specifically expressed in the SAM where it acts as a repressor of flowering. In the inflorescence meristem, presence of *TFL1* inhibit*s* the formation of a floral meristem. Mutants of *tfl1* flower early with a terminal flower phenotype, which is similar to the phenotype of *35S::FT* plants (Kardailsky et al., 1999; Kobayashi et al., 1999). Likely, a single amino acid change on FT (Tyr 85 changed to His) and TFL1 (His 88 changed to Tyr) could re-program either protein function as the other one (Hanzawa et al., 2005). Based on the results of crystal structure, a divergent external three-dimensional loop of segment B on exon four may characterize the antagonistic activity of FT and TFL1 (Ahn et al., 2006). In sugar beet, two paralogs of *FT*, *BvFT1* and *BvFT2*, are both expressed in leaves and have antagonistic functions in determining flowering and growth. *BvFT1* transcription is promoted in SD conditions in autumn and inhibited during winter, whereas *BvFT2* is induced in LD conditions in spring. Three different amino acids in the segment B which forms a loop containing a 14 amino acid stretch determine the reversed function of BvFT1 and BvFT2. *BvFT2* was found to act as a FT-like gene and *BvFT1* was a TFL1-like gene although overall proteins belong to the FT-like clade (Pin and Nilsson, 2012; Pin et al., 2010).

The TFL1-like gene *BFT* has a redundant function with *TFL1*, because in *tfl1* background *bft* mutant plants enhance the termination phenotype of the primary inflorescence (Yoo et al., 2010). *MFT* is a single copy gene of the MFT-like clade in Arabidopsis. Constitutive expression of *MFT* slightly accelerates flowering in LD, but *ft* or *tfl1* mutant phenotypes do not change when combined with a reduced activity of *MFT* (Yoo et al., 2004).

1.2 Transcriptional regulation of *FT*

Expression of *FT* is highly regulated on the transcriptional level by several independent or interdependent inputs (Figure 1). The complex integration of many regulatory networks at *FT* ensures that plants flower at the optimal moment by adapting to the environmental conditions and internal cues.

1.2.1 General background on the transcription machinery

In eukaryotes, the transcription machinery contains three multi-component parts: i) transcription pre-initiation complex (PIC), ii) cofactors and iii) activators. The PIC is assembled by RNA Polymerase II (RNAPII), general transcription factors (GTFs) and a core promoter (Thomas and Chiang, 2006).

RNAPII, as the crucial catalytic enzyme of the PIC, is responsible for the transcription of protein-coding genes in eukaryotes. Yeast and human RNAPII both contain 12 subunits (Young, 1991). In yeast and human, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH have been isolated as GTFs (Hahn, 2004; Thomas and Chiang, 2006).

TATA box binding protein (TBP), a constituent of TFIID, binds to *TATA(A/T)A(A/T)(A/G)* sequences -25nt to -30nt upstream of the transcription start site (TSS as +1nt). The initiation sequence surrounding the TSS (-2nt to +5nt) and downstream promoter sequence (DPS) around +7nt to +30nt downstream of the TSS are associated with TAFs (TBP Associated Factors) which also belong to TFIIDs. Because all those proteins bind to DNA elements at core promoter, there is a basal level of transcription at core promoter (Thomas and Chiang, 2006). Transcriptional activation acts by increasing the efficiency of PIC assembly and transcription initiation (Roeder, 2005).

Cofactors serve as a bridge, linking the PIC and transcriptional activators. One class of cofactors assemble a complex called "Mediator". This complex was first characterized in yeast and found to comprise 20 subunits (Kim et al., 1994). Accumulating evidence suggests that RNAPII, GTFs and other regulatory factors are not recruited to individual gene promoters, but form "transcription factories", physically distributed as sub-nuclear foci and that genes associate with these factories for transcription (Edelman and Fraser, 2012). The concept of transcription factories may easily explain co-expression of two distant genes localized on the same or different chromosomes.

1.2.2 Characterization of the *FT* promoter

Limitations in detecting *FT* transcript by *in situ* hybridization, require the application of β-glucuronidase (GUS) gene, driven by 8.9 kb *FT* promoter, to elucidate the spatial and temporal mRNA signal of *FT* (Takada and Goto, 2003). Promoter truncation results demonstrate that the 5.7 kb sequence upstream of the *FT* translation start codon contains sufficient regulatory elements to mediate spatial and temporal expression of *FT* in LD

conditions. In addition, 4 kb *FT* promoter sequence or shorter failed to drive *GUS* in the phloem of leaves in wild type (WT) and *35S::CO* plants. Furthermore, these promoter constructs combined with a cDNA of *FT* could not rescue the late flowering phenotype of *ft* mutants (Adrian et al., 2010). In sum, a 1.7 kb (from 4 kb to 5.7 kb) promoter region is necessary for *FT* transcription in leaves. Three phylogenetically conserved blocks named *Block C, B* and *A* are identified on the distal, middle and proximal *FT* promoter region respectively (Figure 1). In the background of 5.7 kb *FT* promoter, mutagenesis of conserved sub-blocks of 6 nt to 16 nt, such as *P1/P2* and *S2*, in *Block A* strongly reduce *FT* transcription (Adrian et al., 2010). *Block A* (~ 300bp) contains the *FT* 5´ UTR and conserved sequences corresponding to the minimal promoter and consensus binding sites for several known transcription factors (Adrian et al., 2010; Castillejo and Pelaz, 2008; Kumar et al., 2012; Tiwari et al., 2010). The functions of *Block B* and *Block C* are largely unknown. In addition, it is still unknown how these three conserved blocks communicate with each other.

The *FT* promoter provides a platform of natural variation for flowering time in Arabidopsis. Based on the two Arabidopsis accessions, Est-1 and Col-0, one flowering time Quantitative Trait Locus (QTL) maps region up to 6.7kb upstream of the *FT* TSS (Schwartz et al., 2009). Another independent group analyzed Ull2-5 and Col-0 to find one QTL localized at a 9 kb intergenic region between *FT* and its neighboring gene, *FAS1* (Strange et al., 2011). Taken together, a total of seven independent experiments with nine different *A. thaliana* accessions have been described, where flowering time QTLs map to *FT* locus (Schwartz et al., 2009).
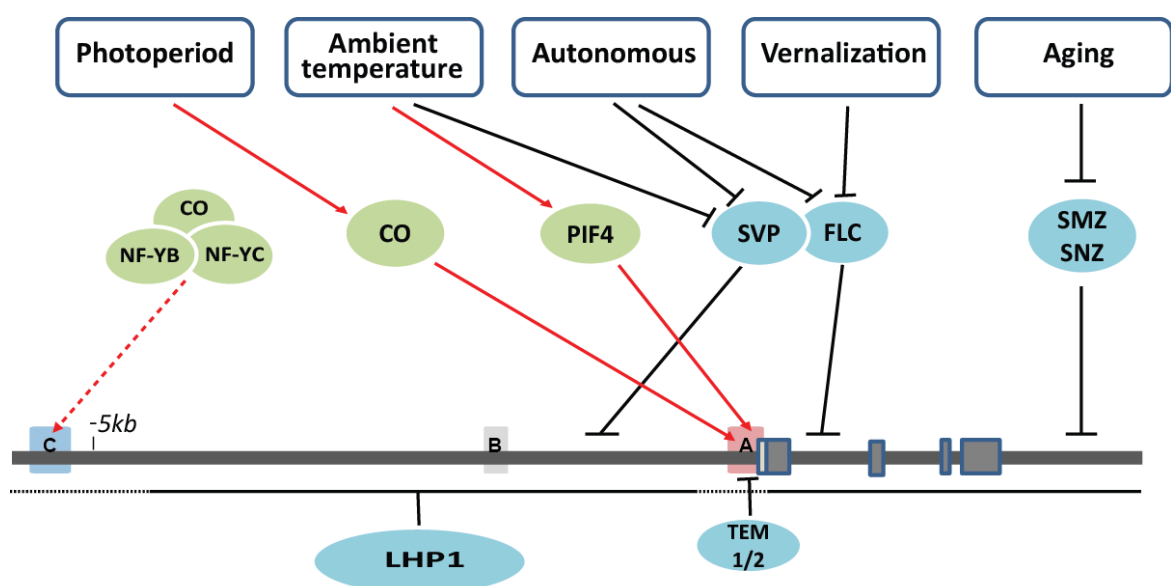

Figure 1. Schematic network of transcription regulation of *FT*.

Transcription of *FT* is regulated by different flowering time pathways and their core components. In the photoperiod pathway, the main activator CO may directly bind to the proximal promoter *FT*. CO interacts with members of the NF-YB and NF-YC families to form a trimetric complex that participates in the photoperiod-dependent regulation of *FT*. In the low ambient temperature and autonomous pathway, SVP represses *FT* through association with *CArG*-boxes in the upstream promoter. In the vernalization pathway, FLC binds preferentially to the first intron of *FT*, which contains a *CArG*-box. *FT* repression by FLC partially depends on SVP. In the aging pathway, SMZ as direct target of *miR172* represses *FT*. SMZ binds to a region downstream of the 3′ UTR. TEM1 represses *FT* and associates with sequences found at the 5′ UTR. PIF4 activated *FT* expression and binds to the 5′UTR region.in the high ambient temperature. The epigenetic repressor LHP1 binds to a large regions spanning the promoter and gene body of *FT* (Dashed line represents the low abundance of LHP1 binding to *FT*).

1.2.3 Activators and repressors of *FT*

Several key components of flowering time pathways have been described as upstream activators or repressors of *FT* (Figure 1).

CO and NF-Y factors: In LD conditions, diurnal rhythmic patterns of *FT* transcript accumulation fit well with the temporal abundance of CO protein. Therefore, *FT* mRNA has a major peak appearing before darkness, and it also shows a minor peak in the morning (Corbesier et al., 2007; Turck et al., 2008).

In LD, CO-mediated activation of *FT* requires the 5.7 kb *FT* promoter and it has been shown that CO directly associates with the *FT* proximal promoter. *In vitro*, CO binds to *TGTG(N2-3)AT(G)* motifs present within *Block A* (Adrian et al., 2010; Tiwari et al., 2010). Furthermore, in a recent ChIP experiment with lines expression a CO tagged HA under the control of the *CaMV 35S* promoter, enrichment of *FT* proximal promoter was shown, although the relative enrichment signals were comparatively low (Song et al., 2012). Further studies are required to fully characterize the binding signals of CO on *FT* and the structural impact of binding *in vivo*.

In transient expression assays, reporter gene signals driven by 1kb and 5.7kb *FT* promoter sequence show similar promoter activity in response to presence of CO (Adrian et al., 2010; Tiwari et al., 2010). However, *in planta,* CO completely fails to drive *FT* when fused to only 1 kb or 4 kb *FT* promoter sequence which suggests that *FT* transcription is influenced by chromatin background (Adrian et al., 2010).

CO interacts with Nuclear Factor YB and YC (NF-YB and NF-YC; also called HEME ACTIVATOR PROTEIN 3 (HAP3) and HAP5) to assemble a trimetric complex *in vitro* and *in planta*. NF-YAs contain a domain with sequences homology to CCT domain proteins, such as CO, CO-Like protein and TOC1 (Wenkel et al., 2006). In general, the NF-YB and NF-YC form a dimer followed by interacting with NF-YA to produce a trimetric complex. The

resulting NF-YA/-YB/-YC trimer is capable of binding to *CCAAT*-motifs with high affinity and participates in transcription activation (Bi et al., 1997; Mantovani, 1999). In yeast, a DNA-binding trimer of HAP2/HAP3/HAP5 requires association of subunit HAP4 to provide the activation function to the whole complex (McNabb and Pinto, 2005). In Arabidopsis, overexpression of NF-YA delays flowering, so CO may replaces of NF-YA position to form a complex with NF-YB and NF-YC (Wenkel et al., 2006).

NF-YB and -YC, both contain histone-fold domains (HFDs) form a dimer which mimics the H2A/H2B core histone dimer. In addition, the trimetric NF-Y complex binds to a 25 nt stretch including one *CCAAT* motif from the *HSP70* promoter. Interestingly, during transcription initiation NF-YB is ubiquitinated at Lys138, similar to H2B, which is mono-ubiquitinated on Lys120 (Nardini et al., 2013). These observations indicate that NF-Y complexes could replace nucleosomes at some loci containing *CCAAT*-motifs.

In contrast to the single copies for NF-YA, -YB, -YC in *yeast* and mammals, there are 10 NF-YAs, 13 NF-YBs, and 13 NF-YCs homologs in Arabidopsis. Moreover, all NF-Ys show variable expression in root, cotyledons, young seedling and flowers. *GUS* driven by promoters of *NF-YA2, -YA6* and *-YB2, -YB3, -YB12* shows phloem-specific expression in leaves. Furthermore, *NF-YC3, -YC4 or -YC9* promoters fused to *GUS* appear strong signals in the phloem and other cell-types of leaves (Kumimoto et al., 2010; Siefers et al., 2009). In LD, *nf-yc3;4;9* triple mutants delay flowering, and overexpression of CO cannot fully accelerate flowering in *nf-yc3;4;9* triple mutants. All three subunits of NF-YC interact with NF-YB2 and NF-YB3 *in vitro*. The results suggest that CO requires NF-YC to activate *FT*.

CIB1 and PIF4: CRYPTOCHROME-INTERACTING BASIC-HELIX-LOOP-HELIX 1 (CIB1) interacts with CRY2 in a blue light-dependent manner. CIB1 associates with the *FT* TSS region and the second intron in ChIP assay, and promotes CRY2-dependent floral transition (Liu et al., 2008). In high red light environment, active PHYB directly interacts with a bHLH domain transcription factor PHYTOCHROME-INTERACTING FACTOR 4 (PIF4), which leads to the ubiquitin-proteasome degradation of PIF4. When PHYB is inactive, accumulation of PIF4 allows shade-avoidance genes expression (Stamm and Kumar, 2010). In response to high ambient temperature in SD, PIF4 directly activates flowering dependent on the sequences found in the 5´ UTR of *FT* (Kumar et al., 2012).

FLC and SVP: The MADS-domain transcription factor FLOWERING LOCUS C (FLC), whose levels decrease gradually during vernalization (Sheldon et al., 2000), represses *FT*

presumably by binding to a *CArG*-box in the first intron in leaves (Lee et al., 2007). In low ambient temperature, another MADS-domain factor, SHORT VEGETATIVE PHASE (SVP), represses *FT* by binding to a 2 kb region upstream of the *FT* promoter. FLC and SVP might form a complex and it seems that they can collectively but also individually repress *FT* (Li et al., 2008) In addition, SVP and FLC also directly repress another floral integrator, *SOC1* in the SAM (Li et al., 2008; Searle et al., 2006).

TEM1 and SMZ: TEMPRANILLO1 (TEM1), is a RAV-family AP2-like protein that counteracts the CO-mediated activation of *FT*. It directly binds to the *FT* 5´ UTR *in vivo*. TEM1 promoter driving *GUS* expression illustrates that the expression levels of *TEM1* decrease gradually when the plant ages. Further, TEM1 expression domains are limited to the distal part of the leaf in older plants (Castillejo and Pelaz, 2008). It has been speculated, that *FT* expression domains expanding to the major veins during aging could partially be due to the reduction of TEM1.

In the aging pathway, *miR156* targets *SQUAMOSA BINDING PROTEIN LIKE 9 (SPL9)* and *SPL10*, which are positive regulators of *miR172* (Wang et al., 2009; Wu et al., 2009). AP2-like transcription factors such as SCHLAFMUETZE (SMZ), SCHNARCHZAPFEN (SNZ), TARGET OF EAT 1 (TOE1), TOE2 and TOE3 are downstream targets of *miR172*. By genetic interaction, SMZ and SNZ are shown to repress *FT* and ChIP-chip data of SMZ show binding occurring approximately 1.5 kb downstream of the *FT* coding region. Interestingly, *TEM1* is also in the binding target list of SMZ, potentially connecting the two independent age-related repressors of *FT* (Mathieu et al., 2009). Accumulation of *miR156* is gradually reduced with aging and results in increment of *miR172* (Wang et al., 2009; Wu et al., 2009).

1.2.4 Polycomb repressive complexes

A nucleosome consists of 147 bp DNA and a histone (H) octamer including two copies each of core Histone 2A (H2A), H2B, H3 and H4. Two nucleosomes are connected by linker DNA and may be further compacted by the binding of Histone 1 (H1) to two adjacent nucleosomes (Bhasin et al., 2006). In eukaryotic nuclei, nucleosomes form the fundamental repeating units of chromatin. Eukaryotic gene transcription requires an open chromatin structure. Therefore, histone modification presents an important layer for controlling transcription. Compared to "naked" DNA in prokaryotes, eukaryotic gene expression is definitely determined by the

chromatin modifications. Dynamic of nucleosome appears to affect all steps of transcription from PIC formation, activators recruitment to elongation (Li et al., 2007).

Tri-methylated lysine of the amino-terminal tail of H3 at position 27 (H3K37me3) is a repressive histone modification mark, and a large number of Arabidopsis genes (~15%) are marked by H3K27me3 (Turck et al., 2007; Zhang et al., 2007a). Several key floral transition genes, such as *FLC*, *FT* and *AG* are enriched with H3K27me3 (Farrona et al., 2008). In *Drosophila melanogaster*, H3K27me3 is deposited by Polycomb Repressive Complex 2 (PRC2). The PRC2 includes four core components: the SET-domain histone methyltransferase *Enhancer of Zeste* (E(Z)) in association with *Suppressor of Zeste* (SU(Z)12) and the WD40-domain proteins *Extra sex combs* (ESC) and *Multicopy suppressor of Ira* (MSI) (Schwartz and Pirrotta, 2007). In Arabidopsis, each subunit with the exception of ESC has a small clade of homologue genes. *MEDEA* (*MEA*), *CURLY LEAF* (*CLF*) and *SWINGER* (*SWN*) are homologs of *E(Z)* methyltransferase gene (Farrona et al., 2008). Mutants of *clf* exhibit curled rosette leaves, earlier flowering and pleiotropic morphology phenotypes (Goodrich et al., 1997). H3K27me3 abundance is dramatically reduced on the *FT* gene body and its promoter in *clf* mutants. *GUS* expression driven by the *FT* promoter is enhanced in the phloem cells of leaves in *clf* grown in LD and SD, but not detected in other cell types. The GUS signal is also restricted to veins in cotyledons of *clf;swn* double mutants. With aging, these double mutants develops to a callus that lack differentiated phloem. It suggests that PRC2 can regulate the transcription levels of *FT* but not alter or control its tissue-specific expression (Farrona et al., 2011).

In Arabidopsis, the distribution of LIKE HETEROCHROMATIN PROTEIN1 (LHP1) shows a high overlap with the distribution of H3K27me3 (Turck et al., 2007; Zhang et al., 2007b). Mutants of *lhp1*, also called *terminal flower 2* (*tfl2*), flower early and exhibit a terminal flower phenotype in LD and SD (Kotake et al., 2003). *GUS* expression data suggests that expression of *CO* and *FT* are restricted to the phloem of leaves. A translational fusion between the genomic *LHP1* region and *GUS* (*gLHP1:GUS*) expresses strongly in the major veins of the proximal leaf, which is opposite to the minor veins of the distal leaf. In addition, expression of *LHP1* is not in a phloem-specific manner. Thus, LHP1 acts antagonistically to CO on *FT* transcription (Takada and Goto, 2003). As mentioned above, 4 kb *FT* promoter sequence is not able to drive *GUS* expression in WT or *35S::CO* background, however, GUS signal are observed in the major veins of *lhp1* mutants. Chromatin structure changes mediated by H3K27me3 and LHP1 negatively regulate *FT* transcription. However, when induction of

CO by treating *35S::CO:GR* plants by Dex, H3K27me3 and LHP1 abundance do not significantly change (Adrian et al., 2010). How to explain that the *FT* expression levels increased without removing the repressors? It could reason that the change of H3K27me3 or LHP1 occurs in the phloem nuclei but not the whole plants with activation of *FT*.

Chromatin plays important role for gene expression. On *FT* locus, how could the conserved blocks especially the distal *Block C* and *Block A* in the proximal promoter communicate with each other? We need the tools to measure their positions in chromatin background.

1.3 Chromosome conformation capture and cell sorting

Chromatin loop occurs when two genome loci physically interact with each other and they are closer than the intervening sequences. Chromatin confirmation capture (3C) and its further developments are powerful methods to detect chromatin loop structure *in vivo*.

1.3.1 3C to measure chromatin interaction

In eukaryotes, chromatin is non-randomly organized and compacted into the nucleus by many folds. This condensation provides a big challenge for transcription of chromosomal DNA (Gondor et al., 2008). Transcriptional activation and repression involve interactions with long-range regulatory elements, such as enhancers or insulators. The model of looping structures to explore such interactions between remote *cis*-elements and promoters has been raised since more than twenty years (Tolhuis et al., 2002). Notably, in 2002, Dekker and colleagues invented the 3C technology. Based on 3C method, a chromosomal looping structure is discovered experimentally in yeast (Dekker et al., 2002). 3C is a powerful method to directly study the chromatin interaction structure between two genomic loci, for example enhancer-promoter looping. In the past ten years, 3C has been widely applied in yeast, animal cells, and a few cases have been reported in plants.

In yeast, a gene loop persisting on the *GAL10* locus is associated with a "transcriptional memory" defined as gene's rapid reactivation kinetics following a cycle of activation and repression (Laine et al., 2009). In mammals, the *Kit* gene is one example during early erythropoiesis. There is a loop formed between a 114 kb-upstream-enhancer and its proximal promoter region to induce *Kit* gene expression. However, *Kit* gene is silenced in mature cells, while the positive enhancer-promoter loop is replaced by another loop formed by a 58 kb-

downstream-*GATA* element and promoter. That means different conformations of chromatin structure determine *Kit* gene's transcription (Jing et al., 2008). Another representative example of looping structure is provided by the beta-globin genes, which contain a cluster of DNase I hypersensitive sites. One hypersensitive region localized ~100 kb upstream of gene, mediates long-range interactions with the proximal promoter as well as several other *cis*-elements on different chromosomes to regulate beta-globin gene transcription (Xu et al., 2010). In addition, looping conformation of promoter-terminator suggests a core-helix model for the topology of active ribosomal RNA (rRNA) genes viewed as "Christmas tree" (Denissov et al., 2011).

In the plant field, the first case of long-range looping has been reported in maize. The *b1* gene has two epialleles *B-I* and *B'*. A hepta-repeat region localized ~100 kb upstream of promoter could physically interact with the transcription initiation region, and the loops are shown to be tissue specific (Louwers et al., 2009). In Arabidopsis, at the *FLC* locus, 5' and 3' flanking regions are involved in a loop formation. This loop structure is disrupted in the early stage of vernalization treatment, but independent of *FLC* expression levels (Crevillen et al., 2013).

Compared Arabidopsis to yeast, the ratio of their genome size is 10 times, and the ratio of gene number is 6 times. But human's genome is 30 times of that in Arabidopsis with a similar number of genes (http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/G/GenomeSizes).

Therefore, chromatin organization in Yeast may help studying and understanding the chromatin looping structure in Arabidopsis. In Yeast, 3C followed by ligation product enrichment globally mapped long-range interactions, which revealed that genome organization significantly links to transcriptional regulation and co-expression of genes during the cell cycle (Tanizawa et al., 2010).

1.3.2 Chromosome organization in the nucleus

3C is used to measure chromosome interactions between the interest region and any other genome loci on a limited locus (one-versus-some). Its derived technology 4C (circular chromosome conformation capture) has been adopted to identify the interacting regions to the H19 imprinting control region during maturation of embryonic stem cells (one-versus-all) (Zhao et al., 2006). Hi-C as an extension version of 3C is capable of detecting long range interactions in an unbiased, genome-wide fashion (all-versus-all). Hi-C has been widely used

in yeast and mammalian cells for modelling the organization of chromosomes (van Berkum et al., 2010).

To investigate the complexity of human chromosomes based on Hi-C method at a resolution of 1 Megabase, a new model called "fractal globule" is established and distinct from the commonly used "globular equilibrium" model. The new 3D model indicates the function of maximally dense packing to prevent easy folding and unfolding at any genomic locus (Lieberman-Aiden et al., 2009). Based on high resolution Hi-C map, local chromatin interaction domains termed "topological domains" (TD) have been characterized on Megabase scale. These domains correlate with regions that constrain the spread of heterochromatin, and are stable across different cell types and conserved across species (Dixon et al., 2012). Recently, Hi-C was carried out in *atmorc1* and *atmorc6* mutants, the whole genomic interaction map revealed that pericentromeric heterochromatin de-condensed as well as the interaction of pericentromeric regions increase. Mutants of *atmorc1/6* cause de-repression of DNA-methylated genes and transposable elements but no losses of DNA methylations or histone methylations (Moissiard et al., 2012). Due to the low resolution of Hi-C map in Arabidopsis, it could be difficult to see the TD structure identified in mammalian cells or small-distance interactions.

In yeast and animal studies of 3C and Hi-C, researchers commonly sample single cell types. However, plant tissue consists of different cell-types and it is not easy to separate single types of cells. Thus, it is a challenge to get pure and enough cells of interest for chromatin structure study in plants.

1.3.3 Cell/nuclei sorting methods

Several methods for purifying cells of one tissue/the same type have been developed such as fluorescence-activated cell sorting (FACS), laser capture microdissection (LCM) and the methods of cultured cell lines. FACS and LCM are widely used for mRNA expression and histone modification profiles *in vivo* (Deal and Henikoff, 2010). In maize, LCM-mediated isolation of epidermal cells or vascular tissues from 6-μm-thick sections of coleoptiles were used for differential gene expression by microarray (Nakazono et al., 2003). By FACS, transgenic plants with cell-specific-promoters fused to *GFP* were created to isolate five individual types of cells from roots for mRNA expression analysis (Birnbaum et al., 2003). However, both LCM and FACS techniques require expensive sorting machinery and careful

operation, but their yields of isolated specific cells are limited. Recently, a new method called INTACT (isolation of nuclei tagged in specific cell types) was applied in Arabidopsis and *Caenorhabditis. elegans* (Deal and Henikoff, 2010; Steiner et al., 2012).

The INTACT system needs a cell/tissue-specific promoter to drive an engineered "nuclear targeting fusion" (NTF) protein. NTF protein is composed of three parts: the WPP domain of Arabidopsis RAN GTPase ACTIVATING PROTEIN 1 which targets the fusion protein to the nuclear envelope, GFP for visualization and the biotin ligase recognition peptide which serves as a substrate for the *Escherichia coli* biotin ligase BirA. *BirA* is constitutively driven by the *ACT2* promoter, and biotinylates NTF in every cell type of interest. Biotin-labeled nuclei are purified with magnetic streptavidin-coated beads (Deal and Henikoff, 2010; Deal and Henikoff, 2011). In plants and animals, INTACT methods are used to isolate pure cell types of interest for gene expression, ChIP and other studies.

2. Aim of this study

Previous studies demonstrated that it is critical to control *FT* transcription in response to environmental changes and endogenous cues. At the *FT* promoter, a 1.7 kb sequence region between 4 kb and 5.7 kb upstream of the transcription start site is necessary for *FT* expression in response to day length. Conserved *Block C* is localized at that 1.7 kb region.

The aim of my study was to unravel the function of conserved *Block C* by identifying novel *cis*-regulatory elements required for the response to inductive long day conditions. A central question of this work was how *Block C* interacts with the proximal promoter to drive *FT* expression. To characterize the communication between the regions and demonstrate a physical interaction between the regulatory regions, genetic experiments as well as new tool 3C were carried out. In addition, to discriminate the chromatin states such as the repressive mark H3K27me3 at the *FT* promoter in phloem compared to the other tissues, the new sorting method INATCT was used in this study.

3 Materials and Methods

Plants growth and flowering time measurement

Seeds were sterilized by 75% ethanol for 5 min followed by 100% ethanol for 5 minutes (min). Sow the seeds on GM media supplemented by 1% sucrose. After stratification at 4°C for 3 days, plants were grown in climate chambers at 22°C in LDs (16 hours light/8 hours dark, 70% light), or SDs (16 hours light/8 hours dark, 70% light). The starting time of growth is regarded as 0-day-old. During collecting samples, ZEITGEBER TIME (ZT) 16 was in the light before switching to darkness and ZT 24 was in the dark before switching back to light condition. For measuring flowering time or collecting plant samples grown on soil, seeds were sowed on soil and treated in 4°C room for 3 days. After that, move the soil pots to the climate chambers or normal greenhouses. Under LDs or SDs, total rosette leaves and cauline leaves were counted to estimate the flowering time.

Bacterial transformation

Competent cells of DH5a or DB3.1 were taken out from -80°C freezer and kept on ice for 5 min until they thawed. Add plasmids, ligation products or BP/LR products into the competent cells, mix a bit and leave the tube on ice for 15 min. Next, the mixture was treated at 37°C for 100 seconds (sec) followed by on ice for 2 min. After adding 800ul LB culture liquid, shake at 200rpm in 37°C chamber for 1 hour. Finally, spread the transformed cells on the LB media with antibiotics, and leave the plate up-down in 37°C chamber overnight. For electronic transformation of agrobacteria, mix the plasmid DNA and competent cells well in the transformation tube and carry out the electronic shock using the transformation apparatus (Bio-Rad). After adding 800ul YEB culture liquid, shake at 200 rpm at 28°C chamber for 2 hours. Finally, spread the transformed cells on the YEB media with antibiotics, and leave the plate up-down in 28°C chamber overnight.

Floral dip and transgenic plant selection

For Arabidopsis, agrobacterium-mediated transformation was carried out according to the floral dip method. After collecting the T1 seeds, they were sowed on the Basta soil tray and selected by spraying with 250 mg/L glufosinate (BASTA®) 2-3 times. T2 seeds were sowed on GM plate with PPT antibiotic to test the 3:1 segregation. T3 seeds can be confirmed as the homozygous or heterozygous genotype according the segregation. T2 or T3 plants are used for flowering time measurement.

Genomic DNA isolation

Isolation methods of plant genomic DNA was according to CTAB method and *DNeasy®
Plant Kit* (Qiagen). DNA was dissolved with dH_2O.

Total RNA isolation

Plant total RNA was isolated by using *RNeasy Plant Mini Kit* (Qiagen). All the stuffs in the
whole procedure should be avoided of RNease contamination.

cDNA synthesis and reverse transcription PCR (RT-PCR)

Total RNAs were treated by DNase I (Ambion) for 2 hours following by inactivating the
DNase I. in order to get liner RNA, 5 µg RNAs was heated up to 75°C for 5 min. During
reverse transcription reaction, mix the RNAs with reaction buffer, dNTPs, oligo dT18 and
transcriptase. Then the reaction was happened at 42°C for 3 hours, and the transcriptase
activity was quenched by 65°C 10 min. The synthesized cDNA can be diluted in 1:3. 2 µl
cDNA was used for qPCR. qPCR reaction includes 2 µl cDNA, 1 µl primers (10 mM), 5 µl
SYBR Green Master Mix (ABI) and 2 µl dH_2O.

Alternatively, while using EVA Green dye reagent (Biotium) in 20 µl PCR reaction, it
included 2 µl cDNA, 1µl primer A and B (10 mM each), 0.2 µl Brown Taq (homemade), 1 µl
20× Eva green, 0.4 µl dNTPs (10 mM), 1 µl 20× basic buffer (200× basic buffer stock
contained 0.8 M KCl, 0.2 M Tris and 0.05 M $MgCl_2$) and 14.4 µl dH_2O. 10 µl or 20 µl qPCR
was performed by Bio-Rad or Roche real-time apparatus respectively.

Plasmid and cloning procedure

To conduct the bombardment assays, *Block C* and 1.7kb region were amplified with *TaKaRa
Ex^TM* Tag DNA polymerase by primers *FT5.7kbFW350/FTpblockC_RE* and
FT5.7kbFW350/FT1.7kb_RE respectively and *8.1kbFTp::FTcDNA-pGREEN* (from Turck
group) was used as the PCR template. To generate entry clones of that promoter, PCR
products were introduced into *pDONR207* vector (Invitrogen) through BP reactions of
GATEWAY method. LR reactions were carried out using the *GW_NOSmin::GreenLUC*
(provided by Samson Simon, MPIPZ) as destination vector. *pCR®2.1-TOPO®* (Invitrogen)
was used to do T-A cloning according to the kit protocols. *BAC* clone F5I14 was amplified in
SW102 cells is 109.590 kb long and contains the full *FT* locus including intergenic
sequences.

GUS histochemical staining

Young seedlings were incubated in 90% Acetone on ice for 30 min, rinsed with 50mM sodium phosphate buffer and incubated at 37°C in GUS staining solution (0.5 mg/ml X-Gluc, 50mM sodium phosphate buffer, 0.5 mM potassium ferrocyanide, 0.5 mM potassium ferricyanide, 0.1% Triton X-100). After overnight incubation, samples were washed with 50mM sodium phosphate buffer for 30 min and 70% ethanol several times until leaves turned white. The GUS staining was visualized and photographed under a stereomicroscope (Leica).

Bombardment

30 µg gold particles were prepared for every 10 bombardments. Firstly, 70% ethanol was used to wash the gold particles. Then gold-ethanol was spun down and washed three times with sterile water. Finally the gold particles were suspended in 500 µl of 50% sterile glycerol.

In each bombardment, 15 µg DNA in total was used including 5 µg *35S::RedLUC-pJAN*, 5 µg *Block C* or *1.7 kb FTp::GreenLUC-pGREEN* and 5 µg *35S::gCO-pBKS* or *pBKS* empty vector. The DNA was mixed with 50 µl prepared gold beads, 50 µl 2.5M $CaCl_2$ and 20 µl 0.1M spermidine. After two washes with ethanol, the DNA-gold mix was suspended in 50 µl 100% ethanol which can be used for two independent bombardments. 5-10 mm Arabidopsis leaves were transformed by the Biolistic^{TM} Particle Delivery System (BIO-RAD, PDS-1000/HE). After 12-24 hours incubation of the samples, 1 mM luciferin was sprayed on the leaves and after one minute the emitted light was immediately measured with the help of a cooled CCD-camera adapted with optical filters to detect Red-LUC and Green-LUC independently. The ratios of GreenLUC/RedLUC signals were calculated with the help of the excel macro Chroma-LUC^{TM} Calculator version 1.0 (Promega).

Western blot

Collect ~100 mg plant sample, and grind it to fine powder. Add 100 µl hot Laemmli buffer, mix immediately and heat it for 15 min at 95-100°C. After centrifuging at 13,000 rpm for 5 min, load 10-20 µl protein per lane onto SDS-PAGE gel (4% stacking upper gel and 12% lower gel). After electrophoresis for 90 min at 200V, protein was electro-transferred from gel to PVDF membrane for 60 min at 100V in 4°C room. Incubate membrane in the blocking buffer TBS-T with 5% milk for 1 hour at room temperature. Wash three times for 5 min with 15 ml of TBS-T. Incubate membrane and primary antibodies (at the appropriate dilution) in 10 ml TBS-T with gentle agitation overnight at 4°C. Remove the primary antibodies and wash three times for 20 min with 15 ml of TBS-T. Incubate membrane with appropriate

HRP-conjugated secondary antibodies with gentle agitation for 1 hour at room temperature. Wash three times for 20 min with 15 ml of TBS-T. After wash once for 20 min with 15 ml of TBS, detect protein signal with ECL kit.

Chromatin Inmunoprecipitation (ChIP)

Seedlings from soil or GM plates were harvested in $1\times$ PBS buffer with 1% formaldehyde. After twice of vacuum 5 min and release 5 min in a vacuum hood, 0.1 M glycine was used for quenching FA. Then grind samples to fine powder in liquid N_2, extract nuclei and get rid of debris through 70 µm and 40 µm nylon mesh. Sonication of the chromatin was carried in a Bioruptor filled with H_2O and a few floating ice. Spin 15 min at full speed at 4°C and transfer the supernatant to a new tube. For inmunoprecipitation of protein-DNA complex, add the specific antibody and incubate in soft agitation in a rotator overnight in the cold room. Then, add protein A or G beads to each sample and continue incubating on rotator for 1-2 hours in the cold room. Next, use magnet to attract the beads and discard the supernatant. Wash the beads with High salt buffer, low salt buffer, LiCl buffer and TE buffer in sequence. Finally, elute the inmunoprecipitated chromatin from beads and add proteinase K to digest protein at 65°C following by extract free DNA by ethanol precipitation.

Chromosome Conformation Capture (3C)

3C procedure in Arabidopsis was carried out as shown in Figure 2. 10-day-seedlings were collected from GM plates or soil and the total weight is around 3 g. Cross-link the seedlings in $1\times$ PBS with 1% formaldehyde. The vacuum procedure was as same as ChIP method. After isolating intact nuclei in Nuclei isolation (NIB) buffer, nuclei were filtered through two layers of nylon meshes (70 µm and 40 µm) following by twice of spinning down and washing in 1 ml NIB buffer. The purified nuclei were re-suspended in 500 µl $1.2\times$ NEB-buffer 4 with a final concentration of 0.15% SDS which functions as destroying the nuclei membrane and loosening the compacted nuclei. Shake at 900 rpm for 60 min at 37°C. In order to sequester SDS, add 125 µl 10% Triton-X100 to a final concentration of 2% in a final 625 µl volume. Shake at 900 rpm for 60 min at 37°C. Next, 50 µl of 500 U restriction endonuclease NlaIII (10 U/µl) and 10 µl $10\times$ NEB-buffer 4 were used for digesting the chromatin DNA in a final volume of ~700 µl for 18 h at 200 rpm at 37°C. To inactivate NlaIII enzyme, 70 µl 10% SDS in a final concentration of 1% was used and incubate for 10 min at 65°C. To establish ligation reaction, 700 µl $10\times$ T4 DNA ligation buffer, 700 µl 10% Triton X-100 and 4.8 ml dH_2O were added to the ~800 µl digested nuclei and the final whole volume was up to 7ml. SDS

was diluted into 0.1%. Incubate at 37°C for 1hour to sequester SDS. Ligation reaction was carried out by adding 400 U T4 DNA ligase (100 U/µl) for 5 h at 16°C following by 45 min at room temperature. De-crosslink was treated with 40 µl 5 mg/ml Proteinase K (PK) over night at 65°C. At last, after purifying 3C ligation products by phenol: chloroform: isoamyl-alcohol (25:24:1), DNA was precipitated ethanol method and the interaction frequency was measured by qPCR.
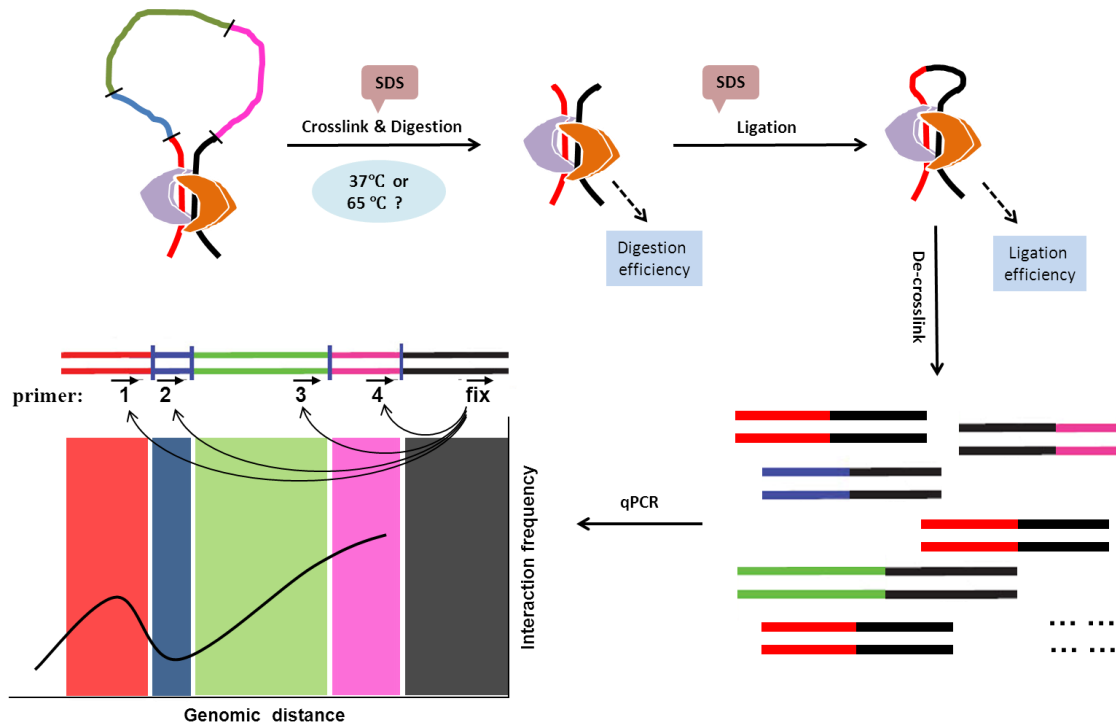
Figure 2. 3C procedure.

Crosslink plant materials and isolate nuclei. Chromatin DNA is digested by restriction enzyme followed by intramolecular ligation. De-crosslink protein and purify DNA. Set 3C primers in one direction and measure interaction frequencies between fixed and any other fragments by qPCR. Digestion and ligation efficiencies are the critical quality controls after digestion and ligation steps respectively. Treatment with 37°C or 65°C before digestion is used to loosen chromatin.

Isolation of Nuclei Tagged in Specific Cell Types (INTACT)

Collect 3-5 g fresh seedlings. Cross-link the seedlings in 1× NPB with 1% formaldehyde and vacuum as shown in ChIP. (Note: when the isolated nuclei will be used for gene expression analysis, skip the formaldehyde crosslink step). Grid seedlings to fine powder and filter the suspension nuclei through the 70 µM and 40 µM cell strainer. Spin down the nuclei and gently resuspend the pellet in 1 ml ice-cold NPBd. Wash M-280 streptavidin-coated Dynabeads with NPB and add it to each sample with $2×10^7$ beads. Rotate the 1.5 ml tubes for

30-60 min at 4°C room. Transfer 1 ml nuclei mixed with Dynabeads to 15 ml tube, and add 9 ml ice-cold NPBt buffer. Nuclei sorting column was set up in cold room. The 10 ml nuclei before sorting were called "Whole" nuclei, and the nuclei bonded with beads which were attracted by the magnet were called "beads" nuclei or sorted nuclei and the nuclei which flowed through the column were named as "flow-through" nuclei. Release the beads nuclei in 1 ml NPBt. Spin down the beads nuclei and resuspend the pellet in 20 μl ice-cold NPB. The sorted nuclei can be used for qRT-PCR or ChIP assays.

Experimental instruments and regents

DNA purification kit: Qiagen

RNA purification kit: Qiagen

cDNA synthesis kit: Fermentas

iQ™ SYBR® Green Supermix: BIO-RAD

Sub-Micron gold particles: BIO-RAD

Eva green dying (20×): Biotium

Luciferin: 50mM in stock

Proteinase K (PK): 5mg/ml in stock

RNase A: 10 mg/mL in stock

MG-132 proteasome inhibitor: Calbiochem

Complete protease cocktail tablets: Roche

Protease inhibitor cocktails: Sigma

70 μm and 40 μm Cell strainer filter: BD Falcon™

MACS Magnet: Miltenyi biotec

IQ5 Real-time cycler: Bio-RAD

384-well Real-time cycler: Roche

PDS-1000/HE Biolistic^{TM} Particle Delivery System: Bio-RAD

Stereo-microscope: Lica

Confocal microscopy: Lica

Enzymes

DNase I: Ambion

Gateway cloning enzymes for BP and LR reaction: Invitrogen

T4 DNA ligase: NEB

Tag DNA ligase: NEB

Tag DNA polymerase: NEB

NlaIII restriction endonuclease: NEB

Vectors

pCR®2.1-TOPO®: Invitrogen, for cloning.

pGEM-T: Promega, for T-A cloning.

pDONR201, *pDONR207*: Invitrogen, for Gateway BP reaction to create entry clone.

35S::REdLUC-pJAN: for Bombardment.

pBKS: for cloning and Bombardment.

GW_GreenLUC-pGREEN: inserting promoter with Gateway for Bombardment.

GW_FTcDNA-pGREEN and *GW_GUS-pGREEN*: for Gateway LR reaction to insert the promoter ahead of *FT* or *GUS* gene.

GW_NOSmin::FTcDNA-pGREEN and *GW_NOSmin::GUS-pGREEN*: for Gateway LR reaction to insert promoter fused to *NOS* minimal promoter

GW_NTF-pBlue: after LR reaction, with Kanamycin resistance in bacterial and plant. This vector was replaced of Basta gene with *NPTII* on the vector of *GW_NTF-pGREEN*.

Antibiotics

Antibioticname	Abbreviation	Final concentration (ug/L)
Ampicillin	Amp	100
Carbenicillin	Carb	100
Gentamycin	Gent	10
Tetracyline	Tet	10

Rifampicin	Rif	100
Spectinomycin	Spec	100
Kanamycin	Kan	50
Phosphinothricin	PPT	12

Bacteria and yeast

The *Escherichia coli* (*E. coli*) strain for cloning and amplification of plasmid DNA are:

DH5a: home-made component cells

DB3.1: home-made component cells for Gateway plasmid carrying *attB* gene.

GV3101: Ago-bacterium with helper plasmid *pSOUP* for amplifying *pGREEN* destination vector.

Transgenic Plants

Plant name	Description	Background
ft-10	T-DNAinsertion	Col-0
5.7kbFTp::FTcDNA	T-DNAinsertion	*ft-10*
4.0kbFTp::FTcDNA	T-DNAinsertion	*ft-10*
5.2kbFTp::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpS1mut::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpS2mut::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpS3mut::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpS4mut::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpCCAATbox1m::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpCCAATbox2m::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpCCAATbox8m::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTpCCAATbox1+8m::FTcDNA	T-DNAinsertion	*ft-10*

dMp::FTcDNA	T-DNAinsertion	*ft-10*
C+A::FTcDNA	T-DNAinsertion	*ft-10*
C+1kb::FTcDNA	T-DNAinsertion	*ft-10*
C+NOSmin::FTcDNA	T-DNAinsertion	*ft-10*
5.7kbFTp::GUS	T-DNAinsertion	Col-0
4.0kbFTp::GUS	T-DNAinsertion	Col-0
5.2kbFTp::GUS	T-DNAinsertion	Col-0
dMp::GUS	T-DNAinsertion	Col-0
C+A::GUS	T-DNAinsertion	Col-0
C+1kb::GUS	T-DNAinsertion	Col-0
C+NOSmin::GUS	T-DNAinsertion	Col-0
35S::CO	T-DNAinsertion	Col-0
co-sail	T-DNAinsertion	Col-0
ACT2::BirA	T-DNAinsertion	Col-0
35S::NTF/ACT2::BirA	T-DNAinsertion	Col-0
CmGAS1::NTF/ACT2::BirA	T-DNAinsertion	Col-0
SUC2::NTF/ACT2::BirA	T-DNAinsertion	Col-0
GL2::NTF/ACT2::BirA	T-DNAinsertion	Col-0

Oligonucleotides

FT5.7kb_fw_350	*GGGGACAAGTTTGTACAAAAAAGCAGGCT CATTTGCTGAACAAAAATCT*
FTproblockC_RE	*GGGGACCACTTTGTACAAGAAAGCTGGGT AAACGTTTGGAAATAGGAAGTATG*
TOPO_Kan_FW	*TGAATGAACTGCAGGACGAG*
TOPO_Kan_RE	*ATACTTTCTCGGCAGGAGCA*

3C_Pri1	*CGCCACTCAAGTTTTGGAGA*
3C_Pri2	*TCCCAAAAGGGTGATAAGA*
3C_Pri3	*GATAGCTTCATACTAAATAATGCGTTT*
3C_Pri4	*AAAGGATTGGATGAGTGCAAA*
3C_Pri5	*TACGGTTTTCTCAGGGCAAT*
3C_Pri6	*CAACGAGATTTGGGGTTAAGG*
3C_Pri7	*CATATTATTTGAGAAGTCGCAATTTT*
3C_Pri8	*CCTCATCCACTTGCCAATCT*
3C_Pri9	*CGGGGAAACCTTCTCAAAG*
3C_Pri10	*GTGGCGGACAATCCATCTAT*
3C_Pri11	*CATAATATGGCCGCTTGTTT*
3C_Pri12	*GGTGGAGAAGACCTCAGGAA*
3C_Prifix	*CGACCCGAGTTAATGCAAAT*
PCR3_FW	*CAAAAACGTGAGACGCAAAA*
PCR3_RE	*TCTGCAACTTAGATTCGCAAAA*
PCR5_FW	*CTGCGACTGCGACCTATTTT*
PCR5_RE	*GCCACTGTTCTACACGTCCA*
PCR6_FW	*TTCCTTTATTTTCCAGTTTGGACAG*
PCR6_RE	*TTGCACGACCAGGATAATTG*
PCR14_FW	*AACTTCGTCCATCGCAAAAA*
PCR14_RE	*GGGAAAATCAACGACCCTT*
AT1g13320Mid_FW	*GCAACCATATAACGCACACG*
AT1g13320Mid_RE	*GCTCTTGGGAAATTGTTGGA*
AT1g13320End_FW	*GGGAGAGCATACCATCTTGC*
AT1g13320End_RE	*CGGAGCCAACTAGGTAAGCA*
PCR3_FW	*CAAAAACGTGAGACGCAAAA*

PCR3_RE	*TCTGCAACTTAGATTCGCAAAA*
PCR5_FW	*CTGCGACTGCGACCTATTTT*
PCR5_RE	*GCCACTGTTCTACACGTCCA*
PCR6_FW	*TTCCTTTATTTTCCAGTTTGGACAG*
PCR6_RE	*TTGCACGACCAGGATAATTG*
PCR14_FW	*AACTTCGTCCATCGCAAAAA*
PCR14_RE	*GGGAAAATCAACGACCCTT*
BlockC_overlapA_RE1	*GTACCGCCAAAAACGTTTGGAAATAGGAAGTATG*
BlockA_overlapC_FW	*TTCCAAACGTTTTTGGCGGTACCCTACTTTTT*
BlockC_overlap1kbRE2	*CAAATACGCAAAACGTTTGGAAATAGGAAGTATG*
1kbp_overlpC_FW	*TTCCAAACGTTTTGCGTATTTGAGTTCGGACA*
1kbFTp-GW-RE	*GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTGATCTT GAACAAACAGGT*
FT_cDNA_RT_FW	*GGTGGAGAAGACCTCAGGAA*
FT_cDNA_RT_RE	*ACCCTGGTGCATACACTGTT*
5.7kb_blockCS1mut_ FW	*AATGCTCGAGGGCTTGTTGTTAAAGATAATGAGATC*
5.7kb_blockCS1mut_RE	*AAGCCCTCGAGCATTATTCTTGATAAACTTCA*
5.7kb_blockCS2mut_ FW	*CTCGCACGCATGTGATGATAGTGAAGTGAGAC*
5.7kb_blockCS2mut_RE	*TCACATGCGTGCGAGTCATACTCTCTCAATAATGTG*
5.7kb_blockCS3mut_FW	*GTCGCGTTCCTGTCGACATCTTGGCCAACATTAGA*
5.7kb_blockCS3mut_RE	*CGACAGGAACGCGACTCAACGTATCTATCTTCATAC*
5.7kb_blockCS4mut_FW	*TAACCACTGCTCCTCATCGGAACTAAAGGATTGGATGAG TGCA*
5.7kb_blockCS4mut_RE	*GTTCCGATGAGGAGCAGTGGTTACAAGATGTCTCACTTC ACTAT*
5.7kb_CCAATbox1m_FW	*CATAAAGGCGGTTATGAGTGCAAAGTATCGAGACG*

5.7kb_CCAATbox1m_RE	*GCACTCATACCGCCTTTATGGAATCTTCTTC*
5.7kb_CCAATbox2m_FW	*ACAAAACGGTTATTGACACATATCTCTTAT*
5.7kb_CCAATbox2m_RE	*TCAATAACCGTTTTGTCAATTGATATTTAGTTG*
5.7kb_CCAATbox8m_FW	*CGGACCGGTTTAGGTATGGACGATGAAAAT*
5.7kb_CCAATbox8m_RE	*ACCTAAACCGGTCCGAACTCAAATACGCAA*
pHISi_gain3.5kfw_NcoI	*CCCCCCATGGTATATGAGTAAACTTGGTCTGACAG*
pHIS_gain3.5kre_NotI	*CCCCGCGGCCGCTCACAGTTCTCCGCAAGAATTGATT*
5.7k_gain3.5kbfw_NcoI	*CCCCCCATGGATGTGTGTTAAGTCTTATAAAGTTA*
5,7k_gain3.5kbre_NotI	*CCCCGCGGCCGCATCTGCAATTAAATAATATAAACCA*
oligo_dT18	*TTTTTTTTTTTTTTTTTT*
T-DNA_LB_RE	*ACCGGCATGCAAGCTGATAA*
T-DNA_RB_FW	*GCTTCCTCGCTCACTGACTC*

Recipes

50× TAE buffer 1 L

242 g Tris base

57.1 mL glacial acetic acid

100 mL 500 mM EDTA (pH 8.0)

Add H2O up to the final volume 1 liter

10× PBS 1 L

80 g NaCl

2 g KCl

14.4 g Na2HPO4

2.4 g KH2PO4

Dissolve into 800 ml ddH2O, Adjust pH to 7.4

Add ddH2O up to 1000 ml and autoclave

10× TBS (concentrated Tris-Buffered Saline) 1 L

24 g Tris base

88 g NaCl

Dissolve in 900 ml distilled water, pH to 7.6 with HCl

Add distilled water to a final volume of 1 liter

1× TBS-T

50 mM Tris

150 mM NaCl

0.05% Tween 20

Adjust pH with HCl to pH 7.6

2× Laemmli Buffer 10ml

1.2 ml 1 M Tris (pH 6.8)

2.0 ml Glycerol

4.0 ml 10% SDS

0.2 ml 1% Bromophenol Blue

2.6 ml dH$_2$O

12% Gel 10 ml

3 ml 40% Acrylamide/Bis (29:1)

2.5 ml 1.5 M Tris (pH 8.8)

4.3 ml ddH2O

100 μl 10% SDS

100 µl 10% APS

10 µl TEMED

4% Stacking Gel 5 ml

0.5 ml 40% Acrylamide/Bis (29:1)

1.25 ml 0.5 M Tris (pH 6.8)

3.2 ml ddH2O

50 µl 10% SDS

50 µl 10% APS

5 µl TEMED

1× NPB

20 mM MOPS (pH 7)

40 mM NaCl

90 mM KCl

2 mM EDTA

0.5 mM EGTA

Sterilize NPB by filtering and stock the buffer at 4°C

Before use add

0.5 mM spermidine

0.2 mM spermine

1× complete protease inhibitors

NPBd and NPBt

NPB containing 2 µg/ml DAPI and

NPB containing 0.1% Triton X-100 respectively

4. Results

4.1 An evolutionary conserved distal enhancer is required for *FT* expression in leaves

4.1.1 *FT* driven by 5.2kb promoter fails to compliment late flowering of *ft* mutant

Regulation of *FT* transcription level is the result of convergence of several critical components such as CO, FLC and SVP (Figure 1). Because *FT* mRNA level is too low to be detected efficiently by RNA *in situ* hybridization, an 8.9kb *FT* promoter (*8.9kbFTp*) fused to GUS was cloned in a previous study to mimic *FT* expression (Takada and Goto, 2003).
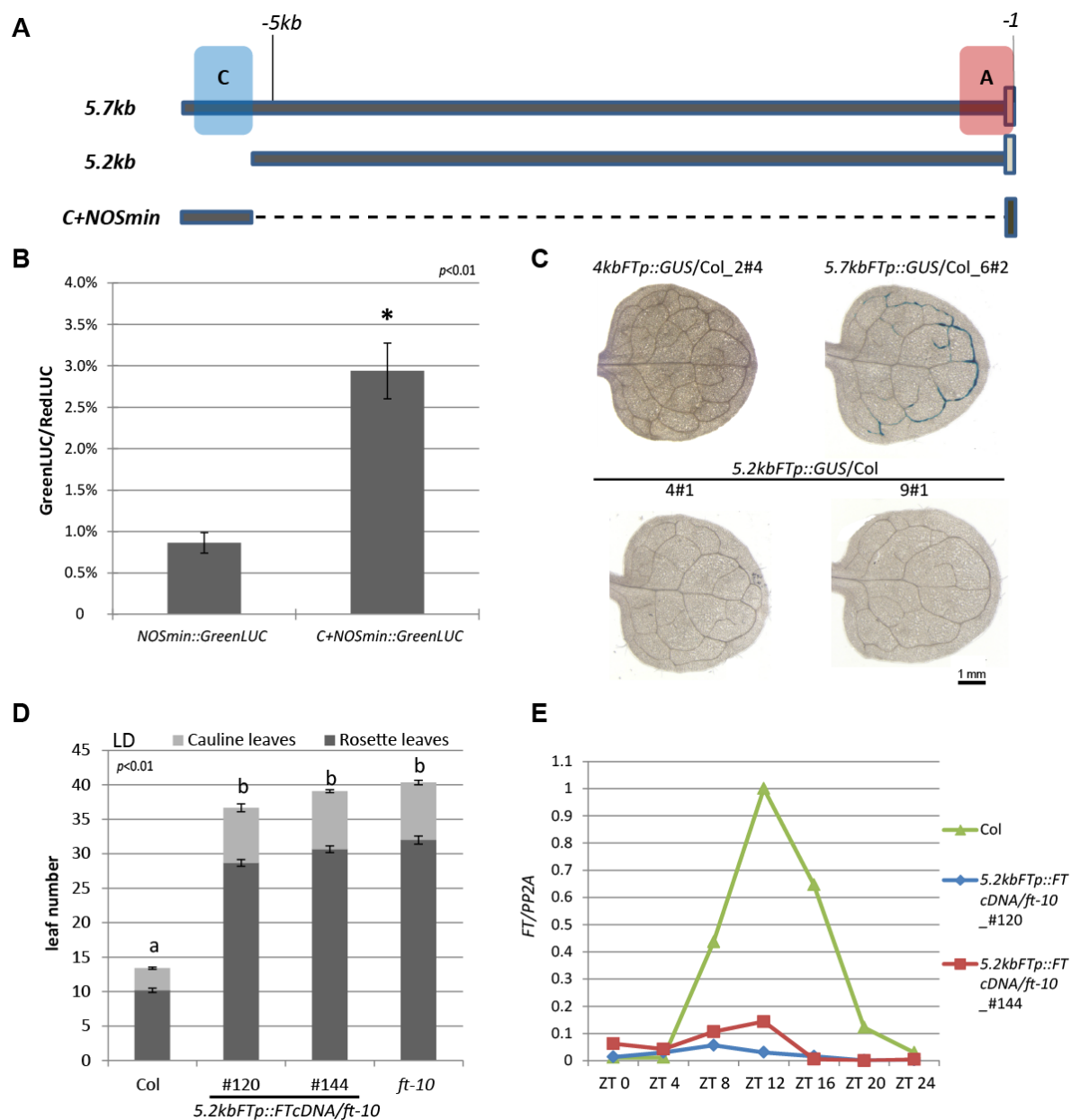

Figure 3. *Block C* as a distal enhancer is necessary for *FT* expression.

(**A**). Schematic representation of *FT* promoter with full length 5.7 kb, 5.2 kb missing *Block C* and *C+NOSmin* (*Block C* fused to *NOS* minimal promoter). The position of nucleotide A of the translation start codon was termed -1. The 5.7 kb upstream promoter sequence was marked -5kb.

(**B**). *Block C* fused to *NOSmin* drives reporter gene *GreenLUC* expression in transient leaf bombardment assays. *35S::RedLUC* co-bombarded with *promoter::GreenLUC* was used as the internal control. Values represent mean relative signal ± SE. The asterisk (*) indicates statistically significant differences relative to the control *NOSmin::GreenLUC*. Statistical significance was determined using the Student's t-test ($p<0.01$).

(**C**). Histochemical localization of GUS activity in the first true leaves of *5.7kbFTp::GUS*, *4.0kbFTp::GUS* and two independent *5.2kbFTp::GUS* transgenic lines. Transgenic plants in Col-0 background were collected at ZT 16 after growth for 10 LDs on GM media in a climate chamber. Bar represents 1 mm for all samples.

(**D**). In LD conditions, flowering time of *ft-10* plants carrying stable T-DNA insertions of a *5.2kbFTp::FTcDNA*. Two independent transgenic lines were shown as line #120 and #144. Col-0 and *ft-10* were tested as controls. Number of rosette and cauline leaves of a representative experiment are shown as mean ± SE. Statistical significance was determined using the Student's t-test ($p<0.01$). Significant differences are indicated by different letters above the bars.

(**E**). Quantitative PCR of *FT* transcripts in *ft-10* seedlings carrying the *5.2kbFTp::FTcDNA* construct. Seedlings were harvested from ZT 0 to ZT 24 on day 10 in LD conditions, and plants were grown on GM media in the climate chamber. Values represent relative fold of *FT* to *PP2A* mRNA levels.

Recently, based on promoter truncation analysis, it was shown that a 5.7 kb *FT* promoter (*5.7kbFTp*) named as the full length promoter of *FT* contains a sufficient amount of *cis*-regulatory elements to initiate *FT* expression in LD. In addition, the study indicated that the region located 4.0-5.7 kb upstream of the translation start codon (TSC) could harbor critical regulatory sequences. A phylogenetic shadowing approach identified a conserved region 5.2-5.6 kb upstream of the *FT* promoter, which was designated as *Block C* (Adrian et al., 2010).

To test the possibility that *Block C* may act as a distal enhancer for *FT* expression, *Block C* was cloned and fused to the minimal promoter of the *Nopaline Synthase* (*NOS*) gene (*NOSmin*) (Figure 3A). *NOSmin* has a length of 105nt (-101 to +4, TSS as -1) and includes a *TATA* box and a *CAAT* box. *NOSmin* has no capability of activating reporter gene expression by itself, but can be used to monitor the effect of positive *cis*-regulatory elements on transcription (Puente et al., 1996). In transient bombardment assays, *NOSmin* fused to the reporter gene *GreenLUC* showed very low background levels of expression compared to a co-bombarded RedLUC reference expressed under the control of the *CaMV 35S* promoter (<1%). When *Block C* was fused upstream of *NOSmin*, the relative signal reached up to 3% (Figure 3B).The results suggest that *Block C* has an enhancing effect on the expression of the reporter gene at least in transient assays.

To answer the question whether *Block C* also functions as an enhancer in stably transformed plants, a 5.2 kb *FT* promoter (*5.2kbFTp*) which lacks of *Block C* was fused to a *GUS* reporter gene (*5.2kbFTp::GUS*) or an *FT* cDNA (*5.2kbFTp::FTcDNA*) and transferred into Col-0 and

ft-10 mutant plants respectively. Histochemical GUS assays showed that no signal was observed in the first true leaf of two independent transgenic lines of *5.2kbFTp::GUS*/Col. As shown in a previous study, a *4.0kbFTp::GUS*/Col line showed no signal, but *5.7kbFTp::GUS*/Col plants showed GUS signal in the minor veins of the distal part of leaf (Figure 3C) (Adrian et al., 2010). In the complementation assay, *ft-10* plants grown in LD did not significantly change flowering time when transformed with the *5.2kbFTp::FTcDNA* construct (Figure 3D). To check whether *FT* mRNA levels match the phenotype of *5.2kbFTp::FTcDNA/ft-10* transgenic plants, *FT* mRNA was measured from two independent lines every four hours from ZT 0 to ZT 24 on day 10 after germination. At ZT 12 and ZT16, when FT expression was highest in Col-0 controls, *FT* mRNA levels were not detectable in line #120 and #144, which fits well with their flowering time similar to that of *ft-10* plants (Figure 3E).

Taken together, in LD, the distal sequences of *Block C* are necessary to complement late flowering phenotype of *ft-10* mutant plants. *Block C* not only promotes reporter gene expression in transient assays, but also functions as a distal enhancer required for *FT* expression in response to long days *in planta*.

4.1.2 Conserved sequences in *Block C* are essential for promoter function

Block C is around 380bp long and largely conserved across the genera Arabidopsis, Arabis, Capsella and Brassica in the Brassicaceae family (personal communication with George Coupland for the *FT* sequence in *A. Alpina*) Until now, no functional *cis*-elements have been characterized in this conserved region and none of the known *trans*-factors of *FT* regulation have been reported to associate with sequences in *Block C*. Several putative motives that have been characterized as transcription factor binding-sites in different context are present in *Block C*. (Adrian et al., 2010). Phylogenetic analysis identified several hyper-conserved sequences called sub-blocks. To access the *cis* regulatory function of the candidate elements in *Block C*, mutagenesis of sub-blocks in the context of the full length promoter was carried out.

Among the conserved sub-blocks, four varying between 8-23 bp in length were chosen. Mutagenesis exchanged all G with T and A with C residues and *vice versa*. Each sub-block was independently mutagenized in the background of the *5.7kbFTp* (Figure 4A). Next,

5.7kbFTpS1-4mut::FTcDNA constructs were transferred into *ft-10* plants to test the complementation.

S1mut : 8 bp AGATCTTT → ctcgaggg
S2mut : 12 bp AGATAGATACGT → ctcgctcgcatg
S3mut : 14 bp TGATAGTGAAGTGA → gtcgcgttcctgtc
S4mut : 23 bp GCCAACATTAGAAGAAGATTCCA → taaccactgctcctcatcggaac

Figure 4. Conserved sub-blocks in *Block C* are required for *FT* expression.
(**A**). Sequence alignment of *Block C*. Schematic representation of *FT* promoter shows position of *Blocks C* and *A*. The sub-blocks *S1*, *S2*, *S3* and *S4* are marked in the alignment of related sequences from five Brassicaceae species. Two accessions of Col-0 and L*er* are chosen from *A. thaliana*. Their original and mutagenized sequences are shown in capital and minor letters respectively.

(**B**). Flowering time of *ft-10* plants carrying constructs to drive *FTcDNA* by *5.7kbFTp* containing mutagenized sub-blocks *S1mut*, *S2mut*, *S3mut* or *S4mut*. Four independent transgenic lines are shown for each mutagenized construct Flowering time was measured in T2 plants in this experiment, and the presence of the transgene was confirmed by PCR. Col-0, *ft-10* and two independent *5.7kbpFTp::FTcDNA/ft-10* lines were tested as controls. Number of rosette and cauline leaves are shown as the mean ± SE. Statistical significance was determined using the Student's t-test (*p*<0.01). Significant differences are indicated by different letters above the bars.

Unexpectedly, all four sub-blocks had strong effects on the full length promoter function. In detail, sub-block *S3* seemed to have the strongest effect, since three of four independent transgenic *5.7kbFTpS3mut::FTcDNA/ft-10* lines flowered as late as *ft-10*. According to their complementation efficiency, mutations in *S1* and *S4* also dramatically influenced full length promoter function. Four independent *5.7kbFTpS2mut::FTcDNA/ft-10* lines flowered intermediately between Col-0 and *ft-10* controls, indicating that the effect of *S2* was not as strong as that of the other three sub-blocks (Figure 4B).

Taken together, conserved sub-blocks in *Block C* play important roles in controlling the distal enhancer function to drive *FT* expression under LD conditions.

4.2 Impact of *CCAAT* boxes on *FT* expression

Block C contains motives such as an *I-box*, a *REalpha* element and a *CCAAT* box but their role in transcription regulation has not yet been characterized (Adrian et al., 2010). The *CCAAT* box is conserved and located close to sub-block *S4* in *Block C* (Figure 4A and 5A). *CCAAT* boxes are confirmed binding sites for NF-Y complexes in yeast and animals (Mantovani, 1999). In Arabidopsis, trimeric NF-Y complexes were shown to bind to *CCAAT* boxes in *in vitro* EMSA assays (Calvenzani et al., 2012). Binding of NF-Y complex to *CCAAT* boxes *in planta* has not yet been demonstrated in plants.

Eight *CCAAT* boxes were found within *5.7kbFTp* and numbered sequentially according to their positions (Figure 5A). Among these *CCAAT* boxes, box 1 is 5295 bp and box 8 is 868 bp upstream of the TSS. Box 1 localizes at *Block C* and box 8 is close to the proximal promoter, but outside of *Block A*. Box 2-7 distribute in the 1-4 kb region from the TSS and the closest distance between two boxes is 191 bp between box 6 and box 7 (Figure 5A).

To access the roles of *CCAAT* box1 in *Block C* as well as other boxes playing on full length promoter, mutagenesis of *CCAAT* boxes 1, 2 and 8 in the background of *5.7kbFTp* was

carried out. Box2 is positioned 3855 bp upstream of the TSS. In this study, *CCAAT* is mutagenized to *AACCG*, and *ATTGG* is mutagenized to *CGGTT*.
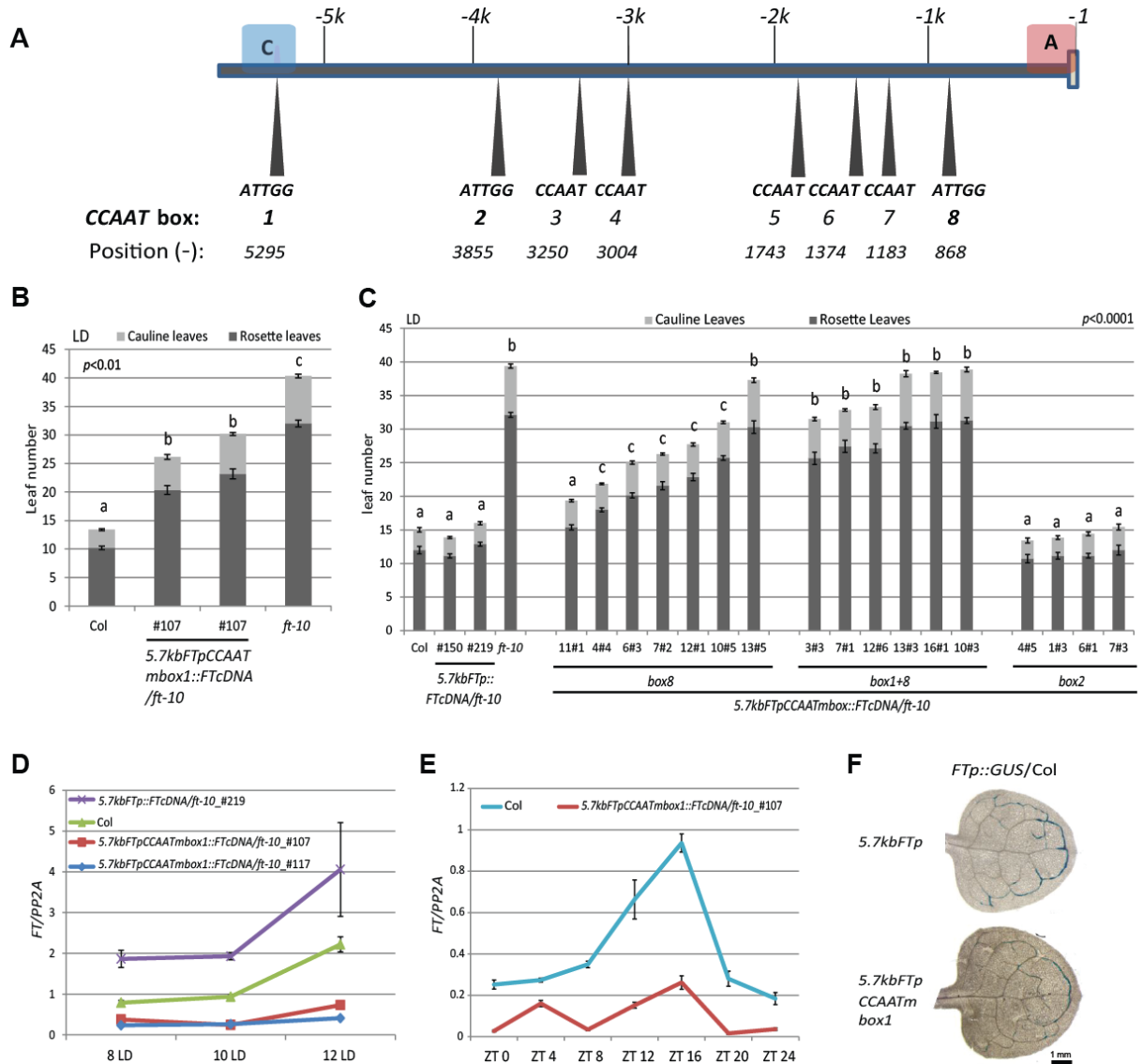

Figure 5. Mutagenesis of *CCAAT* boxes at the full length *FT* promoter.

(**A**). Schematic representation of the position of *CCAAT* boxes on the *FT* promoter. Eight *CCAAT* boxes were identified on *5.7kbFTp*. They were named as *CCAATbox1-8* from the most upstream to the most proximal to the TSS (left to right).

(**B**) and (**C**). Flowering time of *ft-10* plants expressing *5.7kbFTp::FTcDNA* constructs with mutagenesis of *CCAATmbox1*, *CCAATmbox2*, *CCAATmbox8*, or *CCAATmbox1+8*. Col-0, *ft-10* and *5.7kbFTp::FTcDNA/ft-10* were tested as control. Number of rosette and cauline leaves of a representative experiment are shown as the mean ± SE. Statistical significance was determined using the Student's t-test (*p*<0.01 in B and *p*<0.0001 in C). Significant differences are indicated by different letters above the bars. B and C are two independent experiments in the greenhouse.

(**D**). qRT-PCR analysis of *FT* expression in two independent lines of *ft-10* seedlings carrying *5.7kbFTpCCAATmbox1::FTcDNA* constructs (#107 and #117). Seedlings were harvested at ZT 16 on day 8, 10 and 12 after germination in LD, and plants were grown on GM media in the climate chamber. Values represent relative quantitative mean ± SD of *FT* mRNA normalized by *PP2A* mRNA.

(**E**). qRT-PCR analysis of *FT* expression in *ft-10* seedlings carrying *5.7kbFTpCCAATmbox1::FTcDNA* construct (line #107) and Col-0. Seedlings were collected from ZT 0 to ZT 24 on day 10 of growth in LD. Values represent relative quantitative mean ± SD of *FT* mRNA normalized by *PP2A* mRNA.

(**F**). Histochemical localization of GUS activity in the first true leaves of *5.7kbFTpCCAATmbox1::GUS* and *5.7kbFTpro::GUS* transgenic plants in Col-0 background. Material was collected at ZT16 from plants grown 10LD on GM plates in a climate chamber. Bar represents 1 mm for both leaves.

Four independent transgenic lines of *5.7kbFTpCCAATmbox2::FTcDNA/ft-10* plants flowered as Col-0 and *5.7kbFTp::FTcDNA/ft-10* controls indicating that point mutation of *CCAATbox2* did not compromise full length promoter function (Figure 5C). However, when *CCAATbox1* was mutated, flowering time of two independent lines of #107 and #117, showed an intermediate flowering time phenotype between Col-0 and *ft-10*. A similar intermediate flowering time phenotype could also be observed by mutating *CCAATbox8* in *5.7kbFTp* (Figure 5B and 5C). The effect of mutating *CCAATbox8* and *1* were additive since six independent transgenic lines expressing *5.7kbFTpCCAATmbox1+8::FTcDNA/ft-10* showed no difference in flowering time compared to *ft-10* (Student's T-test; *p<0.0001*) (Figure 5C).

At ZT16 on day 8, 10 and 12, the relative amount of *FT* mRNA was much lower in line #107 and #117 of *5.7kbFTpCCAATmbox1::FTcDNA/ft-10* compared to Col-0 control plants (Figure 5D). *FT* mRNA level matched well with the flowering time phenotype of Col-0 and two independent transgenic lines of *5.7kbFTpCCAATmbox1::FTcDNA/ft-10* (Figure 5B and 5D). *FT* mRNA levels from *ft-10* carrying *FT* cDNA driven by *5.7kbFTpCCAATmbox8* or *5.7kbFTpCCAATmbox1+8* promoters will be analyzed in the future.

Time course experiments to compare the diurnal expression pattern showed that *FT* mRNA was obviously down regulated in *5.7kbFTpCCAATmbox1::FTcDNA/ft-10* line #107 compared to Col-0 at any time point from ZT 0-ZT 24 on day 10 of growth in LD (Figure 5E). The results confirmed that mutagenesis of *CCAATbox1* efficiently impaired the full length *FT* promoter function. Histochemical GUS assays performed with *5.7kbFTpCCAATmbox1::GUS/*Col plants further supported this conclusion (Figure 5F).

In sum, several *CCAAT* boxes seem to play critical roles in determining *FT* expression, which was shown for the nearest and farthest box at the full length promoter.

4.3 Distance variation between *Block C* and *A*

4.3.1 Natural variation of distance between *Block C* and *A* in Arabidopsis

Variation in distance between *Block C* and *A* between Col-0 and L*er* has been reported previously, but without an analysis of functional implication (Turck et al., 2007). Considering several studies that mapped flowering time QTL to *FT* promoter in Arabidopsis (Schwartz et al., 2009; Strange et al., 2011), it could be interesting to test the natural variation of distance between *C* and *A* among Arabidopsis accessions and within the Brassicaceae.

To access the distance variation between conserved *Block C* and *A* in Brassicaceae family, phylogenetic analysis was carried out. In *A. thaliana*, the distance was 5 kb in Col-0, 7 kb in *A. lyrata* and in *Arabis alpina* it reached up to 11 kb (Figure 6A).The results demonstrated that there is variation in the distance between these two conserved blocks across species. The distance variation between *Block C* to *A* was detected based on the database from "1001 Genomes Project" (http://www.1001genomes.org/projects/MPICao2010) in Arabidopsis.

Interestingly, among these 80 Arabidopsis accessions, 57 are annotated to contain large deletion regions localized 3 kb upstream of *FT* promoter (Figure 6B). This 1 kb deletion region comprises two DNA fragments, one sharing high similarity to sequence on chromosome 5 (data not shown). The major deletions from these 57 accessions correspond to the L*er* accession, so these 57 accessions are named as short promoter "L*er*-like" accessions. Thus the other 23 accessions likely belong to long promoter "Col-like" accessions. Comparing Col-like and L*er*-like groups, the most likely explanation is that the Col-type originated by one or several insertion events occurring 3 kb upstream of the TSS in L*er*-like accessions.

80 Arabidopsis accessions are sampled from eight different regions in Europe and Asia (Cao et al., 2011) (Figure 7). I analyzed whether *FT* promoter length correlates with the geographic distribution. A border line, which separates eight regions into "West" and "East" group, could be defined as a line passing two cities of Berlin and Athens (Table 1 and Figure 7).

In the "West" group, each region comprises an equal number of L*er*- or Col-like accessions. However, in the "East" group, 34 accessions belong to L*er*-like type and only one is a Col-like accession (Table 1). This suggests that *FT* promoter lengths may act as a molecular marker that correlates with the geographic distribution of Arabidopsis accessions. This observation could be the result of the immigration among Arabidopsis accessions.

Single Nucleotide Polymorphisms (SNPs) and structural variants have been annotated for the 80 accessions (Cao et al., 2011). Notably, the diverged regions appear at *Block C*, *B* and *A* with comparably lower frequencies than that at the neighboring regions (Figure 6B). Very

low SNP frequencies were observed in *Block A*. The SNP frequencies in *Block C*, especially the half part closer to the TSS (5.2-5.5 kb) was low. SNP frequencies in *Block B* were relatively higher than in *Block C* and *A* (Figure 6C). These observations revealed that *Block C* and *A* are conserved in 80 Arabidopsis accessions.
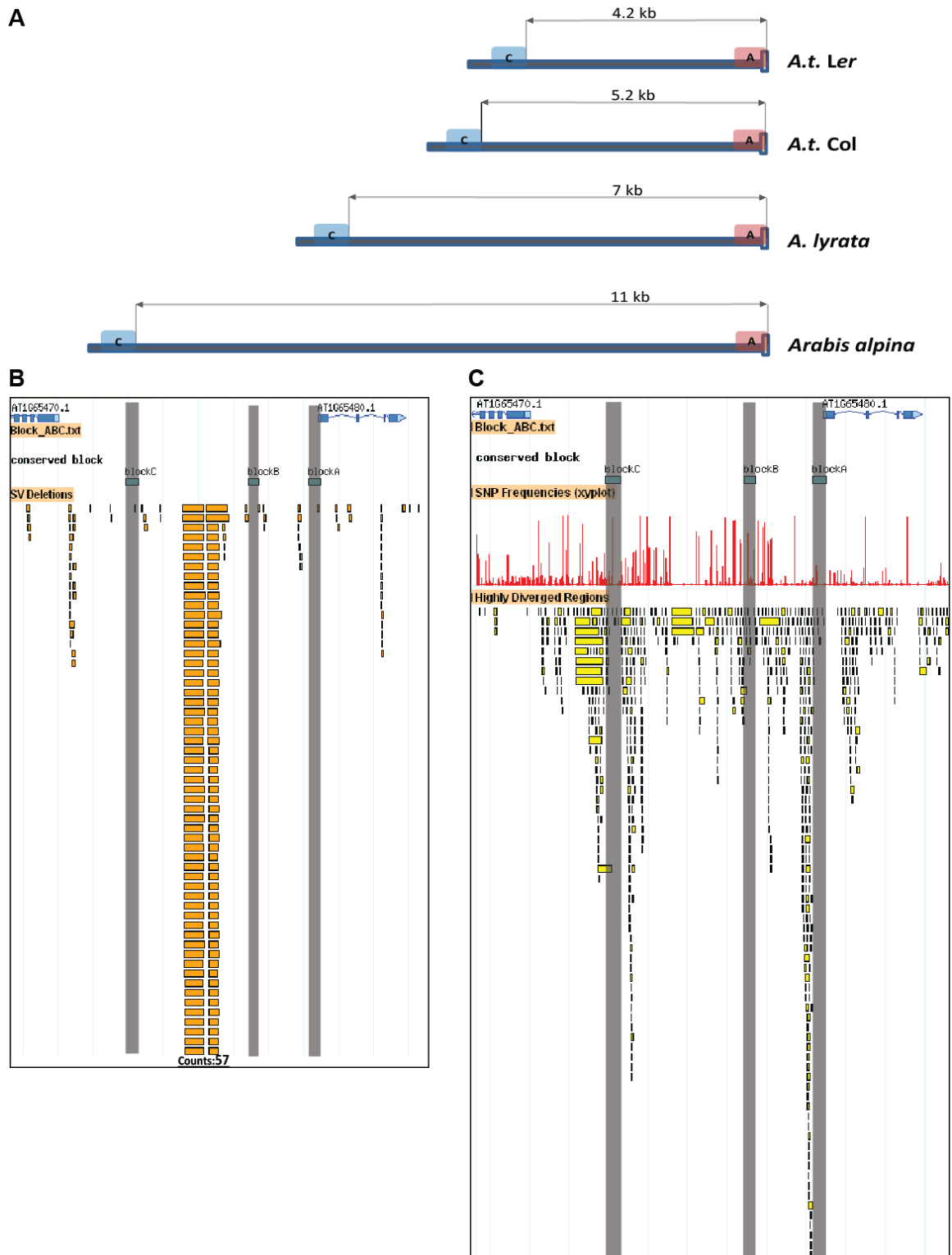

Figure 6. Distance variation between *Block C* and *A*.

(**A**). Schematic representation of different *FT* promoters in the species of *A. thaliana* (L*er* and Col-0), *A. lyrata* and *Arabis alpina*. The distances between conserved *Block C* to *Block A* are shown.

(**B**). Analysis of the deletion regions at *FT* promoter among 80 *A. thaliana* accessions ("1001 Genomes Project").

(**C**). Analysis of the SNPs frequencies and highly diverged regions at *FT* promoter among 80 *A. thaliana* accessions ("1001 Genomes Project").

Table 1. Geographic distribution of L*er*- and Col- type promoters in 80 *A. thaliana* accessions.

Region no.	Group	Region name	Accession no.	Short promoter L*er*-like Accession no.	Long promoter Col-like Accession no.
1	West	Spain and North Africa	12	7	5
2	West	Swabia, Southwest of Germany	10	5	5
3	West	South Tyrol, North of Italy	10	5	5
4	West	Southern Italy	13	6	7
5	East	Eastern Europe	10	10	0
6	East	Caucasus	10	10	0
7	East	Southern Russia	7	6	1
8	East	Central Asia.	8	8	0

Region name and the numbers of accessions in each region are according to the "1001 Genomes Project". The "West" and "East" border is defined as the line passing two cities of Berlin and Athens.
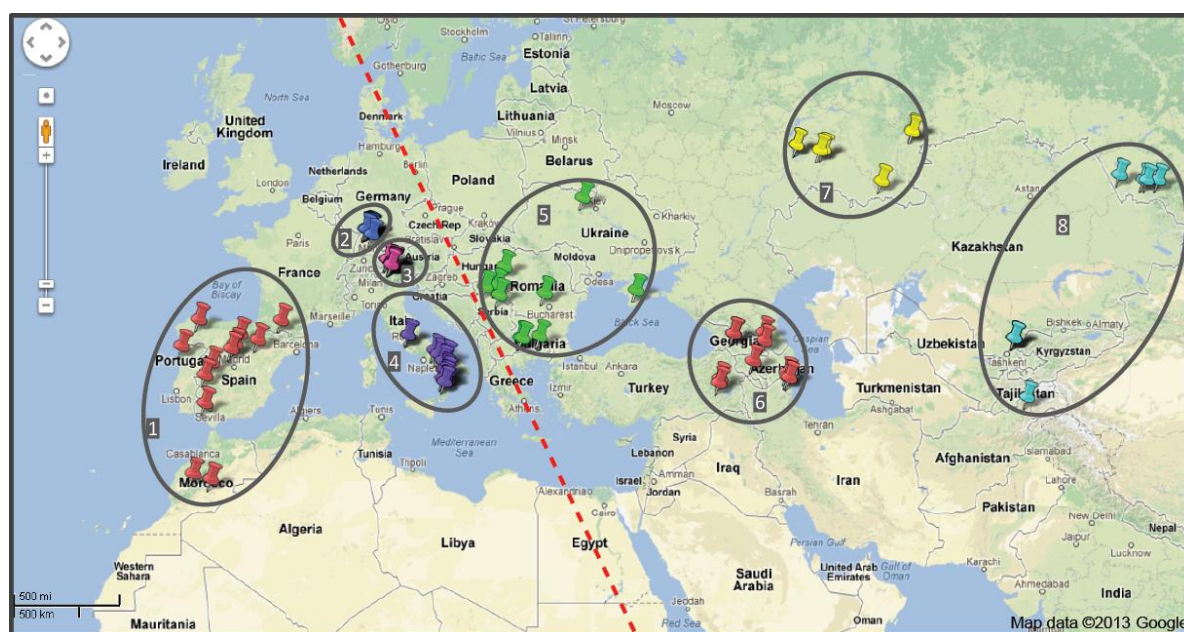

Figure 7. Geographic distribution of 80 *A. thaliana* accessions.

The distribution map of 80 *A. thaliana* accessions was created on by Jun Cao from MPI for Developmental Biology on Nov 8, 2010. The number of regions from 1-8 were described in table 1. The "West" and "East" border is defined as the red line passing two cities of Berlin and Athens.

4.3.2 Effect of changes in distance between *Block C* and *A* at *FT* promoter

To access the biological meaning of distance variation at the *FT* promoter, all necessary and sufficient regulatory sequences on *5.7kbFTp* need to be mapped. A previous study has shown that T-DNA insertions in different positions between *Block C* and *A* slightly delay flowering time compared to WT (PhD thesis of Jessica Adrian). This indicates that *Block C* still functions as an enhancer even if it is located about 9 kb away from the proximal promoter.

In order to define the minimal distance between *Block C* and *A* and observe the effect of shortening the distance, a promoter which missed a 3.5 kb middle part (*dMp*) was fused to *FT* cDNA and *GUS* and transformed to *ft-10* and Col-0 plants respectively. The 2.2 kb long *dMp* was found to mimic the full length *5.7kbFTp* function. Even though one line of *dMp::FTcDNA/ft-10* showed a slightly later flowering time than Col-0, the average flowering time of five independent transgenic lines was similar to Col-0 in LD (Figure 8B). Further, constructs directly fused *Block C* to *A* (*C+A*) to drive *GUS* and *FT* cDNA expression in transgenic plants. *FT* cDNA driven by 1.1 kb *C+A* also effectively accelerated *ft-10* mutants to flower as early as Col-0 in LD (Figure 8B). Conversely, the flowering time of *ft-10* mutants carrying the construct of *C+NOSmin::FTcDNA* was not accelerated to Col-0 levels (Figure 8B).

In *dMp::GUS*/Col transgenic lines, signal appeared in the minor and major veins of distal half of the leaf, which enlarged the expression domain of *5.7kbFTp::GUS*/Col exhibiting signals in the minor veins (Figure 3C and 8C). *C+A::GUS* signal was observed in the minor and major veins of the whole leaf (Figure 8C). This suggests that the middle region between *C* and *A* could contain sequences involved in repressing *FT*. The middle region has been shown to associate with repressive *trans*-factors such as LHP1 and SVP (Farrona et al., 2011; Li et al., 2008).

In addition, GUS signal in true leaves of *C+NOSmin::GUS*/Col was difficult to detected, and very weak signals only were obtained in the major vein of middle leaf (Figure 8C). In the cotyledon of that line, like other promoter GUS lines such as *dMp*, *C+A* and *5.7kbFTp*, GUS signal was obviously detected in the vascular tissues (Figure 3C and 8C). In addition, The *NOSmin* alone failed to drive detectable *GUS* expression in cotyledons (personal communication with Samson Simon in MPIPZ).
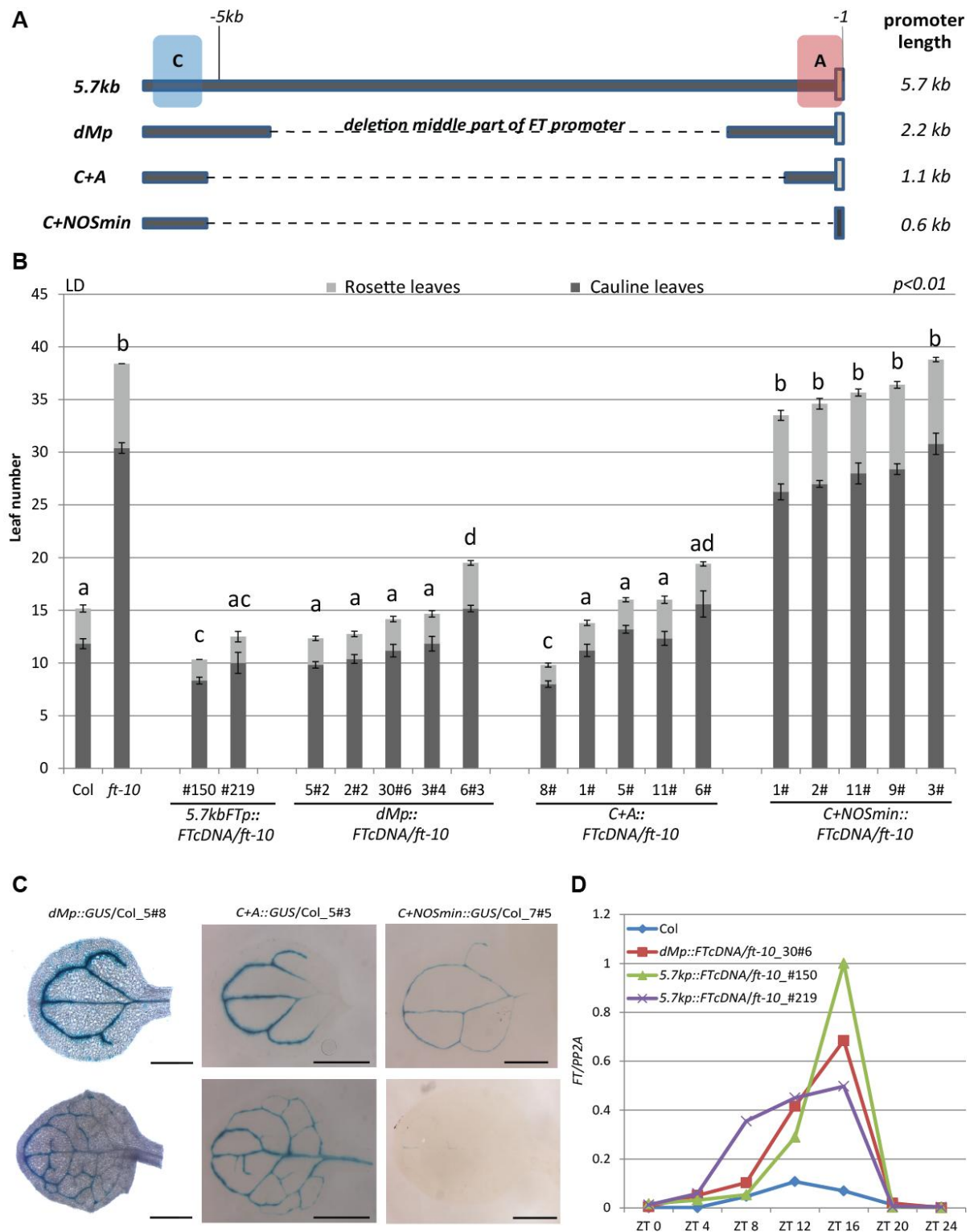
Figure 8. Shortening the distance between *Block C* and *A*.

(**A**). Schematic representation of different *FT* promoters with full length *5.7kbFTp*, *C+NOSmin*, *dMp* deleting 3.5kb middle region of full length promoter and *C+A* keeping the *Block C* and *A*. The length of each promoter is shown on the right.

(**B**). Flowering time of *ft-10* plants carrying the transgenic constructs *FT* cDNA driven by *dMp*, *C+A* or *C+NOSmin* in LD. Five independent transgenic lines are shown for each construct. Col-0, *ft-10* and *5.7kbFTp::FTcDNA/ft-10* were tested as controls. Number of rosette and cauline leaves of a representative experiment are shown as the mean ± SE. Statistical significance was determined using the Student's t-test (*p*<0.01). Significant differences are indicated by different letters above the bars.

(**C**). GUS staining of *C+A::GUS, dMp::GUS* and *C+NOSmin::GUS* in the Col-0 background. The cotyledons and true leaves are shown. Samples were collected at ZT 16 on day 12 in LD. Bar represents 2mm for all samples.

(**D**). qRT-PCR of *FT* expression in *ft-10* seedlings carrying *5.7kbFTp::FTcDNA* and *dMp::FTcDNA* construct and Col-0. Seedlings were collected from ZT 0 to ZT 24 on day 10 in LD. Values represent relative quantitative mean of *FT* mRNA normalized by *PP2A mRNA*.

In complementation assays, *dMp* mimics *5.7kbFTp* function in LD. Furthermore, the rhythm expression pattern of *FT* was similarly if these two promoters in their respective transgenic plants were compared (Figure 8D).Taken together, in LD, *Block C* and *A* are not only required, but also sufficient for *FT* expression.

4.3.3 Altered versions of the *FT* promoter in response to environmental cues

The question of a function of the 3.5 kb middle region between *Block C* and *A* in response to the environmental cues was addressed. When removing the middle region, the flowering time of *ft-10* plants carrying *FT* cDNA driven by different *FT* promoters were measured to test the possibility that *FT* may express differentially under non-inductive conditions or in extreme temperature conditions.

The average flowering time of independent lines containing promoter *C+A* or *dMp* compared to *5.7kbFTp* was slightly earlier in SD conditions than in LD conditions. Although the middle part of the promoter contains putative repressing domains, *C+A* or *dMp* were presumably not sufficient to drive *FT* strong expression in SD (Figure 8B and 9A).

Under LD conditions combined with high temperature of 30°C, flowering time of four lines of *dMp* were similar to that of *5.7kbFTp*, whereas two lines of *C+A* flowered earlier compared to *5.7kbFTp* lines (Figure 9B). Under the same LD conditions, even though flowering accelerated under higher temperature, the flowering time tendency of transgenic plants with different *FT* promoters was similar at 30°C and 22°C (Figure 8B and 9B).

Under MD conditions combined with low temperatures of 16°C, similar as in SD conditions, transgenic *dMp* lines flowered earlier with on average 5 leaves less than the *5.7kbFTp* lines (Figure 9A and 9C). Taken together, these observations suggest that *FT* driven under short promoters could make the difference of flowering time more visible when the promoting factors of *FT* are limited.
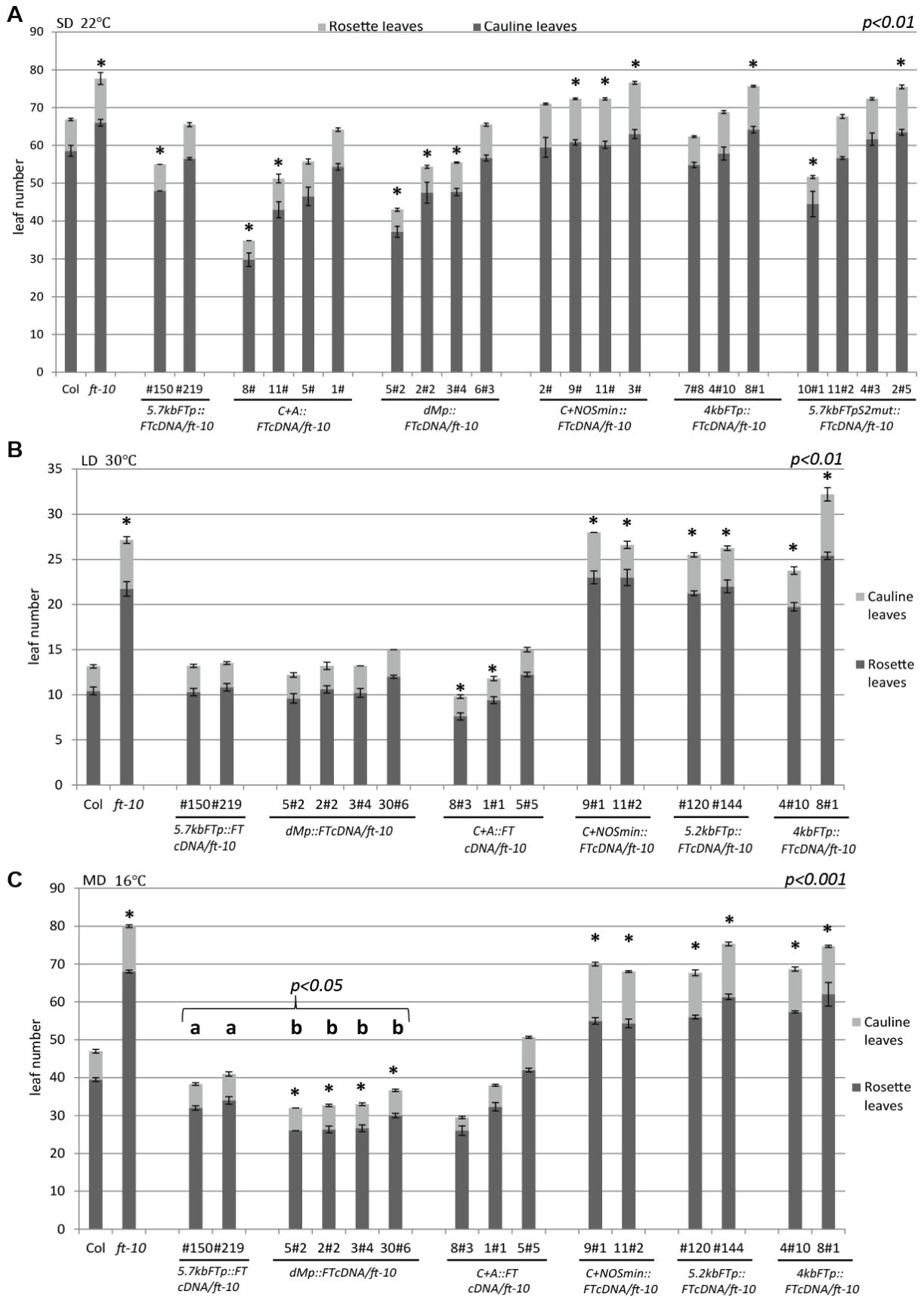
Figure 9. Different length of *FT* promoters in response to day length and temperature.

(**A**). Under SD conditions at 22°C in a climate chamber, flowering time of *ft-10* plants carrying transgenic constructs driving *FT* cDNA by promoters *C+NOSmin*, *C+A*, *dMp*, *4kbFTp* or *5.7kbFTpS2m* were measured. At least three independent transgenic lines are shown for each construct. Col-0, *ft-10* and two independent lines of *5.7kbFTp::FTcDNA/ft-10* were tested as controls. Number of rosette and cauline leaves of a representative experiment are shown as the mean ± SE. Statistical significance was determined using the Student's t-test (*p*<0.01) between any line to Col-0. The asterisk (*****) indicates significant difference.

(**B**). Under LD conditions in a 30°C climate chamber, flowering time determined by different versions of the *FT* promoter (*5.7kbFTp*, *dMp*, *C+A*, *C+NOS*min, *5.2kbFTp* and *4kbFTp*) driving *FTcDNA* in *ft-10* mutant background was determined. Number of rosette and cauline leaves of a representative experiment are shown as the mean ± SE. Statistical significance was determined using the Student's t-test (*p*<0.01) between any line to Col-0. The asterisk (*****) indicates significant difference.

(**C**). Under 12h light /12h darkness day (MD) conditions in a 16°C climate chamber, same transgenic lines as in (B) were tested. Number of rosette and cauline leaves of a representative experiment are shown as the mean ± SE. Statistical significance was determined using the Student's t-test (*p*<0.01) between any line to Col-0. The asterisk (*****) indicates significant difference. Additionally, another statistical significance was determined using the Student's t-test (*p*<0.05) between four lines of *dMp::FTcDNA/ft-10* to line #150 of *5.7kbFTp::FTcDNA/ft-10*. Significant differences are indicated by different letters above the bars.

4.4 *GUS* reporter driven by 4 kb *FT* promoter shows expression in siliques that is independent of photoperiod

To assess the possible function of the middle region of *FT* promoter in different tissues and organs, *GUS* reporter expression patterns driven by *1kbFTp*, *4kbFTp*, *5.7kbFTp* or *8.1kbFTp* were tested in cauline leaves, inflorescence, young siliques and old siliques (Figure 10A-10X). Similar to the GUS pattern in true rosette leaves, *5.7kbFTp* or *8.1kbFTp* could also drive *GUS* in the veins of cauline leaves (Figure 10M and 10S). No signal was observed in the transgenic plants with *4kbFTp* or *1kbFTp* in cauline leaves (Figure 10A and 10G). However, *4kbFTp* could drive *GUS* in the young and old siliques in LD (Figure 10J and 10I) as well as in SD (Figure 10L and 10K). The signals were restricted to the veins in siliques. Similar GUS signals were also observed in *5.7kbFTp* and *8.1kbFTp*, but not *1kbFTp* lines (Figure 10).

In sum, *Block C* is not required for *FT* expression in the veins of siliques and this expression is not dependent on photoperiod.
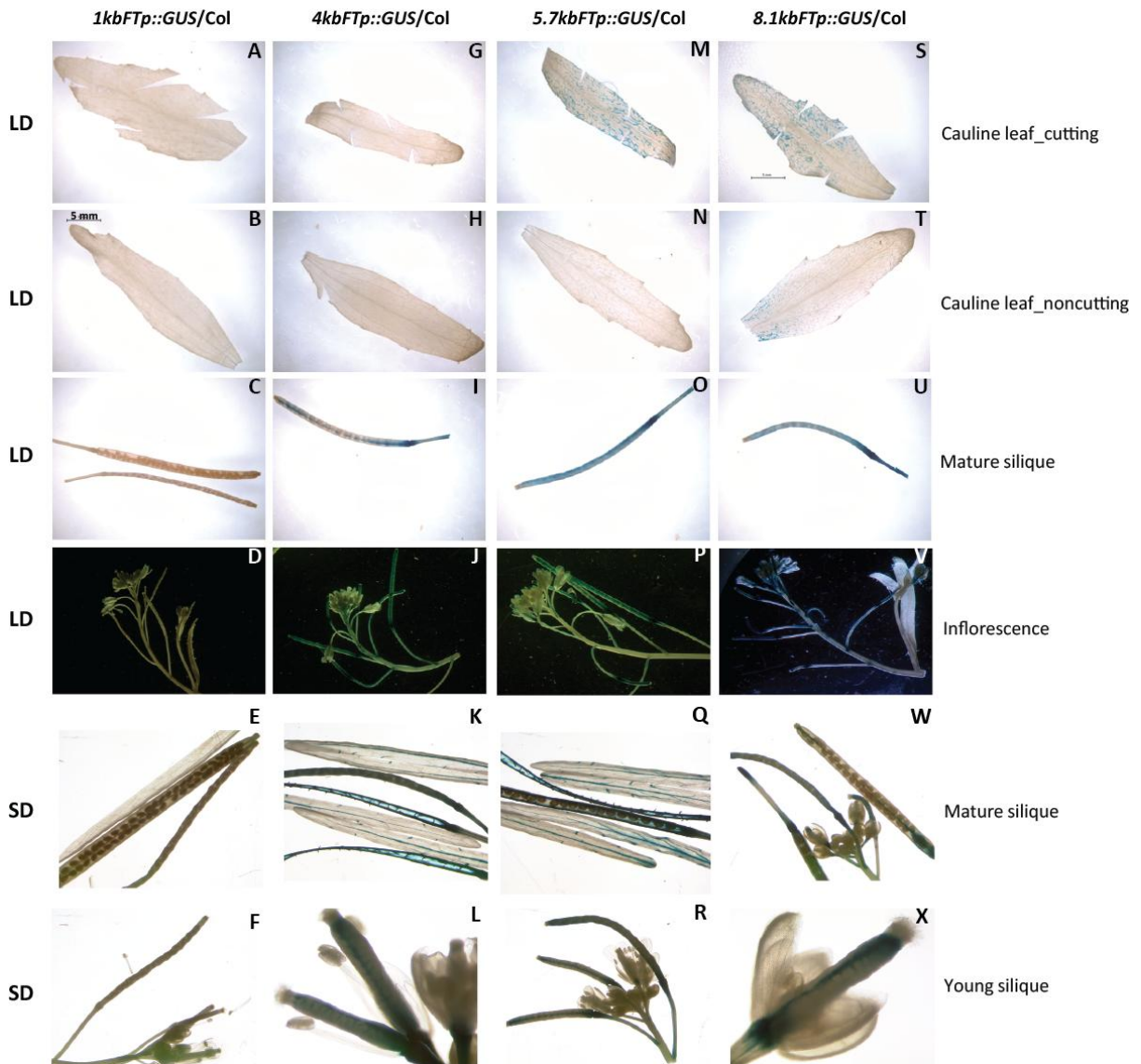
Figure 10. *GUS* **driven by different** *FT* **promoters in cauline leaves and siliques.**
(**A**)-(**X**) Histochemical GUS assay of (**A**)-(**F**) *1kbFTp::GUS,* (**G**)-(**L**) *4kbFTp::GUS*/Col, (**M**)-(**R**)
5.7kbFTp::GUS/Col and (**S**)-(**X**) *8.1kbFTp::GUS*/Col plants in Col-0 background harvested after
bolting. Samples were cauline leaves with cutting margins (**A,G,M,S**), without cutting (**B,H,N,T**),
mature siliques (**C,I,O,U**) and inflorescence (**D,J,P,V**) from plants grown in LD. Mature (**E,K,Q,W**) and
young (**F,L,R,X**) siliques in SD conditions were also sampled. Bar in (**B**) representing 5 mm was the
same bar shared with other cauline leaves in (**A**), (**G**), (**H**). (**M**), (**N**), (**S**) and (**T**). GUS staining buffer
diffuses into the vascular tissue easily by cutting cauline leaf.

4.5 Looping model and 3C method in Arabidopsis

4.5.1 Looping model

Promoter truncation and mutagenesis established that *Block C* and *A* are required and
sufficient to drive *FT* expression in inductive LD. CO could associate with the proximal
promoter, and a CO/NF-Y complex could act as a bridge to form a chromatin loop structure

linking the distal enhancer *Block C* and the proximal promoter (Figure 11). To test this hypothesis, chromosome conformation capture (3C) was established, which allows to detect three-dimensional chromatin structures.
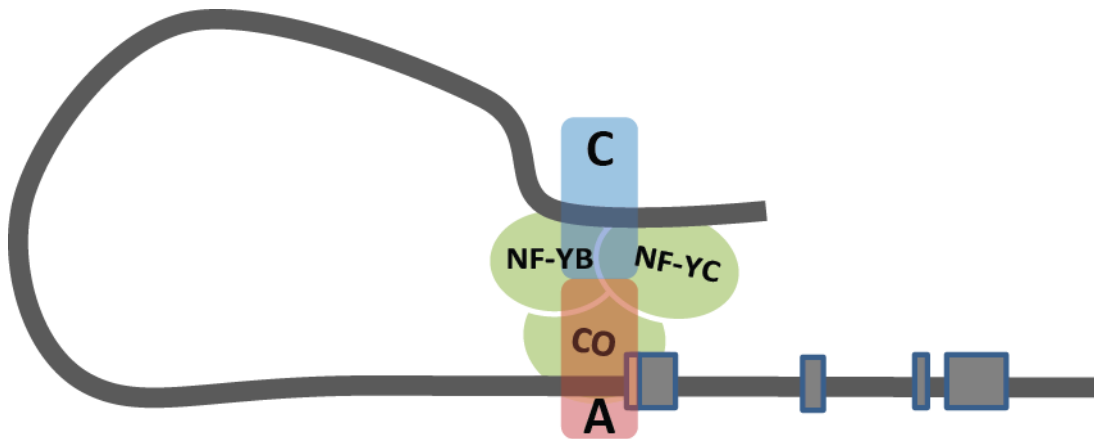

Figure 11. Looping model at *FT* promoter.

Block C acts as a distal enhancer for *FT* expression. At the proximal promoter, *Block A* containing the putative CO binding motifs is also required to initiate *FT* expression. CO interacts with NF-YB and NF-YC to form a trimetric complex. Trimeric NF-Y complexes composed of NF-YA, NF-YB and NF-YC are known to associate with *CCAAT* boxes. A *CCAAT* box localized at *Block C* is critical for full length promoter function. The distance between *Blocks C* and *A* only weakly influenced flowering time in LD. To enable interaction between the two conserved blocks, a chromatin loop may form at the *FT* promoter, which could be mediated by a CO/NF-Y complex in response to long days.

4.5.2 Design of 3C primers

In order to get a high resolution digestion map on the *FT* locus, restriction enzyme NlaIII, which recognizes *CATG* motifs, was chosen after screening most commercially available restriction enzymes. Thirteen NlaIII-digestion-fragments ranging from 169 bp to 1102 bp were selected, a fragment "fix" (739 bp) contained *Block A* and fragment IV (1102 bp) contained *Block C*. Seven fragments were selected for the intergenic region between *Block C* and *A*, while three fragments were selected upstream of *Block C*. The fragment XII, downstream of *Block A*, is located in the gene body (Figure 12A). In theory, any interactions among these thirteen fragments can be measured. However, to test the looping model between *Block C* and *A*, the fragment "fix" was tested against all other fragments.

To test whether 3C primers were functional, *BAC* clone F5I14 was digested by NlaIII and randomly ligated. Ligation product sizes of LI (fix + I) to LXII were in the range of 158-256 bp. Most products only showed a single band (Figure 12B), except for fragments LV, LVI and LXI, where a second, smaller band which may be the self-ligation product by the

fragment was detected. Notably, the smaller fragments of LVI and LXI were harder to detect than the target bands (Figure 12B). Because of variation among different primers in PCR, the *BAC* ligation library and the specific LI-LXII products were used to normalize plant 3C data.
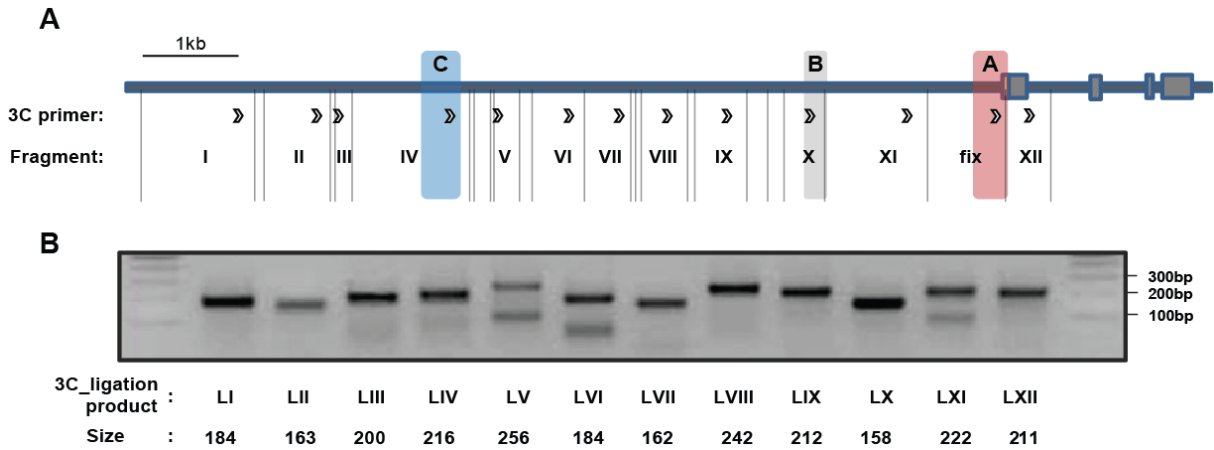

Figure 12. 3C primers and ligation products.

(**A**). On the *FT* locus, phylogenetically conserved blocks *Block A*, *B* and *C* are marked by red, gray and blue boxes respectively. NlaIII digestion sites are indicated by long gray lines. Thirteen restriction fragments were selected, fragments I to XII as well as the fragment "fix", which encompassed *Block A*. On each fragment, a primer (arrows) was designed at a distance of about 100-200 bp to the right restriction site. All 13 primers were oriented in the direction of *FT* transcription.

(**B**). *BAC* clone F5I14 containing the *FT* gene was digested by NlaIII and ligated. PCR was carried out using the *fixed* primer in combination with the other 12 primers to amplify ligation products LI (fix; I) to LXII. The PCR products were analyzed using a 1% agarose gel. Theoretically there are three products per PCR (take fragment fix and I for an example: fix; I, fix; fix and I; I). The size of target product (between fragment fix and I-XII) is indicated below the gel picture

4.5.3 Optimization of SDS concentration in the digestion and ligation

In the 3C procedure, high and appropriate digestion/ligation efficiency is a critical factor. Before digestion, SDS is used to break the nuclear membranes and loosen the chromatin structure. After digestion, SDS functions as an inactivator of NlaIII before the ligation step. Thus, it is critical to find the appropriate SDS concentrations to be added before and after digestion. To determine SDS concentrations, sensitivity experiments of NlaIII and T4 DNA ligase to SDS were carried out in cell free reactions (Figure 13).

Here, we used the 3C ligation product LXII which was amplified by primers annealing to the fragments fix and XII (Figure 11B). LXII only had one NlaIII digestion site and two digestion products were easy to separate by agarose gel electrophoresis. Based on that, we tested the effect of a series of SDS concentrations on digestion and ligation.
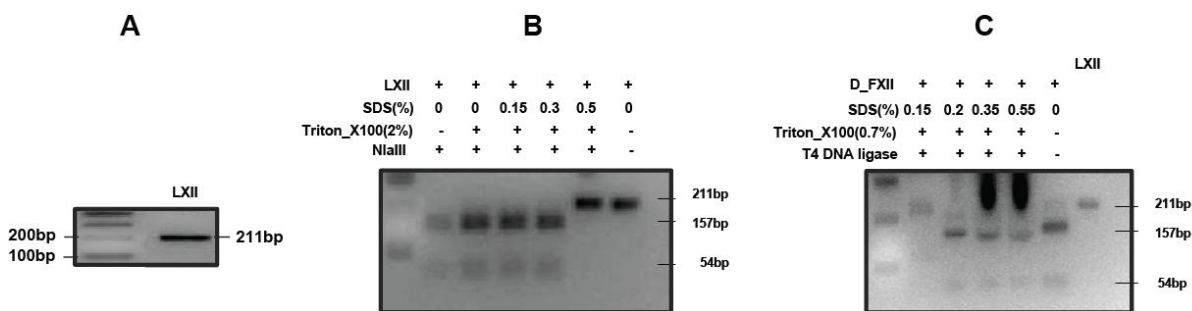
Figure 13. Sensitivity of enzymes NlaIII and T4 DNA ligase to SDS.

(**A**). *BAC*-digestion-ligation product LXII was amplified by PCR and loaded on a 1% agarose gel.

(**B**). Analysis of LXII digestions products by agarose gel electrophoresis. Different concentrations of SDS were added to LXII and the mixture was incubated at 65°C for 10 min following incubation at 37°C for 5 min. Then Triton-X100 was added to a final concentration 2% followed by incubating at 37°C for 15 min. Next, 20U NlaIII were added and digestion was carried out at 37°C overnight. Digestion products were loaded on a 1% agarose gel. LXII size is 211 bp, which can be digested by NlaIII resulting in two bands with the size of 157 bp and 54 bp.

(**C**). Fully digested products from (B) were purified and called digestion-fragments-LXII (D_FXII). SDS was added in different amounts ranging from 0.15 to 0.55% and 0.7% Triton-X100 following by incubation at 37°C for 1 hour. Next, 100U T4 DNA ligase was added and ligation was carried out at 16°C overnight. Ligation products were loaded on a 1% agarose gel. LXII was used as control.

In the digestion test, NlaIII could not cut DNA when SDS concentration was 0.5%. If SDS concentration was lower than 0.3% DNA could be fully digested (Figure 13B). Subsequently, re-ligation of digestion products was used to check the T4 DNA ligase's sensitivity to SDS. Results indicated that re-ligation was carried out with SDS concentration of 0.15%. However, when the SDS concentration was 0.2%, only weak re-legation could be observed and at higher concentrations, ligation was completely inhibited. The strong signals detected at the top of the agarose gel were probably due to high concentrations of SDS (Figure 13C).

According to these results, 0.15% SDS treatment before digestion and 0.1% SDS treatment during ligation were chosen for 3C experiments.

4.5.4 Digestion and ligation controls in 3C

Two protocols for 3C differ in the temperature protocol for an incubation of cross-linked chromatin before the digestion step (Hagege et al., 2007; Hovel et al., 2012). Protocol 1 uses 37°C for 1 hour ("37°C") and protocol 2 uses 65°C for 40 min followed by at 37°C for 20 min ("65°C"). For our 3C experiments, we evaluated both treatments by measuring subsequent digestion and ligation efficiencies. Chromatin prepared from formaldehyde fixed Col-0 and *35S::CO* seedlings were used in these tests.

Digestion efficiency was estimated by the ratio of PCR products amplified from samples incubated in absence and presence of NlaIII cutting and non-cutting PCR products. Results showed that under the same treatment of 37°C or 65°C, Col-0 and *35S::CO* chromatin had similar digestion efficiency efficiencies as measured by the ratio of two PCR products (PCR5/14 or PCR3/6) (Figure 14A).
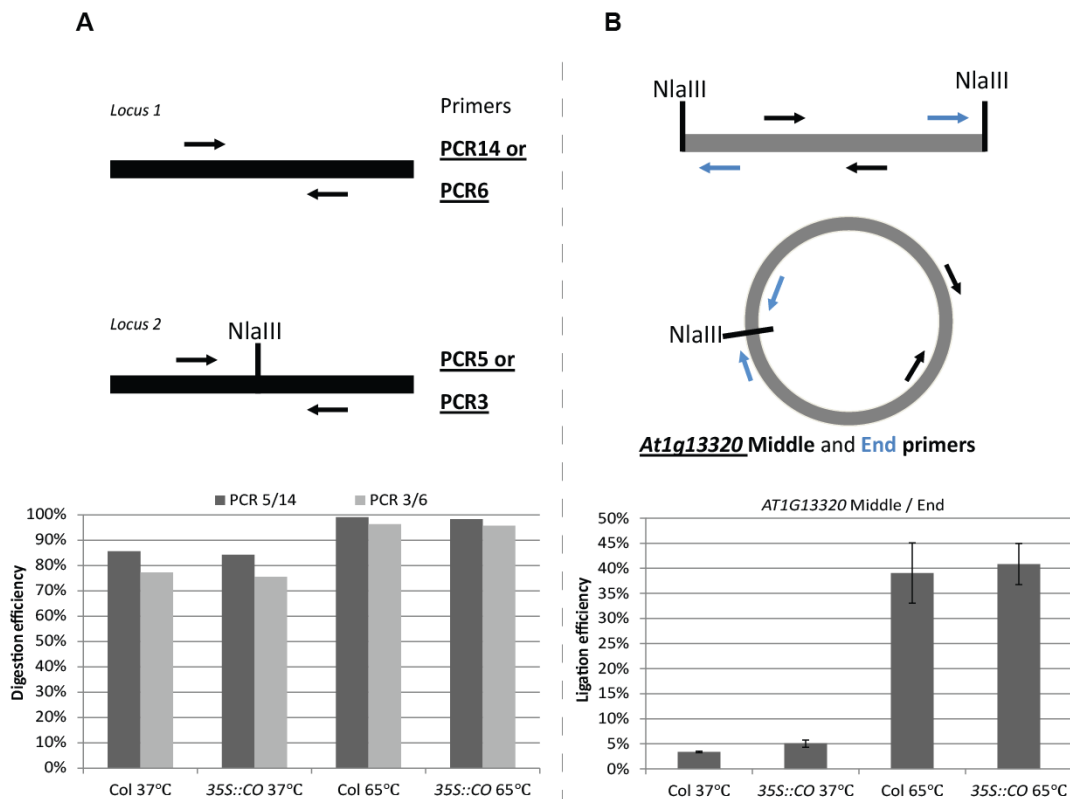

Figure 14. Digestion and ligation efficiency control.
(**A**). 25 µl aliquots of chromatin from pre-digestion and after-digestion were collected and DNA was purified from the samples. On Locus 1, not containing any NlaIII cutting site, PCR products were amplified using primers PCR14 or PCR6. On Locus 2, containing one NlaIII cutting site, the DNA was amplified by primers PCR5 or PCR3 across the cutting site. The ratios of PCR 5/14 or PCR 3/6 products were used to estimate digestion efficiencies. "37°C" and "65°C" stands for the temperature used for chromatin treatment after adding SDS before the digestion step. "37°C": 37°C for 1 hour. "65°C": 65°C for 40 min followed by at 37°C for 20 min.

(**B**). 25 µl aliquots of extract post-digestion and post-ligation were collected and DNA was purified. The *At1g13320* locus, a DNA region ~500 bp between two NlaIII sites was tested for ligation. One pair of primers was used to amplify the total DNA and the other pair positioned outwards at the end amplifies circular ligation products. The ratio of PCR products of *At1g13320* Middle/End was calculated to estimate the ligation efficiency. "37°C" and "65°C" treatment are same as in (A).

In addition, qPCR efficiency of these four primers was comparable (data not shown). However, for both Col-0 and *35S::CO*, the digestion efficiency was above 95% under treatment at 65°C, which was 15% higher than efficiencies reached under treatment at 37°C (Figure 14A). This suggested that high digestion efficiency was coupled with high

temperature treatment. However, at 65°C, 3C ligation products were obvious less than that at 37°C on *FT* promoter (Figure 16). Since the digestion was carried out in the cross-linked chromatin background, high digestion efficiency is not simply equal with high 3C ligation efficiency because high temperatures can reverse the chemical crosslinking by FA. Based on these results, we selected 37°C instead of 65°C treatment before ligation in our 3C assay.

4.5.5 Measurement of primer efficiency

Small variation of primer efficiency is not critical to quantify abundant product that show clear differences in quantity. However, 3C ligation products are at very low abundance. Therefore, primer efficiencies need to be precisely quantified for 3C.
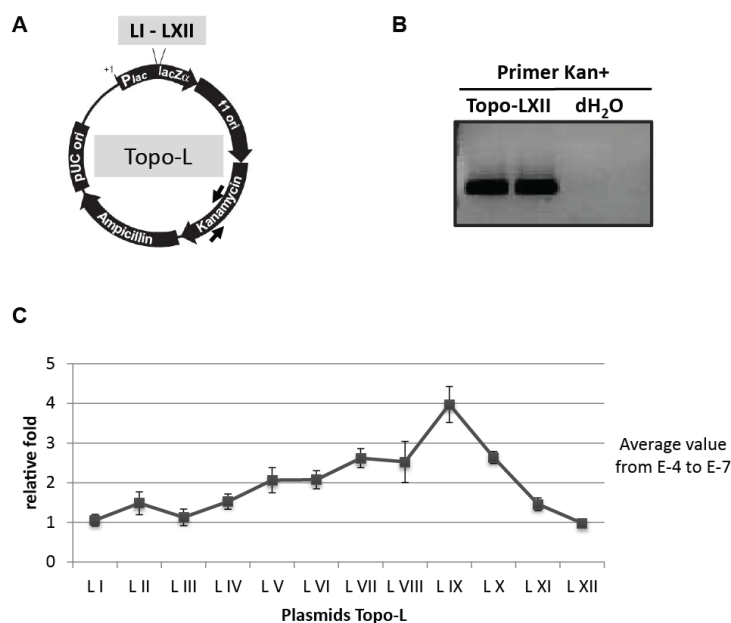
Figure 15. Quantification of 3C primer efficiency based on Topo-L plasmids.
(**A**). Plasmid of Topo-L (Topo vector with insertion of 3C ligation product) was cloned by inserting 3C ligation products LI-LXII into *pCR®2.1-TOPO*. Twelve clones were named Topo-LI to Topo-LXII.
(**B**). Primer Kan+ was used to amplify the shared region of Topo-L localized at *pCR®2.1-Topo* with arrows shown in (A). Topo-LXII was used as a positive control and dH$_2$O was the negative control.
(**C**). Primer Kan+ was used to quantify the twelve plasmids of Topo-LI to LXII by qPCR. 1 µg/µl Topo-L was defined as E-0 (10^0), and 0.0001µg/µl as E-4 (10^{-4}). Topo-L plasmids concentrated at E-4; E-5; E-6 and E-7 were used to normalize primer efficiency. Average value from E-4 to E-7 is shown and the relative concentrations of Topo-LI to Top0-LXI varied 4-fold.

First, we used a primer quantification method using plasmids controls. Twelve plasmids (Topo-LI to Topo-LXII) were cloned and their concentration was measured by optical density. Second, relative quantification of the twelve plasmids was carried out based on qPCR using primers annealing to the shared kanamycin resistance gene (Figure 15A, 15B and

15C). Finally, dilution series from twelve Topo-L plasmids were used as standard curve to quantify primer efficiency during qPCR (Figure 15C).

4.6 3C assays applied at *FT* promoter

4.6.1 3C result based on plasmid control normalization

The expression peak of *FT* mRNA is reached at ZT 16 in LD. In general, *FT* levels are higher in *35S::CO* plants but still peak at ZT 12-16 (Suarez-Lopez et al., 2001). Even in the background of *35S::CO*, GUS signals driven under the control of the full length *FT* promoter is restricted to the vasculature. However, compared to Col-0, the expression domain expands from the minor veins to the major veins in the phloem in the *35S::CO* background (Adrian et al., 2010). Thus, samples were harvested at ZT 16 to carry out 3C.
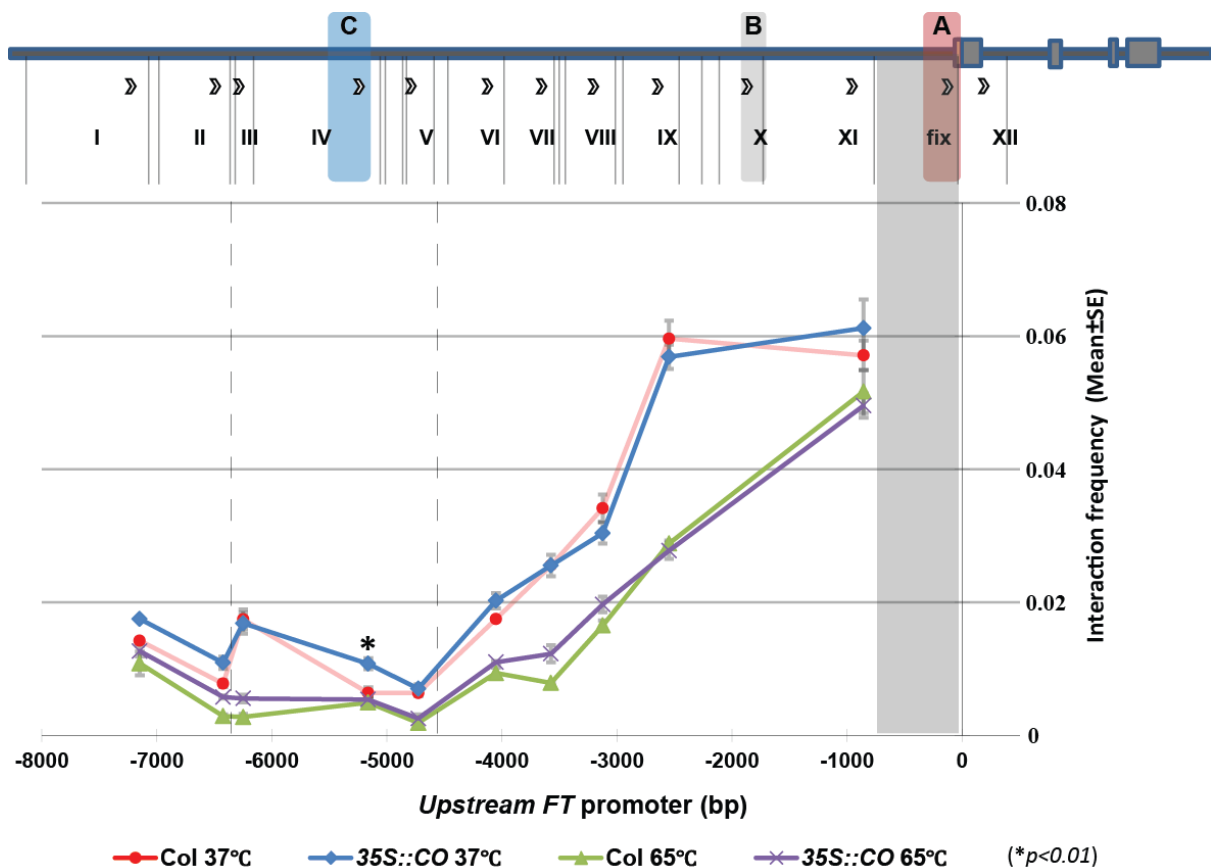

Figure 16. 3C result quantified by plasmids.
10-day-old Col-0 and *35S::CO* plants grown in LD were sampled at ZT 16. After cross-linking, isolation of chromatin, digestion and ligation, 3C interaction efficiencies were calculated after normalization to their corresponding Topo-L plasmids. The effects of treatments with "37°C" and

"65°C" on digestion and ligation efficiencies were shown in Figure 14. For the values at position III, IV and V a statistical analysis within the same treatment of "37°C" or "65°C" was carried out to compare the effect of different genotypes. The asterisk (*) indicates statistically differences. Statistical significance was determined using the Student's t-test ($p<0.01$) based on the technical errors in one experiment.

Once digestion and ligation efficiencies were optimized, 3C ligation products were quantified by each individual Topo-L plasmids using qPCR. Interaction efficiencies were plotted from the proximal *FT* promoter to the 5' upstream region to compare chromatin incubated at different temperatures that was prepared from *35S::CO* and Col-0 plants. All four curves showed tendencies to slope downwardly, with increasing distance from the fixed fragment. This is expected for 3C experiments, which indicated that the technique worked on the *FT* locus (Figure 16).

At all positions of the *FT* promoter, interaction values were higher when samples were ligated after pre-incubation at 37°C than at 65°C, indicating that the 65°C treatment decreased the interaction frequency. In the 37°C treated samples, the interaction frequency between fragments fix and IV was about two times higher in *35S::CO* plants (0.0011 in *35S::CO* compared to 0.0006 in Col-0). Differences were statistically significant between *35S::CO* and Col-0 at position IV, but not at neighbor positions III and V. The interaction frequencies of III and V with fix were comparable between *35S::CO* and Col-0 (Figure 16). The interaction difference at position IV was not detected in samples pretreated at 65°C. The results indicate that at least when applying 37°C treatment, a small peak at position IV, where enhancer *Block C* is localized, can be observed.

4.6.2 3C result based on *BAC* clone control normalization

3C data interpretation relies on precise quantification of different primer efficiencies. The plasmid Topo-L was used as control to quantify most primers efficiencies. However, since twelve plasmid controls from Topo-LI to XII had to be set independently, it produces errors and further takes time to dilute and load controls. Thus, we used another 3C data normalization method based on digestion and ligation of a *BAC* clone, which covers the whole sequence of interest.

BAC clone F5I14 contains the *FT* gene and all upstream and downstream regions. After a random ligation of NlaIII digestion fragments from the *BAC* clone, it was assumed that all the ligation products were present at similar amounts or that small variation among ligation

products could be similar to variation detected for 3C products. We used a serial dilution of
BAC-digestion-ligation products to produce a standard curve to quantify primer efficiency for
each qPCR.
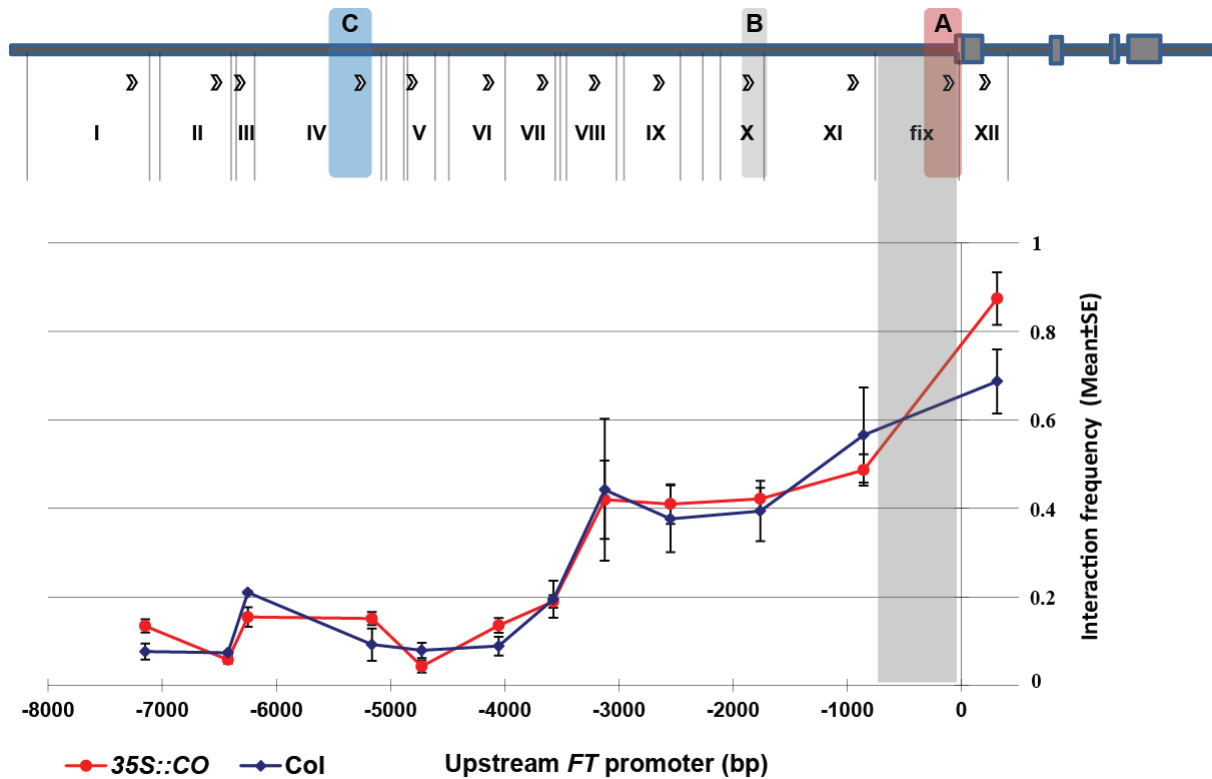
Figure 17. 3C result quantified by the *BAC* clone ligation products.
At ZT 16 in LD conditions, 10-day-old Col-0 and *35S::CO* plants were cross-linked by formaldehyde,
and after chromatin isolation, digestion and ligation, qPCR was carried to amplify ligation products. A
serial dilution of *BAC*-digestion-ligation products was used as standard to quantify products. Each E-4
diluted *BAC* ligation product of LI to LXII was set as interaction frequency 1. The data from a 37°C
pretreatment is shown. *35S::CO* and Col-0 plants are different samples compared to that in Figure 16.

In the 3C assay, a downwardly-sloped interaction frequency curve from *FT* proximal to distal
promoter fragments was observed (Figure 17). As in figure 16, a slightly higher interaction
efficiency was measured at position IV comparing *35S::CO* to Col-0 (Figure 16 and 17). In
addition, the frequencies of nearby fragments III and V with fix were a bit lower in *35S::CO*
(Figure 17). Combining two biological 3C experiments (Figure 16 and 17), a peak at position
IV could be measured.

4.7 INTACT method used for isolation of phloem nuclei

4.7.1 Transgenic INTACT lines

A small peak was observed from two independent 3C experiments at position IV that was enriched in *35S::CO* compared to Col-0, the low signal to background variation made it difficult to conclude that the peak was due to a three-dimensional loop chromatin structure (Figure 16 and 17).

The low amplitude of the peak at position IV could be due to unstable CO protein or because of a transiently established and destabilized loop. Even in *35S::CO* background *FT* is only expressed in phloem companion cells although high levels extend the expression of *FT* to major veins (Adrian et al., 2010). If the chromatin loop is tissue-specific and correlated with expression, it will be difficult to detect in chromatin prepared from whole seedlings, since the percentage of phloem companion cells within the whole leaf cells is low (Table 2).

In mammalian and yeast 3C studies, pure nuclei populations are generated from cell lines. Therefore, isolation of the companion cell or phloem-specific nuclei from the whole seedlings was thought to be beneficial for 3C experiments at *FT*.

In plants, two isolation methods FACS and LCM have been adopted to test the gene expression in tissue-specific manner (Birnbaum et al., 2003; Nakazono et al., 2003). Both of them need expensive sorting machinery and careful operation, but their yields of isolated specific cells were limited.

Table 2. Isolation phloem nuclei by FACS.

Plant material	Sorting channels	Sorting Volume(µl)	Nuclei no.	Percentage(%)
SUC2::H2B:YFP	+	200	4.1×10^4	3.4
	−	1000	8.0×10^5	88.4

In LD conditions, 14-day-old seedlings of *SUC2::H2B:YFP* in Col-0 background (provided by Markus Berns, MPIPZ) were cross-linked by formaldehyde. 10-15 plants were chopped completely and collected for each run of FACS. During sorting, nuclei marked by the yellow fluorescent protein (YFP) were selected through a positive "+" channel, and the majority of the remaining nuclei were selected through a negative "-" channel. Some nuclei were escaped and not collected by both channels. Here, the total nuclei from the positive and negative channels occupied 91.8% of the total nuclei.

Recently, a new nuclei sorting method named INTACT has been used in Arabidopsis and later been applied to Drosophila (Deal and Henikoff, 2010; Steiner et al., 2012). This system needs a cell/tissue-specific promoter to drive an engineered fusion gene named *NTF*.

Firstly, we explored a FACS method to isolate phloem nuclei from transgenic plants carrying a transgene to express a translational fusion of *H2B* and *YFP* under the control of the *SUCROSE-PROTON SYMPORTER 2* promoter (*SUC2::H2B:YFP*). After more than 2 hours of sorting, 4.1×10^4 nuclei with YPF signal were obtained. This number made up a proportion of 3% of the whole nuclei (Table 2). However, interaction signals could not be detected via qPCR, when 2×10^4 nuclei were used in 3C assay (data not shown). The number of nuclei needed for a successful 3C was more than 1×10^7 (Hagege et al., 2007). In conclusion, applying FACS proved to be difficult to enrich enough nuclei for 3C.

In the background of *ACT2::BirA* plants, we transformed *NTF* gene constructs fused to tissue-specific promoters such as the *CmGAS1* (*GALACTINOL SYNTHASE 1* in *Cucumis melo*) promoter specific for the companion cells of the minor veins and the *SUC2*,which drives expression in phloem companion cells of the minor and major veins (Haritatos et al., 2000; Truernit and Sauer, 1995). In *35S::NTF/ACT2::BirA* plants, GPF signal was observed in the nuclei of epidermal cells but also on the plasma membrane (Figure 18A and 18E). In addition, GFP signal in the cell of veins was observed in *35S::NTF/ACT2::BirA* plants (Figure 18B). GFP could only be observed in the nuclei of minor vein cells in *CmGAS1::NTF/ACT2::BirA* plants (Figure 18C and 18F). When using the *SUC2* promoter, the GPF signal expanded to the minor and major vein cells of leaves (Figure 18D). In summary, the phloem specific *CmGAS1* and *SUC2* promoters were able to drive *NTF* in their expected cells or tissues, and the NTF protein was mainly localized in the nucleus.
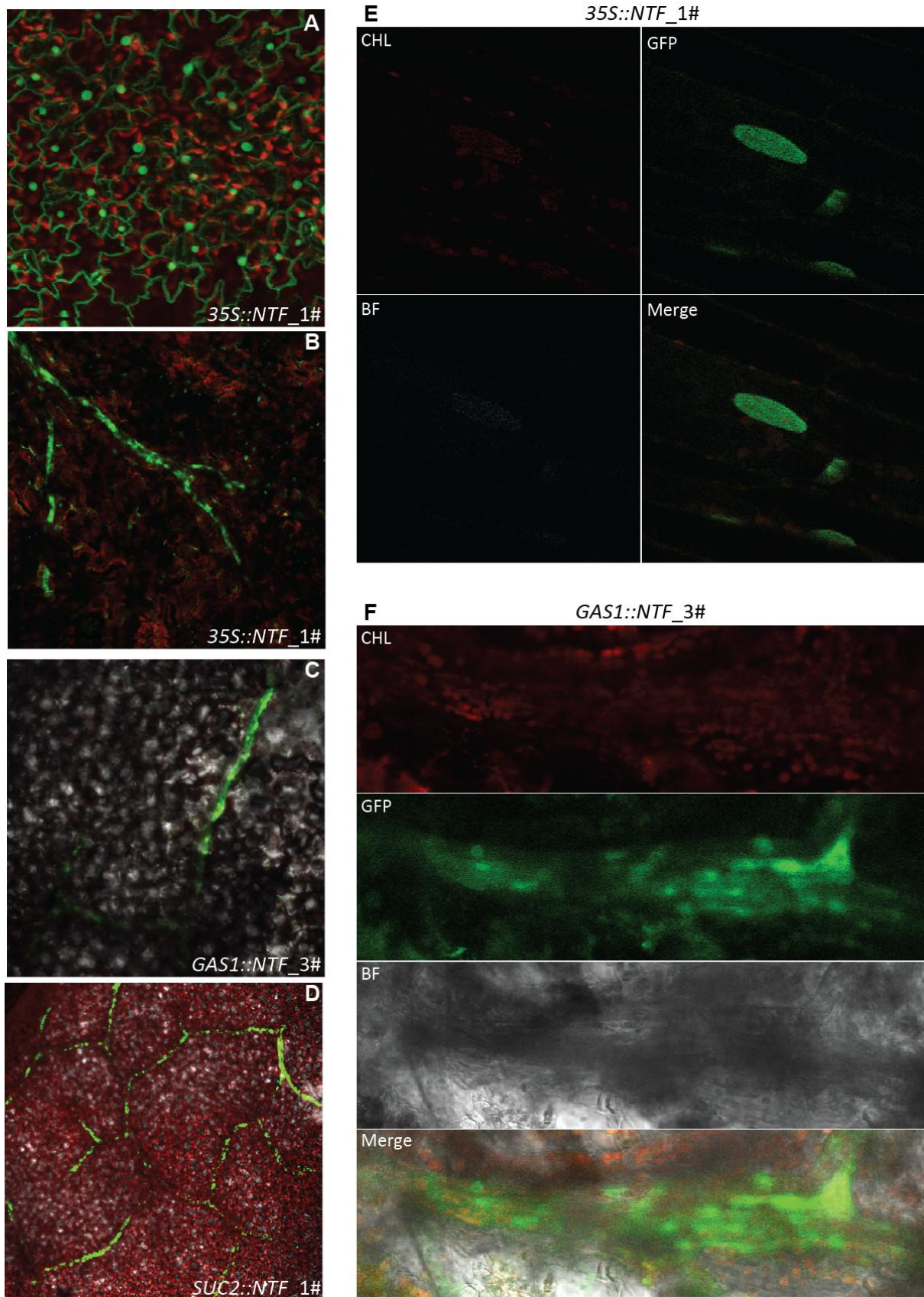
Figure 18. GFP signals in INTACT transgenic plants.

Pictures were taken with a fluorescent confocal microscope of 10-day-old plants of *35S::NTF* epidermal cells (**A**) and leaf veins (**B**), *GAS1::NTF* minor veins (**C**) and *SUC2::NTF* veins (**D**).

Subcellular localization of GFP signals of 25-day-old plants of *35S::NTF* epidermal cells (**E**) and *GAS1::NTF* minor veins (**F**). The background of all transgenic plants is *ACT2::BirA/*Col-0. In (E) and (F),"GFP", GFP fluorescence; "CHL", chlorophyll auto fluorescence; "BF", bright field image; and "Merge", merge of GFP, CHL, and BF.

4.7.2 INTACT assay test on RT-PCR and western blot

After producing stable transgenic plants containing *CmGAS1::NTF/ACT2::BirA*, the purification technique was tested coupled by measuring mRNA levels of marker genes. In addition, GFP and H3 protein levels were measured to confirm whether INTACT was successful.
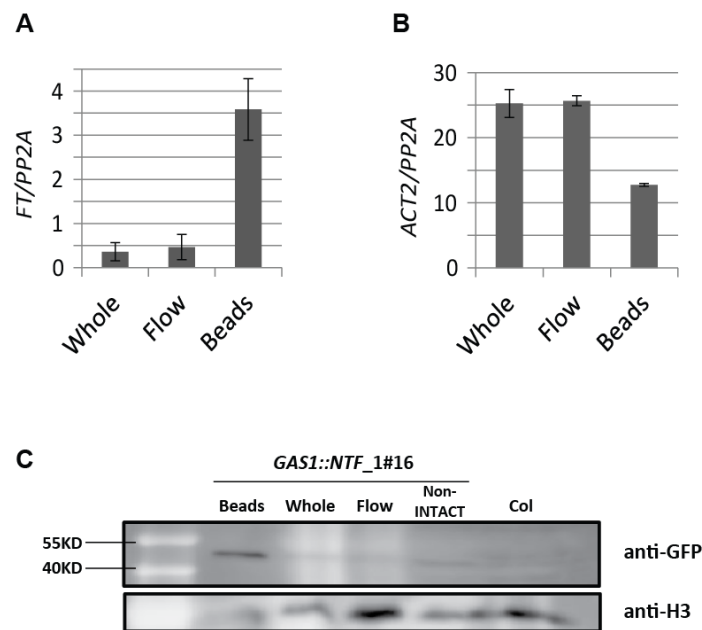
Figure 19. INTACT qRT-PCR and Western blot.

10-day-old plants of *CmGAS1::NTF/ACT2::BirA* grown in LD were sampled at ZT 16. qRT-PCR analysis of *FT* (**A**) and *ACT2* (**B**) transcript levels in the sorted nuclei "beads", flow-through nuclei "flow" and non-sorted nuclei "whole". Relative fold of values were normalized by *PP2A* transcripts. (**C**) The GFP fusion protein levels from the nuclei of "beads", "flow", "whole" and non-INTACT treated samples of *CmGAS1::NTF/ACT2* and Col-0 were tested in a western blot using anti-GFP antibodies and H3 from rabbit was used as internal control.

For each sorting procedure, I collected aliquots of nuclei before sorting as "whole", the nuclei attached by beads as "beads" and the flow-through nuclei as "flow". After INTACT sorting of nuclei from *CmGAS1::NTF/ACT2::BirA* followed by qRT-PCR, transcript levels of *FT*, *ACTIN2* and *PP2A* were measured from nuclei within 'whole', 'flow' and 'beads'. *FT* transcript levels were around eight times enriched in the sorted phloem nuclei compared to non-sorted nuclei (Figure 19A). However, *ACTIN2* mRNA showed an opposite enrichment, with half the amount of transcripts in the sorted nuclei compared to all nuclei (Figure 19B). When using GFP antibodies to test the enrichment of NTF fusion protein in the sorted nuclei,

the band of NTF protein detected by GFP was significantly stronger than in non-sorted nuclei (Figure 19C). The qRT-PCR and western blot analysis suggested that the INTACT technique worked properly with the samples used.

4.8 Tissue specific H3K27me3 enrichment of *FT*

The *FT* locus is enriched for the repressive H3K27me3 chromatin modification on the gene body and the promoter region (Farrona et al., 2011). Based on the ChIP-seq or ChIP-chip experiments performed with the whole Arabidopsis seedlings, H3K27me3 enrichment does not significantly change if *FT* is induced (Adrian et al., 2010).
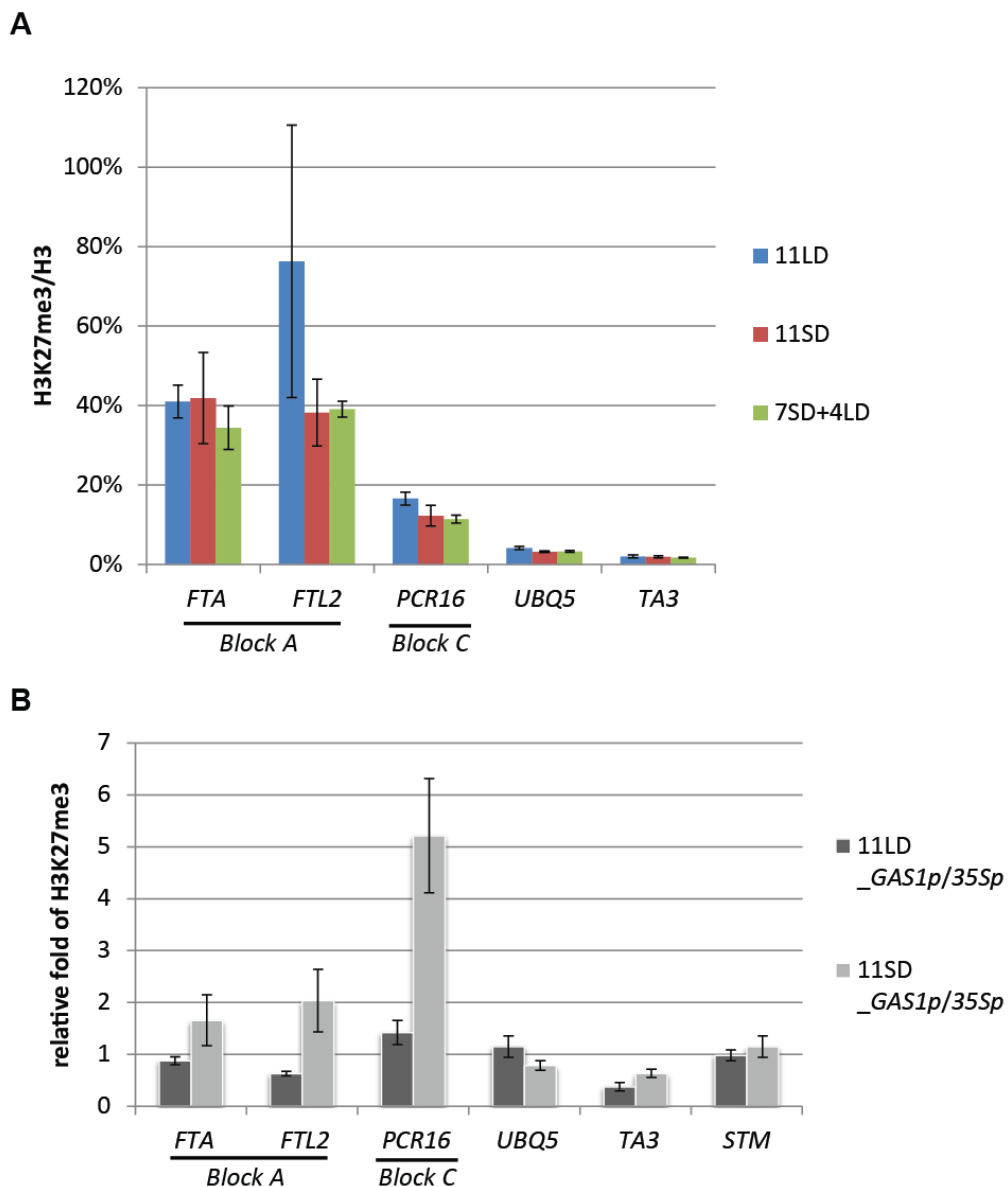

Figure 20. INTACT ChIP of H3K27me3 on *Block A* and *C* of *FT*.

(**A**). 11-day-old *CmGAS1::NTF/ACT2::BirA* plants were grown in LD conditions, SD conditions and 7LD+4SD conditions. H3K27me3 ChIP assay was carried on *FT*, *UBQ5* and *TA3* loci. The regions amplified by primers *FTA* or *FTL2* overlap with *Block A*, and *PCR 16* with *Block C*.

(**B**). 11-day-old *CmGAS1::NTF/ACT2::BirA* and *35S::NTF/ACT2::BirA* plants were grown in LD conditions and SD conditions. Using the sorted nuclei after the INTACT procedure, the enrichment of H3K27me3 on *FT*, *UBQ5*, *TA3* and *STM* genes were measured. Values were given as the relative fold of histone H3K27me3 after normalized to H3 gained from *CmGAS1::NTF/ACT2::BirA* sorted nuclei to *35S::NTF/ACT2::BirA* control. This was the preliminary result from one experiment.

Nuclei sorting by INTACT make it possible to answer the question whether the abundance of H3K27me3 at the same gene locus is different across cell types. The amount of nuclei ($\sim 10^4$) is enough for one ChIP assay (Deal and Henikoff, 2011). In my study, 10^5 phloem nuclei were isolated from *CmGAS1::NTF/ACT2::BirA* plants per INTACT. Thus, before 3C assay, it was easier to test the correlation between H3K27me3 enrichment and *FT* transcript levels in phloem by INTACT-ChIP.

On the *FT* locus, whole seedlings were used as ChIP sample from different culturing conditions such as LD, SD and a shift from SD to LD. *FT* transcripts of Col-0 at ZT 16 are observed in LD and in shifting conditions, but not in SD (Corbesier et al., 2007). As expected, ChIP with H3K27me3 on the *FT* promoter (*Block A* and *C*) showed no difference among these three conditions (Figure 20A).

After purification of nuclei from *35S::NTF* and *GAS1::NTF* expressing plants by the INTACT method, I carried out ChIP with H3K27me3. Material was collected from both lines from plants grown under LD and SD conditions. In lines of *GAS1* and *35S* under LD and SD, their relative enrichment of H3K27me3 was normalized to H3 control. Next, relative fold is normalized by the H3K27me3 signal in phloem cells compared to that in all cells. If the relative fold is one, it suggests that H3K27me enrichment did not change in all cells.

The results suggested that H3K27me3 was relatively enriched at *Block A* and *C* of the *FT* promoter in sorted phloem nuclei from plants grown under SD compared to LD. In addition, H3K27me3 on the control loci such as *UBQ5*, *TA3* and *STM* was not significantly changed between LD and SD (Figure 20B). In sum, compared SD to LD, the higher amount of the repressive mark H3K27me3 in SD could be one of the important features of transcription repression of *FT* in the phloem. In the future, isolated nuclei from phloem companion cells will be used for 3C.

5. Discussion

5.1 A distal enhancer is required for *FT* transcription in the leaf

A 5.7 kb *FT* promoter contains sufficient regulatory sequences for *FT* transcription under inductive long days, but shorter promoters such as 4kb or 1 kb cannot drive *FT* in Col-0 or *CO* overexpressor background. In *A.thaliana*, the 380bp long *Block C*, localized at a distal *FT* promoter region, is conserved across species of the Brassicaceae family (Adrian et al., 2010). Furthermore, a 5.2 kb *FT* promoter with deletion of *Block C* fails to drive *GUS* expression and cannot produce enough *FT* transcripts to complement the late flowering phenotype of the *ft* mutant under LD conditions (Figure 3). Taken together, these results indicate that *Block C*, 5.2 kb upstream of the *FT* translation start site, acts as an enhancer. This distal enhancer is essential for *FT* transcription in phloem cells of leaves before the floral transition.

In transient bombardment assays, *Block C* fused to the *NOS* minimal promoter (*C+NOSmin*) slightly increased *LUC* reporter gene expression (Figure 3B). In stably transformed plants, a weak GUS signal was also observed driven by promoter *C+NOSmin* in the phloem cells of the cotyledon, but the GUS signal almost disappeared in the true leaves (Figure 8C).These results suggest that *Block C* could be a weaker enhancer in transient assays as well as in phloem of cotyledons *in planta*. In true leaves, there could be an induced expression of a *Block C* specific repressor later in development that counteracts enhancing effect of *Block C*. But the proximal promoter seems to overcome that repression effect on *Block C* in phloem of true leaves (Figure 8C).

Strong enhancers are enriched in positive histone marks such as H3K4me1, H3K9ac and H3K27ac, and depleted with repressive marks such as H3K27me3 and H3K36me3 in human cells (Bulger and Groudine, 2011). Enhancers are very often cell type specific, with only few ones showing activity in more than two cell types and a majority being specific to a single cell type (Ernst et al., 2011; Heintzman et al., 2009). A 4 kb *FT* promoter without *Block C* still drives *GUS* reporter gene in siliques independent of photoperiod, but not in rosette and cauline leaves (Figure 10). This demonstrates that *Block C* seems to be a leaf-specific enhancer. On the *FT* promoter, *Block C* as well as the proximal promoter show comparably low H3K27me3 occupancy, whereas the middle intergenic region and gene body have high enrichment by H3K27me3 (Adrian et al., 2010). Genome-wide mapping of DNase I hypersensitive (DH) sites has been widely used to identify *cis*-regulatory elements in human cells (Boyle et al., 2008). In Arabidopsis, a genome-wide high resolution mapping of DH

sites reveal that a significant DH signal peak overlaps with *Block C*. Interestingly, the DH site signal is much stronger in two week old leaves compared to flowers (Zhang et al., 2012). The features of H3K27me3 mark and DH site provides more evidences to the point that *Block C* could serve as an enhancer that is accessible and active in a leaf-specific manner. And this enhancer could be recognized by phloem-specific *trans*-factors.

Neither functional *cis*-elements nor *trans*-factors binding sites have been identified in *Block C*. Several conserved sub-blocks localized at 5.3-5.5 kb upstream of the *FT* TSS were required for full length promoter function (Figure 4B). Among them, sub-block *S1* (*AGATCTTT*) contains a putative motif *AGATCT* which has been identified as DNA-binding site for yeast GATA-class proteins (Badis et al., 2008). The GATA-class proteins are conserved in plants, but so far, no relationship of this family to photoperiod control of flowering has been reported. In the future, *S1* or sequences stretches including *S1-S4* should be used as baits to screen Arabidopsis transcription factors in yeast-one-hybrid assays.

Astonishingly, ~100bp region including *S1-S4* plays important role for *FT* transcription. qRT-PCR did not detect any long RNA with poly A tail generated from *Blcok C* (data not shown), and no experimental evidence has reported possibly non-coding RNA or small RNAs in that region. It is possible that a complex formed by many interacting transcription factors has many *cis* binding sites in *Block C*. *Block C* may mediate a local, inter- or intra-chromosome looping structure. And that three dimensional interactions could be required for the formation of *FT* transcription machinery or guiding *FT* to the transcription factory. Additionally, it is also possible that *Block C* affects chromatin remodeling on *FT* locus.

In *5.2kbFTp::FTcDNA/ft-10* plants, *FT* mRNA levels at ZT 16 were very low compared to Col-0, which was also observed in different developmental stages (Figure 3E and data not shown). However, at ZT 8 or ZT 12, *FT* accumulated a low amount of transcripts. Flowering time data suggest that *FT* mRNA transcribed at non-peaking time such as ZT 8 or ZT 12 could not trigger flowering efficiently, possibly because FT cannot be effectively transported to the SAM at all times during the day (Figure 3C and 3D). FTIP1, as a transporter of FT protein, interacts with FT and probably this complex moves out of companion cells to sieve elements. Transcript levels of *FTIP1* are stable across the whole day and not regulated by photoperiod, GA, and vernalization pathways (Liu et al., 2012). This suggests that FT protein movement form companion cells to sieve elements could be independent of day time. However, further steps in the transport may still be dependent on multiple factors. Thus, whether *FT* transcripts with a morning peak contribute to floral transition is still an

interesting question. To address this, time-point specific as well as phloem specific promoters to drive *FT* in complementation assay could be of great interest.

In sum, a distal conserved enhancer named *Block C* was identified. *Block C,* which is composed of several independent functionally important sub-blocks, is required for *FT* expression specifically in phloem of leaves mediated by the long day pathway.

5.2 *CCAAT* boxes are required for *FT* expression

CCAAT boxes are ubiquitous elements in more than 30% of all eukaryotic promoters. They usually localize 60-100 bp upstream of transcription start sites in the forward or reverse orientation (Bucher, 1990; Mantovani, 1999). The *CCAAT* box is the putative binding cassette of the NF-Y trimeric complex. This complex can bind to *CCAAT* in yeast and humans and probably plants, although binding to *CCAAT* has not yet demonstrated *in vivo*. On the *FT* promoter, two *CCAAT* boxes seem to be functional for *FT* expression, and one is not (Figure 5B and 5C). The remaining five others will have to be tested in future studies.

Recent results from crystal structure suggest that the dimer of NF-YB and NF-YC, which is mediated by their histone-fold domains, behaves in a similar way as a dimer of core histones H2A/H2B (Nardini et al., 2013). This suggests that the trimeric NF-Y complex could interact with *CCAAT* in nucleosome free regions to form a nucleosome-like protein-DNA complex. In the distal enhancer *Block C*, a *CCAAT* box was important for *FT* transcription (Figure 5B). In the leaf, a NF-Y complex could bind to this *CCAAT* in *Block C* region, where it has low nucleosome density is suggested by a detection of a DH site (Zhang et al., 2012).

In Arabidopsis, CO, with homology to NF-YA, interacts with NF-YB and NF-YC to form a trimeric complex in *vivo*, and CO may replace of NF-YA in the NF-Y complex (Wenkel et al., 2006). Recent ChIP experiment confirmed that CO probably activates *FT* through directly binding to its proximal promoter (Adrian et al., 2010; Song et al., 2012; Tiwari et al., 2010). Based on these observations, CO/NF-YB/NF-YC trimeric complex could function as a bridge with one end binding to *CCAAT* in *Block C* and the other end to the proximal promoter.

This relatively simple model is confounded by the result that another *CCAAT* box, 868bp upstream of *FT* TSS at a distance of 4.4 kb from the one in *Block C* also plays a role in activating the full length *FT* promoter. Both boxes have an additive effect for the full length promoter function (Figure 5). They could communicate with each other over a long distance

by more than one NF-Y complex. NF-Y complexes as a bridge may bring the enhancer *Block C* close enough to the proximal promoter. This action may recruit activator or produce a permissive chromatin station for transcription. Based on the results from *FT* promoter truncation, *C+A* can mimic full length promoter function (Figure 8B).The results indicate that the *CCAAT* box near the proximal promoter could be not necessary for *FT* expression in LD conditions when the distance between *Block C* and *A* is reduced.

Taken together, at the *FT* locus, NF-Y complexes and CO as *trans*-factors may associate with a *CCAAT* in *Block C* for transcription regulation of *FT*. To elucidate how two *CCAAT* boxes communicate with each other, data from ChIP experiments of NF-YB, NF-YC and CO are required.

5.3 Distance variation between conserved blocks of *FT* promoter

In the *A. thaliana* Col-0 reference accession, the distance from enhancer *Block C* to *Block A* is 5.2kb, whereas in L*er* it is 4.2 kb (Turck et al., 2007). In the Brassicaceae species *Arabidopsis lyrata* and *Arabis alpina* the distances between *C* and *A* increase up to 7 kb and 11 kb respectively (Figure 6A). In *Brassica rapa* and *napus*, *Block C* is 7.5 kb far away from *A* at *FT-A2* promoter (Wang et al., 2012). Compared *A. thaliana*, *A.lyrata* and *B. rapa*, the distance between *Block C* and *A* may correlate with their evolutionary relationship and genome size (Johnston et al., 2005).

Among 80 accessions available from the "1001 Genomes Project", less diverged regions are observed for *FT* coding sequences, compared to the promoter region. This indicates that the deletions, insertions or SNPs accumulating on regulatory sequences produce frequent natural variation that could have an impact on transcription rates of *FT* and result in differentiated response to environmental variables. The conserved *Block C*, *B* or *A* has low SNPs frequencies and lowly diverged regions (Figure 6C). These observations indicate that mutations occurring in the conserved blocks may produce stronger effect than that in their neighboring regions on *FT* transcription. The main difference between the Col-0 and L*er* promoters is a middle region insertion of two DNA fragments in Col. After alignment of all 80 accessions available to Col-0 at *FT* promoter, 23 accessions display long *FT* promoters similar to Col-0 (Table 1). Interestingly, 22 Col-like accessions geographically distribute in the "West" group and only one strain exists in southern Russia (Which may be wrongly annotated due to sequencing error, personal communication with Korbininan Schneeberger).

Do environmental factors or evolutionary mechanisms make most Col-like accessions with long *FT* promoter habit in the west of Europe but not in Asia? Biogeography studies reveal that the postglacial Arabidopsis accessions in Europe might have emigrated from two parts: from the West Asia and the Iberian Peninsula. Accessions in North America could have been transferred from Europe during the new world discovery (Cao et al., 2011; Nordborg et al., 2005; Sharbel et al., 2000). Based on this hypothesis, I assume that L*er*-like accessions could have ancestry in Asia, and later moved to Europe. Col-like strains may firstly have existed in the Iberian Peninsula and later expanded to middle Europe and North America. L*er*-like and Col-like accessions co-colonized the main land of Europe (eg. Germany). Alpine mountain systems could be the natural border for spreading of Col-like accessions from Iberian Peninsula to East Europe and Asia. In consequence, *FT* promoter length could provide information for Arabidopsis population structure and geographic distribution. More accessions are expected to be available for analysis, which will complete the picture.

To test for selective effects, SNPs frequencies of DNA surrounding the long and short *FT* promoters should be carried out according to the linkage disequilibrium theory which has been applied on the *STM* locus in Arabidopsis (Piazza et al., 2010). In Arabidopsis, flowering time QTL frequently maps to the *FT* promoter (Schwartz et al., 2009; Strange et al., 2011), so the correlation between the flowering time of 80 accessions and *FT* promoter length is worth being investigated.

5.4 Middle region of *FT* promoter contains repressive regulatory elements

How does the promoter length influence *FT* transcription? In LD conditions, accession L*er* flowers earlier than Col-0 (Balasubramanian et al., 2006). Because flowering is determined by multiple factors, we cannot simply reason that L*er* accession flower early because they harbor a short *FT* promoter. The *dMp* lacking 3.5 kb middle part of the *FT* promoter mimics the full length promoter in driving *GUS* in phloem cells, but may extend the expression domain to the major veins (Figure 8C). But, *ft-10* mutants carrying *dMp::FTcDNA* did not show accelerated flowering compared to *5.7kbFTp::FTcDNA* expressing lines in LD (Figure 8B). However, *ft-10* is Col-0 background, and Col-0 is an early flowering accession which requires low threshold of FT protein amount for floral transition under inductive long days. *5.7kbFTp* and *dMp* promoters efficiently drove *FT* cDNA expression to reach the threshold of florigen quantity in the SAM at the same time. This is also true for both promoters under LDs

combined with high temperature, which are thought to accelerate flowering by an additive effect of thermal induction on *FT* expression that is independent of photoperiod (Figure 9B) (Balasubramanian et al., 2006). However, in non-inductive SD conditions, *ft-10* plants carrying *dMp::FTcDNA* were slightly earlier flowering than *5.7kbFTp:FTcDNA/ft-10* plants (Figure 9A). A similar tendency was observed in MD conditions, which were combined with cool ambient temperature (Figure 9C).

One possible explanation for the last observation is that the deleted middle region contains a *CArG* box, which is thought to recruit the transcriptional repressor SVP (Lee et al., 2007). SVP represses flowering in low ambient temperatures. SVP and FLC can form a complex that binds to the first intron and the promoter of *FT in vivo* (Lee et al., 2007; Li et al., 2008; Searle et al., 2006). In our complementation assays, *5.7kbFTp::FTcDNA* over-complements *ft* mutants in SD and LD conditions (Figure 8B and 9A). Missing the repressive sequences in the first intron could also reduce the effect of a SVP/FLC repressive complex. When fused to *FT* genomic DNA, the flowering time difference between *dMp* and full length promoters was much more significant (personal communication with Tingting Ning from Turck group). In the future, comparisons should be carried out between constructs that carry variations of the full genomic *FT* locus.

The region between *C* and *A* is highly enriched with the repressive chromatin mark H3K27me3 and the associated protein LHP1. GUS signal driven by full length *FT* promoter in *lhp1* background spreads to the major veins of true leaves (Adrian et al., 2010; Farrona et al., 2011). Similarly, promoter *C+A*, which removes the binding sequences of LHP1, can drive *GUS* in the major veins in Col-0 (Figure 8C). In the future, the enrichment of H3K27me3 on the proximal promoter and *Block C* region should be tested for the constructs of *C+A*, *dMp* and *5.7kbFTp* in transgenic plants. Previous results demonstrate that PRC2 components repress *FT* transcription, but are not the reason that expression of *FT* is restricted to phloem cells (Farrona et al., 2011). In addition, some unknown activators of *FT* with phloem-specific expression manner could also be of importance for understanding transcriptional regulation of *FT*.

5.5 3C technique normalization in *Arabidopsis*

In vivo, Chromosome conformation capture (3C) is a powerful tool to detect chromatin looping structure (Dekker et al., 2002; Hagege et al., 2007).

3C has been generally applied in yeast and human cells which have single cell type (Hagege et al., 2007; Singh et al., 2009). In Maize, husk tissue where *B-I* has extremely strong expression was sampled (Louwers et al., 2009a). In Arabidopsis no 3C protocol is available before the start of this study. Thus, 3C protocol was optimized in Arabidopsis integrating previous studies from human, yeast and Maize (Hagege et al., 2007; Louwers et al., 2009b; Singh et al., 2009).

3C detects looping between two long-range loci usually in hundred or thousand kb distance. There are very few examples involving the scale of less than 10 kb. In our study, the proposed loop would only cover 5 kb. Therefore, NlaIII, a four-base-cutter, was used to obtain a high resolution of restriction sites. If the region has several close restriction sites, PCR products may not be unique after incomplete digest. So, I designed 3C primers being close to unique cutting site.3C primer specificity is critical because biased amplification will produce a high risk to increase or decrease the actual difference. If the primer's PCR efficiency is too high or too low, they should be discarded.

SDS plays an important role in 3C assays. SDS is used to break the nuclear membrane, loosen chromatin and inactivate restriction enzyme, but at the same time interferes with restriction enzyme and ligase activity. Thus, SDS must be carefully titrated by TritonX-100 not to influence NlaIII and T4 ligase activity.

It is also important to balance the efficiency of digestion and ligation at the control locus. In this study, "65°C" or "37°C" pre-treatments of chromatin were tested before digestion. High temperature produces very high digestion around 98% and 40% ligation efficiency at a control gene. But final 3C products are less with "65°C" than "37°C" treatment. In my experiments, higher enrichment of 3C products relies on relatively high digestion efficiency (~85%) and appropriate ligation efficiency (5%) (Figure13). This reasonable digestion efficiency matches well with other studied suggesting that 80-90% digestion efficiency is critical for successful 3C (Simonis et al., 2007). 3C product amount depends on protein and DNA interaction complex fixation due to crosslinking. High temperature reverses the chemical crosslinks, explaining why lower 3C products were measured. In maize apparently 65°C treatment was needed to get rid of endogenous nuclease activity (maize nature protocol), but it apparently not a problem in Arabidopsis.

A perfect control template to quantify 3C primer efficiency should contain all ligation products of interest in equal amounts. In 3C, control templates are usually made by *BAC* clones that cover the DNA region of interest (Dekker, 2006). We adopted this *BAC* method to normalize primer efficiency (*BAC* control). In parallel, a second method based on cloning all

possible 3C products was also carried out (plasmid control). Both methods worked comparably well in our study, but quantification from the *BAC* clone was much easier, because making the plasmid controls is very time consuming.

In summary, 3C as a new technology was applied in Arabidopsis. After normalization, at *FT* locus, 3C works technically as an expected decrease of interaction with distance to the fixed fragment was detected.

5.6 Physical Interaction of *Block C* with the proximal promoter

Block C fused to *Block A* mimic full length promoter which are required and sufficient for *FT* expression. By which molecular mechanism does *Block C* communicate with *Block A* to regulate *FT* transcription?

Two models for transcription regulation by a distal enhancer are described. In the looping model, a physical enhancer-promoter interaction is mediated by activators or mediator complex. In the tracking model, an enhancer-activator complex scans along the upstream promoter until it encounters the core promoter-RNAPII complex (Bulger and Groudine, 2011). Transgenic lines with T-DNA inserting into the region between *C* and *A* display a flowering time similar to wild type plants under inductive LD conditions (PhD thesis of Jessica Adrian). If the enhancer works in a tracking model, the T-DNA insertion should be deleterious in the scanning scenario.

I favor the looping hypothesis that *Block C* and *A* interacts through a protein complex comprising of CO or CO/NF-YB/NF-YC complex. Enlarging or reducing the distance between *Block C* and *A* did not change flowering time dramatically Additionally, the presence of the *CCAAT* box in *Block C* and the CO-biding sequences in *Block A* may provide the anchoring sites for CO/NF-YB/NF-YC complex.

To test the looping model, 3C was carried out on *FT* locus. A weak interaction between *C* and *A* was detected by 3C in at least two biological replicates (Figure 16 and 17). In *35S::CO* plants within "37°C"treatment, a higher interaction frequency was observed at position IV, which was not observed in *35S::CO* plants within "65°C" treatment (Figure 16). One may argue that CO or CO complex is unstable after high temperature treatment. *In vivo*, CO protein is degraded through the 26S proteasome degradation pathway mediated by SPA1 or PHYB (Laubinger et al., 2006; Valverde et al., 2004). Even *in vitro*, CO protein is also unstable in EMSA assay (Tiwari et al., 2010). The weak interaction between *Block C* and *A*

could due to the instability of CO, thus the looping structure mediated by CO could be difficult to capture.

Another possibility could be due to the problem that interaction correlates with tissue specificity of *FT* expression in a low number of cells. In yeast and human pure cell types are used in 3C. In Maize, at *b1* epialleles a hepta-repeat sequence, 100 kb far from TSS, physically interacts with the proximal promoter in a tissue- and expression level-specific manner (Louwers et al., 2009a). Recently, the formation of a loop by 5' and 3' gene flanking regions has been reported on the *FLC* locus. In this study, samples of 10-day-seedlings were used when *FLC* is expressed in a non-tissue-specific manner. The *FLC* loop did not correlate with expression level (Bastow et al., 2004; Crevillen et al., 2013). For the *FT* expression domain, the total companion cells in the phloem occupy 3% of all leaf cells (Table 2), and I assume that the number of cells expressing *FT* in the proximal major veins even in *35S::CO* plants is very small. Thus, the weak interaction signal could be the result of a dilution effect. Using the sorting method of INTACT, the phloem nuclei express high *FT* mRNA, as well as GFP signal was enriched suggesting that the sorting system using minor vein-specific-promoter *CmGAS1* works well in INTACT (Figure 19). More than 10^7 nuclei are required for a 3C assay. Parallel isolation of phloem nuclei by INTACT will be carried out and used for 3C assays in the future.

Last, weak interaction could also be due to multi-loops formation, which interferes with enhancer-promoter loop at the *FT* locus. MADS-domain proteins SVP and FLC could form a complex binding to the first intron and 2kb upstream region of the *FT* promoter (Lee et al., 2007; Li et al., 2008; Searle et al., 2006). At 6 kb region, it could be an impossible challenge to measure multi-loops in 3C.

5.7 H3K27me3 differentially enriches at *FT* promoter in phloem under LD and SD

H3K27me3 is the repressive histone mark that determines *FT* expression levels and photoperiod dependency (Adrian et al., 2010; Farrona et al., 2011; Takada and Goto, 2003). Under LD conditions, *FT* expresses at a peaking level at ZT 16, but *FT* is repressed under SD conditions (Corbesier et al., 2007). Based on ChIP results from the whole seedlings, the abundance of H3K27me3 mark is not correlated with the *FT* expression level (Adrian et al., 2010) When isolating phloem nuclei for ChIP, H3K27me3 levels at *FT* promoter were significantly enriched in SD than in LD conditions at *Block C* (Figure 20B). Under non-

inductive SD, absence of *FT* expression is partially due to instability of CO, but enrichment of *Block C* and proximal *Block A* with repressing histone mark may add another mechanistic explanation. In LD conditions, presence of CO and reduced H3K27me3 at *Block C* and *A* correlate but it is so far impossible to speculate on causal relationships between this two (Figure 21). Using the INTACT method, it will be interesting to further elucidate the correlation between CO and H3K27me3 in and out of phloem nuclei.
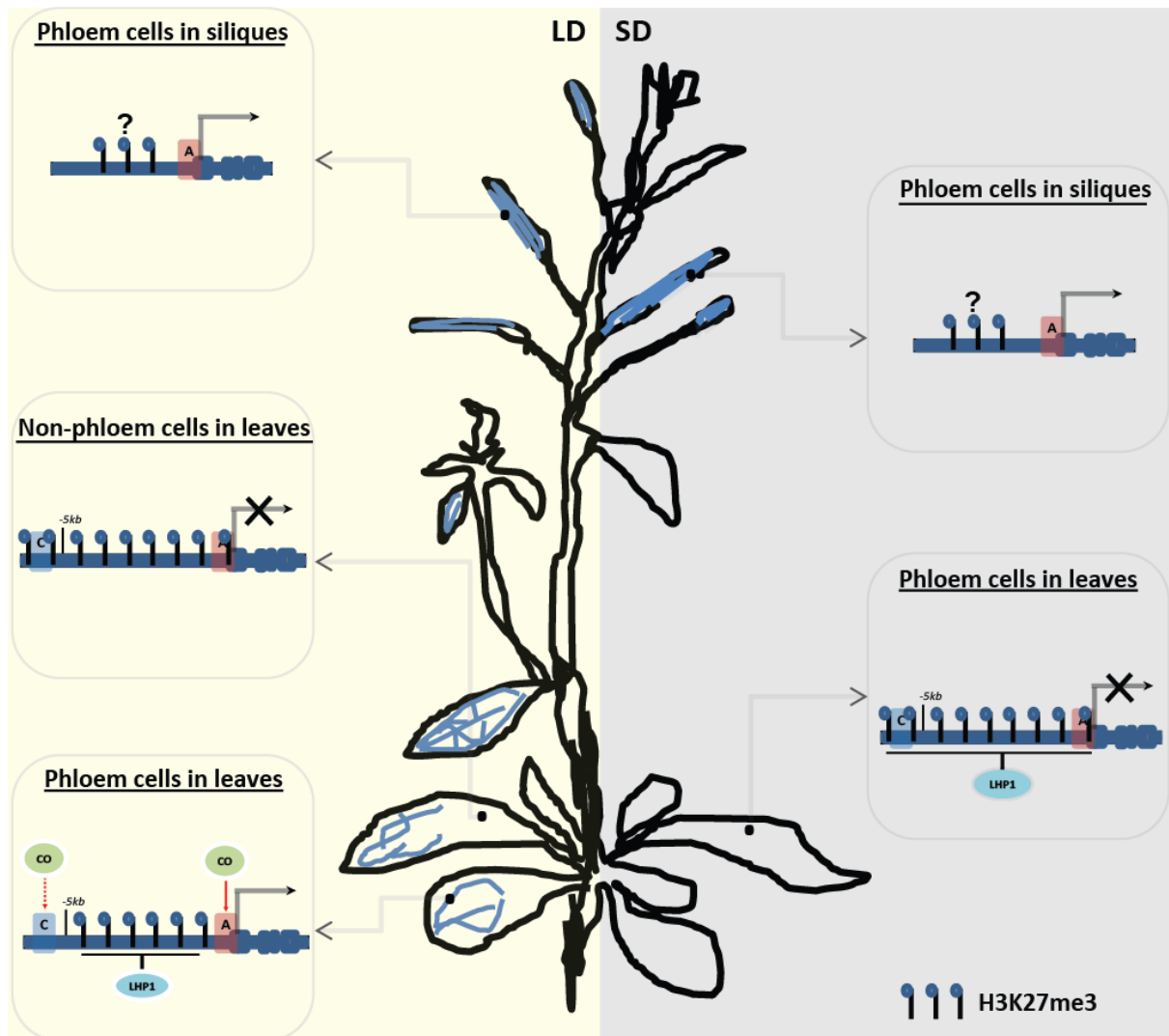
Figure 21. Hypothesis of cell- and tissue- specific expression manner of *FT*.
In LD conditions, in phloem tissue, the enhancer *Block C* and proximal *Block A* are depleted of repressive H3K27me3, as well stable CO could associate with *C* and *A* mediated activation of *FT* transcription. In non-phloem cells of leaves, CO cannot express meanwhile the *Block C* and *A* are not in a permissive chromatin state which make *FT* silent. Under non-inductive SD conditions, CO is degraded, and *FT* locus chromatin is condensed because of enrichment of H3K27me3 at *Block C* and *A* where LHP1 probably binds to. *FT* was repressed. In the phloem of siliques, *FT* expression is independent of photoperiod, and does not require *Block C*. In siliques, *4kbFTp* is sufficient to drive *FT*, but *1kbFTp* fails like in leaves. Blue regions represent the phloem tissue where *FT* expresses.

5.8 *FT* expression in life cycle

FT specifically expresses in companion cells of the phloem in cotyledons, rosette leaves, cauline leaves, sepals and siliques (Figure 21). Apart from the young and old siliques, where a 4 kb promoter without *Block C* drives GUS reporter gene expression, all other tissues require *Block C* as a distal enhancer. This suggests that a different transcriptional regulation of *FT* occurring in siliques, which is also independent of the photoperiod pathway as expression is similar in SD and LD. It could be interesting to test the abundance of H3K27me3 at *FT* in the phloem of siliques because an absence of this mark could explain the loss of photoperiod dependency (Adrian et al., 2010).

FT as part of florigen is famous for triggering the floral transition in plants (Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007; Tamaki et al., 2007). However, after flowering what could be the function of *FT* expression in cauline leaves and siliques? Our recent observation revealed a possible novel function of *FT* in preventing floral reversion under SD conditions (Turck group, personal communication).

SFT, the ortholog of *AtFT* in tomato, influences heterosis in yield as heterozygous sft/+ genotype by affecting the inflorescence architecture (Krieger et al., 2010). In potato, floral and tuberization transitions are controlled by two different FT-like paralogs *StSP3D* and *StSP6A* (Navarro et al., 2011).

In conclusion, it will be interesting to understand the working model of *FT* transcription during the whole life cycle in plants because the results may be relevant for aspects of development beyond the induction of flowering by photoperiod.

6. Literature

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science **309,** 1052-1056.

Adrian, J., Farrona, S., Reimer, J.J., Albani, M.C., Coupland, G., and Turck, F. (2010). cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis. The Plant cell **22,** 1425-1440.

Ahn, J.H., Miller, D., Winter, V.J., Banfield, M.J., Lee, J.H., Yoo, S.Y., Henz, S.R., Brady, R.L., and Weigel, D. (2006). A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. The EMBO journal **25,** 605-614.

Allard, H.A., and Garner, W.W. (1940). Further observations on the response of various species of plants to length of day. Technical Bulletin. United States Department of Agriculture.

An, H., Roussot, C., Suarez-Lopez, P., Corbesier, L., Vincent, C., Pineiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C., and Coupland, G. (2004). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development **131,** 3615-3626.

Andres, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nature reviews. Genetics **13,** 627-639.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., Gebbia, M., Talukder, S., Yang, A., Mnaimneh, S., Terterov, D., Coburn, D., Li Yeo, A., Yeo, Z.X., Clarke, N.D., Lieb, J.D., Ansari, A.Z., Nislow, C., and Hughes, T.R. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Molecular cell **32,** 878-887.

Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D. (2006). Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. PLoS genetics **2,** e106.

Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. Nature **427,** 164-167.

Bhasin, M., Reinherz, E.L., and Reche, P.A. (2006). Recognition and classification of histones using support vector machine. Journal of computational biology : a journal of computational molecular cell biology **13,** 102-112.

Bi, W., Wu, L., Coustry, F., de Crombrugghe, B., and Maity, S.N. (1997). DNA binding specificity of the CCAAT-binding factor CBF/NF-Y. The Journal of biological chemistry **272,** 26562-26572.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science **302,** 1956-1960.

Blazquez, M.A., Ahn, J.H., and Weigel, D. (2003). A thermosensory pathway controlling flowering time in Arabidopsis thaliana. Nature genetics **33,** 168-171.

Borner, R., Kampmann, G., Chandler, J., Gleissner, R., Wisman, E., Apel, K., and Melzer, S. (2000). A MADS domain gene involved in the transition to flowering in Arabidopsis. The Plant journal : for cell and molecular biology **24,** 591-599.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008). High-resolution mapping and characterization of open chromatin across the genome. Cell **132,** 311-322.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of molecular biology **212,** 563-578.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell **144,** 327-339.

Calvenzani, V., Testoni, B., Gusmaroli, G., Lorenzo, M., Gnesutta, N., Petroni, K., Mantovani, R., and Tonelli, C. (2012). Interactions and CCAAT-binding of Arabidopsis thaliana NF-Y subunits. PloS one **7,** e42902.

Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Muller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K.J., and Weigel, D. (2011). Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nature genetics **43,** 956-963.

Castillejo, C., and Pelaz, S. (2008). The balance between CONSTANS and TEMPRANILLO activities determines FT expression to trigger flowering. Current biology : CB **18,** 1338-1343.

Chautard, H., Jacquet, M., Schoentgen, F., Bureaud, N., and Benedetti, H. (2004). Tfs1p, a member of the PEBP family, inhibits the Ira2p but not the Ira1p Ras GTPase-activating protein in Saccharomyces cerevisiae. Eukaryotic cell **3,** 459-470.

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

Crevillen, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. The EMBO journal **32,** 140-148.

Deal, R.B., and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Developmental cell **18,** 1030-1040.

Deal, R.B., and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nature protocols **6,** 56-68.

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nature methods **3,** 17-21.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science **295,** 1306-1311.

Denissov, S., Lessard, F., Mayer, C., Stefanovsky, V., van Driel, M., Grummt, I, Moss, T., and Stunnenberg, H.G. (2011). A model for the topology of active ribosomal RNA genes. EMBO reports **12,** 231-237.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of

chromatin interactions. Nature **485,** 376-380.

Edelman, L.B., and Fraser, P. (2012). Transcription factories: genetic programming in three dimensions. Current opinion in genetics & development **22,** 110-114.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., and Bernstein, B.E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature **473,** 43-49.

Farrona, S., Coupland, G., and Turck, F. (2008). The impact of chromatin regulation on the floral transition. Seminars in cell & developmental biology **19,** 560-573.

Farrona, S., Thorpe, F.L., Engelhorn, J., Adrian, J., Dong, X., Sarid-Krebs, L., Goodrich, J., and Turck, F. (2011). Tissue-specific expression of FLOWERING LOCUS T in Arabidopsis is maintained independently of polycomb group protein repression. The Plant cell **23,** 3204-3214.

Fornara, F., Panigrahi, K.C., Gissot, L., Sauerbrunn, N., Ruhl, M., Jarillo, J.A., and Coupland, G. (2009). Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Developmental cell **17,** 75-86.

Gomez-Mena, C., Pineiro, M., Franco-Zorrilla, J.M., Salinas, J., Coupland, G., and Martinez-Zapater, J.M. (2001). early bolting in short days: an Arabidopsis mutation that causes early flowering and partially suppresses the floral phenotype of leafy. The Plant cell **13,** 1011-1024.

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., and Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature **386,** 44-51.

Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nature protocols **2,** 1722-1733.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nature structural & molecular biology **11,** 394-403.

Hamner, K.C. (1940). Interrelation of light and darkness in photoperiodic induction. Botanical Gazette**,** 658-687.

Hanzawa, Y., Money, T., and Bradley, D. (2005). A single amino acid converts a repressor to an activator of flowering. Proceedings of the National Academy of Sciences of the United States of America **102,** 7748-7753.

Haritatos, E., Ayre, B.G., and Turgeon, R. (2000). Identification of phloem involved in assimilate loading in leaves by the activity of the galactinol synthase promoter. Plant physiology **123,** 929-937.

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature **422,** 719-722.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X., Green, R.D., Lobanenkov, V.V., Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M., and Ren, B. (2009). Histone modifications at human

enhancers reflect global cell-type-specific gene expression. Nature **459,** 108-112.

Hisamatsu, T., and King, R.W. (2008). The nature of floral signals in Arabidopsis. II. Roles for FLOWERING LOCUS T (FT) and gibberellin. Journal of experimental botany **59,** 3821-3829.

Hovel, I., Louwers, M., and Stam, M. (2012). 3C Technologies in plants. Methods **58,** 204-211.

Huang, T., Bohlenius, H., Eriksson, S., Parcy, F., and Nilsson, O. (2005). The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. Science **309,** 1694-1696.

Huijser, P., and Schmid, M. (2011). The control of developmental phase transitions in plants. Development **138,** 4117-4129.

Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A., and Kay, S.A. (2005). FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. Science **309,** 293-297.

Jack, T. (2004). Molecular and genetic mechanisms of floral control. The Plant cell **16 Suppl,** S1-17.

Jaeger, K.E., and Wigge, P.A. (2007). FT protein acts as a long-range signal in Arabidopsis. Current biology : CB **17,** 1050-1054.

Jang, S., Torti, S., and Coupland, G. (2009). Genetic and spatial interactions between FT, TSF and SVP during the early stages of floral induction in Arabidopsis. The Plant journal : for cell and molecular biology **60,** 614-625.

Jing, H., Vakoc, C.R., Ying, L., Mandat, S., Wang, H., Zheng, X., and Blobel, G.A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. Molecular cell **29,** 232-242.

Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J., Hodnett, G., Drabek, J., Lopez, R., and Price, H.J. (2005). Evolution of genome size in Brassicaceae. Annals of botany **95,** 229-235.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., and Weigel, D. (1999). Activation tagging of the floral inducer FT. Science **286,** 1962-1965.

Karlgren, A., Gyllenstrand, N., Kallman, T., Sundstrom, J.F., Moore, D., Lascoux, M., and Lagercrantz, U. (2011). Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. Plant physiology **156,** 1967-1977.

Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell **77,** 599-608.

King, R.W., Hisamatsu, T., Goldschmidt, E.E., and Blundell, C. (2008). The nature of floral signals in Arabidopsis. I. Photosynthesis and a far-red photoresponse independently regulate flowering by increasing expression of FLOWERING LOCUS T (FT). Journal of experimental botany **59,** 3811-3820.

Kobayashi, Y., and Weigel, D. (2007). Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. Genes & development **21,** 2371-2384.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. (1999). A pair of related

genes with antagonistic roles in mediating flowering signals. Science **286,** 1960-1962.

Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. Molecular & general genetics : MGG **229,** 57-66.

Kotake, T., Takada, S., Nakahigashi, K., Ohto, M., and Goto, K. (2003). Arabidopsis TERMINAL FLOWER 2 gene encodes a heterochromatin protein 1 homolog and represses both FLOWERING LOCUS T to regulate flowering time and several floral homeotic genes. Plant & cell physiology **44,** 555-564.

Krieger, U., Lippman, Z.B., and Zamir, D. (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nature genetics **42,** 459-463.

Kumar, S.V., Lucyshyn, D., Jaeger, K.E., Alos, E., Alvey, E., Harberd, N.P., and Wigge, P.A. (2012). Transcription factor PIF4 controls the thermosensory activation of flowering. Nature **484,** 242-245.

Kumimoto, R.W., Zhang, Y., Siefers, N., and Holt, B.F., 3rd. (2010). NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. The Plant journal : for cell and molecular biology.

Laine, J.P., Singh, B.N., Krishnamurthy, S., and Hampsey, M. (2009). A physiological role for gene loops in yeast. Genes & development **23,** 2604-2609.

Laubinger, S., Fittinghoff, K., and Hoecker, U. (2004). The SPA quartet: a family of WD-repeat proteins with a central role in suppression of photomorphogenesis in arabidopsis. The Plant cell **16,** 2293-2306.

Laubinger, S., Marchal, V., Le Gourrierec, J., Wenkel, S., Adrian, J., Jang, S., Kulajta, C., Braun, H., Coupland, G., and Hoecker, U. (2006). Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. Development **133,** 3213-3222.

Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., and Ahn, J.H. (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes & development **21,** 397-402.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell **128,** 707-719.

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., and Yu, H. (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Developmental cell **15,** 110-120.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326,** 289-293.

Lifschitz, E., Eviatar, T., Rozman, A., Shalit, A., Goldshmidt, A., Amsellem, Z., Alvarez, J.P., and Eshed, Y. (2006). The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. Proceedings of the National Academy of Sciences of the United States of America **103,** 6398-6403.

Lin, M.K., Belanger, H., Lee, Y.J., Varkonyi-Gasic, E., Taoka, K., Miura, E., Xoconostle-Cazares, B., Gendler, K., Jorgensen, R.A., Phinney, B., Lough, T.J., and Lucas, W.J. (2007). FLOWERING LOCUS T protein may act as the long-distance florigenic signal in the cucurbits. The Plant cell **19,** 1488-1506.

Liu, H., Yu, X., Li, K., Klejnot, J., Yang, H., Lisiero, D., and Lin, C. (2008). Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis. Science **322,** 1535-1539.

Liu, L., Liu, C., Hou, X., Xi, W., Shen, L., Tao, Z., Wang, Y., and Yu, H. (2012). FTIP1 is an essential regulator required for florigen transport. PLoS biology **10,** e1001313.

Louwers, M., Splinter, E., van Driel, R., de Laat, W., and Stam, M. (2009). Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). Nature protocols **4,** 1216-1229.

Louwers, M., Bader, R., Haring, M., van Driel, R., de Laat, W., and Stam, M. (2009). Tissue- and expression level-specific chromatin looping at maize b1 epialleles. The Plant cell **21,** 832-842.

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

Mathieu, J., Warthmann, N., Kuttner, F., and Schmid, M. (2007). Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. Current biology : CB **17,** 1055-1060.

Mathieu, J., Yant, L.J., Murdter, F., Kuttner, F., and Schmid, M. (2009). Repression of flowering by the miR172 target SMZ. PLoS biology **7,** e1000148.

McNabb, D.S., and Pinto, I. (2005). Assembly of the Hap2p/Hap3p/Hap4p/Hap5p-DNA complex in Saccharomyces cerevisiae. Eukaryotic cell **4,** 1829-1839.

Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J., and Coupland, G. (2005). Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. The Plant cell **17,** 2255-2270.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., Hale, C.J., Feng, W., Michaels, S.D., Frand, A.R., Pellegrini, M., Dekker, J., Kim, J.K., and Jacobsen, S.E. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science **336,** 1448-1451.

Nakazono, M., Qiu, F., Borsuk, L.A., and Schnable, P.S. (2003). Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. The Plant cell **15,** 583-596.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, C., Moras, D., Romier, C., Bolognesi, M., and Mantovani, R. (2013). Sequence-Specific Transcription Factor NF-Y Displays Histone-like DNA Binding and H2B-like Ubiquitination. Cell **152,** 132-143.

Navarro, C., Abelenda, J.A., Cruz-Oro, E., Cuellar, C.A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature **478,** 119-122.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., and Bergelson, J. (2005). The pattern of polymorphism in Arabidopsis thaliana. PLoS biology **3,** e196.

Park, D.H., Somers, D.E., Kim, Y.S., Choy, Y.H., Lim, H.K., Soh, M.S., Kim, H.J., Kay, S.A., and Nam, H.G. (1999). Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. Science **285,** 1579-1582.

Piazza, P., Bailey, C.D., Cartolano, M., Krieger, J., Cao, J., Ossowski, S., Schneeberger, K., He, F., de Meaux, J., Hall, N., Macleod, N., Filatov, D., Hay, A., and Tsiantis, M. (2010). Arabidopsis thaliana leaf form evolved via loss of KNOX expression in leaves in association with a selective sweep. Current biology : CB **20,** 2223-2228.

Pin, P.A., and Nilsson, O. (2012). The multifaceted roles of FLOWERING LOCUS T in plant development. Plant, cell & environment **35,** 1742-1755.

Pin, P.A., Benlloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J.J., and Nilsson, O. (2010). An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. Science **330,** 1397-1400.

Pruneda-Paz, J.L., and Kay, S.A. (2010). An expanding universe of circadian networks in higher plants. Trends in plant science **15,** 259-265.

Puente, P., Wei, N., and Deng, X.W. (1996). Combinatorial interplay of promoter elements constitutes the minimal determinants for light and developmental control of gene expression in Arabidopsis. The EMBO journal **15,** 3732-3743.

Putterill, J., Laurie, R., and Macknight, R. (2004). It's time to flower: the genetic control of flowering time. BioEssays : news and reviews in molecular, cellular and developmental biology **26,** 363-373.

Putterill, J., Robson, F., Lee, K., Simon, R., and Coupland, G. (1995). The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell **80,** 847-857.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS letters **579,** 909-915.

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

Sawa, M., and Kay, S.A. (2011). GIGANTEA directly activates Flowering Locus T in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America **108,** 11698-11703.

Sawa, M., Nusinow, D.A., Kay, S.A., and Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science **318,** 261-265.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nature reviews. Genetics **8,** 9-22.

Schwartz, C., Balasubramanian, S., Warthmann, N., Michael, T.P., Lempe, J., Sureshkumar, S., Kobayashi, Y., Maloof, J.N., Borevitz, J.O., Chory, J., and Weigel, D. (2009). Cis-regulatory changes at FLOWERING LOCUS T mediate

natural variation in flowering responses of Arabidopsis thaliana. Genetics **183,** 723-732, 721SI-727SI.

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A., and Coupland, G. (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes & development **20,** 898-912.

Sharbel, T.F., Haubold, B., and Mitchell-Olds, T. (2000). Genetic isolation by distance in Arabidopsis thaliana: biogeography and postglacial colonization of Europe. Molecular ecology **9,** 2109-2118.

Sheldon, C.C., Rouse, D.T., Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (2000). The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC). Proceedings of the National Academy of Sciences of the United States of America **97,** 3753-3758.

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

Simon, R., Igeno, M.I., and Coupland, G. (1996). Activation of floral meristem identity genes in Arabidopsis. Nature **384,** 59-62.

Simonis, M., Kooren, J., and de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. Nature methods **4,** 895-901.

Singh, B.N., Ansari, A., and Hampsey, M. (2009). Detection of gene loops by 3C in yeast. Methods **48,** 361-367.

Song, Y.H., Smith, R.W., To, B.J., Millar, A.J., and Imaizumi, T. (2012). FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. Science **336,** 1045-1049.

Stamm, P., and Kumar, P.P. (2010). The phytohormone signal network regulating elongation growth during shade avoidance. Journal of experimental botany **61,** 2889-2903.

Steiner, F.A., Talbert, P.B., Kasinathan, S., Deal, R.B., and Henikoff, S. (2012). Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome research **22,** 766-777.

Strange, A., Li, P., Lister, C., Anderson, J., Warthmann, N., Shindo, C., Irwin, J., Nordborg, M., and Dean, C. (2011). Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in Arabidopsis accessions. PloS one **6,** e19949.

Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature **410,** 1116-1120.

Takada, S., and Goto, K. (2003). Terminal flower2, an Arabidopsis homolog of heterochromatin protein1, counteracts the activation of flowering locus T by constans in the vascular tissues of leaves to regulate flowering time. The Plant cell **15,** 2856-2865.

Tamaki, S., Matsuo, S., Wong, H.L., Yokoi, S., and Shimamoto, K. (2007). Hd3a protein

is a mobile flowering signal in rice. Science **316,** 1033-1036.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic acids research **38,** 8164-8177.

Taoka, K.-i., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., Yamaguchi, M., Nakashima, C., Purwestri, Y.A., and Tamaki, S. (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature **476,** 332-335.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Critical reviews in biochemistry and molecular biology **41,** 105-178.

Tiwari, S.B., Shen, Y., Chang, H.C., Hou, Y., Harris, A., Ma, S.F., McPartland, M., Hymus, G.J., Adam, L., Marion, C., Belachew, A., Repetti, P.P., Reuber, T.L., and Ratcliffe, O.J. (2010). The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. The New phytologist **187,** 57-66.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Molecular cell **10,** 1453-1465.

Truernit, E., and Sauer, N. (1995). The promoter of the Arabidopsis thaliana SUC2 sucrose-H+ symporter gene directs expression of beta-glucuronidase to the phloem: evidence for phloem loading and unloading by SUC2. Planta **196,** 564-570.

Tsuji, H., Taoka, K., and Shimamoto, K. (2011). Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. Current opinion in plant biology **14,** 45-52.

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annual review of plant biology **59,** 573-594.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS genetics **3,** e86.

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

van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L.A., Dekker, J., and Lander, E.S. (2010). Hi-C: a method to study the three-dimensional architecture of genomes. Journal of visualized experiments : JoVE.

Wang, J.W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. Cell **138,** 738-749.

Wang, J., Hopkins, C.J., Hou, J., Zou, X., Wang, C., Long, Y., Kurup, S., King, G.J., and Meng, J. (2012). Promoter variation and transcript divergence in Brassicaceae lineages of FLOWERING LOCUS T. PloS one **7,** e47127.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., and Coupland, G. (2006). CONSTANS and the CCAAT box binding complex share a

functionally important domain and interact to regulate flowering of Arabidopsis. The Plant cell **18,** 2971-2984.

Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., and Weigel, D. (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. Science **309,** 1056-1059.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PloS one **2,** e718.

Wu, G., Park, M.Y., Conway, S.R., Wang, J.W., Weigel, D., and Poethig, R.S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. Cell **138,** 750-759.

Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M., and Araki, T. (2005). TWIN SISTER OF FT (TSF) acts as a floral pathway integrator redundantly with FT. Plant & cell physiology **46,** 1175-1189.

Yoo, S.Y., Kardailsky, I., Lee, J.S., Weigel, D., and Ahn, J.H. (2004). Acceleration of flowering by overexpression of MFT (MOTHER OF FT AND TFL1). Molecules and cells **17,** 95-101.

Yoo, S.J., Chung, K.S., Jung, S.H., Yoo, S.Y., Lee, J.S., and Ahn, J.H. (2010). BROTHER OF FT AND TFL1 (BFT) has TFL1-like activity and functions redundantly with TFL1 in inflorescence meristem development in Arabidopsis. The Plant journal : for cell and molecular biology **63,** 241-253.

Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S., and Ahn, J.H. (2005). CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis. Plant physiology **139,** 770-778.

Zhang, W., Zhang, T., Wu, Y., and Jiang, J. (2012). Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. The Plant cell **24,** 2719-2731.

Zhang, X., Germann, S., Blus, B.J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S.E. (2007). The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. Nature structural & molecular biology **14,** 869-871.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS biology **5,** e129.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., and Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nature genetics **38,** 1341-1347.

7. Abbreviations

General abbreviations

:	fused to (in the context of gene fusion constructs)
::	under the control of (in the context of promoter-gene constructs)
-	minus, not present
%	percentage
°C	degrees Celsius
3'	three prime end of DNA fragment
3C	Chromosome Conformation Capture
35S	promoter of the Cauliflower Mosaic virus
5'	five prime end of DNA fragment
5C	Chromosome Conformation Capture Carbon Copy
μ	micro
A	Adenine
Arabidopsis	*Arabidopsis thaliana*
BAC	Bacterial artificial chromosome
bHLH	basic helix loop helix
Bp	base pair
CCT	CONSTANS, CO-like, and TOC1 domain
CC	companion cells
cDNA	complementary DNA
Col-0	*Arabidopsis thaliana* ecotype Columbia-0
ChIP	Chromatin Inmunoprecipitation
DNA	desoxyribonucleic acid
dNTP	deoxyribonucleic triphosphate

Drosohila	*Drosophila melanogaster*
E. coli	*Escherichia coli*
FA	formaldehyde
FACS	fluorescence-activated cell sorting
FAC	florigen activation complex
g	gram
GA	gibberellic acid
GTFs	general transcription factors
GM	½ strength Murashige and Skoog medium
h	hour
H3	histone 3
H3K4me1	mono-methylated lysine 4 at histone 3
H3K4me3	tri-methylated lysine 4 at histone 3
H3K27me3	tri-methylated lysine 27 at histone 3
H3K27ac	acetylated lysine 27 at histone 3
H3K36ac	acetylated lysine 36 at histone 3
HFDs	histone-fold domains
HS	DNase I hypersensitive sites
INTACT	Isolation of Nuclei Tagged in Specific Cell Types
k	kilo
kb	kilobase pair
INTACT	Isolation of Nuclei Tagged in Specific Cell Types
l	liter
LCM	laser capture microdissection
LD	Long day
L*er*	Landsberg *erecta*

IM	inflorescence meristem
M	molar (mol/l)
m	milli
min	minute
mol	mole
mRNA	messenger RNA
n	nano
nt	nucleotide
N-	amino-terminal
NIB	Nuclei isolation buffer
NOSmin	minimal promoter of the *NOS* gene
NTF	nuclear targeting fusion protein
p	pico
PIC	pre-initiation complex
PCR	polymerase chain reaction
PEBP	phosphatidyl ethanolamine binding domain protein
pH	negative logarithm of proton concentration
PPT	Phosphinotricin
PRC	Polycomb repressive complex
rRNA	ribosomal RNA
RNA	ribonucleic acid
RNase	ribonuclease
RT-PCR	reverse transcription PCR
SAM	shoot apical meristem
SD	short day
SE	standard error for statistical analysis

SE	sieve elements in phloem
SUC2p	promoter of the plasma-membrane sucrose-H+ symporter gene
TBP	TATA box binding protein
TD	topological domains
TSS	transcription start site
TSC	translation start codon
T-DNA	transferred DNA
UTR	untranslated region
WT	wild type
ZT	zeitgeber time

Abbreviations of gene and protein names

ACT2	ACTIN2
ATC	ARABIDOPSIS THALIANA CENTRORADIALIS
AP1	APETALA 1
BFT	BROTHER OF FT AND TFL1
C+A	*Block C + Block A*
CCA1	*C*IRCADIAN CLOCK-ASSOCIATED 1
CDF1	CYCLING DOF FACTOR 1
CmGAS1	*GAS1* (*GALACTINOL SYNTHASE 1*) from *Cucumis melo*
CIB1	Cryptochrome-interacting bHLH 1
CLF	CURLY LEAF
CO	CONSTANS
COL	CO-LIKE
COP1	CONSTITUTIVE PHOTOMORPHOGENESIS 1

CRY1	Cryptochrome 1
CRY2	Cryptochrome 2
dMp	deletion of middle part region of *FT* promoter
E(Z)	Enhancer of Zeste *(Drosophila melanogaster)*
ESC	Extra sex combs *(Drosophila melanogaster)*
FKF1	FLAVIN-BINDING, KELCH REPEAT, F-BOX 1
FLC	FLOWERING LOCUS C
FRI	FRIGIDA
FT	FLOWERING LOCUS T
GFP	Green Fluorescent Protein
GI	GIGANTEA
GUS	β-glucuronidase
HAP	HEME ACTIVITOR PROTEIN
Hd1	Heading date 1
LFY	LEAFY
LHP1	LIKE HETEROCHOMATON PROTEIN1 (also known as TFL2)
LUC	Luciferase
MEA	MEDEA
MFT	MOTHER OF FT AND TFL1
MSI	Multicopy suppressor of Ira *(Drosophila melanogaster)*
NF-Y	NUCLEAR FACTOR-Y
NOS	*Nopaline Syntheses*
PcG	Polycomb group genes
PHYA	Phytochrome A
PHYB	Phytochrome B
PIF4	PHYTOCHROME-INTERACTING FACTOR 4

Sce	Sex combs extra *(Drosophila melanogaster)*
SFT	SINGLE FLOWER TRUSS
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1
SPA1	SUPPRESSOR OF PHYA-105
STM	SHOOT MERISTEMLESS
SVP	SHORT VEGETATIVE PHASE
SWN	SWINGER
TEM1	TEMPPANILLO 1
TFL1	TERMINAL FLOWER 1
TFL2	TERMINAL FLOWER 2 (also known as LHP1)
TOC1	TIMING OF CAB EXPRESSION 1
TSF	TWIN SISTER OF FT
YFP	Yellow Fluorescent Protein

Acknowledgements

I started to study in MPIPZ since December 10, 2009. Three years passed, I would like to thank all of those who helped and supported me during my doctoral study. I thank to

Dr. Franziska Turck for her ideas, her discussion, her supervision, her support to trying new technology, her patience and persistence in science that has been a great example to me

Prof. George Coupland for giving me opportunity to do my PhD in his department, his interesting discussion and suggestion and his influence on me how to do good science

Prof. Ute Höcker for being my second examiner and for extending my thesis writing

Prof. Martin Hülskamp for being the head of my thesis committee

Everybody in the Turck group for good times, a great working atmosphere, excellent Thursday group seminar & Friday Journal Club and every-year-Xmas diner especially I want to thank **Jessica Adrian** and **Sara Farrona**, for their great advice during my beginning of PhD study, **Ben Hartwig** for his support and critical reading of my manuscript and his vivid "Dragon" in the stage, **Theo Zografou** for his suggestions, discussions, his unique laughing and great salad, **Tingting Ning** for collaboration and being member of "FT-subgroup" and **Petra Taenzler** for her perfect art of work on cloning and prepare *ft* plants

Elisa De Ansorena for her support of growing transgenic plants and measuring flowering

Till Baar for his bachelor training here and the funny of costume in carnival in MPIPZ

Samson Simon for his critical reading of my thesis and sharing materials

Liron Sarid-Krebs for her help of western blot and providing materials

Dr. Korbinian Schneeberger for nice discussion

Elmon Schmelzer for introducing me to the electron microscopy

Dr. Olof Persson for his great organization of IMPRS retreat three times

Frau Elke Bohlscheid for her help and support

Everybody in the Coupland lab for a great atmosphere and wonderful cake time

Chinese community in MPIPZ for all your discussion, help, wonderful time of tea-table and celebrating our 2013 "sneak" Chinese New Year together

My wife Runan Zou for counting leaves number and looking after plants during her holiday in Kolen, her understanding & support, her correction of my standard language and her love

My family for your help, your support, weekend call and your love always on my side

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit -einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. George Coupland betreut worden.

Köln, 04. März 2013

Liangyu Liu

Lebenslauf

Persönliche Angaben

Name:	Liangyu Liu (刘良玉)
Geburtsdatum:	Dec.30.1984
Geburtsort:	Weifang, China
Nationalität:	Chinesisch
Familienstand:	verheiratet
E-mail:	liu@mpipz.mpg.de or liuliangyu027@gmail.com

Ausbildung

Seit 12/2009: Promotions-Studium an der International Max-Planck Research School (IMPRS) am Max-Planck Institut für Züchtunsforschung (MPIZ) in Köln

07/2007-12/2009: Chinese Academy of Sciences (CAS)

07/2003-06/2007: Huazhong Agricultural University (HZAU), Wuhan, China

07/2000-06/2007: Weifang No.7 middle school