PHYLOGENETIC AND FUNCTIONAL CHARACTERIZATION OF COTTON (*Gossypium hirsutum*)

CENTRORADIALIS/TERMINAL FLOWER1/SELF-PRUNING GENES

Sarah F. Prewitt, A.S., B.A., B.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2017

APPROVED:

Brian Ayre, Major Professor
Rajeev Azad, Committee Member
Kent Chapman, Committee Member
Rebecca Dickstein, Committee Member
Amanda Wright, Committee Member
Art Goven, Chair of the Department of Biological
    Sciences
Su Gao, Dean of the College of Science
Victor Prybutok, Dean of the Toulouse Graduate
    School

Prewitt, Sarah F. *Phylogenetic and Functional Characterization of Cotton (*Gossypium hirsutum*) CENTRORADIALIS/TERMINAL FLOWER1/SELF-PRUNING Genes*. Doctor of Philosophy (Biochemistry and Molecular Biology), December 2017, 171 pp., 14 tables, 20 figures, references, 160 titles.

Plant architecture is an important agronomic trait driven by meristematic activities. Indeterminate meristems set repeating phytomers while determinate meristems produce terminal structures. The *CENTRORADIALIS/TERMINAL FLOWER1/SELF PRUNING* (*CETS*) gene family modulates architecture by controlling determinate and indeterminate growth. Cotton (*G. hirsutum*) is naturally a photoperiodic perennial cultivated as a day-neutral annual. Management of this fiber crop is complicated by continued vegetative growth and asynchronous fruit set. Here, cotton *CETS* genes are phylogenetically and functionally characterized. We identified eight *CETS* genes in diploid cotton (*G. raimondii* and *G. arboreum*) and sixteen in tetraploid *G. hirsutum* that grouped within the three generally accepted *CETS* clades: *FLOWERING LOCUS T* (*FT*)-like, *TERMINAL FLOWER1/SELF PRUNING* (*TFL1/SP*)-like, and *MOTHER OF FT AND TFL1* (*MFT*)-like. Over-expression of *SINGLE FLOWER TRUSS* (*GhSFT*), the ortholog to Arabidopsis *FT*, accelerates the onset of flowering in Arabidopsis Col-0. In mutant rescue analysis, this gene driven by its native promoter rescues the *ft-10* late flowering phenotype. *GhSFT* upstream sequence was used to drive expression of the *uidA* reporter gene. As anticipated, GUS accumulated in the vasculature of Arabidopsis leaves. Cotton has five *TFL1*-like genes, all of which delay flowering when ectopically expressed in Arabidopsis; the strongest phenotypes fail to produce functional flowers. Three of these genes, *GhSP*, *GhTFL1-L2,* and *GhBFT-L2*, rescue the early flowering *tfl1-14* mutant phenotype. *GhSP$_{pro}$:uidA* promoted GUS

activity specifically in plant meristems; whereas, other *GhTFL1*-like promoters predominately drove GUS activities in plant vascular tissues. Finally, analysis of *Gossypium CETS* promoter sequences predicted that *GhSFT, GhSP, GhTFL1-L1, GhTFL1-L2* and *GhBFT-L2* are regulated by transcription factors involved in shoot and flowering development. Analysis of cotton's two *MFT* homologs indicated that neither gene functions to control shoot architecture. Our results emphasize the functional conservation of members of this gene family in flowering plants and also suggest this family as targets during artificial selection of domestication.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AG	AGAMOUS
AP	APETELA
ARF	AUXIN RESPONSE FACTOR
ARR	ARABIDOPSIS RESPONSE REGULATOR
ATC	CENTRORADIALIS
BFT	BROTHER OF FT AND TFL1
bp	base pair
BPC	BASIC PENTACYSTIENE
bZIP	Basic Leucine Zipper
CaMV	Cauliflower Mosaic Virus
CDF	CELL GROWTH DEFECT FACTOR
CETS	CENTRORADIALIS/TERMINAL FLOWER1/SELF-PRUNING
CFU	Colony forming unit
CLCrV	Cotton Leaf Crumple Virus
CO	CONSTANS
CUC	CUP SHAPED COTYLEDON
Dof	DNA-binding One Zinc Finger
DPG	Days past germination
ERF	ETHYLENE RESPONSE FACTOR
EV	Empty vector
FT	FLOWERING LOCUS T

FUL	FRUITFUL
GA	Gibberellic acid
GAS	Galactinol Synthase
GFP	Green Fluorescent Protein
GI	GIGANTEA
GO	Gene ontology
Hd1	Heading date 1
Hd3a	Heading date 3a
ISP	Ion Sphere Particle
JTT	Jones-Taylor-Thornton
kb	kilobase
LD	Long-day
LFY	LEAFY
LiAC	Lithium acetate
MFT	MOTHER OF FT AND TFL1
NAM	NO APICAL MERISTEM
NLP	Nin-like PROTEIN
NJ	Neighbor-Joining
nt	nucleotide
ORF	Open reading frame
PCR	Polymerase chain reaction
PEBP	Phosphatidylethanolamine Binding Protein

PEG	Polyethylene glycol
PGM	Personal Genome Machine
PI	PISTILLATA
RAX	REGULATOR OF AXILLARY MERISTEMS
SAM	Shoot apical meristem
SD	Short-day
SFT	SINGLE FLOWER TRUSS
SOC1	SUPPERSSOR OF OVEREXPRESSION OF CONSTANS1
SP	SELF PRUNING
TCP	TEONSINTE-BRANCHED1/CYCLOIDEA/PFC
TF	Transcription factor
TFBS	Transcription factor binding site
TFL1	TERMINAL FLOWER1
TRV	Tobacco Rattle Virus
TSF	TWIN SISTER OF FT
U.S.	United States
USDA	United State Department of Agriculture
UTR	Untranslated region
VIGS	Virus-Induced Gene-Silencing
VIP	VIRE2 INTERACTING PROTEIN
WOX	WUSCHEL Homeobox
WT	Wild-type

WUS   WUSCHEL

X-gluc   5-Bromo-4-chloro-3-indol-β-D-glucuronic acid

CHAPTER 1

INTRODUCTION

1.1   Cotton is a Perennial Plant Cultivated as an Annual Row Crop

Cotton (*Gossypium* ssp) is the world's principal crop for fiber production in the textiles

industry accounting for around 25 percent of total world fiber use

(www.ers.usda.gov/topics/crops/cotton-wool.aspx). The USDA estimates 2017 worldwide

cotton production to be 115.4 million bales (one bale = 490 lbs, Meyer and MacDonald, 2015).

While hundreds of countries contribute to worldwide cotton trade, the five top producers are

projected contribute 76 percent of total production. The largest producer, India is predicted to

produce 25 percent of total production, or 29 million bales. The U.S. is the third largest

producer and is projected to produce $8.4 billion of fiber in 2017 (19 million bales, 490 lbs/bale,

90 cents/lb)(Meyer, 2017). The U.S. cotton industry produces over 125,000 jobs industry-wide

and provides products and services tallying over $21 billion annually

(www.ers.usda.gov/topics/crops/cotton-wool.aspx). As a crop, cotton is primarily grown for its

fibers whose cell walls are cellulose rich and account for 30 percent of the seed coat formed

around the cotton seed. Besides its fiber value, the embryo produces oils and proteins that are

processed, providing food-grade oils and livestock meal products and increasing the value of

the cotton seed overall. Given this overall view of cotton seed economics, improving the yield

of cotton is likely to be of great significance to the cotton industry.

Gossypium ssp. is a woody, perennial genus originating in tropical regions of the world

and is made up of approximately 50 diploid and tetraploid species (Small and Wendel, 2000).

Cotton domestication of tetraploids *G. hirsutum* and *G. barbadense* has dramatically altered its

architecture offering an advantageous system for studying the genetic and molecular

mechanisms establishing plant architecture. Wild *G. hirsutum* lines are tall, lanky, and short-day

(SD) photoperiodic. Architecture of these plants is driven by the apical dominance of the main

stem during the long-days (LDs) of the growing season, followed by production of a few fruiting

branches late in the growing season when days shorten. Domesticated lines of *G. hirsutum* and

G. barbadense, grown as annual row crops, are shorter, bushier plants in which photoperiodic

control of flowering was lost. These plants produce a first fruiting branch as early as node five

and begin flowering approximately sixty days past germination (DPG) (Oosterhuis, 1990). In

domesticated cotton crops, the main stem of the plant exhibits monopodial growth in which

the shoot apical meristem (SAM) remains indeterminate while leaves, axillary buds, and

internodes form at its flanks throughout the life of the plant. Fruiting branches are produced

from the axillary buds of the main stem from node five and upward. A fruiting branch is a

sympodial, cymose inflorescence (Gore, 1935), whose apical meristem (inflorescence apical

meristem) produces a single subtending leaf and an inflorescence axillary meristem before

converting to a floral meristem, resulting in production of a flower and finally the cotton boll.

This growth pattern is repeated by the newly formed inflorescence axillary meristem and all

subsequent inflorescence axillary meristems of the fruiting branch and is responsible for the

reiterative sympodial growth habit that produces a zig-zag pattern as opposed to the straight

appearance of the main stem (Fig 1.1) (McGarry and Ayre, 2012a).

Both wild and domesticated cotton are perennial plants. They balance reproductive

growth with maintenance of vegetative growth for parental survival. The result of this growth

strategy is asynchronous flowering and fruit set. This growth plan is significantly different from

2

that of annuals that sacrifice parental survival to concentrate end-of-season resources into reproductive growth ensuring next generation success. In developed countries, domesticated cotton is cultivated as an annual row crop, where its perennial nature is less compatible with highly-mechanized harvesting techniques that are better suited for synchronized annual life strategies. Photoperiodism and asynchronous flowering and fruiting complicate breeding and crop management which can compromise fiber yield and quality (Oosterhuis, 1990).

Continued vegetative growth after the onset of flowering can result in excessive vegetative growth resulting in production problems of fruit abortion, delayed maturity, boll rot, and harvesting difficulties (Jost et al., 2006). Due to resource availability, the earliest forming cotton bolls produce the highest quality, longest, and strongest fibers. Primary fruiting positions (i.e. the first fruiting position of the sympodial branch) of nodes seven to twenty account for 60-70 percent of total plant yield. All other bolls produced account for the remaining plant yield and are generally of lower quality fiber than those closer in to the main stem (Oosterhuis, 1990; Jost et al., 2006). Mechanized harvesting practices do not allow separation of low-quality from higher quality fruiting bolls. Furthermore, since the highest quality bolls are produced first on the plant at lower positions of the main stem, quality is reduced as the canopy of vegetative growth above creates an unfavorable humid environment and hinders penetration of insecticides to these lower bolls, leaving them at risk for boll rot and damage by chewing insects, aphid honey dew, or field dust and debris. Growth regulators during season and defoliants in preparation for harvesting are used to control this vegetative growth (Oosterhuis, 1990), and result in increased costs. These treatments also have adverse environmental consequences. Management of cotton's perennial growth habit through genetic manipulation

3

has been a long-standing goal of the breeding community to benefit cotton production by increased yields, reduced costs, and better crop management.

1.2   Plant Architecture is Determined by Activities of Shoot Meristems

Plant architecture, or the shape of the shoot system, varies through the plant kingdom including unicellular, colonial, siphonous, and filamentous-multicellular body plans. Within the group of filamentous-multicellular plants, angiosperms reside as a monophyletic clade having a three-dimensional, tubular shape shared with all vascular plants. Remarkably, from this rather simple body plan, an immense collection of diverse architectures between species evolved (Sussex & Kerk, 2001). In addition to the array of architectures found in this class of plants, individuals can alter their body plans based upon environmental stimuli. Underlying these differences in architecture is the activities of shoot meristems.

Plant meristems, pools of undifferentiated cells from which growth occurs, are either indeterminate or determinate. Indeterminate meristems contain a pool of undifferentiated cells while discriminate tissues and organs (stems, leaves, and buds) develop below and to the flanks of the undifferentiated population producing a monopodial growth pattern. Determinate meristems lose this pool of undifferentiated cells and commonly terminate into inflorescence or floral structures on the shoot. When an apical meristem terminates in this way, the closest axillary meristem is relieved of apical dominance and continues a species-specific body plan. This phenomenon constitutes the sympodial growth pattern.

Plant architecture is an amalgam of agronomically important traits including the position of branches, leaf shape, and the timing and placement of reproductive structures.

Manipulation of mechanisms underlying this trait can have a significant impact on crop plant success. A famous example of this is the large increases in yield upon introduction of semi-dwarf varieties of wheat and rice during the Green Revolution. These crops produce shorter, sturdier stalks than their predecessors, which were susceptible to lodging losses. Grain yields were also increased at the expense of stalk straw biomass (Peng et al., 1999). In the past fifteen years, genes responsible for regulating the fate of meristems have been characterized through extensive research. Since the activities of shoot meristems determine the position and timing of harvestable organs, these genes are targets for manipulation aimed to increase plant productivity.

1.3   FT Promotes Determinate Growth in Response to Photoperiod

Julius Sachs first brought forth the idea of a long-distance signal for floral promotion. In 1865 working with *Tropaeolum majus* and *Ipomoea purpurea*, he concluded that when light-exposed, leaves produce a flower-forming substance (reviewed in Jan A D Zeevaart, 2006). This idea was linked to a photoperiod stimulus through observations that LD spinach was induced to flower when the leaves, but not the plant meristem, was exposed to LDs (Knott, 1934). At the same time Mikhail Chailakhyan showed graft-transmission of this stimulus, broadened this observation to other species, and devised the term florigen ('flowering morphogen') (reviewed in Chailakhyan & Krikorian, 1975). Florigen remained elusive to researchers for many years and was given the moniker: 'Holy Grail of Plant Physiology'. Through interspecies grafting experiments, it was shown that an induced stock could stimulate an uninduced scion across species, genera, and photoperiodic response types (Zeevaart, 1976). This meant that florigen

could cross graft junctions and was probably a universal trigger for floral induction in higher

plants. Identification of photoperiodic mutants in LD Arabidopsis, characterization of the

CONSTANS (*CO*)/*FLOWERING LOCUS T* (*FT*) module as the primary regulator of the

photoperiodic response, and the merging of classical physiological and genetics approaches led

to the emergence of FT as the sole or major component of florigen (Turck et al., 2008; Zeevaart,

2008; McGarry and Ayre, 2012b).

The transition of an indeterminate meristem to a determined meristem that results in

floral production can be controlled through many different pathways. There are six pathways

known to promote this transition in facultative LD Arabidopsis, including: photoperiod

(predominate pathway), ambient temperature, vernalization, aging, autonomous, and

gibberellic acid (GA) pathways (Fornara et al., 2010). The photoperiodic pathway is well-

elucidated and initiates with a perception by leaf photoreceptors that is transmitted to shoot

meristems through the phloem-mobile message of florigen. The molecular basis for this

transmission begins with leaf photoreceptors sensing day-length periods allowing for

GIGANTEA (*GI*) to activate *CO* in the minor veins of leaves (Sawa et al., 2007). In the SDs of

spring, *CO* mRNA is highest at dusk, but the CO protein is quickly degraded in dark conditions

via the ubiquitin pathway (Suárez-López et al., 2001; Jang et al., 2008; Liu et al., 2008).

However, during the LDs of summer, CO protein is stabilized and acts to activate *FT*. This

overlapping of CO protein accumulation and light conditions is known as the external

coincidence model of flowering in photoperiod plants (i.e. CO protein activity coincides with

external stimuli, light). CO drives expression of *FT* in the companion cells of minor veins in

leaves, then its product FT is translocated through the phloem to shoot meristems (Samach et

al., 2000; An et al., 2004; Wigge et al., 2005; Corbesier et al., 2007; Jaeger and Wigge, 2007).

Spatially, it is in shoot meristems where the photoperiodic pathway converges with other

flowering pathways to act upon determinacy factors. FT interacts with Basic Leucine Zipper

(bZIP) transcription factor (TF) FD to activate *SUPPRESSOR OF OVEREXPRESSOR OF CONSTANS*

(*SOC1*), *APETELA 1* (*AP1*), *FRUITFUL* (*FUL*), and *LEAFY* (*LFY*). The activation of the determinacy

factors promotes flowering and the formation of floral structures (Fig 1.2, Abe et al., 2005;

Wigge et al., 2005). Research showing the movement of FT along the phloem to the shoot

meristem, where it promotes flowering, established it as the long-sought florigen (Corbesier et

al., 2007; Jaeger and Wigge, 2007).

The role of FT in LD Arabidopsis is conserved in SD plant rice, but with opposite effect—

namely, promotion of flowering in SDs and repression in LDs. Heading date 3a (Hd3a) is the rice

FT ortholog and has been shown to be a phloem-mobile floral promoter (Tamaki et al., 2007).

Heading date 1 (*Hd1*), the rice *CO* ortholog was the first identified quantitative trait locus for

the timing of flowering among different rice cultivars and similar to *CO*, encodes a B-Box zinc

finger protein (Yano et al., 2000; Izawa et al., 2003; Tsuji et al., 2011). Contrasting the

Arabidopsis *CO/FT* module, *Hd1* mRNA peaks around midnight while *Hd3a* mRNA peaks at

dawn. The dominate model is says Hd1 is an activator of *Hd3a* under SD (long nights), but

converts to a repressor under early exposure to light via a phytochrome interacting pathway.

This model is supported by rice's delayed flowering when exposed to short night breaks (Tsuji

et al., 2011). This and analyses of many *FT* homologues in other species producing floral

induction (discussed further in Chapter 3) gives solid evidence for the conservation of FT as a

floral promoter among angiosperms, but highlights that regulations of the florigenic pathway

can vary greatly.

1.4   The *CETS* Gene Family Evolved in Angiosperms to Regulate Plant Development

FT is a member of the CE<u>N</u>TRORADIALIS/<u>T</u>ERMINAL FLOWER1 (*TFL1*)/<u>S</u>ELF PRUNING

(*CETS*) gene family in angiosperms. Protein products of this gene family contain a

phosphatidylethanolamine binding domain and are alternatively referred to as PEBP proteins.

In plants, CETS family members can be classed in three broad groups: MOTHER OF FT AND

TFL1-like (MFT-like), FT-like, and TFL1-like (Carmel-Goren et al., 2003; Chardon and Damerval,

2005; Carmona et al., 2007; Danilevskaya et al., 2008). The amino acid sequences of CETS

proteins in plants are very similar, yet their functions have diverged after gene duplication

events through angiosperm lineages.

All family members appear to have activities that regulate development. MFT in

Arabidopsis functions mainly in regulation of seed germination through the abscisic acid and GA

signaling pathways (Xi and Yu, 2010). MFT-like proteins are the most ancient of the PEBP

proteins in angiosperms and similar sequences are found also in gymnosperms and lower plants

(Hedman et al., 2009). FT- and TFL-like proteins are angiosperm-specific and evolved to control

plant architecture and flowering. In general, FT-like proteins are promoters of determinate

growth (i.e. flowering); whereas, TFL1 proteins maintain indeterminate growth patterns.

In Arabidopsis TFL1 is highly abundant in shoot meristems, where similar to FT it also

interacts with FD, but in a manner which represses determinacy factors. In this way TFL1 is an

indeterminacy factor which maintains the meristem stem cell population and encourages

vegetative growth (Kobayashi, 1999; Hanano and Goto, 2011). TFL1 function antagonizes FT

(Kardailsky, 1999; Kobayashi, 1999). In Arabidopsis under LD conditions, the *tfl1* mutant

transitions to flowering early, produces fewer rosette leaves, and comprises two to three

solitary flowers that replace the main inflorescence branch (Alvarez et al., 1992). In contrast,

over-expression of this gene causes delay of flowering (Benlloch et al., 2007), and in severe

cases leafy flowers replace a normal floral structure. *TFL1* thus inhibits determinate growth.

Research fusing either a transcriptional activator or transcriptional repressor to TFL1 showed

early or delayed flowering, respectively, compared to wild-type (WT) plants. This indicates that

TFL1 affects the transcription of meristem determinate genes. In an *fd* mutant, these effects

were suppressed. Taken together with bimolecular fluorescent assays in which TFL1 was

observed to associate with FD in the nucleus, it was concluded that TFL1 represses flowering in

an FD-dependent manner (Hanano and Goto, 2011). TFL1, then, represses flowering and

maintains indeterminacy in the shoot meristem by modulating or inhibiting FT activity.

Correspondingly, the *tfl1* phenotype is observed only under inductive LD conditions when FT is

produced. Under SD conditions when FT is absent, there is nothing to modulate and *tfl1*

mutants do not display the early flowering phenotype (Alvarez et al., 1992). TFL1-like proteins

functioning to maintain indeterminate growth are also described in tomato (Pnueli et al., 1998),

soybean (Wang et al., 2015), snapdragon (Bradley et al., 1996), as well as several other plant

species.

CETS have been found to play similar roles in most crop species. Investigations in

tomato have provided the best insight into CETS influence on perennial, sympodial growth in

which vegetative and reproductive growth are balanced throughout the plant life cycle similar

to cotton development. In WT tomato the primary shoot terminates in an inflorescence after approximately nine nodes and is followed by sympodial branching. The first sympodial branch arises from the uppermost axillary meristem of the main stem, sets three compound leaves, then completes in an inflorescence like the main stem. Reiterative sympodial branches likewise, comprise three compound leaves and a terminating inflorescence (Shalit et al., 2009). In this manner, the tomato vine appears as a single axis with sequential fruiting trusses, but is actually a primary shoot followed by reiterating sympodia.

SELF PRUNING (SP), a tomato TFL1 homolog, maintains indeterminacy in shoot meristems. *sp* mutants, originally identified in 1927, are not affected in main stem development and the axis still terminates after approximately nine nodes. However, sympodial growth is quickly reduced causing a compact and determinate phenotype with near homogenous fruit set (Yeager, 1927). Identification of the *sp* determinate phenotype revolutionized the processed tomato industry. Commercially-cultivated WT (indeterminate) tomatoes, comprising high-quality successively ripening clusters of fruits, produce tomato types that are eaten fresh. These indeterminate varieties are cultivated in greenhouses where fruit is continually hand-harvested over extended periods of time and heavily pruned to maximize fruit size and quality (Saltveit, 2005). While the pruning of indeterminate varieties is essential to maintain market fresh quality, it also restricts yield (Peet, 2005). In *sp* determinate varieties, sequential sympodial shoots transition to flowering increasingly earlier. *sp* determinate plants are bushier with near uniform fruit ripening (Pnueli et al., 1998). In this manner, *sp* varieties lend themselves to once-over mechanical harvesting, increasing the yield of these varieties. Since the industry of processed tomatoes (i.e. used to make sauces, pastes, and juices) requires high yields enabled

by mechanical harvesting for economic viability, *sp* varieties have come to control the processed tomato industry (Saltveit, 2005).

SINGLE FLOWER TRUSS (SFT), tomato's FT ortholog, is phloem-mobile and promotes flowering; however, not in a day-length (photoperiod) dependent manner (Pnueli et al., 1998; Carmel-Goren et al., 2003; Lifschitz et al., 2006). Overexpression of *SFT* in WT tomato leads to early primary-shoot termination, but shows no impact on the regularity of sympodial growth patterns (Lifschitz et al., 2006; Shalit et al., 2009). Consequently, high SFT/SP in meristems is realized by either *SFT* over-expression or by *sp* mutation; however, *SFT* over-expression has greater impact on the primary axis while the *sp* mutation has greater effect in sympodial shoots. It is likely that the SFT/SP ratio model for balancing determinate and indeterminate growth is universal in flowering plants. Study of these tomato genes has been used to establish a paradigm that meristem state (indeterminate vs. determinate) is regulated by a ratio of the two factors, SFT (FT)/SP (TFL1). Local meristem ratios of SFT/SP have been hypothesized and experimentally shown to influence shoot development so that vegetative and reproductive growth can be balanced throughout the life cycle of the plant. In this model, a high SFT/SP ratio in the meristem would cause determination and result in a terminating structure (flower), whereas low SFT/SP would confer indeterminacy and continued production of vegetative structures from the meristem (Fig 1.3, Pnueli et al., 1998; Lifschitz et al., 2006; Shalit et al., 2009). SFT (FT) and SP (TFL1) have species-specific variation of expression and it is becoming clear that the expression and balance of these gene products account for the diversity of plant architecture and contribute to crop domestication.

In addition, research has demonstrated that other TFL1-like CETS in Arabidopsis, BFT

and ATC, also repress floral formation, although to a lesser extent than TFL1. *BFT* mRNA

oscillates in a diurnal expression pattern like *FT.* Expression of *BFT* is much lower in Arabidopsis

grown in SD conditions than in LDs. *BFT* over-expression delayed flowering and produced

abnormal floral organ phenotypes similar to *TFL1* over-expression. Loss-of-function mutants

and RNAi lines quickened termination of the apical and axillary inflorescences (Yoo et al., 2010).

These results established that *BFT* is a floral repressor, but its expression pattern is distinct from

either *FT* or *TFL1. ATC* mRNA oppositely shows higher expression in SD rather than LD

conditions (Yoo et al., 2010; Huang et al., 2012). An *atc-2* mutant flowered normally under LDs,

but earlier than WT plants in SD conditions (Huang et al., 2012). Grafting experiments

conducted with *Cauliflower mosaic virus 35S (CaMV35S)::ATC* stocks or WT stocks grafted with

atc-2 scions showed movement of *ATC* mRNA across the graft union suggesting that *ATC* acts

systemically to inhibit flowering (Huang et al., 2012). These results led to the conclusion that

ATC also acts redundantly to *TFL1* to maintain indeterminacy, but primarily in a SD-dependent

manner. This indicates that *ATC* could act as a hypothesized anti-florigen, a signal that

originates in the leaves like that of FT (florigen) and is translocated to promote indeterminate

growth in meristems. More recent evidence for TFL1-like CETS as anti-florigens was shown in

the species *Chrysanthemum seticuspe*. *CsAFT* was shown to be expressed in leaves

predominately under non-inductive LDs, suppress flowering when over-expressed in inductive

SD conditions, and induce late flowering via grafting (Higuchi et al., 2013). It is probable that in

other species anti-florigenic CETS are predominately expressed in leaves in response to

environmental stimuli and translocated to meristems. Additionally, redundant control of

meristem activities by the activities of several CETS broadens the paradigm of FT/TFL ratio in

determining determinate vs. indeterminate meristem state into one in which the ratios of FT-

like/TFL1-like protein ratios influence the state of meristems.

1.5   Manipulation of *CETS* Expression in Cotton Alters Plant Architecture

Efforts were initiated to understand the role of *CETS* genes in cotton plant architecture.

Arabidopsis *FT* and *Gossypium hirsutum SFT* and *SP* transiently expressed or silenced through

virus-based systems (*Cotton Leaf Crumple Virus, CLCrV,* and *Tobacco Rattle Virus, TRV*) have all

altered plant architecture of both wild, photoperiodic plants and domesticated, day-neutral

varieties. *dCLCrV::AtFT* infection in photoperiodic *G. hirsutum* uncoupled flowering from

photoperiod and produced a more determinate plant, including a change of leaf shape from

highly-lobed to lanceolate. The transient expression of *FT* in these experiments also allowed for

successful crosses between the infected ancestral plant and a domesticated day-neutral

accession (McGarry and Ayre, 2012a). In the same study, day-neutral *G. hirsutum dCLCrV::AtFT*-

infected plants carried a highly-compact, determinate architecture with near synchronized

flowering (McGarry and Ayre, 2012a).

More recently, it was shown that ratios of *GhSFT* and *GhSP* activities regulate patterns

of vegetative and reproductive branching architecture in cotton. Gain-of-function *GhSFT*

analysis, a dominate-negative construct (dCLCrV:GhSFTQ139D), and Virus-Induced Flowering

experiments established the role of *GhSFT* as a florigenic compound that regulates flowering in

photoperiodic and day-neutral *G. hirsutum*. V̲irus-I̲nduced G̲ene-s̲ilencing (VIGS) of *GhSP* in

ancestral *G. hirsutum* resulted in termination of the main stem in a floral bud by node five in all

infected plants and the conversion of all axillary meristems including the cotyledonary meristems into terminating floral buds. These results were also found in a day-neutral silenced accession and establish *GhSP* as a powerful inhibitor of sympodial growth in cotton. It was hypothesized then that a combination of *GhSFT* gain-of-function and *GhSP* silencing would synergistically enhance determinate growth in cotton. To this end, plants were co-infected and results were as predicted—severely compacted plants whose main stem quickly converted into a floral bud and all main stem nodes terminating in floral buds (McGarry et al., 2016). These results demonstrate the role of two *G. hirsutum CETS* as regulators of plant architecture and validate the continued study of *CETS* in cotton for manipulation of plant architecture to improve yield and quality in the cotton industry.

1.6   Understanding of *CETS* Function and Gene Expression Will Elucidate Their Regulation of Cotton Plant Architecture

The previously discussed experiments authenticate *GhSFT* and *GhSP* as regulators of shoot architecture in this perennial, sympodial crop, but do not provide information about their regulation, spatial and temporal expression, or how they function to provide architectural control. Also, as shown below in bioinformatics results, there are a total of eight potential *CETS* homologs. What impacts might other *CETS* have on plant architecture? As discussed above, a CETS anti-florigenic compound traveling long-distance to negatively regulate reproduction has been suggested. Might *GhSP* or another *TFL1*-like *CETS* be an anti-florigen compound in cotton contributing to its perennial nature through environmental stimuli such as perception of non-inductive photoperiod conditions? To understand the answers to these questions, and in

general, assess the role of each cotton *CETS* in the regulation of plant architecture, this study

phylogenetically and functionally characterizes cotton *CETS*. Briefly, *Gossypium CETS* homologs

are identified and named according to sequence conservation with described *CETS* genes.

Gossypium CETS genomic structures are determined. Cotton CETS are phylogenetically assigned

into three major subfamilies: MFT-like, FT-like, and TFL1-like. Functional analysis in Arabidopsis

WT and time-of-flowering mutants identified *Gossypium CETS* candidates sharing conserved

gene function with plant architecture regulators *FT* and *TFL1*. Transgenic lines harboring

CETS_{pro}:uidA demonstrate cotton *CETS* promoter's activities in vasculatures and apex tissues.

Promoter sequence analysis postulates conserved regulatory binding factors and *CETS*

placement in developmental and signaling pathways.
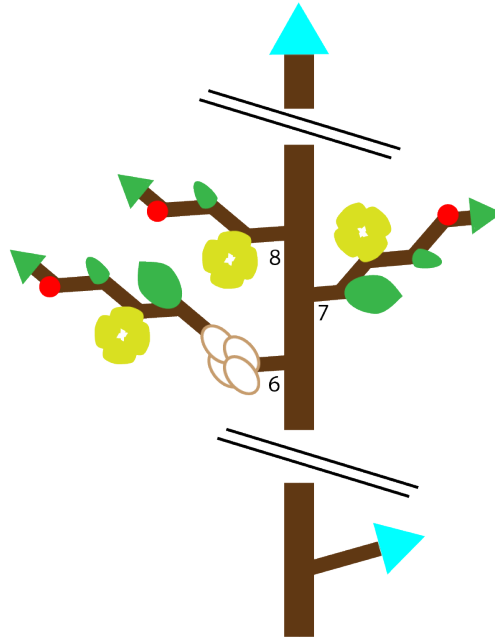
Figure 1.1: Domesticated cotton architecture. A diagram of domesticated cotton plant architecture that highlights asynchronous perennial characteristics that complicate cultivation as an annual row crop. ( ) represents the monopodial bud of the main stem and reiteration of the main stem that can occur at nodes 1-5. These meristems will continue vegetative growth throughout the life of the plant. ( ) represents a growing fruiting branch that will continue reiterative sympodial growth (growth from a series of independent initiations). Asynchronous fruit set complicates harvesting decisions: ( ) mature and open cotton bolls, ( ) developing cotton bolls, ( ) blooming flowers, ( ) immature cotton squares, and ( ) a terminating meristem that will produce fruit. Leaves are not represented. Numbers designate nodes off the main stem.
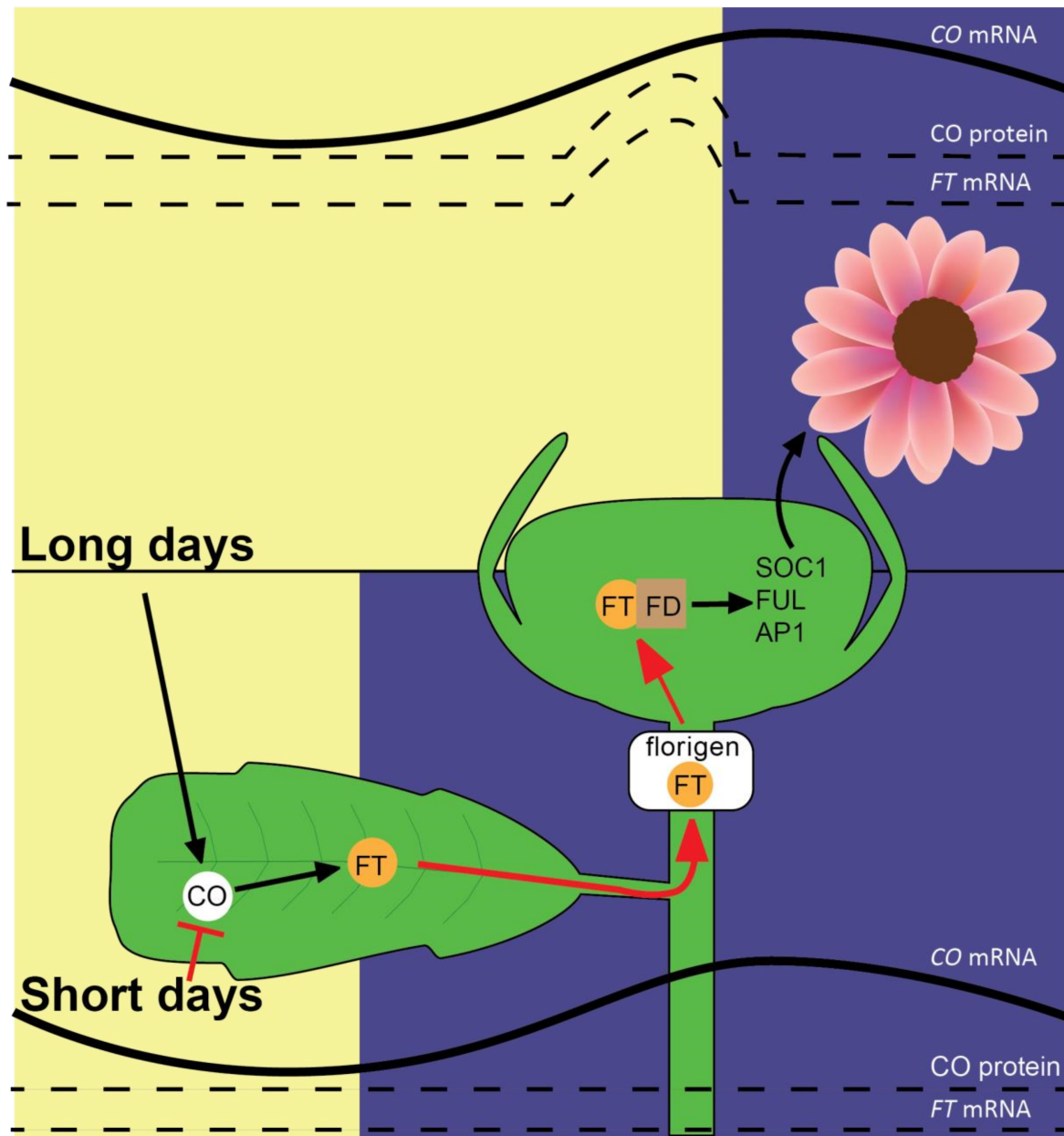
Figure 1.2 The external coincidence model in LD Arabidopsis. The predominate pathway to flowering in facultative LD Arabidopsis is photoperiodic dependent. *CO* mRNA is expressed diurnally; expression peaks late in the day in LDs and light stabilizes the CO protein which promotes expression of *FT* in phloem companion cells. Conversely, in SDs *CO* mRNA expression peaks after dark and CO protein is rapidly degraded via the ubiquitin-ligase pathway. After production in LDs, the FT protein travels via phloem from leaf companion cells into shoot meristems. In meristems, FT interacts with FD to promote the expression of floral meristem identity genes *SOC1, FUL*, and *AP1*. Downstream activations from these genes will stimulate flowering. Adapted from (McGarry and Ayre, 2012b).
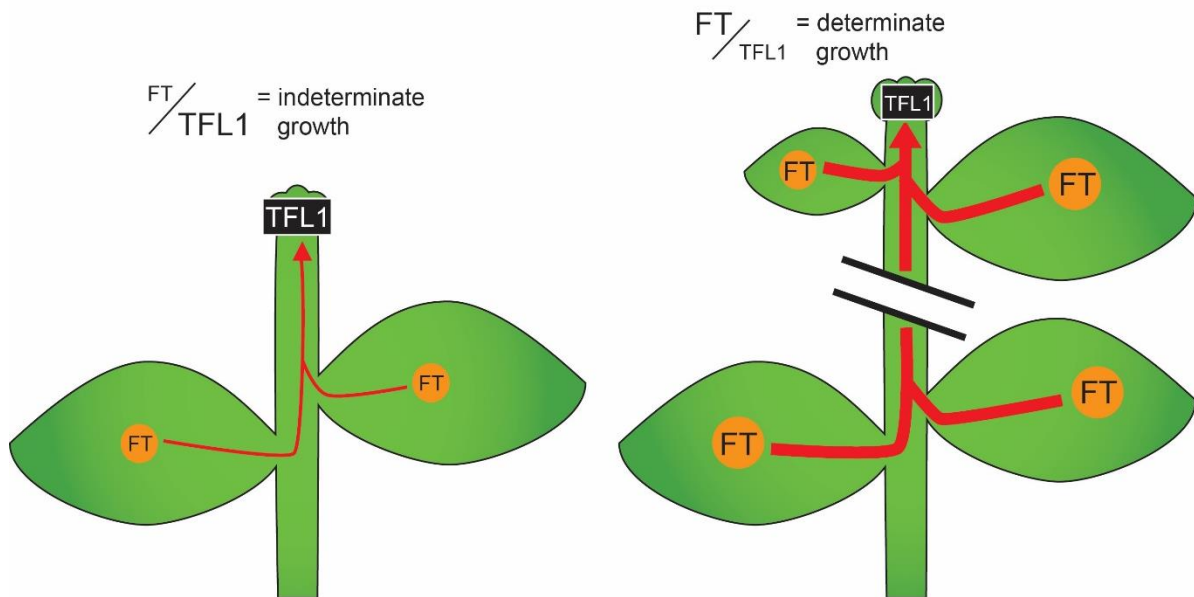
Figure 1.3 Ratios of CETS FT(SFT) and TFL1(SP) regulate meristem activities. (A) In juvenile plants

TFL1 accumulation is high in shoot meristems resulting in vegetative (indeterminate) growth.

Juvenile plants have fewer leaves and therefore FT production and transport is low. The ratio of

meristem FT/TFL1 is in turn low resulting in reiterative vegetative growth. (B) As plants age,

TFL1 accumulation is diminished, while FT production and transport are increased due to the

presence of more leaves and/or induction by environmental cues such as photoperiod. These

changes result in a higher FT/TFL1 ratio. FT function antagonizes TFL1 function and

reproductive (determinate) growth occurs. Adapted from (McGarry and Ayre, 2012b).

18

CHAPTER 2

IDENTIFICATION AND MOLECULAR EVOLUTION ANALYSIS OF THE COTTON *CETS* GENE FAMILY

2.1   Introduction

As discussed in Chapter 1, CETS are important regulators of flowering, thought to act

primarily through competition for interaction in TF complexes in meristems such that ratios of

FT-like/TFL1-like CETS control plant architecture. To test this model in cotton, *CETS* genes

require functional analysis to assess their impact on growth habit. To achieve this goal,

identification and phylogenetic studies of cotton *CETS* are required to make predictive

hypothesis for functional testing. Prior to complete genome sequencing, cotton *CETS*

identification was difficult, partly because EST resources available focused heavily on fiber and

CETS were not represented.

Economically important cotton species, *Gossypium hirsutum* and *Gossypium*

barbadense, are allotetraploids (subgenomes A_tD_t) that were independently domesticated

(Small and Wendel, 2000). Because of close sequence similarity of the A_t and D_t subgenomes,

cotton genomic complexity is rivaled only by *Brassica* in sequenced angiosperms (Paterson et

al., 2012). This genetic complexity hindered the sequencing of *G. hirsutum* or *G. barbadense*

genomes. Prior to sequencing tetraploid cotton, both diploid progenitors, *G. raimondii* (D

genome, Lin et al., 2010) and *arboreum* (A genome, Li et al., 2014), were first sequenced and

made publicly available. In 2015, two assemblies of *G. hirsutum* TM-1 genome and an assembly

of *G. barbadense* cv. Xinhai were released (Li et al., 2015; Yuan et al., 2015; Zhang et al., 2015).

As they became publicly available, the assemblies of *G. raimondii, G. arboreum,* and *G. hirsutum*

were queried for the identification of cotton *CETS*. After identification, sequence analysis and

phylogenetic studies were employed to assess evolutionary relationship and formulate hypotheses for functional testing.

2.2 Materials and Methods

Cotton *CETS* genes were identified by tBLASTn searches using the six known Arabidopsis CETS protein sequences (AtFT, AtTSF, AtMFT, AtTFL1, AtBFT and ATC; accessions included in Table 2.1) as queries against *G. raimondii* (D$_5$ genome, JGI assembly version 2.0, annotation version 2.1) (Paterson et al., 2012), *G. arboreum* (A$_2$ genome, BGI assembly version 2, annotation version 1.0) (Li et al., 2014) and *G. hirsutum* (AD$_1$ genome, NAU-NBI assembly version 1.1, annotation version 1.1; and BGI-CGP assembly version 1.0, annotation version 1.0) (Li et al., 2015; Zhang et al., 2015) assemblies at CottonGen (www.cottongen.org, Yu et al., 2014). Predicted cotton CETS peptide sequences were aligned with CETS proteins from Arabidopsis (*Arabidopsis thaliana*), tomato (*Solanum lycopersicum* and *Solanum pimpinellifolium*), jute (*Corchorus capsularis* and *Capsularis* olitorius), cacao (*Theobroma cacao*), and moss (*Physcomitrella patens*) using Neighbor-Joining (NJ) clustering with Clustal Omega (accessions included in Table 2.1) (Sievers et al., 2011). Cotton, Arabidopsis, tomato, jute, and cacao are classified as eudicots. Tomato belongs to the Asterid clade of plant lineage while cotton, Arabidopsis, jute, and cacao are Rosids and Brassicales-Malvales. Cotton, cacao, and jute are Malvaceae species, and cacao is the closest relative of cotton. Default parameters for multiple sequence alignment were used as follows: matrix Gonnet, gap open 10, gap extension 0.2, gap distance 5, and clustering NJ. A phylogenetic tree based on the multiple sequence alignment was constructed using the Bootstrap test by Maximum likelihood method in Mega 7

(Jones et al., 1992; Kumar et al., 2016). The evolutionary history was inferred by using the

Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model.

The tree with the highest log likelihood (-6296.89) is the one shown in the results. The

percentage of trees in which the associated taxa clustered together is shown next to the

branches. Initial tree(s) for the heuristic search were obtained automatically by applying NJ and

BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, then selecting

the topology with superior log likelihood value. The tree is drawn to scale with branch lengths

measured in the number of substitutions per site. The analysis involved 73 amino acid

sequences. All positions with less than 80 percent site coverage were eliminated. That is, fewer

than 20 percent alignment gaps, missing data, and ambiguous bases were allowed at any

position. There was a total of 170 positions in the final dataset. Tree branches are labeled with

percentages of 1,000 iterations support. The resultant tree was rooted with *Physcomitrella*

patens sequences.

2.3   Results

In *G. raimondii* and *G. arboreum*, eight *CETS* homologs were identified. Sixteen *CETS*

homologs were identified in the NAU-NBI *G. hirsutum* assembly, eight in each of the A_t and D_t

subgenomes corresponding to inheritance from the diploid progenitors. Seventeen homologs

were identified in the BGI-CGP *G. hirsutum* assembly, fourteen orthologous to diploid

progenitor with three *GhSP-D_t* and two *GhBFT-L1-A_t* paralogs. Only the D_t subgenome in this

assembly contained a *GhBFT-L2* homolog (identified genes and accession numbers located in

Table 2.2, McGarry et al., 2016).

Exon/Intron structure of *Gossypium CETS* genes were deduced by aligning genomic and coding sequences. Cotton *CETS* comprise a classical *CETS* genomic structure of four exons and three intervening segments (Fig 2.1). Generally, *Gossypium ssp. CETS* genes have a small structure in the range of 0.7 – 1.8 kilobase (kb). Cotton *SFT*, *FT*-homolog candidate, is the largest of cotton *CETS* in each of the analyzed genomes, having long introns two and three, a structure that is conserved among other *FT*-like homologs (Danilevskaya et al., 2008). Exons two and three of each cotton *CETS* are 60 and 39 base pairs (bps) respectively, showing conservation in size with *CETS* genes in other plant species (Carmel-Goren et al., 2003; Carmona et al., 2007; Danilevskaya et al., 2008; Zhang et al., 2005).

To infer evolutionary relationships of cotton CETS with other plant CETS, sequence analysis and phylogenetic studies were employed. Like in previous CETS phylogenetic studies, highest sequence conservation among CETS was found in the ligand-binding motif that contains a binding pocket and external loop (Figs 2.2 and 2.3). Comparison of *Gossypium ssp.* SFT sequences and Arabidopsis FT at key residues within the binding site were highly conserved, including: Asp-70 (SFT)/Asp-71 (FT), Ala-73/Val-74, Leu-81/82, Tyr-84/85, His-86/87, Glu-108/109, His-117/118, and Gln-139/140 (Fig 2.2). Similarly, *Gossypium ssp.* SP, TFL1-L1, TFL1-L2, BFT-L1 and BFT-L2 sequences were compared to Arabidopsis TFL1 at key residues within the binding site. SP, TFL1-L1 and TFL1-L2 sequences displayed full conservation with TFL1 at these residues while BFT-L1 and BFT-L2 sequences were highly conserved, again including residues: Asp-71 (SP)/70 (TFL1-L1)/67 (TFL1-L2)/71 (BFT-L1 and BFT-L2)/74 (TFL1), Val-74/73/70/Ala-74/Val-77, Leu-82/81/78/82/85, His-85/84/81/85/88, His-87/86/83/87/90, Glu-109/108/105/110/112, His-118/117/114/119/121, Phe-120/119/116/Tyr-121/Phe-123, and

Asp-141/139/136/140/144 (Fig 2.2). BFT-L1 and BFT-L2 sequences non-conserved Tyr-121

residues instead share conservation with AtBFT Tyr-122 (Fig 2.2). Included in these conserved

residues are triads previously reported as responsible for conferring TFL1 vs. FT activity in

Arabidopsis, TFL1 Glu-112-His-88-Asp-144 and FT Glu-109-Tyr-85-Gln-140. *Gossypium ssp.* SFT

sequences shared the critically conserved His-85 residue while SP, TFL1-L1, TFL1-L2, BFT-L1 and

BFT-L2 sequences contained the His-88 residue critical for TFL1 activity (Fig 2.2). Conservation

of these critical residues along with percentage of sequence identity shared with aligned

homologs discussed below contributed to the naming of cotton *CETS.*

A Maximum likelihood phylogenetic tree formed three major clades corresponding to

the generally accepted subfamilies of CETS: FT-like, TFL1-like, and MFT-like (Fig 2.4,McGarry et

al., 2016). Cotton includes one FT-homolog, denoted SFT due to having 88 percent identity to

tomato SFT and 77 percent identity to both AtFT and AtTSF. In contrast, cotton's TFL1-like

subfamily is expanded in comparison to closely related species and includes several members

represented in three subgroups. Cotton SP, TFL1-L1, TFL1-L2, BFT-L1, and BFT-L2 are all

members in the TFL-like subfamily of cotton CETS proteins. Cotton SP shares highest sequences

identity to tomato SP and Arabidopsis ATC, sharing 79 percent identity and 76 percent identity

respectively. SP-like proteins form one of three subgroups within the subfamily of TFL1-like

proteins. Two cotton homologs, TFL1-L1 and TFL1-L2, share greatest sequence identity with

AtTFL1 and are comprised within a second TFL1-like subgroup. Cotton's BFT-L1 and BFT-L2 are

most identical to AtBFT. These proteins are members of the third subgroup of TFL1-like

proteins. Cotton MFT-L1 and MFT-L2 have greatest sequence identity to AtMFT and other MFT

proteins.

Similar to previous CETS phylogenetic reports (Danilevskaya et al., 2008; Shalit et al., 2009; Wang et al., 2015), the gene tree revealed two ancient duplication events (Fig 2.4, branches 1 and 2) giving rise to the three angiosperm branches. Event 1 produced the MFT-like subfamily and the common ancestor of the FT/TFL1-like subfamilies; this division predates the angiosperm lineage. Event 2 further evolved the FT/TFL1-like lineage into two subfamilies, FT-like and TFL1-like. This event occurred after the division of angiosperms from other vascular plants and is common to all angiosperms. Two further events (Fig 2.4, branches 3 and 4) further divided the TFL1-like lineage into three subgroups: TFL1-like, SP-like, and BFT-like. Two further events (Fig 2.4, branches 5 and 6) within the Malvaceae lineage show that *Gossypium* TFL1 and BFT genes experienced duplication after cotton divergence from Malvaceae Jute and Cacao, creating *Gossypium* -specific paralogs TFL1-L1 and TFL1-L2, and BFT-L1 and BFT-L2.

2.4   Discussion

The PEBP protein family is ancient; its members can be found throughout the biosphere. Members in mammalian systems bind lipids, control neuronal development, and regulate several signaling pathways (Corbit et al., 2003; Valle et al., 2008; Yeung et al., 2001). In angiosperms gene family complexity varies. Arabidopsis has six PEBP homologs: FT and TWIN SISTER OF FT (TSF) are FT-like. MFT is the sole MFT-like member. TFL1, BROTHER OF FT AND TFL1 (BFT) and CENTRORADIALIS (ATC) are TFL1-like CETS. The number of CETS, however, varies greatly by species: in *Zea mays*, 23 members; tomato, 6 members; grapevine, 5 members; wheat, 19 members (Carmel-Goren et al., 2003; Chardon and Damerval, 2005; Carmona et al., 2007; Danilevskaya et al., 2008). Monocots tend to have more *CETS* genes than dicots. However

there are exceptions, such as the dicot soybean which was found to have 23 *CETS* gene models

(Danilevskaya et al., 2008; Wang et al., 2015).. *CETS* genes are a part of a family encoding

phosphatidylethanolamine binding proteins. *CETS* are reported to include three major

subfamilies named after encompassed Arabidopsis homologs, FT, TFL1, and MFT (Ahn et al.,

2006; Shalit et al., 2009; Karlgren et al., 2011). *CETS* are important regulators of plant growth

with members of *FT-* and *TFL1*-like subfamilies reported to have significant impact on plant

architecture through the regulation of meristem activities. In Arabidopsis, FT-like proteins FT

and TSF promote determinate growth in meristems, while TFL1-like proteins maintain

indeterminate vegetative growth (Amaya et al., 1999; Kardailsky, 1999; Kobayashi, 1999;

Mimida et al., 2001; Yamaguchi et al., 2005; Yoo et al., 2010). So far, *MFT*-like genes remain

unconnected to the regulation of meristem activities.

Here, publicly available genomic resources were mined to identify members of the *CETS*

gene family in *Gossypium* as an initial study to understand cotton CETS functions. It was

hypothesized that *Gossypium* would have 6 – 10 *CETS* based upon the number of *CETS* in closely

related species. In diploid cottons, eight *CETS* were identified. Sixteen and seventeen *CETS*

were identified from the NAU-NBI and BGI-CGP *Gossypium hirsutum* tetraploid assemblies,

respectively (McGarry et al., 2016). For further phylogenetic studies, both diploid and the

tetraploid *G. hirsutum* NAU-NBI assemblies were analyzed.

Two other studies have reported identification of cotton *CETS* through BLAST searches.

In July 2015, a report identified six cotton *CETS* in the *G. raimondii* assembly (Grover et al.,

2015). This report failed to identify *TFL1-L2* and *MFT-L2*, both located on chromosome nine in

the D_5 genome. Similarly, a 2016 report published after our report showed overlap of *CETS*

identification with our study, but also had notable differences. The study identified *CETS* in both diploids, *G. raimondii* and *G. arboreum*, and the BGI-CGP *G. hirsutum* assembly. *Gossypium BFT-L1* was not identified from any of the assemblies. Additionally, in the BGI-CGP *G. hirsutum* assembly, the study reported only one chromosome D_t-1 *SP* (we report two *SP* genes on chromosome D_t-1) and one D_t-12 chromosome *TFL1-L2* gene as opposed to the two *TFL1-L2* genes we identified, one in each *G. hirsutum* subgenome. The same study also identified two *CETS* denoted as *GhPEBP1* and *GhPEBP2* not identified in our study. These two genes differ highly from classical *CETS* genomic structures having only two introns and are phylogenetically not classified within the three generally accepted angiosperm *CETS* subfamilies (Zhang et al., 2016). Differences in *CETS* identification by the three studies are probably due to variances in search methodologies and genome assemblies. Our analysis used the NAU assembly in addition to the BGI assembly used by Zhang et al. In our study, tBLASTn searches were used to query Arabidopsis CETS proteins against *Gossypium* genome assemblies, while the Grover et al. study employed BLASTp to query Arabidopsis CETS against *G. raimondii* predicted peptides (Grover et al., 2015). Search methods for the 2016 study were not clearly defined (Zhang et al., 2016) and thus cannot be assessed.

SFT is the sole *FT*-like *Gossypium CETS*; this is a reduction in the number of *FT*-like genes in comparison to other species analyzed. Functional studies will show that *SFT* promotes determinate growth. *SFT* is the proposed cotton florigen. *Gossypium*'s *TFL1* subfamily is expanded in comparison to Arabidopsis and other closely related species. This expansion is due to *Gossypium*-specific gene duplications of *TFL1* and *BFT*. Functional studies will show that each cotton *TFL1*-like gene has the potential to maintain indeterminate growth, but their impact on

plant architecture varies. Two *MFT*-like genes were discovered in *Gossypium*. This is an increase

in *MFT* genes in comparison to Arabidopsis' one *MFT,* but shows similarity to other closely-

related species, jute and cacao. In this study, *Gossypium CETS* were named based upon amino

acid residue composition, and the closest homolog as calculated by percent identity.

Genome assembly studies have deduced a cotton specific whole genome duplication

event before speciation of *G. raimondii* and *arboreum* approximately 13-20 million years ago (Li

et al., 2014; Li et al., 2015). It is likely that gene duplication of cotton's *TFL1, BFT,* and *MFT*

arose during this gene duplication event, but these gene duplications might also have been

results of other chromosomal rearrangements in the progenitor of A- and D-genomes.

While sequence and phylogenetic analysis aids hypothesis formation for functional

determination, studies of *CETS* in several species have shown predictions based these

parameters alone can be misleading. Structural analysis of human PEBP, Raf Kinase Inhibitor

Protein, resolved a small globular protein with a putative binding pocket for phosphorylated

ligands (Banfield et al., 1998). Crystal structure determination of the snapdragon CEN

established universal structural features among PEBPs and also detected an unstructured

external loop unique to angiosperm CETS (Banfield and Brady, 2000). Evidence demonstrates

that both the binding pocket and external loop are critical for conferring floral promotion and

repression activities (Hanzawa et al., 2005; Ahn et al., 2006). A Y85H mutation within the

binding pocket or a swap of the FT for TFL1 external loop were both satisfactory for converting

FT's function to floral repression; however, reciprocal modifications failed to alter TFL1

function. New evidence identified four additional amino acids critical for specification of FT's

floral promoting activity (Ho and Weigel, 2014). In domesticated sugar beet, a spontaneous

conversion of a single Y to N in the external loop of an FT-paralog resulted in floral repression

function (Pin et al., 2010a). Additionally, studies of other FT-paralogs and their functions in

differing plant systems demonstrate that specifying amino acids critical for floral promoting or

repressing activities is a complex task (Hecht et al., 2011; Meng et al., 2011), and functional

determination requires additional experimental evidence. In the following chapters, functional

and promoter analysis will further the delineation of cotton *CETS* function.

Table 2.1 Protein accessions used for tBLASTn queries and phylogenetic studies. Arabidopsis protein sequences only were used in tBLASTn queries to identify *Gossypium CETS* homologs, while all listed sequences were used in phylogenetic studies. Accession numbers for Arabidopsis, moss, and jute and tomato's SlSP2G, SlSP3D, SlSP5G, SlSP9D, and SlSP are from NCBI's Protein Database. Cacao protein accessions are from JGI Phytozome 12. Tomato protein accessions SlBFT1-3, SlMFT, SlSP11D, SpSP11c, and SpSP6A are from Sol Genomics Network (solgenomics.net).

protein	species	accession
AtFT	*Arabidopsis thaliana*	AAF03936.1
AtTSF	*Arabidopsis thaliana*	BAA77840.1
AtTFL1	*Arabidopsis thaliana*	AED90661.1
AtCEN	*Arabidopsis thaliana*	NP_180324.1
AtBFT	*Arabidopsis thaliana*	AED97554.1
AtMFT	*Arabidopsis thaliana*	AAD37380.1
PpMFTL1	*Physcomitrella patens*	ACN5453.1
PpMFTL2	*Physcomitrella patens*	ACN54544.1
PpMFTL3	*Physcomitrella patens*	ACN54546.1
PpMFTL4	*Physcomitrella patens*	ACN54547.1
TcTFL1	*Theobroma cacao*	Thecc1EG022560
TcSP	*Theobroma cacao*	Thecc1EG041439
TcBFT	*Theobroma cacao*	Thecc1EG015117
TcSFT	*Theobroma cacao*	Thecc1EG023287
TcMFT-L2	*Theobroma cacao*	Thecc1EG030010
TcMFT-L1	*Theobroma cacao*	Thecc1EG012687
CcMFT-L1	*Corchorus capsularis*	OMO59963.1
CcMFT-L3	*Corchorus capsularis*	OMO68071.1
CcMFT-L2	*Corchorus capsularis*	OMO76632.1
CcSFT	*Corchorus capsularis*	OMO57543.1
CcBFT	*Corchorus capsularis*	OMP11529.1
CcSP	*Corchorus capsularis*	OMO49508.1
CcTFL1	*Corchorus capsularis*	OMO82188.1
CoTFL1	*Corchorus olitorius*	OMO68308.1
CoSP	*Corchorus olitorius*	OMO89182.1
CoBFT	*Corchorus olitorius*	OMO97927.1
CoSFT	*Corchorus olitorius*	OMO64223.1
CoMFT-L2	*Corchorus olitorius*	OMO65020.1
CoMFT-L1	*Corchorus olitorius*	OMO94933.1
SlSP2G	*Solanum lycopersicum*	AAO31791.1
SlSP3D	*Solanum lycopersicum*	AAO31792.1
SlSP5G	*Solanum lycopersicum*	AAO31793.1
SlSP9D	*Solanum lycopersicum*	AAO31795.1
SlSP	*Solanum lycopersicum*	NP_001233974.1
SlBFT-L1	*Solanum lycopersicum*	Solyc01g009560
SlBFT-L2	*Solanum lycopersicum*	Solyc01g009580
SlBFT-L3	*Solanum lycopersicum*	Solyc03g026050
SlMFT	*Solanum lycopersicum*	Solyc03g119100
SlSP11D	*Solanum lycopersicum*	Solyc11g008640
SpSP11C	*Solanum pimpinellifolium*	Solyc11g008660
SpSP6A	*Solanum pimpinellifolium*	Solyc05g055660

Table 2.2 Cotton *CETS* genes identified in *Gossypium ssp.* tBLASTn searches using the six Arabidopsis CETS homologs as query sequences identified cotton *CETS* within four cotton genomes (McGarry et al., 2016).

assigned gene name	*G. raimondii* D5 JGI assembly v2.0	*G. arboreum* A2 BGI-CGP assembly v2.0	*G. hirsutum* AD1 NAU-NBI assembly v1.1 D_t subgenome	A_t subgenome	*G. hirsutum* AD1 BGI-CGP assembly v1.0
SP	Gorai.001G121800.1	cotton_A_09584	Gh_D07G1075	Gh_A07G0997	CotAD_46899 CotAD_43766 CotAD_15834 (D_t ch1) (D_t ch1) (D_t ch13)
TFL1-L1	Gorai.006G155800.1	cotton_A_13428	Gh_D09G1320	Gh_A09G2442	CotAD_02907 CotAD_43979 (D_t ch5) (A_t ch11)
TFL1-L2	Gorai.009G403800.1	cotton_A_31651	Gh_D04G0971	Gh_A04G0520	CotAD_37875 CotAD_57593 (D_t ch12) (D_t ch12)
BFT-L1	Gorai.004G120400.1	cotton_A_39415	Gh_D08G1087	Gh_ Sca15601	CotAD_52730 CotAD_76371 CotAD_75919 (D_t ch4) (A_t ch4) (A_t ch4)
BFT-L2	Gorai.007G010800.1	cotton_A_07540	Gh_D11G0092	Gh_A11G0088	CotAD_02721 (D_t ch9)
SFT	Gorai.004G264600.1	cotton_A_05804	Gh_D08G2407	Gh_A08G2015	CotAD_04102 CotAD_14755 (D_t ch5) (Sca 246.1)
MFT-L1	Gorai.006G192300.1	cotton_A_13046	Gh_D09G1658	Gh_A09G2391	CotAD_03154 CotAD_70215 (D_t ch6) (Sca 4006.1)
MFT-L2	Gorai.009G174600.1	cotton_A_04728	Gh_D05G1586	Gh_ A05G166400	CotAD_55039 CotAD_41263 (Sca 2081.1) (A_t ch7)
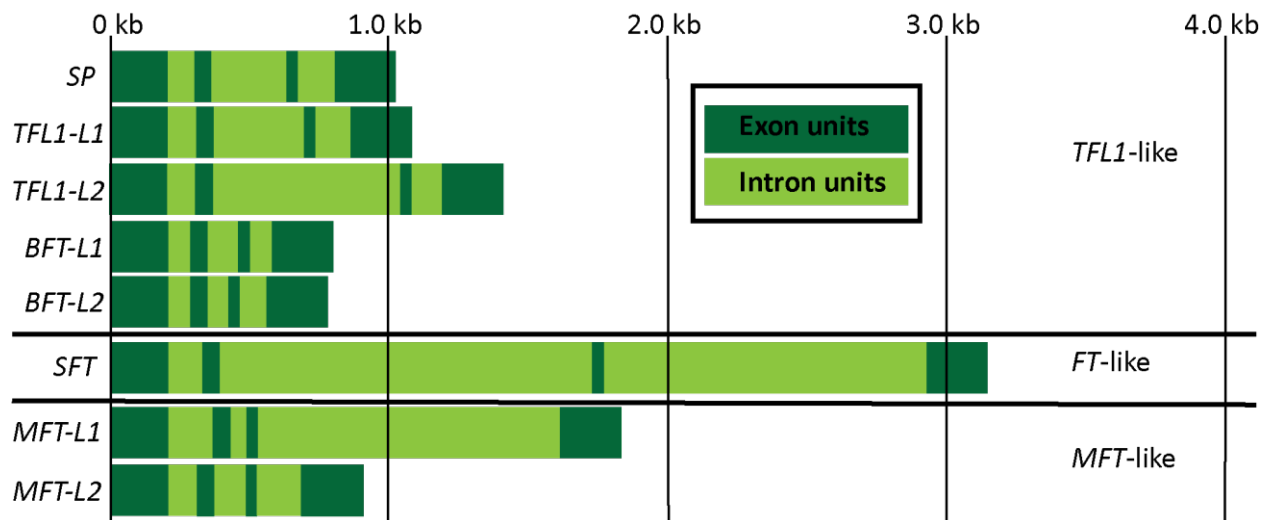
Figure 2.1 Exon/intron structure of *Gossypium ssp. CETS*. *CETS* genomic structure is denoted by dark green boxes as exons and light green boxes as introns. Genes are organized into phylogenetic subfamilies; major subfamily names are designated to the right. Vertical lines can be used to approximate exon and intron size. These genomic structures are representative of *CETS* from analyzed genomes of *G. raimondii* (D-genome), *G. arboreum* (A-genome), and *G. hirsutum* (A_t-and D_t-genomes). Genomic structures are consistent between genomes apart from *GaSP* (not represented here). *GaSP* has a uniquely long intron three in comparison to other *Gossypium SP* sequences with a length of 1,477 bps.

Figure 2.2 Alignment of cotton and Arabidopsis CETS ligand-binding and external loop domains. Cotton *CETS* genes were given descriptive names based upon the deduced polypeptide at variable residues within the ligand binding domain and the external loop. Shown is a multiple sequence alignment of the ligand binding domain and external loop of cotton and Arabidopsis CETS protein sequences. The external loop is boxed. Critical and conserved residues are marked within a cross (+). Critical but variable sites are marked with a pound sign (#). These sites were considered when assigning gene names to cotton *CETS*. His-88 is important for activity as an indeterminate growth factor, while the corresponding Tyr-85 correlates to activity as a determinate growth factor. MFT-like proteins have a Tryptophan at this location. Invariable and similar residues are colored according to biochemical properties.

32

```
PpMFTL3         ----------------------------------------------------MSRSVDP          7
PpMFTL4         ----------------------------------------------------MARSIDP          7
PpMFTL1         ----------------------------------------------------MPRSIDP          7
PpMFTL2         ----------------------------------------------------MARSIDP          7
SlMFT           ----------------------------------------------------MGGKVDP          7
AtMFT           ----------------------------------------------------MAASVDP          7
CcMFT-L3        ----------------------------------------------------MAVSVDP          7
CoMFT-L1        ----------------------------------------------------MAVSVDP          7
CcMFT-L1        ----------------------------------------------------MAVSVDP          7
TcMFT-L1        MPYAWHIRHTNLSHFSPFTPLYKLLLAFLYSTLTPNQTNILISSPLLFALFFSMAVSVDP         60
GaMFT-L1        ----------------------------------------------------MAASVDP          7
GrMFT-L1        ----------------------------------------------------MAASVDP          7
GhMFT-L1-Dt     ----------------------------------------------------MAASVDP          7
GhMFT-L1-At     ----------------------------------------------------MAASVDP          7
SlSP2G          -------------------------------------------------METSARSVDP         10
CoMFT-L2        ----------------------------------------------------MARSVEP          7
CcMFT-L2        ----------------------------------------------------MARSVEP          7
TcMFT-L2        ----------------------------------------------------MARSVEP          7
GrMFT-L2        ----------------------------------------------------MARSVEP          7
GhMFT-L2-Dt     ----------------------------------------------------MARSVEP          7
GaMFT-L2        ----------------------------------------------------MARSVEP          7
GhMFT-L2-At     ----------------------------------------------------MARSVEP          7
SlSP5G          -------------------------------------------------------MPRDP          5
SlSP11D         ----------------------------------------------------MQRERDT          7
AtFT            ---------------------------------------------------MSINIRDP          8
AtTSF           ---------------------------------------------------MSLSRRDP          8
SpSP11C         --------------------------------------------------MSSIRGRDT          9
SpSP6A          ------------------------------------------------------MPRVDP          6
SlSP3D          ---------------------------------------------------MPRERDP          7
CoSFT           ---------------------------------------------------MPRDRDP          7
CcSFT           ---------------------------------------------------MPRDRDP          7
TcSFT           ---------------------------------------------------MPRERDP          7
GaSFT           ---------------------------------------------------MPRDRDP          7
GhSFT-At        ---------------------------------------------------MPRDRDP          7
GrSFT           ---------------------------------------------------MPRDRDP          7
GhSFT-Dt        ---------------------------------------------------MPRDRDP          7
SlSP            ---------------------------------------------------MASKMCEP          8
AtCEN           ---------------------------------------------------MARISSDP          8
CoSP            ----------------------------------------------------MAKLSDP          7
CcSP            ----------------------------------------------------MAKLSDP          7
TcSP            ----------------------------------------------------MAKLSDP          7
GrSP            ----------------------------------------------------MAKLSDP          7
GhSP-Dt         ----------------------------------------------------MAKLSDP          7
GaSP            ----------------------------------------------------MAKLSDP          7
GhSP-At         ----------------------------------------------------MAKLSDP          7
AtTFL1          ------------------------------------------------MENMGTRVIEP         11
SlSP9D          ----------------------------------------------------MARSLEP          7
GaTFL1-L2       -------------------------------------------------------MGEP          4
GhTFL1-L2-At    -------------------------------------------------------MGEP          4
GrTFL1-L2       -------------------------------------------------------MGEP          4
GhTFL1-L2-Dt    -------------------------------------------------------MGEP          4
GrTFL1-L1       ---------------------------------------------------MAREVEP          7
GhTFL1-L1-Dt    ---------------------------------------------------MAREVEP          7
GaTFL1-L1       ---------------------------------------------------MAREVEP          7
GhTFL1-L1-At    ---------------------------------------------------MAREVEP          7
TcTFL1          ----------------------------------------------------MSRAAEP          7
CoTFL1          ---------------------------------------------------MSTRSIEP          8
CcTFL1          ---------------------------------------------------MSTRSIEP          8
AtBFT           ----------------------------------------------------MSREIEP          7
SlBFT-L1        ---------------------------------------------------MSCRDIEP          8
SlBFT-L2        ---------------------------------------------------MSCRDIEP          8
SlBFT-L3        --------------------------------------------------MSSRSTCEP          9
TcBFT           ----------------------------------------------------MSRVPEP          7
GaBFT-L2        ----------------------------------------------------MSRVPEP          7
GhBFT-L2-At     ----------------------------------------------------MSRVPEP          7
GrBFT-L2        ----------------------------------------------------MSRVPEP          7
GhBFT-L2-Dt     ----------------------------------------------------MSRVPEP          7
GhBFT-L1-Dt     ----------------------------------------------------MSRVPDP          7
GrBFT-L1        ----------------------------------------------------MSRVPDP          7
GaBFT-L1        ----------------------------------------------------MSRVPDP          7
GhBFT-L1-At     ----------------------------------------------------MSRVPDP          7
CoBFT           ---------------------------------------------------MSRSVLEP          8
CcBFT           ---------------------------------------------------MSRSVHEP          8
                                                                    :
```

```
PpMFTL3      LVVGRVIGVVIDMFAPSV--DMAVVYTSRKVS-NGCQMKPSATNEAPTVHVTGNNG-DNN      63
PpMFTL4      LVVGKVIGDVIDTFVPSV--DMAIHYSTRQVT-NGCQMMPSATAQAPEIHLSDKSG-GNN      63
PpMFTL1      LIVGKVIGDVIDTFVPRV--DMAIHYSTRQVT-NGCQLKPSATAQAPEIQLSDKSG-DNN      63
PpMFTL2      LVVGKVIGDVIDTFVPSV--DMAIHYSSRQVT-NGCQMKPSATAQAPEIQLSDNSE-GNN      63
SlMFT        LVVGRVIGDVVDMFVPSV--TMSVHYANKHVN-NGCDIKPSIATEPPKIAIGGQPD---E      61
AtMFT        LVVGRVIGDVLDMFIPTA--NMSVYFGPKHIT-NGCEIKPSTAVNPPKVNISGHSD---E      61
CcMFT-L3     LVVGRVIGDVVDMFVPTV--TTSIYYASKHVT-NGCHVKPSIAINPPKVSIDGHPG---H      61
CoMFT-L1     LVVGRVIGDVVDMFVPTV--TMSIYYASKHVT-NGCDVKPSMAINPPKVSIDGHPD---H      61
CcMFT-L1     LVVGRVIGDVVDMFVPSV--TMSIYYASKHVT-NGCDVKPSMAINPPKVSIDGHPD---H      61
TcMFT-L1     LVVGRVIGDVVDMFVPTV--TMSVYYGSRHVT-NGCDIKPSTTINPPKVSINGHSD---E     114
GaMFT-L1     LVVGRVIGDVVDMFVPTV--TMSVYYGSKHVS-NGCDIKPSMAINPPKVAIDGLPD---Q      61
GrMFT-L1     LVVGRVIGDVVDMFVPTV--TMSVYYGSKHVS-NGCDIKPSMAINPPKVAIDGLPD---Q      61
GhMFT-L1-Dt  LVVGRVIGDVVDMFVPTV--TMSVYYGSKHVS-NGCDIKPSMAINPPKVAIDGLPD---Q      61
GhMFT-L1-At  LVVGRVIGDVVDMFVPTV--TMSVYYGSKHVS-NGCDIKPSMAINPPKVAIDGLPD---Q      61
SlSP2G       LVVGKVIGDVLDMFVPVV--DFTVEYASKQISNNGVEIKPAEAAQKPRVHIKGSLH-SNN      67
CoMFT-L2     LVVGRVIGDVLDMFTPAA-TVFTAHYGSKQ-----------------------------      36
CcMFT-L2     LVVGRVIGDVLDMFTPAAATVFTAHYGSKQVT-NGCDIKPSAASDKPHAQILGPPDNSTG      66
TcMFT-L2     LVVGRVIGDVLDIFTPAA--ELTVHYSTKQVH-NGCDIKPSSAADKPHVRILSPVV-SSS      63
GrMFT-L2     LVVGRVIGDVLDMFTPAS--EFIVRYGTKQVT-NGCDIKPSAAADKPHVQILGHPF-SSN      63
GhMFT-L2-Dt  LVVGRVIGDVLDMFTPAS--EFTVRYGTKQVT-NGCDIKPSAAADKPHVQILGHPF-SSN      63
GaMFT-L2     LVVGRVIGDVLDMFTPAS--EFTVRYGTKQVT-NGCDIKPSAAADKPHVQILGHPF-SSN      63
GhMFT-L2-At  LVVGRVIGDVLDMFTPAS--EFTVRYGTKQVT-NGCDIKPSAAADKPHVQILGHPF-SSN      63
SlSP5G       LIVSGVVGDVVDPFTRCV--DFGVVYN-NRVVYNGCSLRPSQVVNQPRVDIDGDDL--RT      60
SlSP11D      LRLARVIGDVLDPFTKSI--NLRVVYN-NKEIRNGCDLRPSMVVNQPRVEVGGDDF--QT      62
AtFT         LIVSRVVGDVLDPFNRSI--TLKVTYG-QREVTNGLDLRPSQVQNKPRVEIGGEDL--RN      63
AtTSF        LVVGSVVGDVLDPFTRLV--SLKVTYG-HREVTNGLDLRPSQVLNKPIVEIGGDDF--RN      63
SpSP11C      LELGGVISDVLDPFTRSI--NLSVVYN-HREVINGTNLRPSQITNQPRVEVGGNDL--ST      64
SpSP6A       LIVGRVIGEVLDPFTRSV--DLRVVYN-NREVNNACVLKPSQVVMQPKVYIGGDDL--RT      61
SlSP3D       LVVGRVVGDVLDPFTRTI--GLRVIYR-DREVNNGCELRPSQVINQPRVEVGGDDL--RT      62
CoSFT        LVVGRVIGDVLDPFTRSI--SLRVSFG-GREVNNGCELKPSQVVNQPRVDIGGEDL--RT      62
CcSFT        LVVGRVIGDVLDPFTRSI--SLRVSFG-GREVNNGCELKPSQVVNQPRVDIGGEDL--RT      62
TcSFT        LVVGRVIGDVLDPFTRSI--SLRVTFA-CREVNNGCELKPSQVVNQPRVDIGGDDL--RT      62
GaSFT        LVVGRVIGDVLDPFTRSI--SLRVTYA-TRDVNNGVELKPSQVVNQPRVDIGGDDL--RT      62
GhSFT-At     LVVGRVIGDVLDPFTRSI--SLRVTYA-TRDVNNGVELKPSQVVNQPRVDIGGDDL--RT      62
GrSFT        LVVGRVIGDVLDPFTRSI--SLRVTYA-TRDVSNGVELKPSQVVNQPRVDIGGDDL--RT      62
GhSFT-Dt     LVVGRVIGDVLDPFTRSI--SLRVTYA-TRDVSNGVELKPSQVVNQPRVDIGGDDL--RT      62
SlSP         LVIGRVIGEVVDYFCPSV--KMSVVYNNNKHVYNGHEFFPSSVTSKPRVEVHGGDL--RS      64
AtCEN        LMVGRVIGDVVDNCLQAV--KMTVTYNSDKQVYNGHELFPSVVTYKPKVEVHGGDM--RS      64
CoSP         LVVGRVIGDVVDAINPCV--KITVTFNSNKQVYNGHEFFPSSVTTKPK------------      53
CcSP         LVVGRVIGDVVDAINPCV--KITVTFNSNKQVYNGHEFFPSSVTTKPK------------      53
TcSP         LVVGRVIGDVDAITPSV--KMTVTFNANKQVYNGHELFPSSVTNKPKVDVHGGDM--RS      63
GrSP         LVLGRVIGDVIDALSPSV--KMSVTFNTNKQVYNGHEFFPSAVTNKPKVEVHGGDM--RS      63
GhSP-Dt      LVLGRVIGDVIDALSPSV--KMSVTFNTNKQVYNGHEFFPSAVTNKPKVEVHGGDM--RS      63
GaSP         LVVGRVIGDVIDALSPSV--KMSVTFNTNKQVYNGHEFFPSAVTNKPKVEVHGGDM--RS      63
GhSP-At      LVVGRVIGDVIDALSPSV--KMSVTFNTNKQVYNGHEFFPSAVTNKPKVEVHGGDM--RS      63
AtTFL1       LIMGRVVGDVLDFFTPTT--KMNVSYN-KKQVSNGHELFPSSVSSKPRVEIHGGDL--RS      66
SlSP9D       LIVGRVIGDVIDSFNPTI--KMSITYN-NKLVCNGHELFPSVVSSRPKVEVQGGDL--RT      62
GaTFL1-L2    LIVGGVVGDVLDSFNPSI--KMSVTFN-NKQVFNGHEFYPSSVATKPRVEIQGGDL--RT      59
GhTFL1-L2-At LIVGGVVGDVLDSFNPSI--KMSVTFN-NKQVFNGHEFYPSSVATKPRVEIQGGDL--RT      59
GrTFL1-L2    LIVGGVVGDVLDSFNPSI--KMSVTFN-NKQVFNGHEFYPSSVATKPRVEIQGGDL--RT      59
GhTFL1-L2-Dt LIVGGVIGDVLDSFNPSI--KMSVTFN-NKQVFNGHEFYPSSVATKPRVEIQGGDL--RT      59
GrTFL1-L1    LMVGRVIGDVMDSFIPSI--KMLVTFN-NKQVFNGHEFYPSTVVTKPRVEVAGGDM--RT      62
GhTFL1-L1-Dt LMVGRVIGDVMDSFIPSI--KMLVTFN-NKQVFNGHEFYPSTVVTKPRVEVAGGDM--RT      62
GaTFL1-L1    LMVGRVIGDVMDSFIPSI--KMSVTFN-NKQVFNGHEFYPSTVVTKPRVEVVGGDM--RT      62
GhTFL1-L1-At LMVGRVIGDVMDSFIPSI--KMSVTFN-NKQVFNGHEFYPSTVVTKPRVEVVGGDM--RT      62
TcTFL1       LVVGRVIGDVLDSFIPSI--TMTVTFN-NKRVFNGHEFYPSTVATKPRVEIEGGDM--RT      62
CoTFL1       LIVGRVIGDVLDSFIPSI--IMTVSFN-NKKVFNGHEFFPSTVASRPRVEIEGGDL--RT      63
CcTFL1       LIVGRVIGDVLDSFIPSI--TMTASFN-NKKVFNGHEFFPSTVAFRPRVEIEGGDL--RT      63
AtBFT        LIVGRVIGDVLEMFNPSV--TMRVTFNSNTIVSNGHELAPSLLLSKPRVEIGGQDL--RS      63
SlBFT-L1     LIVAKVIGEVVDSFNPSV--KMNVTYNGTKQVFNGHELMPLVIASKPRVEIGGEDM--RS      64
SlBFT-L2     LIVARVIGEVVDSFNPSV--KMVIYNGTKQVFNGHELMPLVIASKPRVEIGGEDM--RS      64
SlBFT-L3     LAVGRVIGEVVDSFSPSV--KMKVIYNGRKQVSNGHEIMPAVVATQPRVEIGGEDM--RS      65
TcBFT        LTVGRVIGEVVDNFTPSV--KMTVTYNSNKQVANGHELMPAVIVARPRVEIGGEDM--RA      63
GaBFT-L2     LTVGRVIGEVVDNFTQSV--QMTVTYNPNKQVANGHELMPAVISARPRVEIGGNDM--RD      63
GhBFT-L2-At  LTVGRVIGEVVDNFTQSV--QMTVTYNPNKQVANGHELMPAVISARPRVEIGGNDM--RD      63
GrBFT-L2     LTVGRVIGEVVDNFTPSV--QMTVTYNPNKQVANGHELMPAAISARPRVEIGGNDM--RD      63
GhBFT-L2-Dt  LTVGRVIGEVVDNFTPSV--QMTVTYNPNKQVANGHELMPAAISARPRVEIGGNDM--RD      63
GhBFT-L1-Dt  LIIGRVIGEVVDNFFPSV--KITVTYNSNKQVANGHELMPALITGRPRVEIGGDDM--RP      63
GrBFT-L1     LIIGRVIGEVVDNFFPSV--KITVTYNSNKQVANGHELMPALITGRPRVEIGGDDM--RP      63
GaBFT-L1     LIIGRVIGEVVDNFFPSV--KITVTYNSNKQVANGHELMPALITARPRVEIGGDDM--RP      63
GhBFT-L1-At  LIIGRVIGEVVDNFFPSV--KITVTYNSNKQVANGHELMPALITARPRVEIGGDDM--RP      63
CoBFT        LSIGRVIGEVVDYFTPSV--KLIVTYNSNKQVANGHELMPALISARPRVEIGGDDL--RS      64
CcBFT        LSIGRVIGEVVDYFTPSV--KLIVTYNSNKQVANGHELMPAVISARPRVEIGGDDM--RA      64
             * :. *:. *::                   :
```

```
PpMFTL3       FFTLIMTDPDAPSPSEPSLREWVHW------------IVTDIPGNSSTTTSGQGSKRARE    111
PpMFTL4       LYTLIMIDPDAPSPSEPTLREWLHW------------IVTDIPGNSGGSEMTSGFPRLNE    111
PpMFTL1       YYTLVMTDPDAPSPSEPSLREWLHW------------IVTDIPGNSGGSETNTGFPWLSE    111
PpMFTL2       YYTLVMTDPDAPSPSEPSLREWLHW------------IVTDIPGNSGGSETTSGFSWLQE    111
SlMFT         FYTLVMTDPDAPSPSEPTMREWVHW------------IVTDIPGCSN-------------    96
AtMFT         LYTLVMTDPDAPSPSEPNMREWVHW------------IVVDIPGGTN-------------    96
CcMFT-L3      LYTLVMTDPDAPSPSEPSMREWVHW------------KGDT------------------    90
CoMFT-L1      LYTLVMTDPDAPSPSEPSMREWVHW------------IVCDIPGGTN-------------    96
CcMFT-L1      LYTLVMTDPDAPSPSEPSMREWVHW------------IVCDIPGGTN-------------    96
TcMFT-L1      LYTLVMTDPDAPSPSEPSMREWVHW------------IVSDIPGGTN-------------    149
GaMFT-L1      FYTLVMTDPDAPSPSEPTMREWVHW------------IVSDIPGGTN-------------    96
GrMFT-L1      FYTLVMTDPDAPSPSEPTMREWVHW------------IVSDIPGGTN-------------    96
GhMFT-L1-Dt   FYTLVMTDPDAPSPSEPTMREWVHW------------IVSDIPGGTN-------------    96
GhMFT-L1-At   FYTLVMTDPDAPSPSEPTMREWVHW------------IVSDIPGGTN-------------    96
SlSP2G        LYTLVMADPDAPSPSEPTFREWLHW------------IVTDIPEGGD-------------    102
CoMFT-L2      ----VMVDPDAPSPSEPRLRE---------------IVVDIPHGHD-------------    63
CcMFT-L2      LYTLVMVDPDAPSPSEPRLREWLHW------------IVVDIPHGHD-------------    101
TcMFT-L2      LYTLVMVDPDAPTPSEPRLREWLHW------------IVVDIPEGHD-------------    98
GrMFT-L2      LYTLVMVDPDAPSPSEPRLREWLHW------------IVVDVPEGQD-------------    98
GhMFT-L2-Dt   LYTLVMVDPDAPSPSEPRLREWLHW------------IVVDVPEGQD-------------    98
GaMFT-L2      LYTLVMVDPDAPSPSEPRLREWLHW------------IVVDIPEGQD-------------    98
GhMFT-L2-At   LYTLVMVDPDAPSPSEPRLREWLHW------------IVVDIPEGQD-------------    98
SlSP5G        FYTLIMVDPDAPNPSNPNLREYLHW------------LVTDIPAATG-------------    95
SlSP11D       FYTLVMVDPDAPTPSNPCHKDYLHW------------LVTNIPASTG-------------    97
AtFT          FYTLVMVDPDVPSPSNPHLREYLHW------------LVTDIPATTG-------------    98
AtTSF         FYTLVMVDPDVPSPSNPHQREYLHW------------LVTDIPATTG-------------    98
SpSP11C       FYTLIVVDPDAPSPSNPNLREYLHW------------LVTDIPATTG-------------    99
SpSP6A        FYTLIMVDPDAPSPSNPNLREYLHW------------LVTDIPATTG-------------    96
SlSP3D        FFTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
CoSFT         FYTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
CcSFT         FYTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
TcSFT         FYTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
GaSFT         FYTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
GhSFT-At      FYTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
GrSFT         FYTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
GhSFT-Dt      FYTLVMVDPDAPSPSDPNLREYLHW------------LVTDIPATTG-------------    97
SlSP          FFTLIMIDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    99
AtCEN         FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    99
CoSP          ----VMTDPDVPGPSDPYLKEHLHW------------IVTDIPGTTD-------------    84
CcSP          ----VMTDPDVPGPSDPYLKEHLHW------------IVTDIPGTTD-------------    84
TcSP          FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    98
GrSP          FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    98
GhSP-Dt       FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    98
GaSP          FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    98
GhSP-At       FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    98
AtTFL1        FFTLVMIDPDVPGPSDPFLKEHLHW------------IVTNIPGTTD-------------    101
SlSP9D        FFTLVMTDPDVPGPSDPYMREHLHW------------IITDIPGTTD-------------    97
GaTFL1-L2     FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    94
GhTFL1-L2-At  FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    94
GrTFL1-L2     FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    94
GhTFL1-L2-Dt  FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    94
GrTFL1-L1     FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    97
GhTFL1-L1-Dt  FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    97
GaTFL1-L1     FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    97
GhTFL1-L1-At  FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    97
TcTFL1        FFTLVMTDPDVPGPSDPYLREHIHW------------IVTDIPGTTD-------------    97
CoTFL1        FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    98
CcTFL1        FFTLVMTDPDVPGPSDPYLREHLHW------------IVTDIPGTTD-------------    98
AtBFT         FFTLIMMDPDAPSPSNPYMREYLHW------------MVTDIPGTTD-------------    98
SlBFT-L1      AYTLVMIDPDVPGPSDPYLREHLHW------------IVTDIPGSTD-------------    99
SlBFT-L2      AYTLIMTDPDVPGPSDPYLREHLHW------------IVTDIPGSTD-------------    99
SlBFT-L3      AYTLIMTDPDAPSPSDPYLREHLHW------------IVTDIPGTTD-------------    100
TcBFT         AYTLILTDPDAPSPSDPYLREHLHCPQSKLQTNISSRMVTDIPGTTD-------------    110
GaBFT-L2      AYTLIMTDPDAPSPSGPFLREHLHW------------MVTDVPGTTD-------------    98
GhBFT-L2-At   AYTLIMTDPDAPSPSGPFLREHLHW------------MVTDVPGTTD-------------    98
GrBFT-L2      AYTLVMTDPDAPSPSGPFLREHLHW------------MVTDVPGTTD-------------    98
GhBFT-L2-Dt   AYTLIMTDPDAPSPSDPFLREHLHW------------MVTDVPGTTD-------------    98
GhBFT-L1-Dt   SYTLIMTDPDAPSPSDLYLREHLHW------------MVTDIPGTTD-------------    98
GrBFT-L1      SYTLVMTDPDAPSPSDPYLIEHLHW------------MVTDIPGTTD-------------    98
GaBFT-L1      SYTLIMTDPDAPSPSDPYLREHLHW------------MVTDIPGTTD-------------    98
GhBFT-L1-At   CYTLIMTDPDAPSPSDPYLREHLHW------------MVTDIPGTTD-------------    98
CoBFT         AYTLIMTDPDAPSPSDPYLREHLHW------------MVTDIPGTTD-------------    99
CcBFT         AYTLIMTDPDAPSPSDPYLREHLHW------------MVTDIPGTTD-------------    99
              ::  ***.* **        :
```

35

```
PpMFTL3        PASSAKQPNVERKKKGPAASTTDKELPSAADQGAAKPRTSGKEVVPYVGPCPPIGIHRYI    171
PpMFTL4        -------------------------------LIAPSKSCGRELVPYMGPRPPVGIHRYI    139
PpMFTL1        ------------------------------QATSTSSSGRELVPYIGPRPPIGIHRYI    139
PpMFTL2        ------------------------------QVTHTSSSGRELVPYMGPRPPIGIHRYI    139
SlMFT          ------------------------------------VGRGKEVLGYVGPRPPVGIHRYI    119
AtMFT          -----------------------------------PSRGKEILPYMEPRPPVGIHRYI    119
CcMFT-L3       -------------------------------------------TIHGARPPVGIHRYI    105
CoMFT-L1       ---------------------------------PTQGKEILPYMGPRPPVGIHRYI    119
CcMFT-L1       ---------------------------------PTQGKEILPYMGPRPPVGIHRYI    119
TcMFT-L1       ---------------------------------PTRGKEILVYMGPRPPVGIHRYI    172
GaMFT-L1       ---------------------------------PTRGKEILAYMGPRPPVGIHRYI    119
GrMFT-L1       ---------------------------------PTRGKEILAYMGPRPPVGIHRYI    119
GhMFT-L1-Dt    ---------------------------------PTRGKEILAYMGPRPPVGIHRYI    119
GhMFT-L1-At    ---------------------------------PTRGKEILAYMGPRPPVGIHRYI    119
SlSP2G         ---------------------------------ASQGREMVEYMGPKPPAGIHRYV    125
CoMFT-L2       ---------------------------------ATKGRELVPYMGPCPPTGIHRYI    86
CcMFT-L2       ---------------------------------ATKGRELVPYMGPCPPTGIHRYI    124
TcMFT-L2       ---------------------------------ATKGKEMVPYMGPQPPTGIHRYI    121
GrMFT-L2       ---------------------------------ATKGRELVAYMGPQPPTGIHRYI    121
GhMFT-L2-Dt    ---------------------------------ATKGRELVAYMGPQPPTGIHRYI    121
GaMFT-L2       ---------------------------------ATKGRELVAYMGPQPPTGIHRYI    121
GhMFT-L2-At    ---------------------------------STKGRELVAYMGPQPPTGIHRYI    121
SlSP5G         ---------------------------------ATFGNEVVGYESPRPSMGIHRYI    118
SlSP11D        ---------------------------------VTFGNEVVSYECPRPTMGIHRLV    120
AtFT           ----------------------------------TTFGNEIVCYENPSPTAGIHRVV    121
AtTSF          ---------------------------------NAFGNEVVCYESPRPPSGIHRIV    121
SpSP11C        ---------------------------------VTFGNEVICYESPRPSMGIHRIV    122
SpSP6A         ---------------------------------TRFGNEIVCYENPTPTMGIHRFV    119
SlSP3D         ---------------------------------SSFGQEIVSYESPRPSMGIHRFV    120
CoSFT          ---------------------------------ATFGQEVVCYESPRPTVGIHRFA    120
CcSFT          ---------------------------------ATFGQEVVCYESPRPTVGIHRFT    120
TcSFT          ---------------------------------ASFGQEVVCYESPRPTVGIHRFL    120
GaSFT          ---------------------------------ASFGQEVVCYESPRPTVGIHRFV    120
GhSFT-At       ---------------------------------ASFGQEVVCYESPRPTVGIHRFV    120
GrSFT          ---------------------------------ASFGQEVVCYESPRPTVGIHRFV    120
GhSFT-Dt       ---------------------------------ASFGQEVVCYESPRPTVGIHRFV    120
SlSP           ---------------------------------CSFGREVVGYEMPRPNIGIHRFV    122
AtCEN          ---------------------------------VSFGKEIIGYEMPRPNIGIHRFV    122
CoSP           ---------------------------------ASFGREVVNYEMPRPNIGIHRFV    107
CcSP           ---------------------------------ASFGREVVNYEMPRPNIGIHRFV    107
TcSP           ---------------------------------ATFGREVVNYEMPRPNIGIHRFV    121
GrSP           ---------------------------------ATFGREMVNYEMPRPNIGIHRFV    121
GhSP-Dt        ---------------------------------ATFGREMVNYEMPRPNIGIHRFV    121
GaSP           ---------------------------------ATFGREMVNYEMPRPNIGIHRFV    121
GhSP-At        ---------------------------------ATFGREMVNYEMPRPNIGIHRFV    121
AtTFL1         ---------------------------------ATFGKEVVSYELPRPSIGIHRFV    124
SlSP9D         ---------------------------------ATFGRELVSYETPRPNIGIHRFV    120
GaTFL1-L2      ---------------------------------ATFGREVVNYEIPRPDIGIHRFV    117
GhTFL1-L2-At   ---------------------------------ATFGREVVNYEIPRPDIGIHRFV    117
GrTFL1-L2      ---------------------------------ATFGREVVNYEIPRPDIGIHRFV    117
GhTFL1-L2-Dt   ---------------------------------ATFGREVVNYEIPRPDIGIHRFV    117
GrTFL1-L1      ---------------------------------ATFGREVVSYENPKPNIGIHRFV    120
GhTFL1-L1-Dt   ---------------------------------ATFGREVVSYENPKPNIGIHRFV    120
GaTFL1-L1      ---------------------------------ATFGREVVSYENPKPNIGIHRFV    120
GhTFL1-L1-At   ---------------------------------ATFGREVVSYENPKPNIGIHRFV    120
TcTFL1         ---------------------------------ATFGREVVSYEIPRPNIGIHRFV    120
CoTFL1         ---------------------------------ITFGREVVSYEIPRPNIGIHRFV    121
CcTFL1         ---------------------------------ATFGREVVSYEIPRPNIGIHRFV    121
AtBFT          ---------------------------------ASFGREIVRYETPKPVAGIHRYV    121
SlBFT-L1       ---------------------------------ASFGREIISYVNPKPVIGIHRYV    122
SlBFT-L2       ---------------------------------VSFGKEIVSYESPKPVIGIHRYV    122
SlBFT-L3       ---------------------------------ISFGREIVCYETPKPVIGIHRYV    123
TcBFT          ---------------------------------ASFGREVVSYETPKPTVGIHRYV    133
GaBFT-L2       ---------------------------------VSFGRELISYEAPNPAVGIHRYV    121
GhBFT-L2-At    ---------------------------------VSFGREVVSYEAPNPAVGIHRYV    121
GrBFT-L2       ---------------------------------VSFGREVVSYETPNPAVGIHRYV    121
GhBFT-L2-Dt    ---------------------------------VSFGREVVSYETPNPAVGIHRYV    121
GhBFT-L1-Dt    ---------------------------------ASFGREVVSYETPKPTVGIHRYV    121
GrBFT-L1       ---------------------------------ASFGREVVSYETPKPTVGIHRYV    121
GaBFT-L1       ---------------------------------ASFGREVISYETPKPTVGIHRYV    121
GhBFT-L1-At    ---------------------------------ASFGREVISYETPKPTVGIHRYV    121
CoBFT          ---------------------------------ASFGREVVGYETPKPTVGIHRYV    122
CcBFT          ---------------------------------ASFGREVVGYETPKPIVGIHRYV    122
                                                            *   ****
```

```
PpMFTL3      FVLFKQPTGK-PLLVTAPS---VRNNFNTRTFAVEHGLGFPVAATYFNAAKEPGSRRR--       225
PpMFTL4      FVLFRQPL-T-PFHITPPT---VRSNFNTRYFAAQCGLGLPVAATYLNAQKEPGSRRR--       192
PpMFTL1      FVLFKQPS-Q-SFLISPPA---ARNNFSTRNFAAYYGLGLPVAATYCNSQKEPASRNR--       192
PpMFTL2      FILFKQPS-T-PFLISPPT---VRNNFSTRNFASHYGLGLPVAATYCNAQKEPGSRRR--       192
SlMFT        LVLFRQNAPM-QEIFQAPV---ARAHFRTRMFAHQLDLGVPVATVYFNAHKEPANRKR*-       173
AtMFT        LVLFRQNSPV-GLMVQQPP---SRANFSTRMFAGHFDLGLPVATVYFNAQKEPASRRR--       173
CcMFT-L3     FVLFQQNSPM-GTAVQPPA---SRANFNTRLFADHLNLGLPVATVYINAQKEPISRRR--       159
CoMFT-L1     FVLFQQNGPM-GTAVQPPA---SRANFNTRLFADHLNLGLPVATVYFNAQKEPISRRR--       173
CcMFT-L1     FVLFQQNGPM-GTAVQPPA---SRANFNTRLFADHLNLGLPVATVYFNAQKEPISRRR--       173
TcMFT-L1     LVLFQQKGPL-GQ-VQQPA---SRANFSTRLFAQHLNLGQPVATVYFNAQKEPVSRRR*-       225
GaMFT-L1     LVLFQQKGPL-GA-VQQPA---TRANFSTRFFADHLNLGLPVATVYFNAQKEPVSRRR--       172
GrMFT-L1     LVLFQQKGPL-GA-VQQPA---TRANFSTRFFADHLNLGLPVATVYFNAQKEPVSRRR--       172
GhMFT-L1-Dt  LVLFQQKGPL-GA-VQQPA---TRANFSTRFFADHLNLGLPVATVYFNAQKEPVSRRR*-       172
GhMFT-L1-At  LVLFQQKGPL-GA-VQQPA---TRANFSTRFFADHLNLGLPVATVYFNAQKEPVSRRR*-       172
SlSP2G       FTLFRQKEAE-QVPHKPPQ---GRSNFKTRQFASDNGLDLPVAALYFNSQKEHAAHH---       178
CoMFT-L2     LALFKQERAAAAGGIQLPN---GRANFNTRQFAAQNGLGLPVAALYFNSHKEPALKKR--       141
CcMFT-L2     LALFKQGRAAEAGGIQLPN---GRANFNTRQFAAQNGLGLPVAALYFNSQKEPALKKR--       179
TcMFT-L2     LVLFKQERAT-EGGCQLPD---ARANFSTRQFAAQNSLGLPVAAVYFNSQKEPAVKKR*-       175
GrMFT-L2     LALFKQEGAM-EGRIQVAD---ARANFSTRRFAAQSRLGLPVAAVYFNSQKEPAAKKR--       175
GhMFT-L2-Dt  LALFKQEGAM-EGRIQVAD---ARANFSTRRFAAQNRLGIPVAAVYFNSQKEPAAKKR--       175
GaMFT-L2     LALFKQEGAM-EGRIQVAD---ARANFSTRRFAAQNRLGLPVAAVYFNSQKEPAAKKR--       175
GhMFT-L2-At  LALFKQEGAM-EGRIQVAD---ARANFSTRRFAAQNRLGLPVAAVYFNSQKEPAAKKR*-       175
SlSP5G       FVLYRQ-LGC-DAID-APDIIDSRQNFNTRDFARFHNLGLPVAAVYFNCNREGGTGGRRL       175
SlSP11D      LVLFRQ-LRR-EIIY-APE---NRQNFDTREFAKLYNFGLPVAAVYFNCQRENGTGGRRI       174
AtFT         FILFRQ-LGR-QTVY-APG---WRQNFNTREFAEIYNLGLPVAAVFYNCQRESGCGGRRL       175
AtTSF        LVLFRQ-LGR-QTVY-APG---WRQQFNTREFAEIYNLGLPVAASYFNCQRENGCGGRRT       175
SpSP11C      FSLFRQ-LGR-ETVY-APN---WRQNFNTRQFAELYNLGLPVAAVYFNCQRENGTGGRRC       176
SpSP6A       LVLFRQ-LGR-ETVY-PPG---WRQNF-------------------------------       140
SlSP3D       FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGSGGRRR       174
CoSFT        FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGSGGRRR       174
CcSFT        FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGSGGRRR       174
TcSFT        FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGSGGRRR       174
GaSFT        FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGGSGGRRT       174
GhSFT-At     FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGSGGRRT       174
GrSFT        FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGSGGRRT       174
GhSFT-Dt     FVLFRQ-LGR-QTVY-APG---WRQNFNTRDFAELYNLGLPVAAVYFNCQRESGSGGRRT       174
SlSP         FLLFKQ-KKR-QTISSAPV---SRDQFSSRKFSEENELGSRVAVYFNCQRETAARRR--       175
AtCEN        YLLFKQ-TRR-GSVVSVPS---YRDQFNTREFAHENDLGLPVAAVFFNCQRETAARRR--       175
CoSP         FLLFKQ-KRR-QTVRFIPT---SRDQFNTRKFAEDNELGLPVAAVYFNAQRETAARRR--       160
CcSP         FLLFKQ-KRR-QTVRFTPT---SRDQFNTRKFAEDNELGLPVAAVYFNAQRETAARRR--       160
TcSP         FLLFKQ-KRR-QAVISTPS---SRDHFNTRKFAAENELGLPVAAVYFNAQRETAARRR*-       174
GrSP         FLLFKQ-KGR-QTVRSIPS---SRDRFYTRKFAEENELGVPVAAVYFNAQRETAARRR--       174
GhSP-Dt      FLLFKQ-KGR-QTVRSIPS---SRDRFDTRKFAEENELGVPVAAVYFNAQRETAARRR*-       174
GaSP         FLLFKQ-KGR-QTVRSIPS---SRDRFDTRKFAEENELGVPVAAVYFNAQRETAARRR--       174
GhSP-At      FLLFKQ-KGR-QTVRSIPS---SRDRFDTRKFAEENELGVPVAAVYFNAQRETAARRR*-       174
AtTFL1       FVLFRQ-KQR-RVIFPNIP---SRDHFNTRKFAVEYDLGLPVAAFFNAQRETAARKR--       177
SlSP9D       FVLFKQ-KSR-SSVS-QPT---SRDHFNTRNFAQENNLEQPVTAVFFNAQRETAARRR--       172
GaTFL1-L2    FVLFKQ-KRR-QVIR-SPS---SRDNFNTRDFAAENDLDLPVAAVYFNARRETAARRR--       169
GhTFL1-L2-At FVLFKQ-KRR-QVIR-SPS---SRDNFNTRDFAAENDLDLPVAAVYFNARRETAARRR*-       169
GrTFL1-L2    FVLFKQ-KRR-QVIR-SPS---SRDNFNTRDFAAENDLGLPVAAVYFNARRETAARRR--       169
GhTFL1-L2-Dt FVLFKQ-KRR-QVIR-SPS---SRDNFNTRDFAAENDLGLPVAAVYFNARRETAARRR*-       169
GrTFL1-L1    FVLFKQ-KRR-QIIK-SPC---SRDNFNTRRFASENDLGLPVAAVYFNAQRETAARRR--       172
GhTFL1-L1-Dt FVLFKQ-KRR-QIIK-SPC---SRDNFNTRRFASENDLGLPVAAVYFNAQRETAARRR*-       172
GaTFL1-L1    FVLFKQ-KRR-QIIK-SPC---SRDNFNTRRFAFENDLGLPVAAVYFNAQRETAARRR--       172
GhTFL1-L1-At FVLFKQ-KRR-QIIK-SPC---SRDNFNTRRFASENDLGLPVAAVYFNAQRETAARRR*-       172
TcTFL1       FVLFKQ-KRR-QMIT-SPS---SRDNFSTRGFAAENDLGLPVAAVYFNAQRETAARRR*-       172
CoTFL1       FVLFKQ-KRR-QIIK-PPS---SRDNFSTRDFAAENDLGLPVAAVYFNAQRETAARRR--       173
CcTFL1       FVLFKQ-KRR-QIIK-PPS---SRDNFSTRDFAAENDLGLPVAAVYFNAQRETAARRR--       173
AtBFT        FALFKQ-RGR-QAVKAAPE---TRECFNTNAFSSYFGLSQPVAAVYFNAQRETAPRRRPS       176
SlBFT-L1     FVLYKQNRGR-QT-VKPSV---SRDHFNTRKFAVENGLGSPVAAVYFNAQRETAARRR*-       175
SlBFT-L2     FILYKQNRGR-QT-VKPPV---SRDHFNTNARKFAVENGLGSPVAAVYFNAQRETAARRR*-       175
SlBFT-L3     FLLYKQ-RGR-QT-VRAPA---TRDQFNTRSFSAENGLGSPVAAVYFNAQRETAARRR*-       175
TcBFT        FILFKQ-RGR-QT-VRPPT---SRDYFNTRRFSQENGLGLPVAAVYFNAQRETAARRR*-       185
GaBFT-L2     FILFKQ-RGR-RT-VKSPS---SRDYFNTRRFSAENGLGLPVAAVYFNAQRETAARRR--       173
GhBFT-L2-At  FILFKQ-RGR-RT-VKSPS---SRDYFNTRRFSAENGLGLPVAAVYFNAQRETAARRR*-       173
GrBFT-L2     FILFKQ-RGR-RT-VKSPS---SRDYFNTRRFSAENGLGLPVAAVYFNAQRETAARRR--       173
GhBFT-L2-Dt  FILFKQ-RGR-RT-VKSPS---SRDYFNTRRFSAENGLGLPVAAVYFNAQRETAARRR*-       173
GhBFT-L1-Dt  FVLFKQ-RGR-KT-VRPPS---SRDCFNTRRFSADNGLGLPVAAVYFNAQRETAARSRR*       174
GrBFT-L1     FVLFKQ-RGR-QT-VRPPS---SRDCFNTRRFSADNGLGLPVAAVYFNAQRETAARSRR-       174
GaBFT-L1     FVLFKQ-RGR-QT-VRPPS---SRDCFNTRRFSADNGLGLPVAAVYFNAQRETAARSRR-       174
GhBFT-L1-At  FVLFKQ-RGR-QT-VRPPS---SRDCFNTRRFSADNGLGLPVAAVYFNAQRETAARSRR*       174
CoBFT        FVLFKQ-RGR-QT-VRPPS---SRDYFNTRSFSEENGLGLPVAAVYFNAQRETAARRR--       174
CcBFT        FVLFKQ-RGR-QT-VRPPS---SRDYFNTRSFSEENGLGLPVAAVYFNAQRETAARRR--       174
             *::*                  *   *
```

```
PpMFTL3        --- 225
PpMFTL4        --- 192
PpMFTL1        --- 192
PpMFTL2        --- 192
SlMFT          --- 173
AtMFT          --- 173
CcMFT-L3       --- 159
CoMFT-L1       --- 173
CcMFT-L1       --- 173
TcMFT-L1       --- 225
GaMFT-L1       --- 172
GrMFT-L1       --- 172
GhMFT-L1-Dt    --- 172
GhMFT-L1-At    --- 172
SlSP2G         --- 178
CoMFT-L2       --- 141
CcMFT-L2       --- 179
TcMFT-L2       --- 175
GrMFT-L2       --- 175
GhMFT-L2-Dt    --- 175
GaMFT-L2       --- 175
GhMFT-L2-At    --- 175
SlSP5G         --- 175
SlSP11D        *-- 174
AtFT           --- 175
AtTSF          --- 175
SpSP11C        E*- 177
SpSP6A         --- 140
SlSP3D         SAD 177
CoSFT          --- 174
CcSFT          --- 174
TcSFT          *-- 174
GaSFT          --- 174
GhSFT-At       *-- 174
GrSFT          --- 174
GhSFT-Dt       *-- 174
SlSP           --- 175
AtCEN          --- 175
CoSP           --- 160
CcSP           --- 160
TcSP           --- 174
GrSP           --- 174
GhSP-Dt        --- 174
GaSP           --- 174
GhSP-At        --- 174
AtTFL1         --- 177
SlSP9D         --- 172
GaTFL1-L2      --- 169
GhTFL1-L2-At   --- 169
GrTFL1-L2      --- 169
GhTFL1-L2-Dt   --- 169
GrTFL1-L1      --- 172
GhTFL1-L1-Dt   --- 172
GaTFL1-L1      --- 172
GhTFL1-L1-At   --- 172
TcTFL1         --- 172
CoTFL1         --- 173
CcTFL1         --- 173
AtBFT          Y-- 177
SlBFT-L1       --- 175
SlBFT-L2       --- 175
SlBFT-L3       --- 175
TcBFT          --- 185
GaBFT-L2       --- 173
GhBFT-L2-At    --- 173
GrBFT-L2       --- 173
GhBFT-L2-Dt    --- 173
GhBFT-L1-Dt    --- 174
GrBFT-L1       --- 174
GaBFT-L1       --- 174
GhBFT-L1-At    --- 174
CoBFT          --- 174
CcBFT          --- 174
```

Figure 2.3 Alignment of CETS proteins. *G raimondii* (Gr)*, G. arboreum* (Ga)*,* and *G. hirsutum* (Gh,

NAU-NBI) CETS protein sequences are aligned with CETS from Arabidopsis (*Arabidopsis*

thaliana, At), jute (*Corchorus olitorius* and *Corchorus capsularis*, Co and Cc), cacao (*Theobroma cacao*, Tc), tomato (*Solanum lycopersicum* and *Solanum pimpinellifolium*, Sl and Sp), and moss (*Physcomitrella patens*, Pp) using Clustal Omega. (*) denotes identical residues, (:) indicates conserved residues (scoring >0.5 in the Gonnet PAM 250 matrix), and (.) denotes conservation between amino acids with weakly similar properties (scoring <0.5 in the Gonnet PAM 250 matrix).
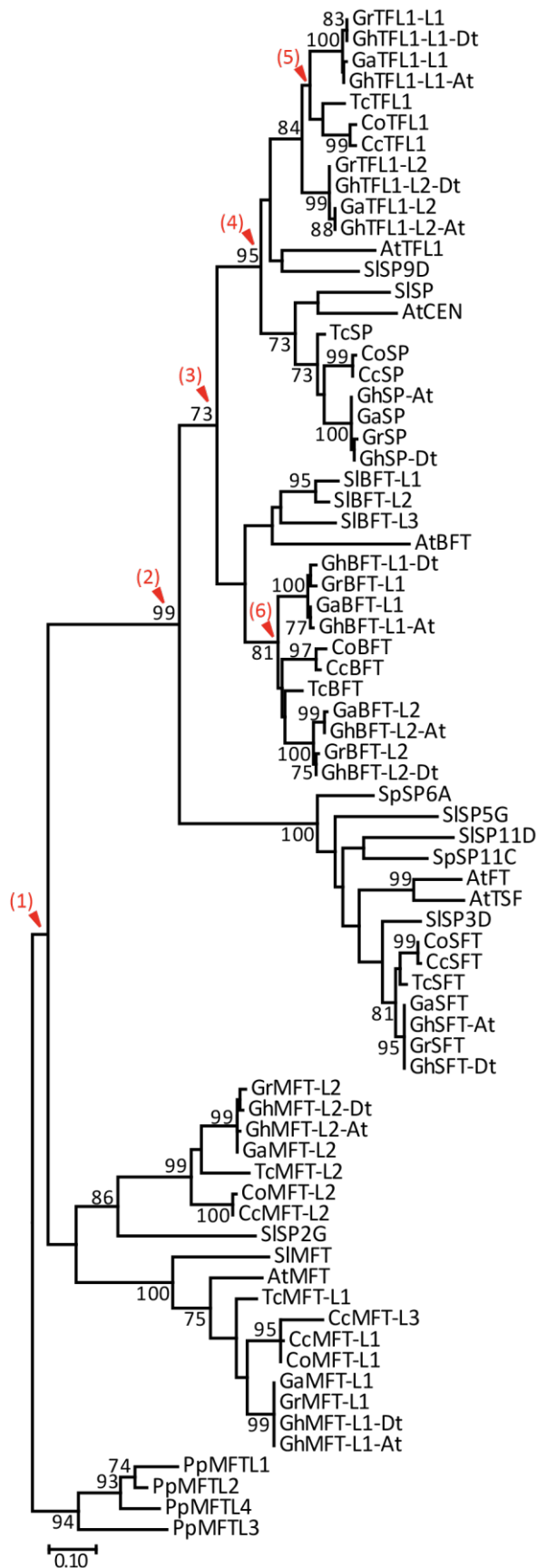
Figure 2.4 *Gossypium* CETS form three major clades. Shown is a phylogenetic tree constructed from the predicted polypeptide sequences of the putative *G. raimondii* (Gr), *G. arboreum* (Ga), and *G. hirsutum* (Gh, NAU-NBI assembly) homologs with Arabidopsis (*Arabidopsis thaliana*, At*),* tomato (*Solanum lycopersicum*, Sl), cacao (*Theobroma cacao*, Tc), jute (*Corchorus capsularis* and *Capsularis olitorius*, Cc and Co), and moss (*Physcomitrella patens*, Pp) CETS. The scale bar represents amino acid substitution frequency determined by the Poisson correction method. Duplication events in the evolution of CETS are noted by red arrows and parenthetical numbers. Evolutionary analyses were conducted in MEGA7.

CHAPTER 3

OVER-EXPRESSION OF COTTON *CETS* IN MODEL SPECIES *ARABIDOPSIS THALIANA* ALTERS PLANT

ARCHITECTURE

3.1   Introduction

Chapter 1 introduced the concept of florigen, discussed its role in both LD and SD

photoperiodic systems and presented evidence demonstrating *FT* homologs as the long-sought

florigen signal. A hallmark of the florigen concept, though, is its conservation even among

species whose flowering pathway is not determined by day-length. This conservation requires

four summary conditions. First, varied light conditions activate florigen in leaves, which moves

into apical meristems, where it promotes flowering. Secondly, while the environmental or

endogenous stimulus for florigen may vary between species, the signal is universal and

transferrable between systems. Third, both florigen and its response in meristems are

quantitative. Lastly, anti-florigenic agents exist which balance the activity of florigen in

meristems (Zeevaart, 1976; Lifschitz et al., 2014).

Evaluation of the florigen systems in varied species has demonstrated its universality

along with species-specific functions. Plants are characterized as LD, SD, or day-neutral

determined by their flowering response to photoperiod. Other classifications include those

based on life cycles including annuals, biennials, and perennials, or growth patterns as

monopodial or sympodial systems. The conservation of the *CO/FT* component in regulation of

LD and SD models, Arabidopsis and rice respectively, was addressed in Chapter 1. To

demonstrate the universality of the florigen system beyond photoperiodic systems into day-

neutral species, several experiments were performed using expression of the tomato *FT*

homolog *SFT* demonstrating its graft-transmissibility and substitution for light doses in tomato and photoperiods in both LD Arabidopsis and SD Maryland Mammoth tobacco (Lifschitz et al., 2006). In the biennial onion, three *FT* homolog were established in regulation of both flowering and bulb formation via photoperiod and cold exposure stimuli (Lee et al., 2013). In perennial *Arabis alpina*, *TFL1* homolog, *AaTFL1,* was shown to regulate flowering in subsequent years vernalization induction (Wang et al., 2011). Deciduous perennial, apple, maintains principally sympodial branching and controls flowering in lateral branches by a combination of endogenous and exogenous cues, expression of at least one apple *FT* homolog was demonstrated in the promotion of flowering through these cues (Vent, 2008; Kotoda et al., 2010; Haberman et al., 2016).

Despite this overall conservation of the florigen system and *CETS* function in plants, several studies in different systems stress that gene function should not be assigned based upon sequence conservation alone, especially when genome duplication events lead to paralogous *CETS* in a single system. Sugar beet comprises two *CETS* that fall into the *FT*-like subclade of *CETS* genes. When over-expressed in transgenic Arabidopsis, *BvFT2* accelerated flowering while *BvFT1* resulted in repressed flowering. Importantly, both sugar beet proteins carry peptide residues identified as functionally important FT residues; however, their activities were shown to be opposite of each otherwhen constitutively expressed in transgenic sugar beet and Arabidopsis (Pin et al., 2010b). In Sorghum, there are thirteen *CETS* homologs within the *FT*-subclade. Of these thirteen, six were functionally tested in Arabidopsis based upon differences in endogenous expression patterns. Transgenic Arabidopsis carrying hypothesized sorghum florigens, *SbFT1, SbFT8,* or *SbFT10* were drastically affected by transgene over-

expression. *SbFT1* expression driven by *2xCaMV35S~pro~*caused an onset of flowering 5 times

earlier control plants under LD conditions while heterologous expression of *SbFT8* or *SbFT10*

genes were more drastic. *SbFT8* plants flowered without forming an extended inflorescence

and comprised two curled cauline leaves; these plants died without setting seed. *SbFT10*

expression driven by the *2xCaMV35S~pro~* failed to produce transformed Arabidopsis for analysis,

and a weaker promoter was necessary to study its effects. These phenotypes demonstrate

varied levels of determinate growth functionality of these genes. On the other hand, three

other Sorghum *FT*-homologs, *SbFT2, SbFT6,* and *SbFT9,* had no impact on flowering when

constitutively expressed in transgenic Arabidopsis (Wolabu et al., 2016). Finally,

characterization of a soybean *BFT* paralog through over-expression in Arabidopsis displayed

unexpected floral promoting activity by the gene as opposed to a hypothesized floral repressing

function (Wang et al., 2015).

Study of gene function in cotton lags behind other model systems because of a lack of

mutant populations and cotton's recalcitrance to stable transformation. Cotton tissue culture is

a laborious task and total experimental time from transformation to regeneration of a new

plant and collection of seed can take 12 - 18 months. For these reasons, based on the

underlying assumption that cotton *CETS* will display conserved cross-species function,

functional characterization was studied through introduction of cotton genes into the

laboratory model species, *Arabidopsis thaliana.*

Gene function analysis usually begins by gain- or loss-of-function studies. Our aim was

to address the role of *CETS* genes in regulating plant architecture. Through constitutive

expression of the genes in Arabidopsis, insight was gained into the CETS protein function. For

this study, two copies of the constitutive *2xCaMV35S$_{pro}$* was used to drive expression of cotton

CETS genes in the Columbia-0 (Col-0) background to observe the effect of constitutive cotton

CETS gene expression.

3.2 Materials and methods

3.2.1 Vector Construction

Total RNA was extracted and purified using a combine hot borate protocol (Wan and

Wilkins, 1994) and column purification using Quick-RNA Miniprep (Plus) System according to

manufacturer's instruction (Zymo Research, Irvine, CA, USA). RNA was reverse transcribed using

oligo dT$_{18}$ and Superscript III (Invitrogen, Carlsbad, CA, USA). *CETS* coding sequences were PCR

amplified from cotton cDNA (oligonucleotides listed in Table 3.1), column purified using Wizard

SV Gel and PCR Clean-up System (Promega, Madison, WI, USA), digested with EcoRI and XbaI

(NEB, Beverly, MA, USA) and cloned into the same sites of overexpression vector pART7. The

expression cassette was released with digestion by NotI and cloned into the same site of binary

vector pART27 resulting in pART27-*2xCaMV35S$_{pro}$:GhCETS* constructs.

For *BFT-L1* and *TFL1-L2* genes, the coding sequence of these genes were synthesized

based on the known *G. raimondii* sequence before the release of the *G. hirsutum* genome

rather than being amplified from isolated *G. hirsutum* DNA; therefore, these constructs are

more accurately described as pART27-*2xCaMV35S$_{pro}$:GrCETS*.

3.2.2 Plant Material and Growth Conditions

pART27-*2xCaMV35S*$_{pro}$*:GhCETS* constructs were electroporated with the Bio-rad Gene Pulser Electroporation System using manufacturer's instruction (Bio-rad Laboratories, Hercules, CA, USA) into *Agrobacterium tumefaciens* strain GV3101-MP90 and transformed into Arabidopsis Col-0 CS7000 (Arabidopsis Biological Resource Center, Columbus, OH, USA) by floral dip method (Clough and Bent, 1998). T_1 transformants were selected on medium containing half-strength Murashige and Skoog nutrients (PhytoTechnologies Laboratories, Shawnee Mission, KS, USA), 1% sucrose, 2.5 g/L gelrite, and 100 μg/mL kanamycin. WT Col-0 seed were grown alongside on plates without kanamycin selection for comparison of phenotypes. Seeds were stratified for two days at 4°C then transferred to a 12-hour day, 22°C/18°C day/night temperature regime for germination and growth. After ten days, plants showing resistance to kanamycin were transferred to Fafard 3B Grower Mix soil (Sun Gro/Fafard, Agawam, MA, USA) and growth was continued in the same photoperiod and temperature regimes.

3.2.3 Flowering Time Assessment

Days to inflorescence was measured when the inflorescence measured 1 cm in height. Inflorescence to flowering was measured as days from inflorescence to observance of white petals. Rosette and cauline leaves were separately counted as additional measures.

3.2.4   Photography

Plant phenotypes were photographed with a Cannon SureShot A360 and a SPOT Insight 2 CCD camera (Diagnostic Instruments, Inc., Sterling Heights, MI, USA) mounted on a Nikon SMZ1500 stereomicroscope (Nikon, Melville, NY, USA).

3.3   Results

Gossypium hirsutum L. (A_tD_t subgenomes) has sixteen *CETS* genes, eight from each diploid background. To initiate functional characterization of this gene family, coding sequences for the eight *CETS* were cloned behind two copies of the *CaMV35S* enhancer and transcriptional start site and introduced into Arabidopsis Col-0. T_1 plants were grown under twelve-hour semi-LDs. Heterologous expression of the cotton *FT* ortholog, *GhSFT*, accelerated the transition to reproductive growth with the primary inflorescence evident by 20±2 DPG and 5±1 rosette leaves produced (*n*=18 plants, Fig 3.1A, D-E). The *2xCaMv35S_{pro}:GhSFT* inflorescence has 2±1 cauline leaves (*n*=18 plants, Fig 3.1A, F). In comparison, WT Arabidopsis produced an inflorescence by 36±5 DPG with 10±2 rosette leaves (*n*=18 plants, Fig 3.1B, D-E). The WT inflorescence had 3±1 cauline leaves (*n*=18 plants, Fig 3.1B, F). Five of eighteen independent *GhSFT* lines produced homeotic terminal flowers lacking sepals, petals, and stamens and harboring three unfused carpels surrounding an inner fused carpel (Fig 3.1C). Flowers preceding the terminus were normal. This terminal flower phenotype is consistent with overexpression of *AtFT* in WT Arabidopsis (Kardailsky, 1999; Kobayashi, 1999). This analysis demonstrates that cotton's *FT*-homolog, *GhSFT*, has conserved function with Arabidopsis *FT* and acts to promote determinate growth in Arabidopsis.

The transition to reproductive growth in Arabidopsis can be considered with respect to progression through three phases: vegetative rosette (V-phase), a primary inflorescence bearing cauline leaves subtending axillary branches (I1-phase), and a primary inflorescence bearing flowers (I2-phase) (Ratcliffe et al., 1998). A novel I1* phase was also described in Arabidopsis overexpressing *AtTFL1.* This phase is intermediate to I1 and I2 in which after the I1 phase, the primary inflorescence harbors axillary shoots that are not subtended by cauline leaves before transition into the I2 stage of development (Ratcliffe et al., 1998).

In our experiments, over-expression of each of the five cotton *TFL1*-like genes (*GhSP, GhTFL1-L1, GhTFL1-L2, GhBFT-L1,* and *GhBFT-L2*) in Col-0 extended V- and I1- phases. *GhSP, GhTFL1-L1,* and *GhTFL1-L2* exhibited the greatest impact on the extension of these phases, producing 2 to 3 times the number of rosette and 6-8 times the number of cauline leaves and transitioning to reproduction 4-15 days later than WT control plants (*n*=18 plants/construct, Fig 3.2A-C, F, K-M). Over-expression of *GhBFT-L1* and *GhBFT-L2* also extend these phases, but to a lesser degree, having ~1.5 times the number of rosette and 2-3 times the number of cauline leaves and producing an inflorescence 3-5 days later compared to WT controls (*n*=18 plants/construct, Fig 3.2D-F, K-M).

Over-expression of each cotton *TFL1*-like gene also resulted in the production of the intermediate I1* phase in transformed plants (pictured *GhTFL1-L1,* Fig 3.2G). Because these were T_1 generation plants, variation was great among individual transformed lines in this phase, with some plants never producing functional flowers. In individual lines that eventually formed flowers, I1* nodes became progressively more flower-like, having floral clusters surrounded by whorled leaf-like structures at the uppermost nodes (pictured *GhTFL1-L2,* Fig 3.2H). At times,

I1* structures produced abnormal flowers with floral buds originating from unfused carpels in the inner whorl (pictured *GhTFL1-*L2, Fig 3.2I). Furthermore, some *GhSP* (n = 8), *GhTFL1-L1* (n = 4), *GhTFL1-L2* (n = 2) and *GhBFT-L2* (n = 1) plants were still in the I1*-phase at experimental termination of 90 DPG (pictured *GhSP*, Fig 3.2J).

To fully understand the indeterminate effect of cotton *TFL1-*like genes during the reproduction, a fourth metric, 'days: inflorescence to flowering' in which flowering was defined as the sight of white petals above sepals on any flower, was collected. In WT Arabidopsis, flowering occurs two days after plants produce an inflorescence (Fig 3.2N). Comparatively, over-expression of all cotton *TFL1-*like genes significantly delayed floral production by 10 (*GhBFT-L1*) to 32 (*GhTFL1-L2*) days (Fig 3.2N). Moreover, much like some plants never shifted from I1* phase to I2, for each *TFL1-*like construct, several plants failed to produce normal flowers: *GhSP* (n = 8), *GhTFL1-L1* (n = 4), *GhTFL1-L2* (n = 8), *GhBFT-L1* (n = 1), and *GhBFT-L2* (n = 6)

Our results are similar to reports of *AtTFL1* overexpression as well as ectopic expression of other *TFL1* homologs from ryegrass, citrus, apple, sugarcane and soybean in Arabidopsis (Ratcliffe et al., 1998; Jensen et al., 2001; Pillitteri et al., 2004; Kotoda and Wada, 2005; Coelho et al., 2014; Baumann et al., 2015; Wang et al., 2015) and establish that each cotton TFL1-like protein promotes indeterminate growth. In cotton, the indeterminate activities of each cotton *TFL1-*like gene might contribute to the overall perennial nature of the plant and the maintenance of indeterminacy of cotton's monopodial main stem.

Ectopic expression of cotton *MFT-*like genes introduced into Arabidopsis had far less impact than overexpression of *FT-* and *TFL1-*like genes. In the vegetative phase, overexpression

of *GhMFT-L1* and *GhMFT-L2* marginally accelerated developmental progress (Fig 3.3A-E).

However, overexpression of these genes extended the I1-phase. While WT controls produced

3±1 cauline leaves, *GhMFT-L1* and *GhMFT-L2* generate 4±1 cauline leaves (Fig 3.3 A-C, F). The

acceleration of flowering time by *GhMFT* genes is consistent with a moderate acceleration of

flowering with *AtMFT* overexpression; in that study the total number of leaves was used as the

measure of accelerated flowering and plants over-expressing *AtMFT* flowered on average three

leaves earlier than WT controls (Yoo et al., 2004). A literature search did not reveal any other

independent analysis of *MFT* over-expression in the control of plant architecture. Interestingly,

while expression of all cotton *FT-* and *TFL1*-like genes in Arabidopsis produced homeotic flower

phenotypes, expression of *GhMFT*-like genes in Arabidopsis did not cause floral abnormalities.

This analysis suggests that cotton *MFT*-like genes effect in shoot meristems of cotton is weak

and that it is unlikely these genes significantly contribute to cotton's shoot architecture.

3.4   Discussion

In this study, cotton *CETS* coding sequences were expressed from a *2xCaMV35S$_{pro}$*,

leading to gene expression throughout transformed Arabidopsis. Overexpression studies are

important tools of molecular biology that aid in the dissection of gene function. An early over-

expression study involving over-expression of yeast genes affecting mitotic chromosomal

segregation highlights the value of over-expression when recessive mutants are unavailable as

is the case in cotton. In that study, over-expression of yeast genomic libraries aided in the

identification of important cell cycle regulators that hadn't previously been identified although

great efforts to isolate mutants had been made (Meeks-Wagner et al., 1986). In our study,

heterologous constitutive expression of cotton *FT-* and *TFL1*-like genes in Arabidopsis resulted in major shifts in the plant architecture of transformed plants. These findings are consistent with many studies over-expressing *CETS* from a variety of species. While, as described above, it is also known that minor changes in amino acid sequence can change the function of apparent homologs in the *CETS* gene family, this does not appear to be the case in *Gossypium* and demonstrates that the function of studied genes in Arabidopsis correlates with phylogenetic relationships identified in Chapter 2. This study highlights the conservation of function of *CETS* in cotton and suggests that cotton *FT-* and *TFL1*-like genes regulate cotton's perennial growth habit.

FT-like gene function for the most part can be characterized as promoting determinate growth, but variations in this function as described above exist. Over-expression of cotton's sole *FT*-ortholog, *GhSFT*, in Arabidopsis drastically hastened the transition to flowering, cutting the production of vegetative growth by half in this monopodial model indicating functional conservation with *FT* genes in other species. Since *SFT* is the sole *FT*-like homolog in cotton, this result suggests that *SFT* is likely cotton's florigen and an integral regulator of plant architecture. Virus-based mis-expression of *SFT* in photoperiodic and day-neutral varieties of cotton have confirmed this conserved role of *SFT* in the regulation of flowering and as cotton's florigenic component (McGarry et al., 2016). Silencing of *GhSFT* using a TRV construct in day-neutral, DeltaPine 61, resulted in significantly late flowering and increased vegetative growth. In line with the concept that *SFT*(*FT*) controls other aspects of growth determination, infected plants had larger main stem leaves and elongated petioles compared with controls. The florigenic function of *GhSFT* was tested using gain-of-function *dClCrV:GhSFT* constructs introduced into

photoperiodic TX701 and day-neutral cotton DeltaPine 61. Infected photoperiodic TX701 plants

flowered under noninductive long days. These plants had floral branches emerging as early as

node five as seen in day-neutral varieties and displayed prolific flowering. Again, compacted

internode length and reduced subtending leaves size in these plants suggest that *SFT* function

reaches beyond promotion of flowering into controlling determination of more growth aspects.

SFT overexpression in day neutral DeltaPine 61 caused more determinate growth of plants with

features like those detected in TX701. Plants transitioned to flowering early at node 4, leaves

were smaller, internodes were shorter and stems were thinner. These findings validate that

GhSFT encodes a florigenic signal that is not limited to photoperiod induction in photoperiod

sensitive lines; its activity is preserved and central to the early flowering and compact growth

habit desired in day-neutral varieties. In addition, the consistency of these finding in cotton

with phenotypes observed in transgenic Arabidopsis expressing *GhSFT* validates the idea of

using Arabidopsis as a surrogate for testing cotton *CETS* function.

Over-expression of each cotton *TFl1*-like homolog in Arabidopsis delayed the onset of

flowering and produced a bushy, highly vegetative architecture. These results are similar to the

demonstration of delayed flowering with all three Arabidopsis *TFL1*-like homologs. *AtTFL1,*

AtBFT, and *ATC* over-expression all delay the onset of flowering and cause floral abnormalities.

AtTFL1 but not *AtBFT* nor *ATC* overexpression is reported to cause a novel growth stage,

denoted I1*, similar to the observation in plants over-expressing cotton's *TFL1*-like genes (Yoo

et al., 2010; Hanano and Goto, 2011; Huang et al., 2012). This phenotype is also described in

Arabidopsis plants over-expressing *TFL1* homologs from citrus, *Lotus japonica* and other species

(Pillitteri et al., 2004; Guo et al., 2006).

Some T$_1$ plants over-expressing *GhSP*, *GhTFL1-L1*, *GhTFL1-L2* or *GhBFT-L2* developed highly branched structures, failed to produce flowers before experimental termination at 90 days, and were over 2 feet tall at the time of termination. This is the first description of such drastic phenotypes with *TFL1* over-expression; although, this might be due to termination of other experiments before the realization of this degree of phenotype rather than the lack of phenotype in other *TFL1*-like over-expression experiments.

Similar to virus-based experiments with *GhSFT*, VIGS construct *TRV:GhSP* demonstrated *GhSP*'s function in maintaining indeterminate growth in photoperiodic and day-neutral cotton systems (McGarry et al., 2016). Photoperiodic TX701 plants harboring *TRV:GhSP* flowered significantly early under noninductive LD conditions in comparison to controls. Plant main stems, which normally remain indeterminate in all cotton systems, terminated with a floral bud by node five. Additionally, all axillary meristems, including those subtended by the cotyledons, generated determinate floral buds directly on the main stem, abolishing branching and additional vegetative growth in these plants. These substantially determinate characteristics were also observed in *TRV:GhSP*-infected TX701 plants grown under inductive SD condition, and in *TRV:GhSP*-infected day-neutral DP61 grown under non-inductive LD conditions. Thus, silencing *GhSP* resulted in more synchronous flowering and confirmed that *GhSP* is needed to maintain indeterminate growth in both sympodial and monopodial branch systems. Similar experiments designed to target other cotton *TFL1*-like genes for silencing are underway, but for now their function in maintaining indeterminate growth that was observed in Arabidopsis has yet to be validated in cotton systems.

In 2003, a study found that *AtMFT* over-expression resulted in a slight acceleration of flowering as measured by a count of total rosette and cauline leaves under 16-hour LD conditions (Yoo et al., 2004). *MFT*-like genes are reported to predominately have effects on seed germination and embryo development. In this study, two cotton *MFT*-like genes were investigated with respect to their ability to affect plant architecture. Rosette and cauline leaves were assessed as separate measures, as opposed to total leaf count. Over-expression of either *GhMFT-L1* or *GhMFT-L2* slightly accelerates determinate growth as measured by days to inflorescence and number of rosette leaves. However, after reproductive transition *GhMFT-L1* and *GhMFT-L2* plants produced on average one more cauline leaf than WT control plants, which is interpreted as a delay in development during this stage of growth. If total number of leaves is considered in our study, *GhMFT-L1* and *GhMFT-L2* plants also flower on average one leaf earlier than WT controls similar to the report of *AtMFT* overexpression and the statistical significances of the separate measures on plant architecture is lost. This difference in measurement findings highlights that *GhMFT-L1* and *GhMFT-L2* effects on plant architecture are minimal in Arabidopsis and that *GhMFT*-like genes are not predicted to regulate flowering in cotton.

Finally, while heterologous expression studies using a constitutive promoter are useful for determining gene function, they do not inform on natural spatial, temporal or environmental regulation. Hence, while this study demonstrates that the gene products of cotton *FT*- and *TFL1*-like genes have the potential for regulating cotton plant architecture, cotton *CETS* activities are dependent on where and when they are expressed in the plant and further studies are required to validate *CETS* function. In the following studies, the use of

genomic clones and promoter studies allow for the consideration of *CETS* regulatory elements

impact on cotton *CETS* ability to regulate plant architecture.

Table 3.1 Oligonucleotides used for vector cloning in heterologous overexpression experiment.

Primer	Sequence
GhSP EcoRI ATG fwd	5' ctcgtggaattcatggcaaaactgtcagatcctctt
GhSP XbaI STOP rev	5' ctcgtgtctagattaggcgtcttctagcagctg
GhBFT-L1 EcoRI ATG fwd	5' ctcgtggaattcatgtcaagagtccccgaccca
GhBFT-L1 XbaI STOP rev	5' ctcgtgtctagatcatcttcttgatcttgcgg
GhSFT EcoRI ATG fwd	5' ctcgtggaattcatgcctagagatagagatcctttg
GhSFT XbaI STOP rev	5' ctcgtgtctagatcatgtcctacggccacc
GhTFL1-L1 EcoRI ATG fwd	5' ctcgtggaattcatggcaagggaagtagagcctc
GhTFL1-L1 XbaI STOP rev	5' ctcgtgtctagatcaacgtcttcttgcagctg
GhMFT-L1 EcoRI ATG fwd	5' ctcgtggaattcatggctgcctccgttgatcctc
GhMFT-L1 XbaI STOP rev	5' ctcgtgtctagatcaacgccttcggctgacg
GhBFT-L2 EcoRI ATG fwd	5' ctcgtggaattcatgtcaagggtccctgagccac
GhBFT-L2 XbaI STOP rev	5' ctcgtgtctagattatcttcttcttgcagcagtt
GhMFT-L2 EcoRI ATG fwd	5' ctcgtggaattcatggcccggtccgttgaacca
GhMFT-L2 XbaI STOP rev	5' ctcgtgtctagactaacgtttcttagctgctgg
GhTFL1-L2 EcoRI ATG fwd	5' ctcgtggaattcatgggagagcctctcattgttg
GhTFL1-L2 XbaI STOP rev	5' ctcgtgtctagattagcgtctccttgcagcag
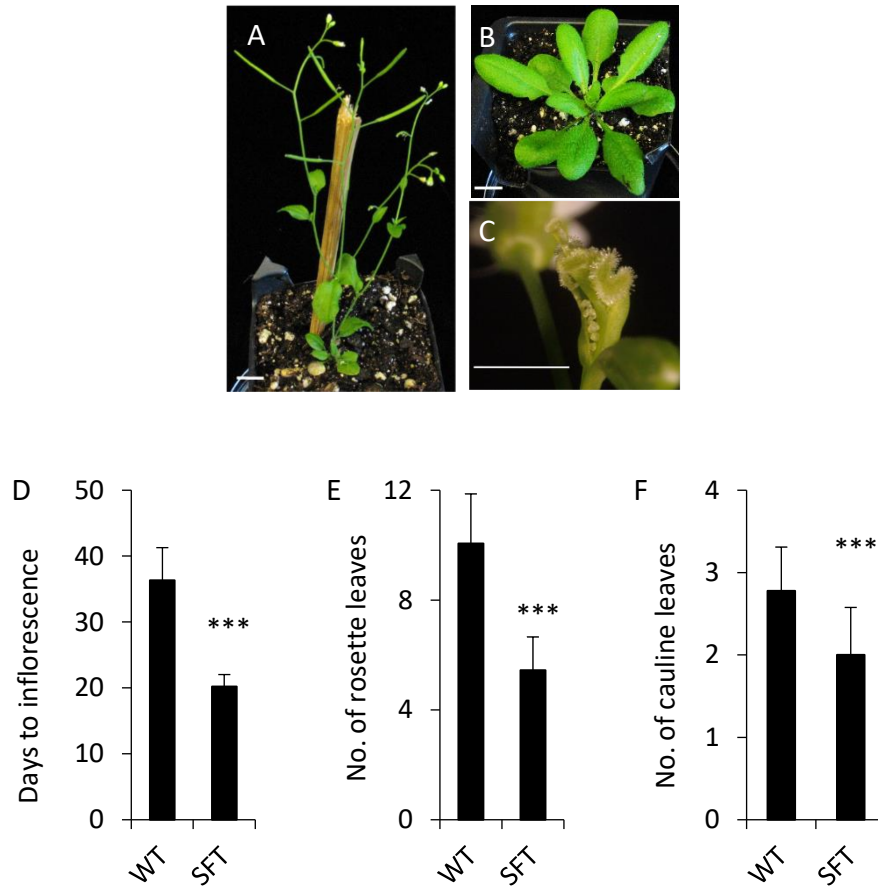
Figure 3.1 Accelerated determinate growth phenotype of transgenic Arabidopsis transformed with *2xCaMV35S:GhSFT*. (A) Early flowering 5-week-old *GhSFT* compared to (B) 5-week-old untransformed WT plant. Scale bars = 1 cm. (C) Homeotic terminal flower phenotype of *GhSFT*. Scale bar = 1 mm. Flowering time was assessed using (D) days to inflorescence, (E) number of rosette leaves, and (F) number of cauline leaves. Error Bars, ±SD. Significant differences from untransformed plants indicated with asterisks ($p<0.05$, *; $p<0.01$, **, $p<0.001$, ***) were determined using a two-tailed Student's t-Test.
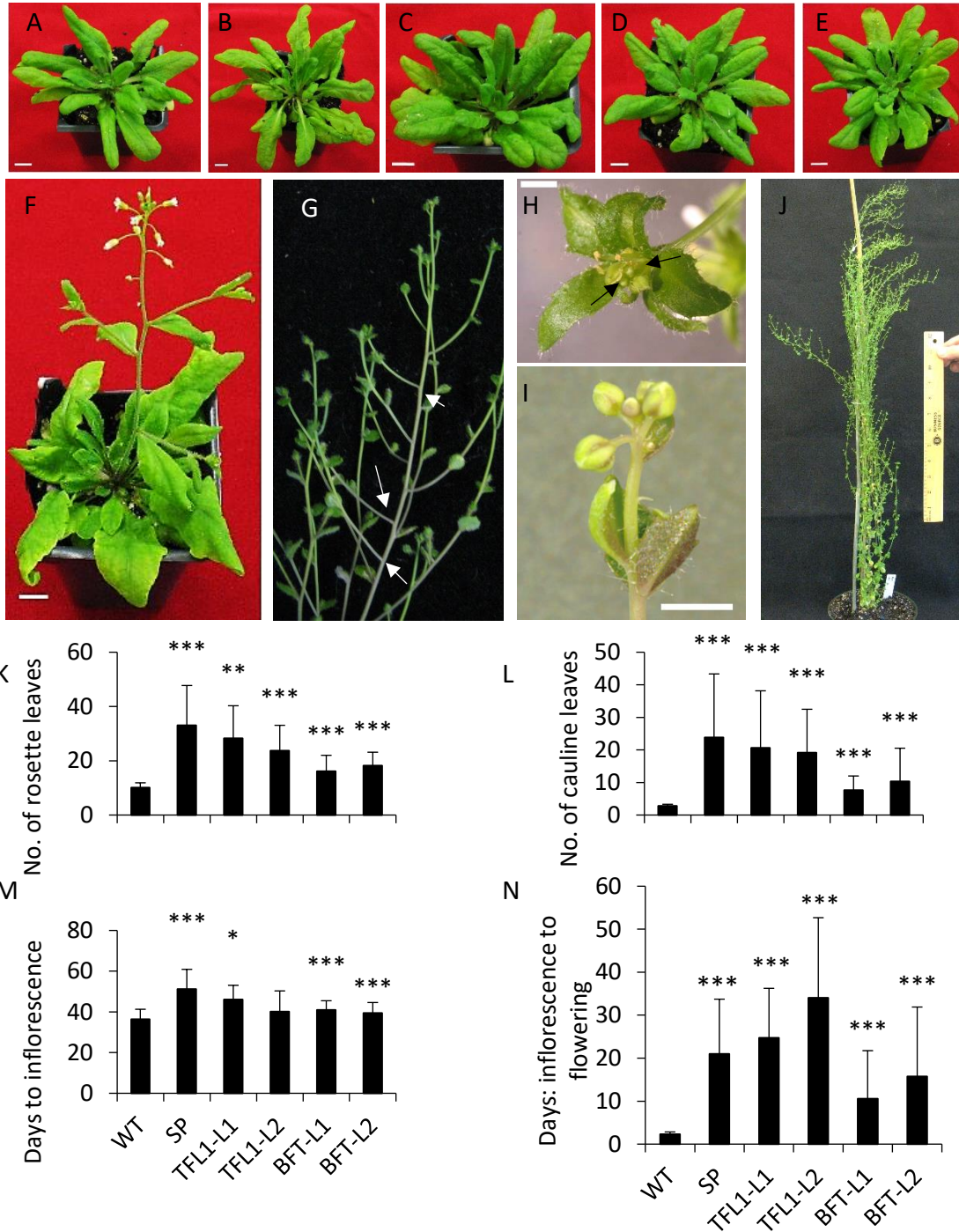
Figure 3.2 Phenotypes of transgenic Arabidopsis transformed with *2xCaMV35S:GhTFL1*-like

CETS. 7-week-old *GhSP* (A), *GhTFL1-L1* (B), *GhTFL1-L2* (C), *GhBFT-L1* (D), and *GhBFT-L2* (E)

delayed flowering phenotypes in comparison to flowering WT control (F). I1* phenotypes

produced by over-expression of cotton *TFL1*-like genes: (G) arrows point to a few of many the

I1* axillary branches of a *GhTFL1-L1* plant, (H) *GhTFL1-L2* I1* floral structure with abnormalities

and floral buds originating from the inner whorl of unfused carpels (black arrows), (I) *GhTLF1*

I1* branch showing a cluster of flowers surrounded by whorled leaf-like organs, and (J) *GhSP*

plant in I1* phase at 16-weeks old. Flowering time was assessed using number of rosette (K)

and cauline(L) leaves, days to inflorescence (M) and days:inflorescence to flowering (N). (A-G)
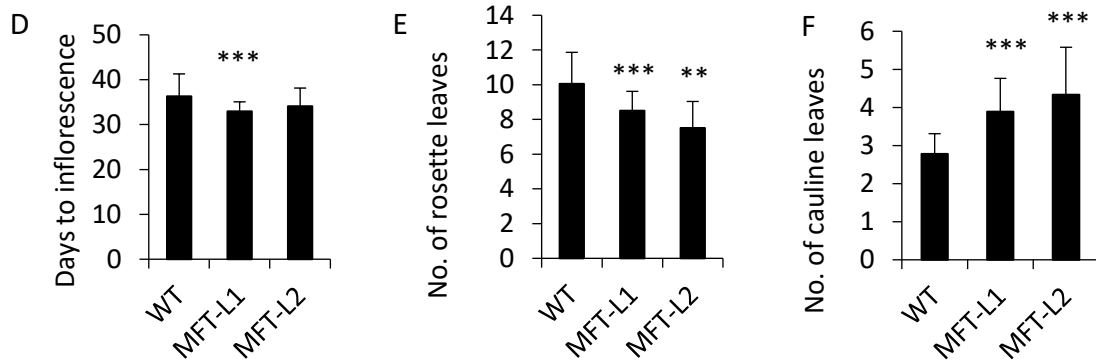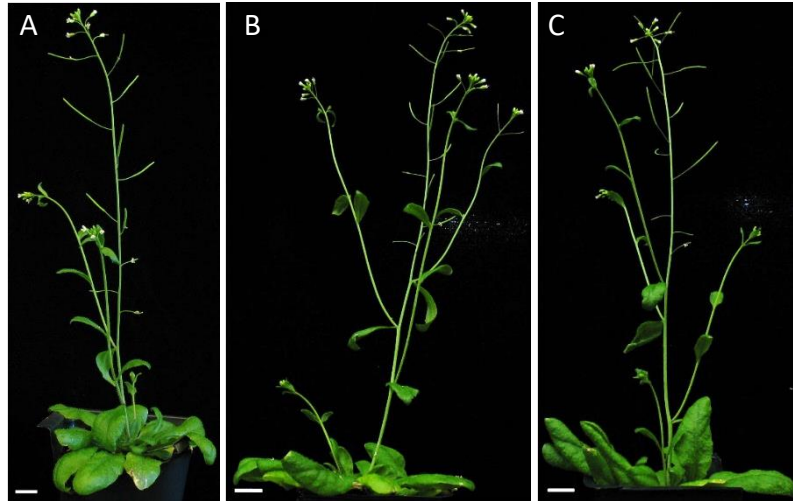
Scale bars = 1 cm. (H-I) Scale bars = 1 mm.

Figure 3.3 Phenotypes of transgenic Arabidopsis transformed with *2xCaMV35S:GhMFT*-like

CETS. (A) 6-week-old WT control compared to (B) *GhMFT-L1* and (C) *GhMFT-L2* of the same age.

Flowering time was assessed using days to inflorescence (D) and number of rosette (E) and

cauline (F) leaves.

CHAPTER 4

UNDERSTANDING THE ROLE OF COTTON *CETS* IN FLOWERING TIME REGULATION

THROUGH RESCUE ANALYSIS OF ARABIDOPSIS FLOWERING TIME MUTANTS WITH *Gossypium*

hirsutum GENOMIC CONSTRUCTS

4.1   Introduction

To test gene function, it is useful to introduce large genomic fragments that include the

whole gene of interest and surrounding regulatory elements, including the promoter, intron

sequences and transcriptional terminating sequences into a mutant system. This allows for an

analysis of temporal and spatial regulations placed on the gene of interest, along with gained

knowledge of the translated polypeptide's function. Introduction of a gene under native

regulation into WT plants can provide information, but full impact of the construct can also be

masked by the functional, endogenous gene. Alternatively, introduction into a known mutant

lacking the hypothesized function can be used in complementation or rescue studies to assess a

level of mutant functional restoration to WT level.

For Arabidopsis many flowering-time mutants are available along with several of each

CETS gene. *ft-10* is an insertional, loss-of-function mutant of *FT* having a T-DNA insert in the first

intron of the gene. The mutant plant is severely delayed in determinate growth under LD

conditions, flowering after 40.6 leaves in comparison to 15 leaves in WT Arabidopsis (Yoo et al.,

2005). This mutant has an indeterminate phenotype, producing many inflorescence branches.

In contrast, *TFL1* mutant, *tfl1-14,* is a strong early-flowering mutant allele. *tfl1-14* harbors a

single nucleotide substitution in its first exon causing missense mutation T69I. Under LD

conditions this mutant flowers early, has reduced plant height, a determinate primary

59

inflorescence with a terminal flower, and outgrowth of many axillary inflorescence, resulting

from the loss of apical dominance imposed by the primary inflorescence. The terminal region of

the primary inflorescence consists of a single or 2-3 clustered flowers. Often, the three outer

whorls of the terminating flower are missing or mosaic. Given these alterations in transitioning

to determinate growth, these mutants are valuable backgrounds for investigating cotton *CETS*

function under the control of their native regulatory sequences.

Generating genomic clones is often an arduous task because the size of the required

fragments reduces the probability of success in amplification and cloning. The options exist to

amplify smaller fragments of the desired amplicon (e.g. promoter, gene and terminator

individually) or if available, clone the entire sequence out of a genomic library (for instance,

from a BAC library). In both cases, traditional cloning is often restricted or made tedious by

limitations of PCR, limited unique restriction sites and many sub-cloning steps to produce a final

construct containing the full sequence of interest. Furthermore, manipulations to incorporate

desired tags such as a fluorescent molecule (Green Fluorescent Protein, GFP) or other reporter

modules (*uidA* gene) to follow expression patterns and protein localization is also complicated

by restriction sites and subsequent cloning steps. An alternative to bypass complex cloning

steps is to employ techniques of synthetic biology to accomplish one-step cloning tasks.

Synthetic biology tools, aimed at assembling DNA parts into organisms for useful

reprogramming, offers a faster and more straightforward path for organism modification.

Synthetic biology in yeast and microbes has seen fast advancement. After the spread of 'omics'

sciences and the sequencing of *Saccharomyces cerevisiae* and *Escherichia coli* genomes,

synthetic biology initiated in the first years of the twenty-first century. One of the first reports

of molecular engineering to reprogram cells for designed purpose was the building of a genetic toggle in *E. coli* in which repressible promoters drove mutually inhibitory gene products so that gene expression oscillated based upon external cues (Collins et al., 2000). A few years later, the precursor pathway for the antimalaria drug Artemisinin was engineered into *E. coli.* (Martin et al., 2003). Other notable advances in microbial synthetic biology include biofuel production using amino acid metabolism in *E. coli,* the description of Gibson Assembly cloning, the beginning of MIT's iGEM and the engineering of synthetic yeast chromosome arms (Atsumi et al., 2008; Gibson et al., 2009; Gibson, 2009). iGEM's BioBrick repository contains 35,000 biological components. However less than 1% of those components are plant-specific and the ability to effectively organize and assemble large DNA fragments for plant transformation is a major obstacle of plant synthetic biology.

Yeast homologous recombination has shown to be a powerful tool for molecular cloning and synthetic biology in microbial genomes (Larionov et al., 1997; Raymond et al., 2002; Kouprina and Larionov, 2008; Gibson, 2009; Shao et al., 2009). Here, to expand the success of yeast homologous recombination toward plant synthetic biology, a 4-system shuttle vector allowing for growth and selection in *E. coli*, *Saccharomyces cerevisiae*, and Agrobacterium and T-DNA transfer into plant cells was created. DNA fragments are assembled into the 4-system shuttle vector through overlapping regions of end homology that allow for *in vivo* crossover recombination. Because multiple DNA fragments can be concurrently assembled, yeast homologous recombination provides a rapid and robust system for plant synthetic biology. Using this system, a toolbox of clones containing cotton *CETS* genomic sequences for studying expression and function was also created. Cotton *CETS* genomic clones assembled through

yeast homologous recombination were introduced into Arabidopsis flowering-time mutants *ft-10* and *tfl1-14* to understand how *CETS* regulatory elements contribute to the genes' role in regulating plant architecture.

4.2   Materials and Methods

4.2.1   Plasmid Construction by Yeast Homologous Recombination

The 4-system shuttle vector, pSFP100, was created through yeast homologous recombination. For construct replication and selection in yeast, ARS-CEN-*HIS3* sequences were PCR amplified as a single 1,861 bp product from pRS313 (Sikorski and Hieter, 1989) using a Phusion/Phire polymerase mix (NEB) and a following touchdown PCR protocol: denaturation for 3 minutes at 98 °C followed by 12 cycles of denaturation at 98 °C for 5 seconds, annealing at 72-60 °C (-1 °C/cycle) for 20 seconds, and extension at 72 °C for 40 seconds, then 30 cycles of denaturation at 98 °C for 5 seconds, annealing at 60 °C for 20 seconds and extension at 72 °C for 40 seconds then a final extension of 72 °C for 10 minutes. For selection of transgenic plants harboring pSFP100 constructs, overlap extension PCR was employed to create gene cassette *NOS$_{pro}$::BAR:CaMV35SpA* for selection with glufosinate ammonia. The *NOS* promoter from pGPTV-BAR (Becker et al., 1992) and the *BAR:CaMV35S* site from pMDC123 (Curtis and Grossniklaus, 2003) were separately PCR amplified such that they overlapped by 20 bp at the ends to be fused. These products were used as templates for overlap extension PCR with only the outer primers to create a gene cassette. All three reactions were carried out using a Phusion/Phire polymerase mix. The *NOS* promoter from pGPTV-BAR was amplified with the following protocol: denaturation for 3 minutes at 98 °C followed by 30 cycles of 98 °C for 5

seconds, 66 °C for 20 seconds and 72 °C for 20 seconds then a final extension at 72 °C for 10 minutes. The *BAR:CaMV35S* site from pMDC123 was amplified by a touchdown PCR protocol: denaturation for 3 minutes at 98 °C followed by 12 cycles of denaturation at 98 °C for 10 seconds, annealing at 65 – 54 °C (-1 °C/cycle) for 20 seconds, and extension at 72 °C for 20 seconds, then 30 cycles of denaturation at 98 °C for 10 seconds, annealing at 55 °C for 20 seconds and extension at 72 °C for 20 seconds then a final extension of 72 °C for 10 minutes. Fusion PCR used to create the final gene cassette, *NOS_{pro}::BAR:CaMV35SpA* used: denaturation for 3 minutes at 98 °C followed by 30 cycles of denaturation at 98 °C for 5 seconds, annealing at 66 °C for 20 seconds and extension at 72 °C for 20 seconds then a final extension at 72 °C for 10 minutes. Both ARS-CEN-*HIS3* and *NOS_{pro}::BAR:CaMV35SpA* products were inserted into binary vector pCAMBIA0390 by yeast homologous recombination. Primer sequences are listed in Table 4.1 and Figure 4.1 diagrams homologous recombination overlaps and the final construct. To confer end homology for homologous recombination, reverse primer 'pRS313 nt 1189 rev' included a forty nucleotide 5' overhang homologous to a region overlapping the pUC9 multiple cloning site within the T-DNA of binary pCAMBIA0390. Similarly, reverse primer 'pMDC123-pA35 rev' comprised a forty nucleotide 5' overhang homologous to the region just inside and overlapping the left border of pCAMBIA0390. Forward primers 'pRS313 nt 4360 fwd' and 'pGPTV-BAR pNos fwd' each included twenty nucleotides of 5' overhang reciprocally homologous to 5' end of the PCR products to create forty nucleotides of homology between these PCR product ends. Both inserts replaced most of T-DNA sequence between the left and right borders of pCAMBIA0390 leaving an abbreviated multiple cloning site and *NOS* polyadenylation signal just inside the right border.

For homologous recombination in yeast, pCAMBIA0390 was linearized with HindIII (NEB) and all three linear fragments were introduced into yeast strain PJ694a (*MAT**a** trp1-901 leu2-3, 112 ura3-52 his3-200 gal4Δ gal80Δ LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacz*) using a Lithium Acetate (LiAC)/ssDNA/Polyethylene glycol (PEG) yeast transformation method (Gietz and Woods, 2006). A single PJ694a colony from a fresh YPD plate was inoculated into 10 mL YPD broth and incubated at 30 °C and 250 rpm overnight. The starter culture was inoculated into 40 mLs of pre-warmed YPD broth. Incubation at 30°C and 250 rpm continued 4-5 hours until an OD of $2x10^7$ cells/mL was reached. Cells were harvested by centrifugation at 3,000 x g for 5 minutes. The supernatant was discarded and cells were washed with 25 mL sterile water followed by 1 mL 100 mM LiAC then resuspended in 100 mM LiAC to a final concentration of 2 x 10^9 cells/mL. $1x10^8$ cells were used per transformation. A transformation mix of 240 μL 50 % PEG, 36 μL 1 M LiAC and 50 μL ssDNA (2 mg/mL) was added to the cells, followed by transforming DNA resuspended in a final volume of 34 μL sterile water. 500 ng of linearized pCAMBIA0390 and 300 ng PCR inserts were used as transforming DNA. Cells mixtures were heat-shocked at 42°C for thirty minutes with agitation every 2 minutes. Cell pellets were harvested by centrifugation at 6,000 rpm for 15 seconds. Supernatant was removed and cells were resuspended in 1 mL of sterile water. $2x10^7$ and $8x10^7$ cells of each transformation were plated separately onto yeast synthetic dropout plates lacking histidine and grown 3-5 days at 30 °C. Yeast DNA was harvested for analysis using the Harju 'Bust n Grab' protocol (Harju et al., 2004). Harvesting consisted of disrupting yeast cells by cycling exposure to extreme cold (liquid nitrogen) then extreme hot (95°C water bath), extracting globular proteins and cellular waste with chloroform extraction, and ethanol precipitation of DNA. DNA samples were resuspended

in 50 µL of water and 3 µL was used to transform XL1 Blue *E. coli* cells by electroporation as

previously described in Chapter 3.2.2 (Bio-rad Laboratories). Cells were plated onto LB medium

containing 50 µg/mL Kanamycin for selection. Selected *E. coli* transformants were transferred

into 2 mL of selective broth medium and cultured for DNA isolation. DNA was isolated from

bacterial cultures using an alkaline lysis miniprep protocol (Birnboim and Doly, 1979; Sambrook

et al., 1989) and analyzed by restriction digest. Two samples were Sanger sequenced through

the recombination region (MWG Operon, Louisville, KY, USA) for verification of correct

assembly. A pSFP100 clone was re-sequenced by semiconductor sequencing technology on a

Personal Genome Machine (PGM, Life Technologies, Grand Island, NY, USA). Details for Ion

Torrent semi-conductor sequencing on the PGM is described below.

pSFP100-*GhCETS* and pSFP100-*GhCETS-GFP* vector series were similarly assembled by

yeast homologous recombination. 2.0 kb promoter, gene (exon and intron) and 1.0 kb

terminator sequences were separately PCR amplified. Template genomic DNA was isolated by a

CTAB extraction protocol (Weigel and Glazebrook, 2002). 10-fold dilutions of genomic DNA

were used as templates in amplification reactions. Template source material, polymerase, and

cycling protocols used in *CETS* amplification reactions are listed in Table 4.2. Oligonucleotide

sequences for *CETS* genomic amplifications are listed in Table 4.3. Figure 4.2 depicts a diagram

of the homologous recombination strategy. To provide forty bps of end homology for

homologous recombination of the pSFP100-*GhCETS* series: 1.) Promoter-forward primers

'GrCETSXp fwd' carried forty nucleotides of homology overlapping the 3' end of *HIS3* yeast

selection marker in pSFP100. 2.) Promoter-reverse 'GrCETSpX rev' and gene-forward 'GrCETSx

fwd' primers were forty nucleotides in length and complements of one another. 3.) Primers

'bGrCETSX fwd' and 'bGrCETSX rev' were used to create dsDNA 'bridges' of homology between the 3' end of the gene PCR product (which excluded the stop codon) and the 5' end of the terminator PCR product. In this manner, bridges with stop codon or gene tags (GFP, GUS, c-Myc, epitopes, etc.) designed with the same flanking homologous sequences could be inserted to make the native protein or carboxy-in-frame fusions between reporter genes and genomic sequences. 4.) Terminator-reverse primers 'GrCETSXt rev' carried forty nucleotides of 5' overhang homologous to a region of the pSFP100 backbone in front of the *NOS* polyadenylation signal just inside the right border. Gene-reverse 'GrCETSX rev' and terminator-forward 'GrCETSXt fwd' primers contained only sequence-specific nucleotides. Gene-to-terminator 100 bp dsDNA 'bridges' were synthesized by annealing two single-stranded oligonucleotides, 'bGrCETSX fwd' and 'bGrCETSX rev', having twenty bp overlap in the presence of Phusion DNA polymerase and dNTPs in an abbreviated PCR protocol: denaturation at 98°C for 30 seconds followed by 15 cycles of 98°C for 5 seconds, 57°C for 10 seconds and 72°C then a final extension at 72°C for 10 minutes. Oligonucleotide sequences for dsDNA bridge fragment synthesis are listed in Table 4.4.

To assemble pSFP100-*GhCETS-GFP* vectors, meGFP was PCR amplified from p13ADAGLC_G (a synthetic monomeric enhanced GFP construct codon optimized for Arabidopsis) using oligonucleotides comprising homology to *CETS* genomic and terminator sequences to create in-frame fusions. Oligonucleotide sequences for meGFP amplification are listed in Table 4.5. Sequences were amplified using a Phusion/Phire polymerase mix and the following touchdown PCR protocol: denaturation for 3 minutes at 98 °C followed by 12 cycles of denaturation at 98 °C for 5 seconds, annealing at 72- 60 °C (-1 °C/cycle) for 20 seconds, and

extension at 72 °C for 40 seconds, then 30 cycles of denaturation at 98 °C for 5 seconds, annealing at 60 °C for 20 seconds and extension at 72 °C for 40 seconds then a final extension of 72 °C for 10 minutes. All PCR products were gel or column purified using Wizard SV Gel and PCR Clean-up system (Promega). DNA concentrations were determined either spectrophotometrically using a NanoDrop 2000 (Thermofisher Scientific, Richardson, TX, USA) or by gel electrophoresis resolution. In preparation for homologous recombination, plasmid pSFP100 was linearized with BamHI and EcoRI (NEB).

For each construct, linear DNA fragments were transformed into yeast PJ694a using the LiAc/ssDNA/PEG method (Gietz and Woods, 2006) described above. For analysis, DNA was isolated from either individual yeast colonies or total yeast scraped from transformation plates. When individual colonies were used for analysis, single colonies were picked and cultured in 2 mL of selective media overnight at 30°C. DNA was isolated from cultures using the Harju 'Bust n' Grab' protocol described above. When yeast DNA was harvested from total colonies scraped from transformation plates, 1 mL of sterile water was added to the transformation plate to create a slurry of yeast. Yeast was pelleted from the slurry by centrifugation and DNA was isolated following the Harju 'Bust n' Grab' protocol as described. Yeast DNA was electroporated into XL1 Blue *E. coli* cells and transformants were selected by growth on LB plates containing Kanamycin at 50 μg/mL. Samples were screened by restriction digest for appropriate size and two samples containing plasmids of interest were sequenced on a PGM (Life Technologies). Plasmids constructed for mutant complementation are: pSFP100-*GhSFT,* pSFP100-*GhSP*, pSFP100-*GhTFL1-L1*, pSFP100-*GhTFL1-L2,* pSFP100- *GhBFT-L1*, pSFP100-*GhBFT-L2*, pSFP100-*GhMFT-L1*, and pSFP100-*GhMFT-L2*.

4.2.2   Next Generation Sequencing on the Ion Torrent PGM Platform

In preparation for sequencing large plasmids, DNA samples were fragmented using

NEBNext Fast DNA and Library Preparation (NEB). Miniprep DNA was column purified using

Wizard SV Gel and PCR Clean-up system (Promega). Purified DNA was assessed for quantity and

quality spectrophotometrically at 260/280 and 260/230 using a NanoDrop 2000 (Thermofisher

Scientific). 1 µg of DNA in 15.5 µL was combined with 2 µL of NEBNext DNA Fragmentation

Reaction Buffer, 1 µL 100 µM of $MgCl_2$ and 1.5 µL NEBNext DNA Fragmentation Master Mix.

Fragmentation reactions were incubated in a thermal cycler for 20 min at 25°C then 10 min at

70°C for deactivation.

Barcoded adaptors (BIOO Scientific, Austin, TX, USA) were ligated onto fragmented DNA

following the NEBNext Fast DNA and Library Preparation Kit protocol (NEB). Fragmented

samples were mixed with 5 µL barcode adaptors, 5 µL P1 adaptor, 4 µL T4 DNA ligase buffer, 1

µL of Warmstart DNA polymerase and 4 µL of T4 DNA ligase. Ligation reactions were incubated

in a thermal cycler for 15 min at 25°C for ligation followed by 5 minutes at 65°C for

deactivation.

Adaptor-ligated DNA samples were size selected for 310 – 370 bps using AMPure XP

Bead-based Dual Bead Size Selection for 200 bp reads. 60 µL of 0.1X TE buffer was added to

adaptor-ligated DNA. To remove unwanted large DNA fragments, 90 µL of AMPure XP beads

were mixed with DNA and incubated at room temperature for 5 minutes. Beads containing

unwanted large fragments were collected on tube walls by incubation on a magnetic rack for 5

minutes. To remove unwanted small fragments of DNA, 15 µL of AMPure XP beads were

combined with the cleared supernatant in a fresh microcentrifuge tube and incubated at room

temperature for 5 minutes followed by incubation on a magnetic rack for 5 minutes. In this repetition, the supernatant comprising unwanted small fragments is discarded. Target DNA bound to AMPure XP beads was twice washed with 500 µL 80 % ethanol. DNA-bound beads were air dried for 5 minutes after removal of ethanol. Target DNA was eluted from beads by mixing of 45 µL 0.1X TE buffer. Beads were re-collected by incubation on a magnetic rack and 40 µL of cleared adaptor-ligated, size-selected DNA transferred to a clean microcentrifuge tube.

Size-selected, adaptor-ligated DNA was amplified by combination with 10 µL Equalizer Primers and 50 µL NEBNext High-Fidelity 2X PCR Master Mix and thermal cycling conditions: denaturing at 98°C for 30 sec followed by 6 cycles for 98°C for 10 sec, 58°C for 30 sec, and 72°C for 30 sec then a final extension of 72°C for 5 min.

Amplified DNA libraries were normalized using the Ion Library Equalizer kit (Life Technologies). 3 µL Equalizer beads/sample were combined with 6 µL of Equalizer Wash Buffer. Beads were pelleted on a magnetic rack for 3 min and supernatant discarded. Off the magnetic rack, beads were resuspended in 6 µL of Equalizer Wash Buffer. 10 µL of Equalizer Capture Solution was mixed with amplified DNA libraries and incubated at room temperature for 5 min. 6 µL of resuspended washed beads was combined to DNA solution; reactions were incubated at room temperature for 5 min. Reactions were incubated on the magnetic rack for 2 min for bead pelleting. Supernatant was removed and discarded. Beads were twice washed with 150 µL of Equalizer Wash Buffer. For elution of equalized DNA libraries, 100 µL of Equalizer Elution Buffer was mixed with bead pellets, beads were cleared by incubation on a magnetic rack and supernatants containing equalized 100 pM libraries were transferred to clean microcentrifuge

tubes. Barcoded, normalized libraries were then combined for continued sequencing preparation.

Combined libraries were prepared for emulsion PCR and enrichment using either Ion PGM Template OT2 200 kit or Ion PGM Template OT2 400 kit (Life Technologies). Combined DNA libraries were mixed with the following prepared amplification reagents: 500 μL Ion PGM Template OT2 200/400 Reagent Mix, 285 μL ion PGM OT2 200/400 PCR Reagent B, 50 μL ion PGM Template OT2 200/400 Enzyme Mix, and 40 μL Ion PGM Template OT2 200/400 Reagent X. 100 μL of resuspended Ion Sphere Particles (ISP) was added to the amplification reaction. Emulsion PCR to clonally amplify DNA libraries onto ISPs was performed on the Ion OneTouch 2 instrument (Life Technologies). To recover template-annealed ISPs, after the run, recovery tubes are centrifuged to pellet ISPs, ISPs are washed with 500 μL of Ion OneTouch Wash Solution and recollected by centrifugation, supernatant less 100 μL is removed and discarded and ISPs are resuspended in remaining wash solution. Prepared template-positive ISPs containing clonally amplified DNA were enriched for 'live' ISPs on the Ion OneTouch ES instrument (Life Technologies). A fresh 8 well strip was prepared for enrichment by adding: 100 μL of ISP sample to well 1, 130 μL of prepared Dynabeads MyOne Streptavidin C1 beads to well 2, 300 μL of ion OneTouch Wash Solution to each of wells 3-5, a 300 μL of melt-off solution (125 mM NaOH and 87% Tween) to well 7. Wells 6 and 8 remain empty. A fresh 0.2 mL PCR tube containing 10 μL of neutralization solution was loaded onto the instrument and the enrichment protocol was started.

The following parameters were used to create a planned run on the Ion Torrent server: application type—amplicon, run type—forward, template type—Ion PGM Template OT 200 or

400 kit (as appropriate), Sequencing kit—Ion PGM sequencing 200 or 400 kit (as appropriate),

flows—500 (for 200 bp reads), barcode set: IonXpress.

After PGM initialization, sequencing reactions were prepared using either Ion Torrent

Sequencing 200 or 400 kits (Life Technologies). 5 µL of Control ISPs were added to half of the

enriched library to be sequenced. ISPs were pelleted by centrifugation and all but 15 µL of

supernatant was removed. 12 µL of sequencing primer was added. The sequencing sample was

incubated in a thermal cycler for 2 min at 95 °C followed by 2 min at 37 °C, then set at room

temperature during chip check. A fresh 314 Chip was checked via the PGM chip check protocol.

After chip check, 3 µL of Ion PGM Sequencing 200/400 Polymerase was mixed to ISPs, the

sequencing reaction was loaded into the loading port of checked 314 chip and mixed twice.

Additional fluid was removed from the chip by centrifugation. To perform sequencing the

planned run was selected on the PGM and the ISP loaded chip was loaded into the PGM

instrument.

4.2.3   Analysis of Ion Torrent Next Generation Sequencing

Assemblies were first built *de Novo* from barcoded reads in Newbler *de Novo* Assembler

(Roche, Basel, Switzerland) then reassembled to *de Novo* built scaffold using Consed software

(Gordon et al., 1998) for clean-up. Sequence assemblies were aligned to *in silico* plasmid

designs using Clustal Omega (Sievers et al., 2011) to examine for correct assembly.

4.2.4 Plant Transformation and Growth Conditions

Sequence-verified plasmids were electroporated into *Agrobacterium tumefaciens* strain GV3101 MP90 as described in Chapter 3.2.2 (Bio-rad Laboratories) and transformed into Arabidopsis flowering-time mutants *tfl1-14* and *ft-10* using the floral dip method (Clough and Bent, 1998). T_1 seeds were sown on Fafard 3B potting soil (Sun Gro/Fafard), stratified for 2 days at 4°C, then moved to growth chambers for germination. Plants in the *ft-10* mutant background were transferred into 12-hour day, 22°C/18°C day/night temperature regime. Because the *tfl1-14* phenotype is most distinct under LD conditions, *tfl1-14* background plants were transferred into 16-hour day, 22°C/18°C day/night temperature regime. Five dpg, T_1 plants were spray-selected with 20 mg/mL glufosinate ammonia (Finale, Farnam Companies, Phoenix, AZ). Spray selection continued every other day for one week until healthy transformed plants could clearly be distinguished from dying untransformed siblings.

4.2.5 Flowering Time Assessment

T_1 plants were analyzed for restoration of WT plant architecture characteristics. Number of days to inflorescence was measured when inflorescence was 1 cm in height. Number of rosette leaves, cauline leaves and siliques were used as measured of flowering time. Pictures were captured with a Cannon SureShot A360.

4.3   Results

4.3.1   Assembling a 4-System Shuttle Vector for Plant Transformations

To investigate the value of yeast homologous recombination as an effective means for

assembling multiple large fragments of DNA for transformation, a 4-system shuttle vector,

pSFP100, was assembled through yeast homologous recombination. A $NOS_{pro}::BAR:CaMV35SpA$

fusion PCR product was inserted just inside the left border of pCAMBIA0390 to confer BASTA

resistance for simple transformant selection on soil (pCAMBIA0390 lacks plant selection). Since

transfer of T-DNA into plant genomes initiates at the right border, placement of the selectable

marker at the left border ensures that selected transformants contain the entire T-DNA insert.

The *NOS* promoter was chosen for expression of the plant selectable marker since it does not

contain strong enhancers known to influence expression of distant genes. The ARS4-CEN6-*HIS3*

sequences incorporated into pSFP100 provide for replication and selection in yeast. Forty

nucleotides of end homology were created through 5' overhangs of amplification primers for

crossover recombination of flanking fragments in the assembly of pSFP100. Transformation of

yeast cells with overlapping DNA fragments to produce the 4-system binary shuttle vector

pSFP100 resulted in a transformation efficiency of 1.63 x 10^4 as measured in number of colony

forming units (CFUs)/µg of transforming backbone DNA. Comparatively, transformation of

intact autonomously replicating plasmids by the LiAC/ssDNA/PEG method results in an

efficiency of 1 x 10^6 CFUs/µg DNA as reported by (Gietz and Woods, 2006). This indicates that

the process of homologous recombination reduces transformation efficiency by 100-fold.

Negative controls designed to establish the ability of separate fragments to confer

prototrophy to histidine-auxotrophic PJ694a were performed. When PCR product ARS-CEN-

HIS3 was transformed alone, the efficiency of transformation was 1.32 x 10^4 CFUs/µg DNA.

Linear DNA is not functional in yeast prototrophy; this high efficiency may have resulted from:

1.) residual circular plasmid in the PCR mix, 2.) circularization of the ARS-CEN-*HIS3* fragment, 3.)

integration into the yeast genome, or 4.) spontaneous reversion. Negative controls having no

DNA input or a mix of digested pCAMBIA0390 and *NOS$_{pro}$::BAR:CaMV35SpA* PCR product

produced 140 and 135 colonies, respectively (Table 4.6). These colonies had a white, dry flaky

morphology that is a known class of false positives produced at a low rate in transformation of

yeast strain PJ694a (Philip James, news group communication, 14 November 1997).

Ten pSFP100 yeast colonies were selected for analysis. To avoid known false positives,

colonies selected were creamy (not flaky) and slightly pink owing to the *ade2* mutation. After

yeast DNA isolation and transformation into *E. coli*, samples were analyzed for correct assembly

by restriction digest analysis. Restriction digest analysis confirmed correct assembly of pSFP100

in all analyzed samples as evidenced by identical banding patterning (Fig 4.3). This

demonstrates that forty bps of end homology between flanking DNA fragments is sufficient for

the assembly of two inserts into a linear backbone vector. Sanger sequencing of 3.8 kb from

two DNA samples revealed seamless recombination of the inserts into pCAMBIA0390.

Semiconductor sequencing of one clone concurred with these results.

Available plasmid maps of pRS313 from Addgene (catalog number 77142) and NCBI

(accession number U03439.1) describe the orientation of the yeast replication and selection

units in the 5' to 3' direction as CEN6 (nt 4395 – 4511), ARS4 (nt 4512 -4855), then *HIS3* (nt 503

– 1162). Primers used to amplify these sequences for assembly into pSFP100 were positioned

outside the sequences of interest, amplifying nt 4360 – 1189. Each sequencing of pSFP100

clones demonstrated the 5' – 3' orientation of these element to be ARS4 (nt 4395 – 4738),

CEN6 (nt 4739 – 4855), then *HIS3* (nt 503 – 1162), indicating that ARS4 and CEN6 units are

opposite of described orientation in available maps. PGM semi-conductor sequencing of

pSFP100 and pRS313 agreed with the 5'–ARS4-CEN6-3' orientation rather than published 5'-

CEN6-ARS4-3' publicly available sequence data.

4.3.2   Results of Assembling Cotton *CETS* Genomic Inserts into pSFP100

Simultaneous cloning of multiple inserts into a single construct for plant transformation

is invaluable for a variety of projects such as linking gene modules (e.g. promoters, 5' UTR, ORF,

3' UTR, terminator, etc.) or metabolic pathway engineering where a pathway needs to be

introduced. To demonstrate that our 4-system shuttle vector pSFP100 was suitable for one-step

cloning of multiple fragments into a binary vector for plant transformations, homologous

recombination was used to assemble pSFP100-*GhCETS* and pSFP100-*GhCETS-GFP* vector series.

These vector series were created to analyze the functions of genomic *GhCETS* genes under

native control and observe protein localization in Arabidopsis.

To assess the effect of input DNA on transformation efficiency, various amounts of

linearized plasmid (pSFP100) and insert DNA were transformed into PJ694a. Transformation

with 500 ng of linearized backbone, and PCR inserts and dsDNA bridges in equimolar ratio to

linearized backbone resulted in transformation efficiencies of 2.3 x 10^3 – 1.2 x 10^4 CFUs/µg

backbone DNA. To test whether lower DNA inputs and higher insert/backbone ratios would

increase efficiency, conditions of: 1.) 250 ng linearized backbone, PCR inserts at 2-fold molar

excess and bridge fragments at 5-fold molar excess; and 2.) 200 or 100 ng of linearized

75

backbone and equal volumes of PCR inserts and dsDNA bridges were analyzed. When less

plasmid DNA and greater ratios of insert/plasmid concentration were used for transformation,

efficiencies increased (as calculated in CFUs/µg of linearized plasmid) with highest efficiencies

observed when 200 ng of linearized backbone and equal volumes of inserts were used to fill the

volume of transforming DNA aliquot (34 µL of DNA is used to transform cells)(Table 4.7). 8 µL of

each insert were used in these transformations at concentrations ranging from 10 ng/µL – 75

ng/µL of PCR products, resulting in 80 – 600 ng of inserts used for transformations. These

amounts put the molar excess of inserts to linearized pSFP100 at 4-fold or higher, indicating

that higher transformation efficiencies can be realized by driving the homologous

recombination mechanisms forward via providing excess of the smaller molecules. This is

analogous to providing excess substrates to accelerate enzymatic reactions.

Vector series pSFP100-*GhCETS* and pSFP100-*GhCETS-GFP* comprised of cotton *CETS*

promoter, genomic and terminator sequences separately amplified and these sequences with

an in-frame fusion of *GFP* with the genomic sequence, respectively, were assembled. 5'

overhang sequences in cloning primers were designed to provide forty nucleotides of homology

between flanking construct fragments. When one-fifth of total transformed cells were plated

onto synthetic dropout media lacking histidine, 200 – 4,500 colonies (depending on quantities

of transforming DNA as discussed above) formed after 2-3 days incubation at 30 °C (Table 4.7).

To analyze the efficiency of homologous recombination, yeast DNA was isolated either from

individual colonies or a slurry of total transformants harvested from selection plates with sterile

water, transformed into *E. coli* for plasmid bulking then samples of plasmid DNA isolated from

E. coli were screened by restriction. A sample was assessed as correctly assembled if the

expected banding pattern was observed after digestion. In most cases, these digests showed that more than one species of plasmid was created by homologous recombination during the transformation event. However, the correct banding pattern was always observed during analysis and usually in most samples (Table 4.8, Fig 4.4A). Efficiency of homologous recombination of the assemblies was calculated as number of samples showing correct banding pattern/number of total analyzed samples; efficiency ranged from 37 – 100 %. This percentage was generally higher when total yeast DNA rather than individual colonies were screened (Table 4.8), but also correlated with the use of molar excess of inserts in comparison to linearized backbone. In constructing vector pSFP100-*GhSP*, three banding patterns were observed in an initial restriction digest analysis. Sequencing two of the species revealed that the difference in banding pattern was due to incorporation of homeologous terminator sequence in different clones, *GhSP-A$_t$* terminator sequence into clone #1 and *GhSP-D$_t$* into clone #2, rather than a problem with the homologous recombination vector (Fig 4.4C). Therefore, it is likely that inaccuracies in engineering (or assessing) a correct construct via homologous recombination can be a result of unexpected *in vitro* chemistry rather than a defect of yeast homologous recombination. These experiments demonstrate that yeast homologous recombination is an effective method for assembling multiple fragments into a single construct for plant transformation.

For each pSFP100-*GhCETS* assembly, pSFP100-*GhSP-GFP,* and pSFP100-*GhBFT-L2-GFP*, two samples showing the correct banding pattern were sequenced using Ion Torrent next-generation sequencing technology. Barcoding of assembly samples allowed for up to 16 samples to be sequenced in one run providing greater sequencing efficiency when it is

necessary to analyze multiple samples. Ion Torrent sequencing provided an abundance of

sequence information for each clone generating greater than an average of 200 sequence read

depth (Fig 4.4B). Each sample analyzed showed seamless and accurate recombination at each

of five sites of homologous recombination (Fig 4.4C). Within the pSFP100 plasmid region,

sequencing assemblies demonstrated a perfect match to the expected *in silico* designs,

representing the reliability of the Ion Torrent technology. On occasion mismatches of

sequencing assemblies to expected *in silico* pSFP100-*GhCETS* designed was observed within

cloned *CETS* genomic regions. This is represented in (Fig 4.4C) for pSFP100-*GhSP* sequenced

clones #1 and #2. *GhSP-D$_t$* promoter and gene sequences were incorporated into both clones,

while clone #1 contains the *GhSP-A$_t$* terminator sequence and clone #2 contains the *GhSP-D$_t$*

sequence as discussed above. Within their promoter sequences both clones share three single

nt mismatches with the expected *GhSP-D$_t$* sequence (Fig 4.4C). These mismatches may be

representative of sequence deviation between the *G. hirsutum* TM-1 genome reference

sequence (phytozome.net) and *G. hirsutum* DeltaPine 61 from which I isolated genomic

sequences or sequencing errors in the *G. hirsutum* TM-1 assembly. Clone #2 also carried one

additional single nucleotide mismatch to the expected *GhSP-D$_t$* sequence within its promoter

sequence which is probably from PCR error. Sequencing assemblies for both clones were

perfect matches to the expected *GhSP-D$_t$* sequence within their exon and intron regions (Fig

4.4C). Similarly, the sequence assembly for clone #1 completely matched the expected *GhSP-A$_t$*

terminator sequences and the assembly for clone #2 has sequence identity with the *GhSP-D$_t$*

terminator sequence (Fig 4.4C)*. Based upon the results of these experiments, Figure 4.5

describes a general workflow for homologous recombination into binary vector pSFP100 for use

in plant transformations, and incorporates technical aspects learned empirically for most

efficient workflow.

4.3.3   Results of Arabidopsis Mutant Rescue with Cotton *CETS* genomic clones

To further gauge cotton *CETS* regulation of flowering, we tested the ability of genomic

clones to rescue Arabidopsis flowering-time mutants *ft-10* and *tfl1-14*. Introduction of genomic

clones into the Arabidopsis mutants allowed us to consider temporal and spatial regulations

placed on the gene of interest along with gained knowledge of the translated polypeptide's

function. Genomic clones were comprised of 2.0 kb sequences upstream of the ATG start

codon, the full genomic sequence (exons and introns) and 1.0 kb downstream of the stop

codon. Each genomic clone was introduced into both Arabidopsis flowering-time mutants.

Arabidopsis *ft-10* is a loss of function mutant severely delayed in determinate growth,

flowering very late and producing a 'bushy', highly vegetative architecture in comparison to WT

Arabidopsis (Yoo et al., 2005). In our experiments, plants were studied in the T_1 generation and

grown in 12-hour days. *ft-10* (*n*=20) and *ft-10*, pSFP100 empty vector (EV) (*n*=13) plants were

significantly more indeterminate than WT plants, producing an inflorescence 15 days later and

generating 2 and 6 times the number of rosette and cauline leaves, respectively, (Fig 4.6A-C). A

GhSFT genomic clone (*n*=11) fully rescued the *ft-10* mutant during vegetative development,

producing an inflorescence a day later and generating a similar number of rosette leaves as WT

plants (Fig 4.6A-B, D). However, rescue during the I1 phase of reproduction was incomplete in

GhSFT plants; these plants produced a number of cauline leaves intermediate to WT and the *ft-*

10 mutant (Fig 4.6C-D). These results indicate that *GhSFT*'s role in promoting determinate

growth in these plants was equivalent to the role of *AtFT* during vegetative growth, but that its activity was diminished as plants aged, and full rescue was not realized. Of the other seven cotton *CETS* genomic clones tested in the *ft-10* mutant background, none demonstrated any level of rescue (Fig 4.6A-C).

In contrast to late-flowering *ft* mutants, *tfl1-14* is a strong early flowering mutant. Cotton *CETS* genomic clones were introduced into the *tfl1-14* mutant background and studied the T_1 generation under LD (16/8) conditions to test their ability to rescue loss of *TFL1* function. In our hands, *tfl1-14* plants behaved as previously described. The untransformed mutant generated an inflorescence five days earlier and produced six fewer rosette leaves than WT controls (*n* = 12, Fig 4.7A-B, E and G). Without exception, *tfl1-14* plants became determinate, producing only four flowers on the flanks of the primary inflorescence before consumption of the primary SAM (Fig 4.7C-E). EV control plants flowered slightly later than *tfl1-14* plants (*n*=12, Fig 4.7A-C, E-F); the delay in these plants may be due to slow early development in the selection of transgenic plants, which WT and untransformed mutant controls were not exposed to. All EV control SAMs became determinate producing terminal flowers. *GhTFL1-L2, GhSP,* and *GhBFT-L2* genomic clones partially rescued the early flowering *tfl1-14* phenotype. A *GhTFL1-L2* genomic clone displayed the highest level of rescue with plants forming an inflorescence a similar number of dpg as WT (*n*=12, Fig 4.7A, G-H). *GhTFL1-L2* plants produced twice as many rosette and cauline leaves than the mutant and EV plants (Fig 4.7B-C, E-F, H), but not as many as WT (*n*=11, Fig 4.7B-C, G-H). The inflorescence SAM of plants expressing *GhTFL1-L2* also remained indeterminate longer than *tfl1-14* and EV control; these plants set on average of two and a half times more the number of siliques from the primary inflorescence before SAM

consumption (Fig 4.7D). Taken together, this evidence indicates that the *GhTFL1-L2* genomic

clone function is most comparable to that of *AtTFL1* in Arabidopsis. *GhBFT-L2* and *GhSP*

genomic clones also demonstrated partial rescue of the mutant phenotype. Plants harboring a

GhBFT-L2 genomic clone generated twice as many rosette leaves as EV counterparts and two

times more siliques than *tfl1-14* and EV plants before SAM termination (n=12, Fig 4.7A-D, E-G,

I). These results demonstrate that *GhBFT-L2* promotes indeterminate growth in the

heterologous system, although its activity is not robust enough to fully rescue the absence of

AtTFL1 activity. Similarly, *GhSP* had a partial rescuing effect on the mutant phenotype.

Interestingly, its rescue was only observed during reproduction where *GhSP* expression resulted

in the production of more cauline leaves and a SAM that remained indeterminate for

significantly longer than mutant and EV controls; these plants generated on average twice as

many siliques before SAM termination (n=12, Fig 4.7C-F, J). The remaining cotton *TFL1*-like

genes failed to produce any significant rescuing results, including close *AtTFL1* homolog

GhTFL1-L1 (Fig 4.7A-D).

A *GhSFT* genomic clone introduced into the *tfl1-14* background significantly accelerated

the terminal flower phenotype of *tfl1-14*. These plants produced fewer cauline leaves and

siliques than mutant and EV controls (n=12, Fig 4.7C-F and K). This supports *GhSFT*'s role as a

determinate growth factor.

4.4 Discussion

4.4.1 Yeast Homologous Recombination is an Effective Tool for Large-Scale DNA Construction

Plant synthetic biology trails behind microbial synthetic biology largely due to intractability of plant systems. Cloning and assembly of DNA fragments through traditional methods that rely on site-specific digestion and ligation can unnecessarily impede plant synthetic biology during these steps. To address this problem, a 4-system shuttle vector for one-step, large-scale DNA assembly via yeast homologous recombination for plant transformations was constructed. Unlike traditional methods, homologous recombination aligns complementary sequences and allows replacement between homologous fragments by crossover recombination in yeast cells. Yeast homologous recombination is more efficient than in other organisms and has been an important tool for cloning and mutagenesis in the synthetic biology community (Längle-Rouault and Jacobs, 1995; Oldenburg et al., 1997; Raymond et al., 1999; Raymond et al., 2002; Anderson and Haj-Ahmad, 2003). Notably, the method was used for assembly of biosynthetic pathways in yeast (Shao et al., 2009). Here, yeast homologous recombination was used to assemble large constructs containing *G. hirsutum CETS* genomic sequences for plant transformation. As proof of concept, the shuttle vector itself was successfully assembled through yeast homologous recombination. This shuttle vector was created by the replacement of most of the T-DNA sequence of plant binary vector pCAMBIA0390 with a BASTA resistance plant selection cassette and yeast replication and selection sequences. This vector system uniquely allows for one-step assembly of multiple DNA fragments directly into a plant binary vector, avoiding multiple cloning processes that involve sequential insertion, and thereby shortening the cloning process time.

Unlike related *in vitro* methods, like Gibson Assembly, that require enzymatic treatment with exonuclease, DNA polymerase and DNA ligase before transformation, cloning by yeast homologous recombination requires only linear DNA preparation via PCR or restriction digestion, and one-step yeast transformation, and yields high assembly efficiency. In these experiments assembly of four fragments using forty bps of end homology between flanking fragments yielded efficiencies of 40 – 100 % with higher efficiencies correlating with increases in the amount of insert fragments used in comparison to linearized backbone DNA. These results correlate with assembly of biochemical pathways into yeast using the same method and fifty bps of flanking DNA fragment overlap for the assembly of five inserts in a linear backbone (Shao et al., 2009). While we did not try using different lengths of overlap, that report observed higher efficiencies using longer overlaps between flanking inserts when more than five inserts were assembled. However, as discussed here, manipulating the ratio of inserts to linear plasmid also increased efficiency using fifty bps of overlap for greater number of inserts (Shao et al., 2009). This indicates that higher efficiencies in the method can be realized using: 1.) longer regions of end homology or 2.) increased ratios of inserts to plasmid. However, the assembly of even greater inserts into a single construct may require a combination of both strategies for correct assembly of constructs.

Improvements could be made to our vector system. First, in our design, yeast sequences were inserted inside the left and right T-DNA borders so that during plant transformations with *Agrobacterium*, these yeast sequences are also transferred into the plant genome. A major societal concern of GE crops is the insertion of foreign DNA, into plant genomes (Blancke et al., 2015). Most regulatory authorities prefer the removal of selectable markers and other

sequences not directly related to the trait of interest from plant genomes for genetically engineered crop approval (Vigani et al., 2014). For this reason, a binary vector in which yeast replication and selection sequences are outside the left and right borders would be preferred over the current construct.

Second, pSFP100 lacks standardization with the broader plant synthetic biology community. BioBricks is presently the most popular synthetic biology standardization method. The BioBrick repository catalogs over 35,000 BioBrick parts comprising elements such as protein coding, regulatory, ribosome binding sites, and terminators which follow a restriction enzyme assembly standard. Each BioBrick part is standardized by the flanking of standardized prefix and suffix sequences and are used in the design and assembly of synthetic biology circuits through combination of individual parts. This system still requires multiple parallel pairwise assemblies when more than two BioBricks are required in an assembly.

MoClo cloning based upon Golden Gate cloning was introduced to relieve some limitations of BioBrick-based assembly. The Golden Gate system takes advantage of Type IIS restriction enzymes which cleave outside the recognition sequence and allow DNA fragments with compatible ends to efficiently assemble without leaving behind restriction enzyme recognition sequence 'scars' that can complicate subsequent cloning. The method allows the creation of multigene constructs to be assembled in three sequential cloning steps (Weber et al., 2011).

jStack uses yeast homologous recombination in the cloning of MoClo and Golden Gate fragments into a plant binary vector for greater flexibility than can be achieved by MoClo assembly (Shih et al., 2016). jStack strategy employs successive rounds of cloning to pair

community standardization with the ease of yeast homologous recombination Importantly, as a first step in jStack, modules that preserve MoClo standardized fusion sites for promoters, UTRs, signal peptides, coding sequences, and terminators are created. Additionally, linker modules are created for providing overlap sequences used in homologous recombination. In a second cloning step, modules containing a linker-promoter-coding sequence-terminator-linker unit are formed using Type IIS restriction enzymes like MoClo reactions. In step two of the assembly, any number of step one modules can be assembled in parallel in preparation for cloning of multiple gene cassette into a single construct by yeast homologous recombination in step three. Next, step two linker-promoter-coding sequence-terminator-linker are released from their modules by flanking rare cutter restriction sites and a binary vector containing yeast ARS4-CEN6 and Leu selection sequences is linearized. Linearization releases a URA3 dropout cassette such that selection for leucine prototrophy and counter selection with 5-FOA dramatically reduces background from incomplete vector digestion. In a third cloning step, all linear fragments are introduced into a yeast system for homologous recombination and correct constructs are selected with Leu dropout media. Similar to the method employed here, linker sequences provide homology to flanking DNA fragments, but 200 bps of homology was used in the jStack strategy rather than forty. This system improves on the 4-system shuttle vector created here by exclusion of yeast sequences from the final T-DNA insert, DNA part end compatibility with current plant synthetic biology communities, and the use of the URA3 dropout for negative selection. However, while this compliance with Golden Gate standardization allows for compatibility with the plant synthetic biology community, it also requires multiple rounds of cloning; whereas, our system, although non-compliant, achieves

one-step cloning which can expedite the cloning process. Notably, design of our yeast homologous recombination initiated in 2014 while the jStack method was published in 2016 indicating that both projects were probably proceeding concurrently.

4.4.2 Cotton *CETS* Genomic Clone Rescue of Arabidopsis Flowering-Time Mutants

Cotton *FT-* and *TFL1-* like genes regulate the timing of reproductive phase change when constitutively expressed in Arabidopsis from a *2xCaMV35S$_{pro}$*. To further elucidate the strength of effect each cotton *CETS* gene has upon this physiological process, we used *ft-10* and *tfl1-14* flowering time mutants transformed with *CETS* genomic clones in a mutant rescue. The inclusion of native promoter and regulatory sequences provides a finer layer of detail to the study of gene function that cannot be realized using the *35S CaMV* constitutive promoter which lacks the temporal and spatial resolution allowed by inclusion of native regulatory sequences. *ft-10* and *tfl1-14* mutants are severely shifted from normal timing of reproductive phase change, either flowering considerably late or very early, respectively, in comparison to WT lines with the same Col-0 background.

As expected, *GhSFT* is the only cotton *CETS* genomic clone that rescued the late flowering *ft-10* mutant. In this study a 1,947 bp sequence upstream of the ATG start codon of *GhSFT* was used to drive expression of *GhSFT*. Analysis of regulatory elements in a heterologous host can be an effective assay since *cis-* and *trans*-acting factors are conserved across species. For example, Galactinol synthase (GAS) is the first committed enzyme leading to the synthesis of raffinose and stachyose. In *Cucumis melo,* the synthesis of these sugars is integral to phloem loading and *CmGAS1* is expressed specifically in minor veins. Both Arabidopsis and tobacco

differ from melon in the anatomy of leaf venation with tobacco having higher orders of venation. These species also use different mechanisms of phloem-loading, and while Arabidopsis synthesizes a small amount of galactinol in comparison to melon, tobacco does not appear to synthesis galactinol at all. However despite these anatomical and biochemical differences, the *CmGAS1* promoter was also able to drive the minor-vein specific expression of reporter gene *uidA* in both of the distant species demonstrating conservation gene expression regulation across species (Haritatos et al., 2000). However, heterologous expression assays also have limitations since genome sequences may not have conservation. Also, it is well known that regulatory elements can be distant from the gene of interest and may not be captured in our T-DNA constructs. For example there are regulatory elements 5.7 kbs upstream *AtFT* required to provide any level of complement of *ft-10* by an *AtFT$_{pro}$:FT* construct (Adrian et al., 2010). Our demonstration of partial rescue by the *GhSFT* genomic clone containing only a 2.0 kb promoter sequence implies that regulation of *GhSFT* expression does not require the presence of these more distally located promoter elements and may suggest different regulatory mechanisms, for example epigenetic effects, controlling the expression of *AtFT* and *GhSFT*.

To understand if cotton *CETS* genomic clones would rescue the early flowering *tfl1-14* mutant, we introduced cotton *CETS* genomic clones into the mutant background and studied T$_1$ generation plants under LD conditions where the mutant phenotype can be detected. *GhTFL1-L2, GhBFT-L2* and *GhSP* genomic clones each partially rescued the *tfl1-14* phenotype. *AtTFL1* ortholog *GhTFL1-L2* provided the highest level of rescue showing functional conservation of these distant orthologs. *GhBFT-L2* and *GhSP* genomic clones also partially rescued the *tfl1-14* mutant phenotype. *GhBFT-L2* plants transitioned to reproductive growth later and retarded the

terminal flower phenotype, but failed to maintain meristem indeterminacy. This is comparable

to *AtBFT* overexpression studied in the *tfl1-20* mutant background. *AtBFT* driven by the

CaMV35S$_{pro}$ delayed the early flowering of *tfl1-20*, but it also failed to rescue the terminal

flower phenotype (Yoo et al., 2010). Additionally, much like *GhBFT-L2, AtBFT* over-expression in

the WT background drastically delayed flowering and produced homeotic flowers. Comparing

these results in the two different background suggests that, similar to *AtBFT*, *GhBFT-L2*

functions in maintaining meristem indeterminacy, but that this function is probably redundant

to other indeterminate factors such as *GhTFL1-L2 or GhSP*. A *GhSP* genomic clone rescue of *tfl1-*

14 is limited to the reproductive phase of Arabidopsis growth. Accordingly, *ATC* has been

shown to act redundantly to *AtTFL1* primarily in a SD-dependent role (Huang et al., 2012).

Interestingly, transient silencing of *GhSP* in cotton systems has robust effects on cotton plant

architecture while this effect has yet to be demonstrated in efforts to silence *GhTFL1-L1* or

GhTFL1-L2 underscoring that plants of varied architecture probably utilize different pathways

and mechanisms to control meristem states (McGarry et al., 2016 and unpublished data). While

in Arabidopsis maintenance of meristem indeterminacy primarily relies on *TFL1* activities to

appropriately maintain indeterminacy, cotton's indeterminacy thus far appears to rely chiefly

on *GhSP* activity. Neither *GhTFL1-L1* and *GhBFT-L1* genomic clones introduced into *tfl1-14*

resulted in phenotypes different than EV controls implying these genes under native promoter

control fail to restore the loss of *TFL1* activity. In our phylogenetic analysis, both genes are

Gossypium-specific and paralogs of *AtTFL1* ortholog *GhTFL1-L2* and *AtBFT* ortholog *GhBFT-L2*

(orthology was determined using reciprocal blastp queries, Fig 2.3). This suggests that these

paralogs are redundant and have lost the capacity to regulate plant architecture. However,

there may be other explanations such as gene silencing or a failure to capture the full gene

regulatory elements in my genomic constructs.

Table 4.1 Oligonucleotides used in construction of the 4-system shuttle vector (pSFP100). Colored nucleotides are 5' overhangs providing homology for homologous recombination cloning with binary vector pCAMBIA0390, while black nucleotides are sequence-specific for amplifying inserts used to create the shuttle vector.

| primer	sequence 5'
pRS313 nt 4360 fwd	gtcgtttcccgccttcagttacatttccccgaaaagtgcc
pRS313 nt 1189 rev	ccatggtggtggactcctcttagaattcccggggatccgtcgactcct gcaggtttaaataatcggtg
pGPTV-BAR pNOS fwd	ggcacttttcggggaaatgtaactgaaggcgggaaacgacgatcatga gcggagaattaagg
pGPTV-BAR pNOS rev	gcgaaacgatccagatccggtgcag
pMDC123-BAR fwd	gcaccggatctggatcgtttcgcatgagcccagaacgacg
pMDC123-pA35 rev	ggcaggatatattgtggtgtaaacaaattgacgcttagacagacaaac ttaataacacattgcggacg
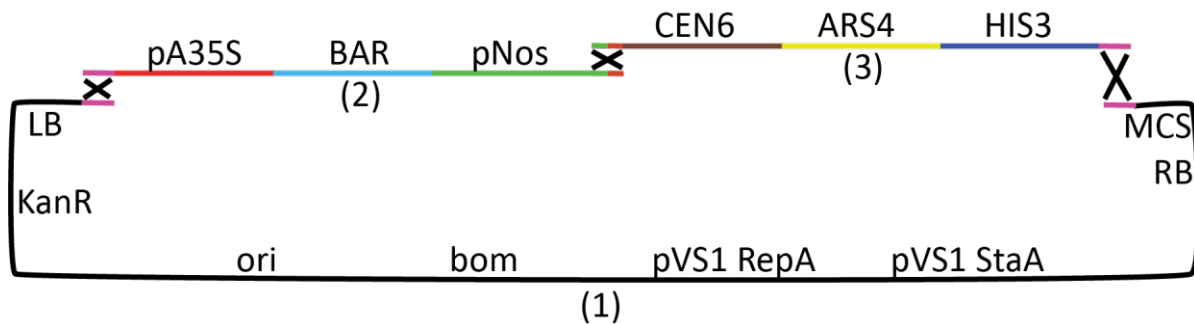
Figure 4.1 pSFP100, a four-system shuttle vector for plant transformation created using homologous recombination. Shown is the assembly strategy for creating shuttle vector, pSFP100. (1) Linearized binary vector, pCAMBIA0390, was used as a backbone for the construction of pSFP100 and contains sequences for selection and replication in *E. coli* and Agrobacterium as well as binary vector T-DNA sequences for plant transformation. (2) Fusion PCR fragment, *NOSp::BAR:CaMV35SpA*, allows for selection of transgenic plants. (3) PCR fragment, CEN6-ARS4-*HIS3*, allows for yeast growth and selection during homologous recombination assembly. Coloring at both ends of the PCR fragments indicate sections of 40 bp end homology to neighboring assembly fragments that was engineered by PCR primer design. X's represent regions of homology overlap for crossover recombination.

Table 4.2   Reagents and protocols used for PCR amplification of cotton *CETS* genomic fragments.

Genomic Fragment	Genomic DNA Template	Polymerase	Cycling protocol			
GhSPp GhSP GhSPt	*G. hirsutum* DeltaPine61	Kappa HiFi	Denaturation	98 °C	5 minutes	1 cycle
			Denaturation	98 °C	20 seconds	35 cycles
			Annealing	59 °C	15 seconds	
			Extension	72 °C	30 seconds	
			Extension	72 °C	10 minutes	1 cycle
GhBFT-L1p GhBFT-L1t GhSFTt GhTFL1-L1p GhBFT-L2p GhBFT-L2t	*G. hirsutum* DeltaPine61	Phusion/Phire	Denaturation	98 °C	3 minutes	1 cycle
			Denaturation	98 °C	5 seconds	12 cycles
			Annealing	66-54 °C (-1 °C/cycle)	10 seconds	
			Extension	72 °C	1 minute	
			Denaturation	98 °C	5 seconds	30 cycles
			Annealing	54 °C	10 seconds	
			Extension	72 °C	1 minute	
			Extension	72 °C	10 minutes	1 cycle
GhBFT-L1 GhBFT-L2 GhTFL1-L2p GhTFL1-L2 GhTFL1-L2t GhMFT-L1t	*G. hirsutum* DeltaPine61	Phusion/Phire	Denaturation	98 °C	3 minutes	1 cycle
			Denaturation	98 °C	5 seconds	12 cycles
			Annealing	72-60 °C (-1 °C/cycle)	20 seconds	
			Extension	72 °C	40 seconds	
			Denaturation	98 °C	5 seconds	30 cycles
			Annealing	60 °C	20 seconds	
			Extension	72 °C	40 seconds	
			Extension	72 °C	10 minutes	1 cycle
GhSFTp GhSFT GhTFL1-L1 GhTFL1-L1t GhMFT-L2	*G. Barbadense* K101	Phusion/Phire	Denaturation	98 °C	1 minutes	1 cycle
			Denaturation	98 °C	5 seconds	12 cycles
			Annealing	72-60 °C (-1 °C/cycle)	15 seconds	
			Extension	72 °C	1 minute	
			Denaturation	98 °C	5 seconds	30 cycles
			Annealing	60 °C	15 seconds	
GhMFT-L2t	*G. hirsutum* DeltaPine61		Extension	72 °C	1 minute	
			Extension	72 °C	5 minutes	1 cycle
GhMFT-L1	*G. hirsutum* DeltaPine61	Phusion	Denaturation	98 °C	1 minutes	1 cycle
			Denaturation	98 °C	30 seconds	12 cycles
			Annealing	72-60 °C (-1 °C/cycle)	30 seconds	
			Extension	72 °C	2 minutes	

91

			Denaturation	98 °C	30 seconds	
			Annealing	60 °C	30 seconds	30 cycles
			Extension	72 °C	2 minutes	
			Extension	72 °C	10 minutes	1 cycle
GhMFT-L1p GhMFT-L2p	*G. hirsutum* DeltaPine61	OneTaq	Denaturation	94 °C	1 minute	1 cycle
			Denaturation	94 °C	15 seconds	
			Annealing	59 °C	40 seconds	35 cycles
			Extension	68 °C	2.5 minutes	
			Extension	68 °C	10 minutes	1 cycle

Table 4.3 Oligonucleotides used for yeast homologous recombination assembly for mutant

rescue experiments. Colored nucleotides are 5' overhangs used to generate end homology for

yeast homologous recombination cloning in 4-system shuttle vector pSFP100. Black nucleotides

are sequence specific for amplifying insert fragments used to create the shuttle vector.

primer	sequence 5'
GrSPp hr fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatgggtatggcatgagaaatcaccatgtatc
GrSPp hr rev	ggatctgacagttttgccatcccacaaactaatataacactgg
GrSP hr fwd	ccagtgttatattagtttgtgggatggcaaaactgtcagatcc
GrSP hr rev	ttacccggggcgtcttctagcagctgtttccc
GrSPt hr fwd	gagctcaacataaagtggttcaccaatggatc
GrSPt hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattcttcaactgggttttttgttcttc
GrBFT-L1p hr fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatgtcaatttgacgatcaatgtcg
GrBFT-L1p hr rev	gggtcggggactcttgacatgatatatatattttttagctaatgaatattgc
GrBFT-L1 hr fwd	gcaatattcattagctaaaaatatatatatcatgtcaagagtccccgaccc
GrBFT-L1 hr rev	tcacccgggtcttcttgatcttgcggcag
GrBFT-L1t hr fwd	gagctcctctgccacttccataataatatatac
GrBFT-L1t hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattcccctcaagatagctagattaagc
GrSFTp fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatgcctaaaaatcagctaccctacg
GrSFTp rev	ggatctctatctctaggcatgatatcgctatttggtcttac
GrSFT hr fwd	gtaagaccaaatagcgatatcatgcctagagatagagatcc
GrSFT hr rev	tgtcctacggccaccggatccac
GrSFTt hr fwd	gagctcaataaataatattgttgttgttgatc
GrSFTt hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattccacatttattcaatttggtctc
GrTFL1-L1p fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatgcattgattgaacttacttgctc
GrTFL1-L1p rev	ggctctacttcccttgccatttgaggagttctgaatgaaagaaagag
GrTFL1-L1 hr fwd	ctctttctttcattcagaactcctcaaatggcaagggaagtagagcc
GrTFL1-L1 hr rev	tcacccgggacgtcttcttgcagctgtttctc
GrTFL1-L1t hr fwd	gagctcttaacctgcacaaaagtatatctg
GrTFL1-L1t hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattcaggtgagtttgtgcctgattt
GrMFT-L1p hr fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatggatttgtattacctctatc
GrMFT-L1p hr rev	ggatcaacggaggcagccatgggagaaagaggagtgggggtgcagt
GrMFT-L1 hr fwd	actgcacccacccccactcctctttctcccatggctgcctccgttgatcc
GrMFT-L1 hr rev	tcacccgggacgccttcggctgacgggctc

GrMFT-L1t hr fwd	gagctcatctcccacacacacttctctc
GrMFT-L1t hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattccatctaccttttaattggaag
GrBFT-L2p hr fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatgagaatagttaccaaattaaggatccaaagagtg
GrBFT-L2p hr rev	ggctcagggacccttgacatgatgaacaagacgatatgtatg
GrBFT-L2 hr fwd	catacatatcgtcttgttcatcatgtcaagggtccctgagcc
GrBFT-L2 hr rev	ttacccgggtcttcttcttgcagcagtttctc
GrBFT-L2t hr fwd	gagctcctatggctgccccatagacattaag
GrBFT-L2t hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattcggcatgcttaattgggtagg
GrMFT-L2p hr fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatggcttgatttaaaatctccg
GrMFT-L2p hr rev	gtggttcaacggaccgggccatagtgtgttggactagacctgcg
GrMFT-L2 hr fwd	cgcaggtctagtccaacacactatggcccggtccgttgaaccac
GrMFT-L2 hr rev	ctacccgggacgtttcttagctgctggctcc
GrMFT-L2t hr fwd	gagctcctcatagcttacagtgcattatttgg
GrMFT-L2t hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattcccaatgtataggagtgaagc
GrTFL1-L2p hr fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatgattacttagtatattattcag
GrTFL1-L2p hr rev	actcccccaacaatgagagg
GrTFL1-L2 hr fwd	ttcattcagtgtcaccaagaatgggagagcctctcattgttgg
GrTFL1-L2 hr rev	ttacccggggcgtctccttgcagcag
GrTFL1-L2t hr fwd	gagctcttaaaccttcaaaactcaagaaag
GrTFL1-L2t hr rev	gtggtggtggctagcgttaacactagtcagatctaccatgggaattcgagaggttccttaaacgttcag
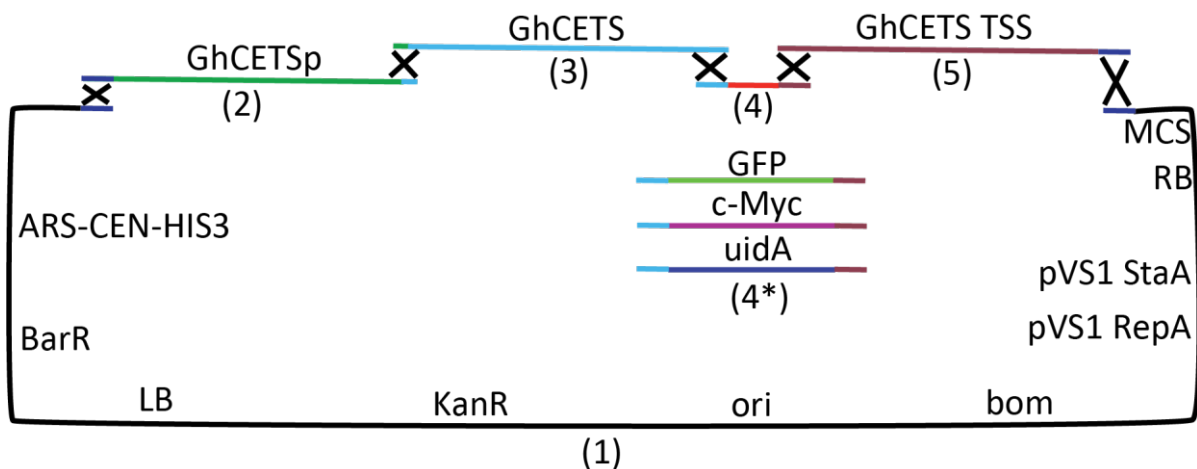
Figure 4.2 Yeast homologous recombination was employed to create constructs containing *GhCETS* genomic clones. Shown is a schematic of the homologous recombination strategy used to create the pSFP100-*GhCETS* series of vectors. Fragment (1) pSFP100 contains replication and selection sequences for *E. coli* (ori, *KanR*), Agrobacterium (StaA, RepA) and yeast (ARS-CEN-*HIS3*) along with selection (*BarR)* and binary vector T-DNA border (LB, RB) sequences for plant transformation. Fragments (2,3 and 5) are 2 kb promoters, *GhCETS* genomic sequences and 1 kb translation terminating sequences (TSS) respectively. These sequences were separately amplified with end homology to flanking DNA fragments built into primers as indicated by alternate coloring of PCR product ends. (4) dsDNA oligonucleotide bridge fragments of 100 bps provide the required homology between 3' gene sequences and 5' TSS sequences plus a stop codon (red). Use of the bridge allows for assembling these genomic sequences into vectors with an N-terminal molecular tag without the necessity of re-amplifying the fragments with different primers. Fragments (4*) represent possible alternative fragments to be used to create in frame fusions with genomic sequences.

Table 4.4 Oligonucleotides used for the synthesis of dsDNA bridge fragments utilized in the assembly of pSFP100-*GhCETS* vectors. Colored nucleotides are 5' overhangs used to generate end homology for homologous recombination cloning in 4-system shuttle vector pSFP100. Black nucleotides provide complementation for bridge synthesis and create an XbaI-Stop codon-SacI sequence between the 3' end of a *CETS* gene and the 5' end of its terminator sequence in the final genomic clone.

oligonucleotide	sequence 5'
b*GrSP* fwd	ctgtctatttcaatgctcaaagggaaacagctgctagaagacgcccgggtaagcggccgcgag
b*GrSP* rev	gggcaaacatcattgatccattggtgaaccactttatgttgagctcgcggccgcttac
b*GrBFT-L1* fwd	gcagtttacttcaatgcccagagagaaactgccgcaagatcaagaagacccgggtgagcggccgcgag
b*GrBFT-L1* rev	cagtagtaaattatgtatgtatatattattatggaagtggcagaggagctcgcggccgctcag
b*GrSFT* fwd	ctgccagagggagagtggatccggtggccgtaggacacccgggtaggcggccgcgag
b*GrSFT* rev	cttgaaatcaaacatgatcaacaacaacaatattatttattgagctcgcggccgcctac
b*GrTFL1-L1* fwd	ctgtgtatttcaatgctcaaagagaaacagctgcaagaagacgtcccgggtgagcggccgcgag
b*GrTFL1-L1* rev	caagaacaccccatcacagatatactttgtgcaggttaagagctcgcggccgctcac
b*GrMFT-L1* fwd	ctgtctatttcaacgcccaaaaagagcccgtcagccgaaggcgtcccgggtgagcggccgcgag
b*GrMFT-L1* rev	ccagtttttatatttatatattttatatatagagagagagagagagaagtgtgtgtgggagatgagctcgcggccgctcac
b*GrBFT-L2* fwd	cttcaatgcccagagagaaactgctgcaagaagaagacccgggtaagcggccgcgag
b*GrBFT-L2* rev	ccttcattttattatacttaatgtctatggggcagccataggagctcgcggccgcttac
b*GrMFT-L2* fwd	gcagtgtatttcaattctcaaaaggagccagcagctaagaaacgtcccgggtaggcggccgcgag
b*GrMFT-L2* rev	gtattacagaacctaccaaataatgcactgtaagctatgaggagctcgcggccgcctac
b*GrTFL1-L2* fwd	gctgtttatttcaatgcacgaagagaaactgctgcaaggagacgccccgggtaagcggccgcgag
b*GrTFL1-L2* rev	ccaatccaaatggaacgttctttcttgagttttgaaggtttaagagctcgcggccgcttac

Table 4.5 Oligonucleotides used to amplify meGFP sequence in the assembly of pSFP100-

GhCETS-GFP vectors. Colored nucleotides are 5' overhangs used to generate end homology to

CETS genomic and terminator sequences for in-frame fusions. Black nucleotides are sequence

specific for an 8xAlanine-meGFP sequences.

oligonucleotide	sequence 5'
GrSP-8xAla-GFP fwd	ctatttcaatgctcaaagggaaacagctgctagaagacgcgctgctgcagcggccgcggctgccatggtg
GrSP-GFP rev	gggcaaacatcattgatccattggtgaaccactttatgtttcacttgtacagctcgtccatgccgtg
GrBFT-L1-8xAla-GFP fwd	cttcaatgcccagagagaaactgccgcaagatcaagaagagctgctgcagcggccgcggctgccatggtg
GrBFT-L1-GFP rev	cagtagtaaattatgtatgtatatattattatggaagtggcagagtcacttgtacagctcgtccatgccgtg
GrSFT-8xAla-GFP fwd	taactgccagagggagagtggatccggtggccgtaggacagctgctgcagcggccgcggctgccatggtg
GrSFT -GFP rev	cttgaaatcaaacatgatcaacaacaacaatattatttatttcacttgtacagctcgtccatgccgtg
GrTFL1-L1-8xAla-GFP fwd	gtatttcaatgctcaaagagaaacagctgcaagaagacgtgctgctgcagcggccgcggctgccatggtg
GrTFL1-L1-GFP rev	caagaacaccccatcacagatatactttgtgcaggttaatcacttgtacagctcgtccatgccgtg
Gr MFT-L1-8xAla-GFP fwd	ctatttcaacgcccaaaaagagcccgtcagccgaaggcgtgctgctgcagcggccgcggctgccatggtg
Gr MFT-L1-GFP rev	ccagttttatatttatatattttatatatagagagagagagagaagtgtgtgtgggagattcacttgtacagctcgtccatgccgtg
GrBFT-L2-8xAla-GFP fwd	gtacttcaatgcccagagagaaactgctgcaagaagaagagctgctgcagcggccgcggctgccatggtg
GrBFT-L2-GFP rev	ccttcattttattatacttaatgtctatggggcagccatagtcacttgtacagctcgtccatgccgtg
GrMFT-L2-8xAla-GFP fwd	gtatttcaattctcaaaaggagccagcagctaagaaacgtgctgctgcagcggccgcggctgccatggtg
GrMFT-L2-GFP rev	gtattacagaacctaccaaataatgcactgtaagctatgagtcacttgtacagctcgtccatgccgtg
GrTFL1-L2-8xAla-GFP fwd	ttatttcaatgcacgaagagaaactgctgcaaggagacgcgctgctgcagcggccgcggctgccatggtg
GrTFL1-L2-GFP rev	ccaatccaaatggaacgttctttcttgagttttgaaggtttaatcacttgtacagctcgtccatgccgtg

Table 4.6 Results from a yeast transformation experiment performed to create plant transformation vector, pSFP100. Positive

transformation control was circular pRS313. Negative controls were: 1.) ARS-CEN-*HIS3* PCR product alone, 2.) *HindIII*-pCAMBIA0390

and *NOS_{pro}::BAR:CaMV35SpA* PCR product together, and 3.) water alone. Strain PJ694a has a high transformation efficiency, but is

also known to spontaneously generate background colonies with a distinct flaky appearance. 1/5 volume of transformed cells was

plated onto plate #1 while the remaining 4/5 volume was plated onto plate #2. The experimental transformation to create plasmid

pSFP100 was performed in duplicate. Transformation efficiency is calculated in CFUs/µg of plasmid DNA.

transformation	input DNA	no. of CFUs (plate #1)	no. of CFUs (plate #2)	transformation efficiency
pSFP100	1. 500 ng *HindIII*-pCAMBIA0390 2. 300 ng ARS-CEN-HIS 3. 300 ng *NOS_{pro}::BAR:CaMV35SpA*	3,032	5,144	1.63×10^4
pSFP100 (duplicate)	1. 500 ng *HindIII*-pCAMBIA0390 2. 300 ng ARS-CEN-HIS 3. 300 ng *NOS_{pro}::BAR:CaMV35SpA*	3,672	6,144	1.97×10^4
positive control	1. 500 ng pRS313 (yeast vector)	lawn	lawn	
negative control #1	1. 300 ng ARS-CEN-*HIS3*	776	1,864	1.32×10^4
negative control #2	1. 500 ng *HindIII*-pCAMBIA0390 2. 300 ng *NOS_{pro}::BAR:CaMV35SpA*	49	86	
negative control #3	water control	50	90	
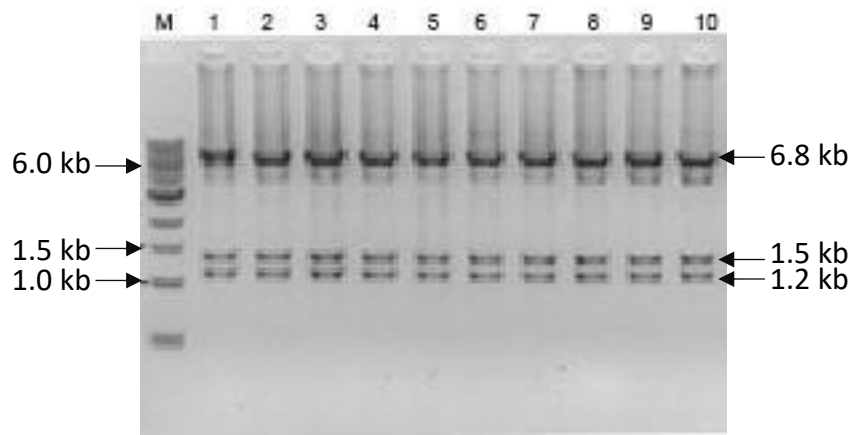
Figure 4.3 Correct assembly of 4-system shuttle vector pSFP100 by yeast homologous

recombination. Physical characterization by NotI digestion confirming proper construction

of pSFP100 in all screened samples. pSFP100 is a 9,647 bp circular plasmid with three NotI

restriction sites. NotI digestion produces three restriction fragments of 6.8, 1.5, and 1.2 kbps

in length.

Table 4.7 Transformation efficiencies of assembling pSFP100-*GhCETS* and pSFP100-*GhCETS-GFP* vector series. Linearized plasmid, *GhCETS* genomic PCR products and 100 bp bridges or *GFP* PCR products were introduced into yeast strain PJ694a. Various amounts of linearized plasmid and inserts were used for yeast transformations. Efficiency was slightly higher when less linearized plasmid and more inserts were introduced into the transformation mix. Transformation efficiency is calculated in CFUs/µg of plasmid DNA.

linearized backbone (ng)	PCR inserts/bridge fragments (picomolar ratios to linear backbone)	no. of cells plated	no. of colonies formed	transformation efficiency
500	equimolar	2×10^7	230 - 1,232	$2.3 \times 10^3 – 1.2 \times 10^4$
250	2x excess PCR inserts 5x excess dsDNA bridges		440 - 500	$8.8 \times 10^3 – 1.0 \times 10^4$
200	all inserts in equal volume to		892 – 4,500	$2.2 \times 10^4 – 1.1 \times 10^5$
100	final volume of 34 µL and in >4x excess		332 - 825	$1.7 \times 10^4 – 4.1 \times 10^4$

Table 4.8 Efficiency of homologous recombination in assembly of pSFP100-*GhCETS* and pSFP100-*GhCETS-GFP* vector series.

Individual or total yeast DNA transformed into *E. coli* was screened by restriction digest before sequencing confirmation. During transformation, the correct construct was not always the only species found in isolated yeast DNA. Typically, a few different banding patterns were observed in restriction digest analysis. % of homologous recombination efficiency is calculated as (no. of samples showing correct banding pattern)/(total no. of analyzed samples).

method of yeast isolation	efficiency of assembly
individual colonies	37 – 100 %
total transformants	60 – 100 %

A

M #1 #2 #3 #4 #5

10.0 kb
6.0 kb
4.0 kb
2.0 kb

9.6 kb
uncut plasmid
4.0 kb

M #6 #7 #8 #9 #10

10.0 kb
6.0 kb
4.0 kb
2.0 kb

9.6 kb
uncut plasmid
4.0 kb

B

Sample	Sequenced bases	Coverage	No. of Reads Sequenced	No. of Reads Assembled	Average Depth of Reads
#1	342,271	25x	1,918	1,633	20.9±8.0
#2	3,415,551	249x	23,415	21.724	234.7±113.5

C

pSFP100-*GhSP-Dt*
#2
#1
pSFP100-*GhSP-At*

2,000   4,000   6,000   8,000   10,000   12,000

promoter   exons and introns   terminator

GhSP-Dt

#2

#1

GhSP-At

10,000   12,000

Figure 4.4 Representative analysis of assembled pSFP100-*GhCETS* constructs. (A) Physical

characterization of ten clones isolated in assembly of pSFP100-*GhSP* by restriction digestion.

Clones were digested with SbfI and NcoI. The correct clones exhibit two bands of 9.6 and 4.0 kb.

Residual uncut plasmid can be seen as a band with fainter intensity slightly above 4.0 kb in some samples. Clones #1 and #2 were chosen for sequence verification using Ion Torrent NGS sequencing technology. (B) Results of sequencing pSFP100-*GhSP* clones #1 and #2. The amount of data recovered for clone #1 was abnormally low in comparison with sequencing of all other assemblies. Typically, sequencing clones resulted in greater than 200x coverage. (C) MAFFT alignment of expected *in silico* pSFP100-*GhSP-Dt* and pSFP100-*GhSP-At* designs with consensus sequences generated from the sequencing assemblies of clones #1 and #2. Identical sequence is indicated by gray whereas red coloring shows mismatches and white indicates gaps in sequences. *GhSP-At* and *GhSP-Dt* are similar throughout their coding regions, but differ in non-coding genomic regions. Lavender rectangles highlight site of recombination and show seamless accurate integration. The inset below zooms into the genomic region of *GhSP*. Both clones comprise the *GhSP-Dt* promoter and terminator sequences; however, clone #1 has the terminator sequences for *GhSP-At* incorporated while clone two contains the *GhSP-Dt* terminator sequence.

1. Assembly Design
   - DNA inserts with 40 nucleotides of overlap to flanking fragments (PCR primers or small, synthesized dsDNA
   - longer overlaps may be required when assembling larger or more of inserts

2. DNA Preparation
   - Linearize pSFP110 (REases)
   - insert amplifications
   - insert digests
   - insert synthesis

3. PEG/LiAc/ssDNA Transformation
   - 34 uL final DNA volume
   - 100 ng linear pSFP110
   - Inserts in molar excess (3:1 or greater)
   - Small inserts (<= 100 bp) may require greater excess (5:1 or 10:1)

4. Selection
   - Histidine prototrophy selection (*HIS3* is only expressed in a circular construct)

5. Analysis
   - Total yeast isolation: modified Bust-n-Grab method
   - Transformation into *E. coli*
   - Restriction digests for correct banding pattern
   - PCR to confirm assembly site
   - Sequencing confirmation (NGS)

6. Plant Tranformation
   - Transformation in Agrobacterium and plant inoculation
   - Biolistic bombardment

Figure 4.5 A flowchart for plant transformation via yeast homologous recombination using 4-system shuttle binary vector pSFP100. Steps of the cloning process from assembly design to plant transformation are described.

A

Days to inflorescence

90
80
70
60
50
40
30
20
10
0

A   C D   C   A   C D E   B C   D E   B   C D E   C D E   E

Col-0 ft-10 EV GhSFT GhSP GhTFL1-L1 GhTFL1-L2 GhBFT-L1 GhBFT-L2 GhMFT-L1 GhMFT-L2

B

No. of rosette leaves

60
50
40
30
20
10
0

A   B C   B C   A   B C   B   C   B   B C   B C   C

Col-0 ft-10 EV GhSFT GhSP GhTFL1-L1 GhTFL1-L2 GhBFT-L1 GhBFT-L2 GhMFT-L1 GhMFT-L2

C

No. of cauline leaves

16
14
12
10
8
6
4
2
0

A   B   B   A   B   B   B   B   B   B

Col-0 ft-10 EV GhSFT GhSP GhTFL1-L1 GhTFL1-L2 GhBFT-L2 GhMFT-L1 GhMFT-L2

D
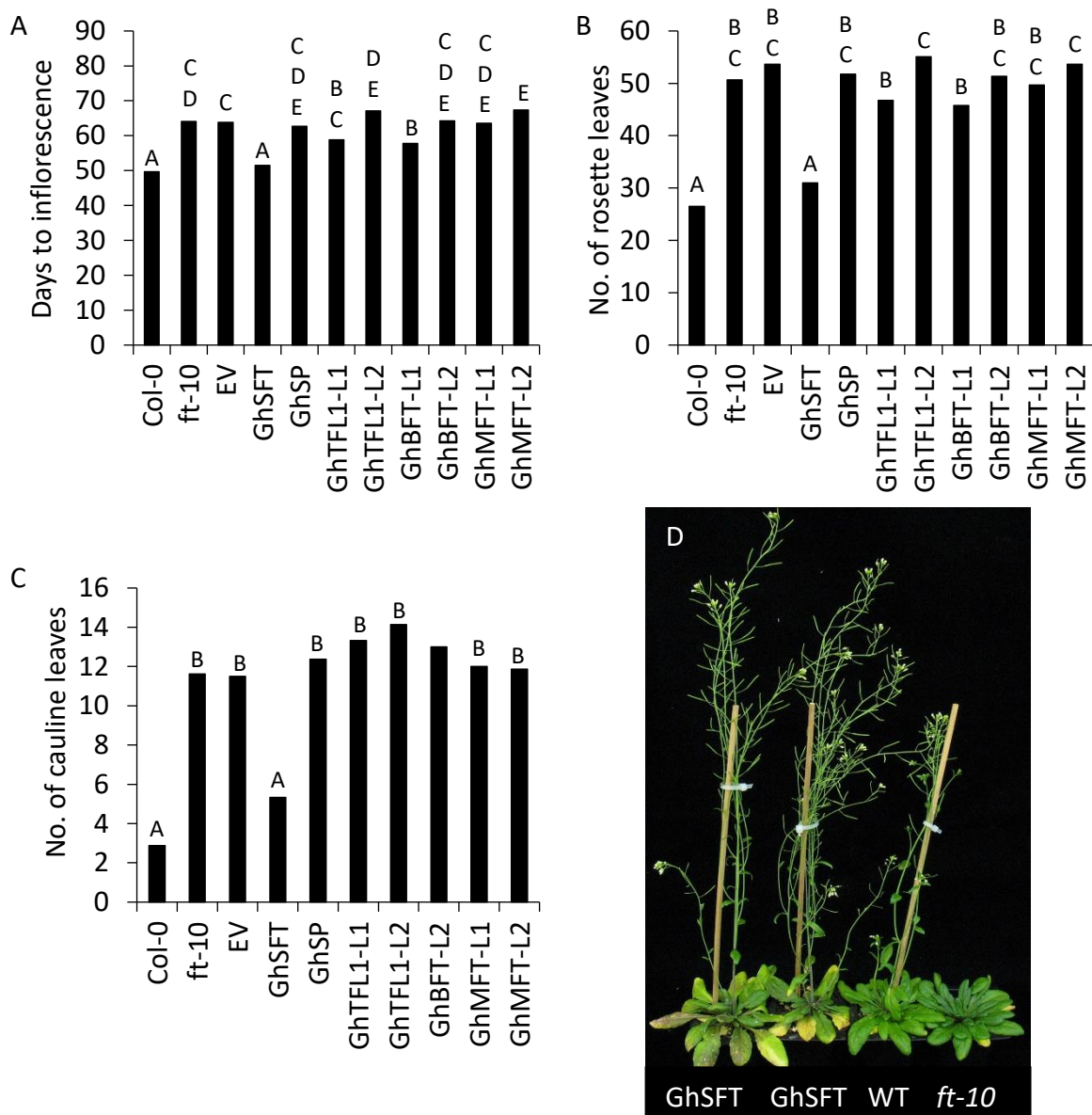
GhSFT   GhSFT   WT   *ft-10*

Figure 4.6 Phenotype of Arabidopsis *ft-10* mutant transformed with cotton *CETS* genomic clones. Rescue of the mutant phenotype was determined by assessing days to inflorescence (A), number of rosette (B) and cauline (C) leaves. Plants were grown under 12/12 day/night photoperiod. Different letters represent significant differences based on univariate ANOVA with Tukey's H-S-D ad hoc analysis at the 0.05 level. Note *GhSFT* transformed plants are always in subgroup A along with WT Col-0. (D) Representative rescue of the *ft-10* mutant by *GhSFT*; all pictured plants are 60 DPG. Col-0, *ft-10* and *GhSP*; n = 20. EV, n = 14. *GhSFT,* n = 11. *GhTFL1-L1,* n = 15. *GhTFL1-L2,* n = 16. *GhBFT-L1,* n = 4. *GhMFT-L1* and *GhMFT-L2,* n = 17.

A

Days to inflorescence

C (Col-0 ~19)
A (tfl1-14 ~14), B, B, B, C, A B, B, B, A B, B

B

No. of rosette leaves
12, 10, 8, 6, 4, 2, 0

D (Col-0 ~10.5)
A, A, A B, A B, C, A B, B, A, A, A B

C

No. of cauline leaves
3.0, 2.5, 2.0, 1.5, 1.0, 0.5, 0.0

D (Col-0 ~2.7)
A B, B C, C, A B C, C, A B C, B C, A, A B C, B C

D

No. of siliques
40, 35, 30, 25, 20, 15, 10, 5, 0

D (Col-0 ~33)
A B, A B, B, A B, C, A B, A B, A, A B, A B

tfl1-14, Col-0, EV, GhSP, GhTFL1-L1, GhTFL1-L2, GhBFT-L1, GhBFT-L2, GhSFT, GhMFT-L1, GhMFT-L2
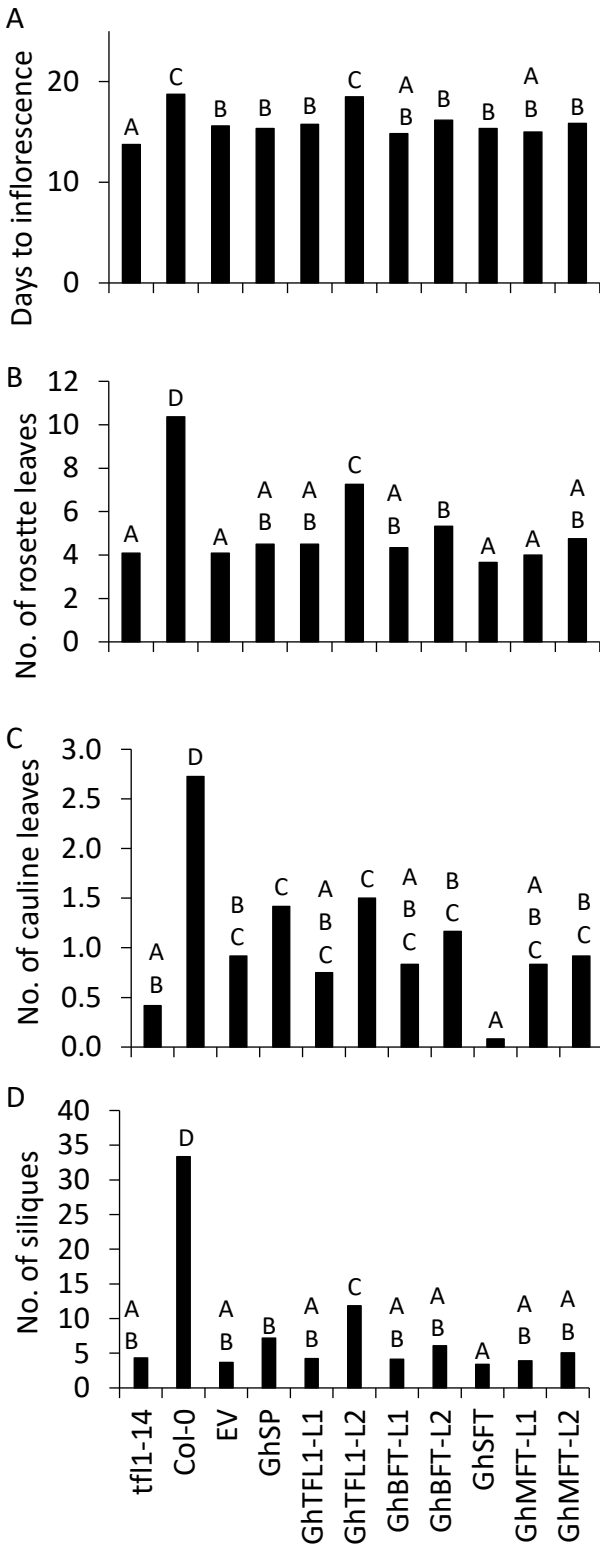
E

F

G

H

I

J

K

Figure 4.7 Phenotype of transgenic Arabidopsis *tfl1-14* mutant transformed with cotton *CETS* genomic clones. *tfl1-14* rescue was assessed by determining days to inflorescence (A), number of rosette (B) and cauline (C) leaves, and number of siliques (D). Plants were grown under 14- hour LDs. Different letters represent significant differences based on univariate ANOVA with Tukey's H-S-D ad hoc analysis at the 0.05 level. Depictive 18-day-old plants displaying levels of rescue in the *tfl1-14* background: (E) *tfl1-14* early flowering mutant phenotype, (F) EV control plant, (G) non-flowering WT plant, (H) non-flowering *GhTFL1-L2* plant, (I) *GhBFT-L2* plant showing partial rescue of the early-flowering phenotype, (J) *GhSP* plant displaying an indeterminate inflorescence (red box). (K) transgenic *GhSFT* accelerated the early-flowering mutant phenotype during the reproductive phase; the pictured plant lacks cauline leaves. n = 12 for all genotypes.

CHAPTER 5

PROFILING *GOSSYPIUM HIRSUTUM CETS* GENE FAMILY EXPRESSION PATTERNS BY GUS

ANALYSIS

5.1   Introduction

Current evidence indicates that CETS act non-cell autonomously over long or short

distances to determine shoot meristem identity. Studies in gene expression patterns, protein

localization, and interactions have significantly contributed to this understanding. In

Arabidopsis *FT* expression is driven by CO activity in the companion cells of leaf vasculature

(Samach et al., 2000; An et al., 2004; Wigge et al., 2005) and FT, through interaction with bZIP

TF FD, acts in the shoot meristem to promote flowering (Kardailsky, 1999; Kobayashi, 1999; Abe

et al., 2005; Wigge et al., 2005). Further studies exploring accumulation of *FT:GFP* transgene

mRNA and FT:GFP protein accumulation revealed that it is the FT protein, not mRNA, that is

translocated via the phloem into the shoot meristem (Corbesier et al., 2007). Similarly, *TSF*

expression analyzed by RT-PCR and *gTSF:GUS* studies observed expression in shoot vasculature

tissues (hypocotyl, cotyledon and leaf), but not shoot meristem tissue (Yamaguchi et al., 2005).

While TSF has been found to interact with FD (Kim et al., 2016), grafting experiments using a

CaMV*35S:TSF:T7* fusion stock and *ft tsf* scion did not rescue late flowering of the *ft tsf* scion and

immunohistochemical probing of the T7 reporter failed to show TSF:T7 accumulation in scion

tissues (Jin et al., 2015b). This led to the conclusion that TSF ability to move through phloem is

lower than that of FT. Studies in rice, pumpkin, poplar, tomato and other species also

demonstrate that *FT* homologs in these species have predominate leaf expression and protein

products that travel to promote determinate growth in shoot meristems (Kojima, 2002;

Böhlenius et al., 2006; Lifschitz et al., 2006; Lin et al., 2007; Tamaki et al., 2007; Shalit et al., 2009).

While these discussed FT homologs work distal from the site of synthesis to promote determination of shoot meristems, Arabidopsis TFL1 is synthesized locally in meristem cells and has been established to act in a non-cell autonomous manner through plasmodesmata connections to maintain meristem indeterminacy (Conti and Bradley, 2007). Although cell-to-cell movement may be a general hallmark of CETS proteins, TFL1 movement is more localized than *FT*-like florigens. However, *TFL1* homolog expression is varied. For instance in tomato, *SP* expression is not restricted to meristems, but is more generally expressed throughout the plant (Pnueli et al., 1998). Several species such as poplar, citrus trees and pea comprise *TFL1* homologs that are predominately expressed in apices and others expressed primarily in leaves (Foucher et al., 2003; Esumi et al., 2005; Mohamed et al., 2010).

In Arabidopsis *TFL1*-like *ATC* is expressed specifically under non-inductive SD conditions. In transgenic plants, phloem-specific expression using the $SUC2_{pro}$ to drive *ATC* expression was sufficient to suppress flowering under inductive LD conditions. Moreover, Arabidopsis inflorescence grafting experiments with an *ATC* over-expressing stock and WT or *atc-2* scions demonstrated that ATC was graft transmissible and delayed flowering in comparison to control grafts (Huang et al., 2012). More recently, a *Chrysanthemum seticuspe* gene expressed in leaves, CsAFT, suppressed flowering in SD inductive conditions and also delayed flowering via graft transmission (Higuchi et al., 2013). These results support the hypothesis of CETS anti-florigens that are produced in leaves in response to environmental or endogenous signals and transported long-distance to shoot meristems to maintain indeterminate growth.

Here cotton *CETS* gene expression patterns were analyzed using *CETS~pro~*:*uidA* fusions in Arabidopsis to further our knowledge of *CETS* regulation of plant architecture through examination of their promoter activities. Most cotton *CETS* promoters drove GUS activity in vasculature tissues. *GhSP~pro~* promoted GUS activity primarily in meristems. Additionally, to infer pathway regulation on *CETS* activity, *Gossypium CETS* promoter sequences were computationally analyzed for conserved motifs associated with known TFs.

5.2 Materials and Methods

5.2.1 Plasmid Construction

CETS~pro~:*uidA* constructs were created using traditional cloning methods. Genomic DNA was isolated from snap-frozen leaf tissue using a CTAB extraction protocol (Weigel and Glazebrook, 2002). Two to three kb promoter sequences were amplified from isolated *Gossypium hirsutum* Acala Maxxa DNA or from a pSFP100-*CETS* construct containing the promoter of interest. Promoter sequences were amplified using Phusion HS II polymerase or Onetaq polymerase (NEB); oligonucleotide sequences used for amplification are listed in Table 5.1. *GhSP~pro~*, *GhSFT~pro~*, *TFL1-L~pro~*, and *BFT-L2~pro~* PCR products and binary vector pGPTV-BAR (Becker et al., 1992) were digested with XmaI and SbfI (NEB). PCR digestion reactions were column cleaned using Wizard SV Gel and PCR Clean-up System (Promega). Products were ligated into the binary vector using T4 DNA ligase (NEB). Recovered plasmids were sequenced by semiconductor sequencing on the Ion Torrent PGM (Life Technologies) according to the protocol described in Chapter 4.2.2 *GhBFT-L1~pro~*, *GhMFT-L1~pro~*, *GhMFT-L2~pro~*, and Gh*TFL1-L2~pro~* PCR products were sub-cloned into PCR8/GW/TOPO by TA cloning (Thermofisher Scientific) and

subsequently cloned into pGPTV-BAR by restriction digest and ligation as described above. Constructs were verified by Sanger sequencing (MWG Operon).

5.2.2 Plant Transformations and Growth Conditions

Plasmids were electroporated into *Agrobacterium tumifaciens* strain Gv3101-MP90 as previously described. Arabidopsis Col-0 ecotype were transformed by floral dip (Clough and Bent, 1998). T_1 seeds were harvested, and transformants selected on plates containing ½-strength Murashige -Skoog nutrients (PhytoTechnologies Laboratories), 1% sucrose, 2.5 g/L gelrite and 10 μg/mL glufosinate ammonium. After selection, 7 – 10-day-old plants were transferred to Fafard 3B potting soil (Sun Gro/Fafard) for continued growth in 12 hour light, 12-hour dark lighting conditions.

5.2.3 GUS Staining

For histochemical staining, plant tissues were harvested, vacuum infiltrated in staining buffer (50mM Na-phosphate buffer, pH 7; 0.2% triton; 2mM X-gluc (5-bromo-4-chloro-3-indol-β-D-glucuronic acid)), and incubated overnight at 37°C. Plants were then flushed of chlorophyll by incubation in 70% ethanol for until clear for visualizing staining.

5.2.4 Image Capture

Stained plants were viewed and imaged with a Nikon SMZ1500 stereomicroscope (Nikon) equipped with a SPOT Insight 2 CCD camera (Diagnostic Instruments, Inc.).

5.2.5   Computational Promoter Analysis

2.0 kb promoter sequences of orthologous cotton *CETS* genes from *G. raimondii* (D-genome), *G. arboreum* (A-genome), and *G. hirsutum* (A_t- and D_t-subgenomes) were analyzed for conserved regulation using the regulation predication tool at PlantRegMap (http://plantregmap.cbi.pku.edu.cn/). A p-value of 1x 10^{-5} was used for a binding site prediction threshold. The set of *G. raimondii* TFs predicted to target all orthologous promoter sequences was assessed for significantly over-represented GO terms using the GO enrichment tool at PlantRegMap (http://plantregmap.cbi.pku.edu.cn/); a 0.01 p-value threshold was used for assessing significance of GO enrichment. Over-represented GO terms were further refined and visualized using REVIGO (http://revigo.irb.hr/). Parameters used for REVIGO analysis were: allowed similarity, medium (0.7); database with GO term size, Arabidopsis thaliana; and semantic similarity measure, SimRel. Q-values from GO enrichment analysis were provided.

5.3   Results

5.3.1   GUS Activity Driven by Cotton *CETS* promoters

To establish the patterns of cotton *CETS* expression, qualitative GUS assays were performed to demonstrate promoter activity. Two to three kb of sequence upstream of *CETS* ATG start codons were fused to the *uidA* gene and constructs containing these fusions were introduced into WT Arabidopsis. Because cotton *CETS* exhibit developmental regulation, T_1 generation plants were assayed for GUS activity at three developmentally distinct time points: young rosette (under 10 DPG), mature rosette (20-30 DPG) and flowering plant (45-65 DPG).

GUS activity from *GhSFT~pro~* was weak and observed predominately in root apical meristems during the young rosette stage. 40% of plants with the most intense staining also showed activity in the apical minor veins of expanding leaves (Fig 5.1A). Activity was absent in mature rosette and flowering plants screened.

GhSP~pro~ promoted GUS activity in all plant meristems (SAM, axillaries, leaf lateral, RAM and root lateral) in 20 of 32 young rosette plants and all shoot meristems of 11 of 12 mature rosettes stained (roots were not stained in older plants, Fig 5.1B and Fig 5.2A). In 5 of 8 stained *GhSP~pro~* flowering plants, GUS activity was observed in immature floral buds, but was absent from flowers and siliques (Fig 5.2B).

In *GhTFL1-L1~pro~* plants, X-gluc staining was observed in all leaf vasculature of fully expanded leaves, in minor veins at the apex of expanding leaves, and was absent in unexpanded leaves. This pattern is consistent with phloem maturation as leaves undergo the source to sink transition and was observed in 22 of 28 *GhTFL-L1~pro~* young rosettes stained (Fig 5.1C). *GhTFL1-L1~pro~* also drove robust GUS activity in the vasculature of the hypocotyl and roots of these plants (Fig 5.1C). This activity decreases in *GhTFL-L1~pro~* plants over time. No leaf vasculature staining was observed in 12 mature rosette and 8 flowering *GhTFL1-L1~pro~* plants stained.

GUS activity driven by *GhTFL1-L2~pro~* was identical to the patterning seen with *GhTFL1-L1~pro~* plants in young rosette plants; 11 of 14 plants analyzed demonstrated source-to-sink leaf, hypocotyl and root system vasculature staining (Fig 5.1D). Again, like GUS activity driven by the *GhTFL1-L1~pro~*, GUS activity promoted by *GhTFL1-L2~pro~* decreased as plants aged. X-gluc staining was observed in leaf vasculature in a source-to-sink pattern and in the hypocotyl vasculature of

112

8 of 15 mature rosettes analyzed (Fig 5.2C); however, no staining was observed in any tissues of the 16 flowering stage plants examined.

When young rosette plants of *GhBFT-L2$_{pro}$* plants were examined, 24 of 28 plants also showed leaf vasculature X-gluc staining in a source-to-sink pattern and hypocotyl vasculature staining, but no staining in the vasculature of the root system (Fig 5.1E). GUS activity driven by the *GhBFT-L2$_{pro}$* was robust and persisted throughout the life cycle of transformed plants. 24 of 28 mature rosette plants examined demonstrated both source-to-sink leaf vasculature as well as hypocotyl vasculature GUS activity. Additionally, 22 of 24 flowering plants analyzed showed intense leaf vasculature staining (Fig 5.2D-E). In 14 of the 22 *GhBFT-L2$_{pro}$* flowering plants that showed leaf vasculature X-gluc staining, GUS activity was also observed in sepal vasculature of open flowers (Fig 5.2F). *GhBFT-L1$_{pro}$* failed to promote GUS activity in any plants stained at any developmental stage indicating a lack of promoter activity, at least in Arabidopsis.

In plants harboring *GhMFT-L1$_{pro}$* or *GhMFT-L2$_{pro}$* constructs, weak activity was observed in petioles and mid-ribs of fully expanded leaves (Fig 5.1F-G). This expression was observed in young and mature rosettes, but was absent in flowering plants. In 40% of *GhMFT-L2$_{pro}$* plants observed, variable GUS activity was observed in filament vasculature (Fig 5.2E-F).

5.3.2   Computational Predictions of Cotton *CETS* Regulation

To predict elements responsible for the observed expression patterns, the orthologous *CETS* promoters from *G. raimondii* (D-genome), *G. arboreum* (A-genome) and *G. hirsutum* (A$_t$- and D$_t$-subgenomes) were analyzed for conserved regulation by *G. raimondii* TFs. In multicellular organisms, gene expression is regulated through the compound interaction of TFs

with *cis* regulatory elements or TF binding sites (TFBSs). Computationally identifying true TFBSs in promoter sequences is complex because many TFs bind to short degenerate sequence motifs (6-12 bp in length). These sequences occur frequently in genomes and predictive analysis of promoter sequences can include a high rate of false positives. Previously, it has been shown that using comparative analysis of orthologous promoter sequences (DNA footprinting) combine to TFBS prediction considerably reduces false positive predictions and enhances functionally relevant predictions (Ovcharenko et al., 2005). For this study, 2.0 kb promoter sequence upstream of ATG start codons of cotton *CETS* orthologs were analyzed for conserved regulation by *G. raimondii* TFs using the regulation prediction tool at PlantRegMap (Jin et al., 2015a; Jin et al., 2017). This tool allows both orthologous and non-orthologous promoter sequences to be analyzed for potential conserved or co-regulation by TFs by predicting all regulatory interactions between TFs in a database of and each input gene, then finding TFs overrepresented in the input gene set. While the tool does not incorporate DNA footprinting, in this study TFs predicted to bind all *CETS* orthologous promoters used as input sequences, in general, bound to the same location of each *CETS* ortholog, mimicking the goal of DNA footprinting in prediction of conserved regulation. In addition, most predicted TFs showed clustered patterns of binding to promoters, which is expected for interactions in protein complexes (Fig 5.3A-H and Table 5.2). Sets of *Gossypium raimondii* TFs predicted in regulation of all four orthologous promoters were then analyzed for gene ontology (GO) term enrichment (Supek et al., 2011; Jin et al., 2017) to detect biological processes highly-associated with the predicted regulation of cotton *CETS*. Excluding cotton *BFT-L1* promoter sequences, all cotton

CETS promoters were predicted to be regulated by TFs involved in developmental processes. This strengthens the supposition that cotton *CETS* are involved in development.

Cotton *SFT* promoters were predicted to have conserved regulation by sixteen *G. raimondii* TFs. Several Dof family TFs were predicted to bind cotton *SFT* promoters at -700 bp and -400 bp from their ATG start codons (Fig 5.3A and Table 5.2). Dof TFs have been established to control gene expression in vasculature tissues (Baumann et al., 1999; Papi et al., 2002). These regulation predictions suggest that these elements in the *SFT* promoters might be at least partially responsible for the pattern of vasculature X-gluc staining observed in *GhSFT_{pro}:uidA* transgenic plants. MADS-box family TFs PISTILLATA (PI) and SOC1 were also predicted to bind *Gossypium SFT* promoters. In Arabidopsis PI forms a heterodimer with AP3; together these TFs bind to CArG-box sequences (consensus $CC(A/T)_6GG$) and are responsible for normal formation of petals and stamens in floral development (Krizek and Meyerowitz, 1996; Riechmann et al., 1996). SOC1 is a major integrator of flowering pathways; its expression is activated through photoperiod, vernalization, autonomous, and hormone-induced flowering induction (Koornneef et al., 1998; Lee et al., 2000; Samach et al., 2000; Moon et al., 2003). While both predicted MADS-box TFs are normally considered downstream of FT (SFT) in floral development, regulation of *SFT* by either might suggest a positive feedback loop that acts to promote meristem determinacy.

Gossypium SFT promoters also showed predicted regulation around -300 bp from ATG start codons by TEONSINTE-BRANCHED1/CYCLOIDEA/PFC14 (TCP14). TCP proteins are known to have versatile functions in several aspects of plant development including branching and vegetative growth. Because of their possible importance in fiber development, TCPs in both *G.*

raimondii and *G. arboreum* have been identified and characterized (Ma et al., 2014; Ma et al., 2016). TCP14 in *G. raimondii* is most abundantly expressed in leaf tissue (Ma et al., 2014)placing it near where *SFT* is expressed in leaf vasculature.

Over-represented GO terms of these discussed and other conserved TFs predicted to regulate cotton *SFTs* include responses to gibberellin and salicylic acid, and developmental regulations including meristem development, reproductive structure development, and seed coat development (Fig 5.3A). In Arabidopsis, gibberellic acid contributes to flowering in a pathway separate from the dominate photoperiodic pathway through floral integrators including *FT* (Wilson et al., 1992; Gómez-Mena et al., 2001). Since *GhSFT* has been shown to play a role in regulation of flowering in both photoperiodic and day-neutral cotton varieties (McGarry et al., 2016), predicted regulation of cotton *SFT* genes by putative TFs responsive to gibberellins suggest that the GA pathway may contribute to flowering in cotton.

MADS-box SOC1 was also predicted to regulate cotton *TFL1*-like homologs, *SP*, *TFL1-L1*, and *TFL1-L2* (Figure 5.3B-D and Table 5.2). Again, SOC1 promotes flowering, and it is reasonable to speculate negative regulation of *TFL1*-like genes as an aspect of promoting meristem determinacy.

Two BASIC PENTACYSTIENE (BPC) proteins, BPC2 and BPC6 are predicted to regulate *SP* and *TFL1-L1* promoter sequences (Fig 5.3B-C and Table 5.2). BPC TFs in Arabidopsis, barley, and soybean have been shown to bind $(AG)_n$ sequence repeats (Sangwan and O'Brian, 2002; Santi et al., 2003; Meister et al., 2004; Kooiker et al., 2005). In Arabidopsis BPCs are a family of functionally redundant TFs; triple and quadruple knockouts of Arabidopsis homologs show reduced ethylene sensitivity and pleotropic effects on vegetative and reproductive growth,

including changes in architecture such as small curled leaves, short internodes, hypocotyls and siliques, dwarfed inflorescence, and floral defects including unopened flowers. Unopened flowers sometimes contained aborted floral organs inside and general defectiveness in formation of sepals, petals and stamens (Monfared et al., 2011). These mutant phenotypes are more determinate and the predicted regulation of cotton *SP* and *TFL1-L1* promoter sequences by BPC TFs suggests positive regulation of these genes could contribute to maintenance of meristem indeterminacy.

TCP TFs were predicted to bind *SP* promoters at three locations approximately -700 bp from ATG start sites, two locations in *TFL1-L1* promoters approximately -400 bp from ATG start sites, and *TFL1-L2* promoters at two locations -300 bp upstream from the ATG start sites (Fig 5.3B-D and Table 5.2). In general, *SP*-binding TCPs are more abundantly expressed in bud tissue compared to other tissues (Ma et al., 2014), matching the X-gluc activity seen in Arabidopsis transformed with *GhSP$_{pro}$:uidA* (Figs 5.1A and 5.2A-B)*. TFL1-L1$_{pro}$*- and *TFL1-L2$_{pro}$*-binding TCPs were reported to be more abundantly expressed in leaf and shoot tissues (Ma et al., 2014), corresponding to the vasculature X-gluc staining observed in these *promoter:uidA* transformed plants (Figs 5.1B-C and 5.2C). Similar to cotton *SFT$_{pro}$*, *TFL1-L1$_{pro}$*, and *TFL1-L2$_{pro}$* sequences were predicted to be bound by multiple Dof TFs, suggesting that vasculature expression seen with *TFL1-L1$_{pro}$:uidA* and *TFL1-L2$_{pro}$:uidA* could be influenced by these factors (Fig 5.3C-D and Table 5.2).

Among the sixteen conserved regulators of cotton *SP$_{pro}$* sequences, leaf and reproductive structure development, ethylene response, meristem maintenance and aging GO terms were enriched (Fig 5.3B). Thirteen TFs predicted to regulate cotton *TFL1-L1* genes were

enriched in shoot system and reproductive structure development, and ethylene, gibberellin, and red or far red light response GO terms (Fig 5.3C). The predicted TFs associated with response to red or far-red light is interesting since leaf vasculature X-gluc staining was observed in *TFL1-L1$_{pro}$:uidA* plants. Together, the data suggest photoperiodic regulation of *TFL1-L1*. Meristem and reproductive structure development and meristem maintenance were also over-represented GO terms of the twelve TFs conserved in regulation of cotton *TFL1-L2* promoter sequences (Fig 5.3D).

Over-represented GO terms of six predicted regulators of cotton *BFT-L1* orthologs does not include terms associated with development (Fig 5.3E and Table 5.2). Phylogenetic analysis of CETS polypeptide sequences that include CETS from Malvales, Jute (*Corchorus olitorius* and *Corchorus capsularis* assemblies) and cacao (*Theobroma cacao*) (Argout et al., 2011) show that cotton is the sole Malvid (Brassicales-Malvales) with a sequenced genome to have two BFT homologs (Fig 2.3). The absence of X-gluc staining in *GhBFT-L1$_{pro}$:uidA* plants analyzed along with a lack of predicted TFs associated with development in GO enrichment analysis alludes to loss of developmental function of this duplicated gene through changes in its promoter sequence.

Cotton's other *BFT* homologs, *BFT-L2* genes, are predicted to be regulated by six TFs with enriched GO terms for flower and shoot system development and responsiveness to auxin (Fig 5.3F and Table 5.2). Similar to other cotton *CETS* promoters, *BFT-L2$_{pro}$* sequences were predicted to be bound by two MADS-box TFs (orthologs to AtAP1 and AtAP3) approximately -1,000 bp from ATG start sites (Fig 5.3F and Table 5.2). This binding prediction may reflect negative regulation of an indeterminate factor to accomplish determinate growth activities.

In contrast to *FT-* and *TFL1*-like cotton homologs, cotton's *MFT*-like genes are not predicted to be regulated by TFs involved in shoot system development. Twenty-three putative TFs showing conserved regulation of cotton *MFT-L1* promoter sequences were over-represented in cell wall biogenesis, xylem development and response to abscisic acid GO terms (Fig 5.3G and Table 5.2). Cotton *MFT-L2* promoter sequences have predicted conserved regulation by nineteen TFs with over-representation of pigment biosynthesis, signaling and response to lipid GO terms (Fig 5.3H and Table 5.2).

5.4   Discussion

To evaluate cotton *CETS* promoter activity, *CETS~pro~:uidA* constructs were introduced into Arabidopsis and GUS staining was evaluated in the T_1 generation. Given that *CETS* characteristically control developmental aspects of plants, GUS activity was assessed at three developmental stages, including early rosette, late rosette and flowering. *GhSFT~pro~* promoted weak GUS activity, primarily observed in the root apical meristem, but also visible within the minor veins at the tips of expanding leaves in 40% of plants in which staining was observed. This activity was only observed during the early rosette stage and GUS activity was absent in all *GhSFT~pro~:uidA* plants observed during later developmental stages; although, roots were not assayed at later developmental stages. These results correspond to the partial rescue of the *ft-10* mutant with the *GhSFT* genomic clone in which plants mimicked WT growth during vegetative growth, but where more intermediate between WT and EV controls after the transition to reproductive growth occurred (Chapter 3, Fig 3.1). In general, florigen homologs are proposed to be expressed in leaf vasculature in response to stimuli and transported into

shoot meristems to promote determinate growth. As discussed in Chapter 4, complementation

of *ft* by an Arabidopsis genomic clone requires that *FT* be driven by at least 5.7 kb upstream of

its translational start site (Adrian et al., 2010). While our 1.8 kb *SFT_{pro}* drove weak GUS activity

in leaf vasculature, a 4.0 kb of *FT_{pro}* has no vasculature GUS activity. Additionally, plants

transformed with *4.0kbFT_{pro}:FT* show no complementation of ft-10 (Adrian et al., 2010). These

differences seem to indicate different upstream regulation requirements for gene expression of

GhSFT and *FT* and may contribute to their differing flowering responses to photoperiod. The

weakness of staining in the shoot system and partial (as opposed to full) rescue of *ft-10* by the

GhSFT genomic clone might reflect: 1. that cotton's *SFT_{pro}* has diverged significantly from that of

FT such that it has developed differing regulations or 2. that the 1.8 kb *SFT_{pro}* used for these

studies might not include *SFT*'s full regulatory elements.

In Chapter 3, cotton *TFL1*-like *CETS* with constitutive promotion acted to maintain

indeterminate growth in Arabidopsis. However, introduction of genomic clones of these genes

into the early flowering *tfl1-14* in Chapter 4 showed that only *GhSP, GhTFL1-L2* and *GhBFT-L2*

were able to rescue the mutant, while *GhTFL1-L1* and *GhTFL1-L2* could not. This suggested that

while their gene products were capable of functioning in regulating the transition from

indeterminate to determinate growth, differences in their regulatory elements led to functional

divergence of these paralogous genes. To further explore *TFL1*-like *CETS* regulatory elements,

promoter:uidA constructs were introduced into Arabidopsis and 2.0 kb of promoter sequence

was computationally analyzed. *GhSP_{pro}* was found to promote GUS activity specifically in

meristem throughout Arabidopsis. Full analysis of *GhSP* in our studies indicate that *GhSP* likely

acts locally to maintain indeterminate growth; although, in cotton *GhSP* was found to be

expressed in all tissue types suggesting that there are probably different mechanisms in cotton regulating *GhSP* expression (McGarry et al., 2016). Additionally, *Gossypium SP* promoter sequences were predicted to have conserved regulation by TFs involved in developmental processes, reinforcing evidence that *SP* acts developmentally to maintain indeterminate growth.

In contrast, *GhTFL1-L1_{pro}:uidA* and *GhTFL1-L2_{pro}:uidA* promoted GUS activity primarily in vasculature tissues throughout the Arabidopsis plant. This activity may indicate these genes act as anti-florigens: indeterminate factors produced in leaves and transported into meristems where they act to inhibit determinate growth, or that their action could be more local controlling other developmental aspects such as leaf development. Unpublished spatial expression of these genes as determined by RT-qPCR, demonstrate that both are mostly expressed in source leaves in photoperiodic sensitive TX701 under non-inductive LD conditions, while this expression is absent for *TFL1-L1* or very low for *TFL1-L2* under SD conditions (Roisin McGarry, personal communication). This expression pattern correlates with GUS activity driven by these promoters in leaf vasculature of Arabidopsis. Down-regulation under inductive photoperiod conditions suggests that these genes might be anti-florigenic components that are expressed specifically in response to photoperiod during non-inductive conditions.

GhBFT-L2_{pro}, but not *GhBFT-L1_{pro}* promoted GUS activity when introduced into Arabidopsis. Similarly, a *GhBFT-L2,* but not a *GhBFT-L1,* genomic clone rescued the *tfl1-14* mutant. A report on flowering time genes in *G. raimondii* also failed to locate expression of *GhBFT-L1* in several tested tissues of photoperiodic sensitive *G. hirsutum* race Yucatenense

(Grover et al., 2015). Additionally, unpublished spatial expression data shows only very low expression of *GhBFT-L1* in the monopodial main stem of photoperiodic sensitive TX701 specifically under inductive SD conditions. However, in day-neutral line DP61, *GhBFT-L1* expression was low but found in both vegetative and reproductive apices and sink tissues under both LD and SD conditions with higher expression found under inductive SD conditions (Roisin McGarry, personal communication). This might indicate that while *GhBFT-L1$_{pro}$* is inactive in photoperiodic systems, in lines selected for day-neutrality *GhBFT-L1$_{pro}$* activity has been regained.

In this study, *GhBFT-L2$_{pro}$* promoted intense GUS activity in Arabidopsis shoot vasculature tissues throughout the life of the plants. Leaf vasculature GUS activity was observed in a source to sink pattern. GUS activity was also observed in hypocotyl and sepal vasculature. Correlating with this abundant promotion of GUS activity, a *GhBFT-L2* genomic clone was able to rescue the early flowering *tfl1-14* mutant. Taken together, these results suggest that GhBFT-L2 is involved in maintaining indeterminate growth and might act as a distal signal being produced in shoot vasculature tissues and transported into meristems to regulate meristem activities. Spatial expression of *GhBFT-L2* in cotton has yet to be evaluated by RT-qPCR (Roisin McGarry, personal communication).

Cotton *MFT*-like genes expression under native or constitutive regulation did not greatly alter the architecture in any of the three Arabidopsis genetic background tested. Their gene promoters weakly drove GUS activity in the vasculature of petioles and filaments in the case of *GhMFT-L2$_{pro}$*. Computational analysis of these promoters also did not predict regulation by TFs

involved in control of plant architecture. Total evidence indicates that cotton's *MFT*-like genes are not active in the regulation of plant architecture.

Finally, computational analysis for conserved regulation of orthologous *CETS* promoters in four analyzed cotton genomes predicted that cotton *FT-* and *TFL1*-like genes which impact plant architecture when expressed in Arabidopsis predicted regulation by TFs involved in shoot system development and the timing of transition from indeterminate to determinate growth. In this approach orthologous promoter sequences from four cotton genome assemblies were analyzed for conserved regulation by *G. raimondii* TFs. While the results of this analysis 1. predicted binding by TFs involved in biological processes that are consistent with phenotypes observed in our function analysis and 2. predicted clusters of binding regulations similar to what is expected in the integration of several TFs interacting in the control gene expression, this analysis is limited in scope. For instance, the binding motifs of *G. raimondii* TFs are not tested, but instead these motifs have been transferred from binding motifs of experimentally tested orthologs in different plant species such as Arabidopsis and rice. Therefore, while the predictions might be true they would need experimental validation. This task could be accomplished by analysis of gene promoter deletions fused to *uidA* and introduced into Arabidopsis similar to the studies presented here or through CRISPR-Cas9 targeted deletions of promoter elements of interest in cotton.

Table 5.1 Oligonucleotides used for cloning of *promoter:uidA* constructs.

| Primer	Sequence
GhSPp SbfI nt -2031 fwd	aagcttcctgcaggggggtatggcatgagaaatcacc
GhSPp XmaI nt -1 rev	cagtttcccgggcccacaaactaatataacactgg
GrBFT-L1p nt -2013 fwd	ggtgttcttatgtagtgacaccgattatttaaacctgcaggatgtcaatttgacgatcaatgtcg
GhBFT-L1 XmaI nt -1 rev	ctcgtgcccggggatatatatattttttagctaatg
GhSFTp SbfI nt -1806 fwd	ctcgtgcctgcaggcctaagcctaaaaatcagctaccctac
GhSFTp XmaI nt -1 rev	atctctcccggggatatcgctatttggtcttactgtg
GhTFL1-L1p SbfI nt -2076 fwd	aagcttcctgcagggagacttgtataggttttgc
GhTFL1-L1p XmaI nt -1 rev	ttccctcccgggttgaggagttctgaatgaaag
GhMFT-L1p SbfI nt -1923 fwd	atatctcctgcaggtttagatctctaactgagttggtgagatg
GhMFT-L1p XmaI nt -1 rev	tcccgggtctagagggagaaagaggagtgggggtggggtgca
GhBFT-L2p SbfI nt -1959 fwd	aagcttcctgcaggggatccaaagagtgatttaacc
GhBFT-L2p XmaI nt -1 rev	gaccctcccggggatgaacaagacgatatgtatg
GhMFT-L2p SbfI nt -2231 fwd	ctcagacctgcaggcctttgaagccctcttcctttt
GhMFT-L2p XmaI nt -1 rev	ggaccgcccgggagtgtgttggactagacctg
GhTFL1-L2p SbfI nt -2124 fwd	ctcgtgcctgcaggtgcaaaattttggaggctacaact
GhTFL1-L2p XmaI nt -1 rev	ctcgtgcccgggtggtgacactgaatgaaaagaagaga
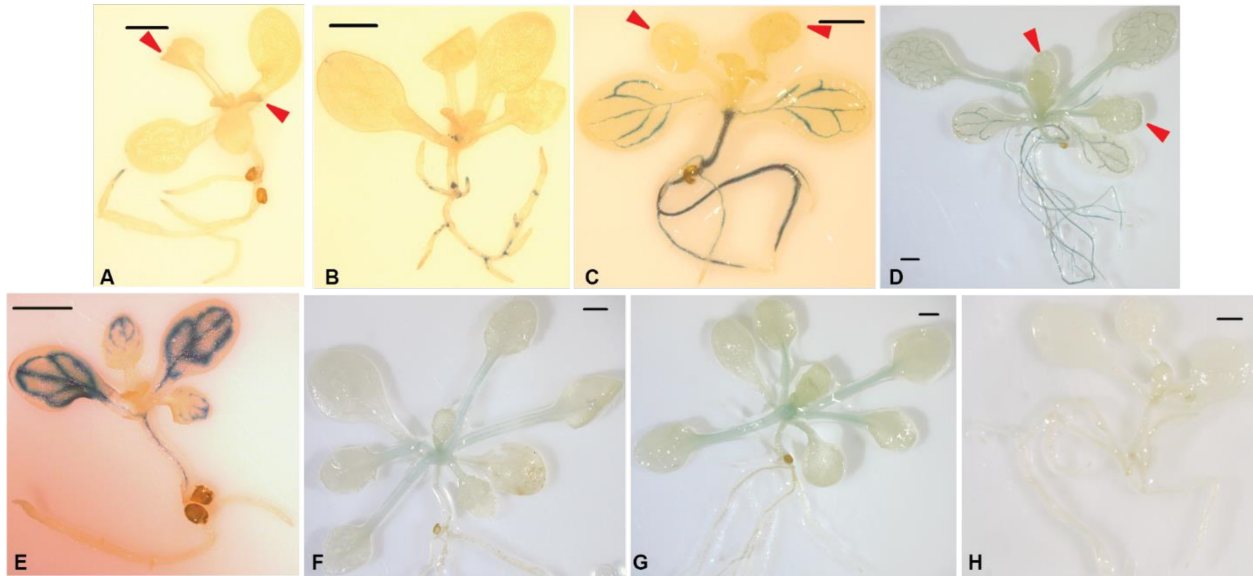
Figure 5.1 GUS activity in transgenic Arabidopsis carrying *GhCETS_pro:uidA* during early development. (A) Staining in 9 dpg *GhSFT_pro:uidA* plants appeared predominately in the RAM; in 40 % of plants with visible staining, X-gluc was also observed in a source to sink pattern in leaf vasculature. (B) X-Gluc staining is detected in SAM, axillary meristems, hydathodes, RAM and root lateral meristems of 9 dpg plants harboring *GhSP_pro:uidA.* (C) Staining in 9dpg *GhTFL1-L1_pro:uidA* plants is intense in root and hypocotyl vasculature with fainter staining of leaf vasculature in source to sink patterning. (D) 10 dpg *GhTFL1-L2_pro:uidA* plants similarly show intense staining of root and hypocotyl vasculature and source to sink pattern staining in leaf vasculature. (E) X-gluc staining of 9 dpg *GhBFT-L2_pro:uidA* plants appeared most intensely in a source to sink pattern in leaf vasculature; staining was also evident in hypocotyl vasculature, but absent in the root. (F) X-gluc staining was faint in the petiole vasculature of fully expanded leaves of 10 dpg *GhMFT-L1_pro:uidA* plants. (G) Similarly, 10 dpg *GhMFT-L2_pro:uidA* plants stained in petiole vasculature of fully expanded leaves; staining in *GhMFT-L2p:uidA* plants was slightly more intense than observed in *GhMFT-L1p:uidA* plants. (H) EV pGPTV:BAR 10 dpg plants treated with X-gluc did not stain. Red arrows in (A), (C), and (D) point to faint staining in the distal veins of expanding leaves. Scale bars = 1 mm.
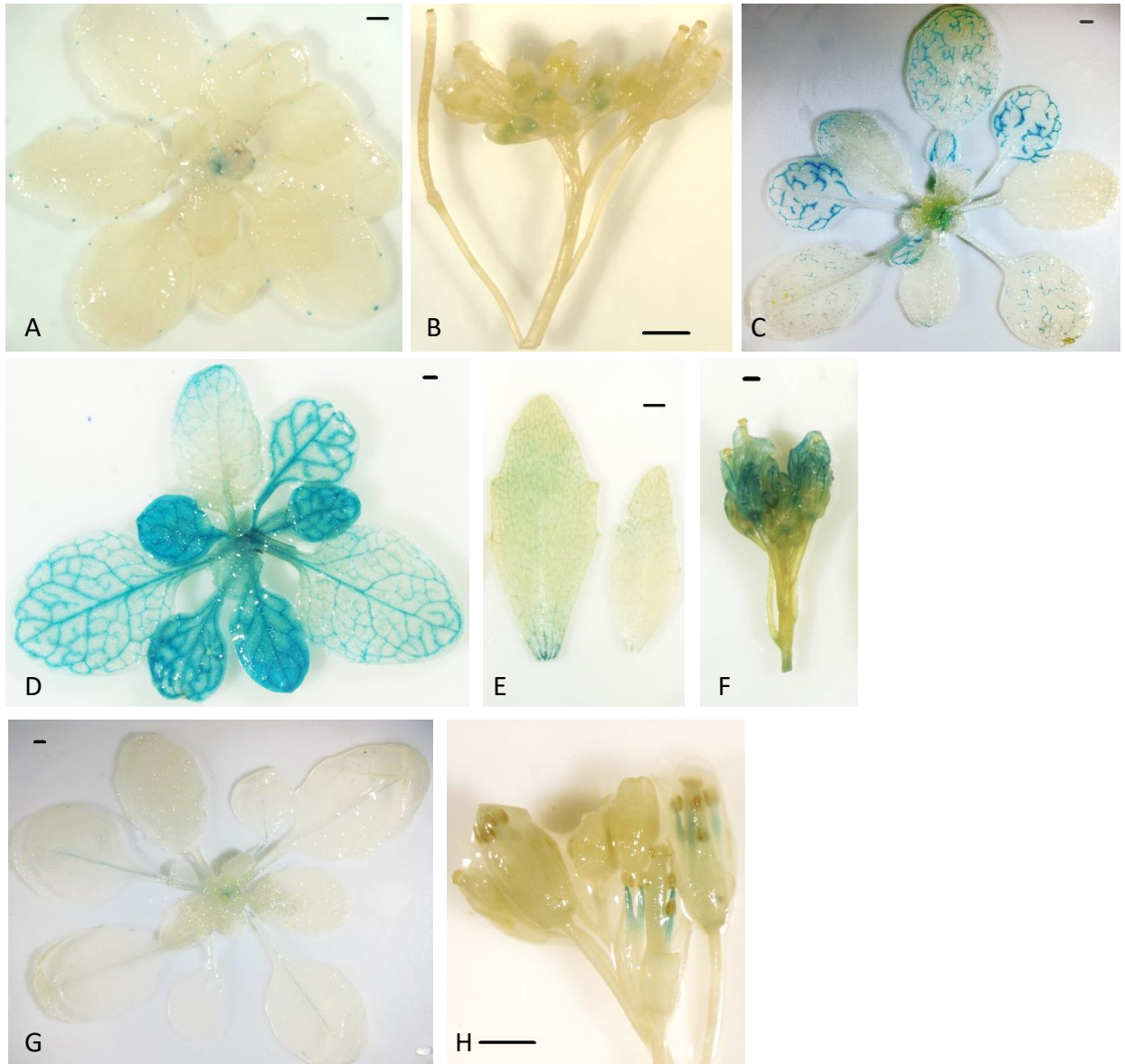
Figure 5.2 GUS Activity of mature rosette and flowering transgenic Arabidopsis harboring

GhCETS_{pro}:uidA. (A) SAM, axillary meristem, and hydathode X-gluc staining in 26 dpg

GhSP_{pro}:uidA plants. (B) X-gluc staining in immature floral buds of 47 dpg *GhSP_{pro}:uidA* plants.

(C) Source to sink pattern staining in the leaf vasculature of 20 dpg *GhTFL1-L2_{pro}:uidA* plants. (D)

Intense source to sink staining pattern in leaf vasculature of 26 dpg *GhBFT-L2_{pro}:uidA* plants. (E-

F) X-gluc staining in leaf and sepal vasculature in 47 dpg *GhBFT-L2_{pro}:uidA* plants. (G) Mid-rib X-

gluc staining in 20 dpg plants harboring *GhMFT-L2$_{pro}$:uidA*. (H) Filament staining in 62 dpg

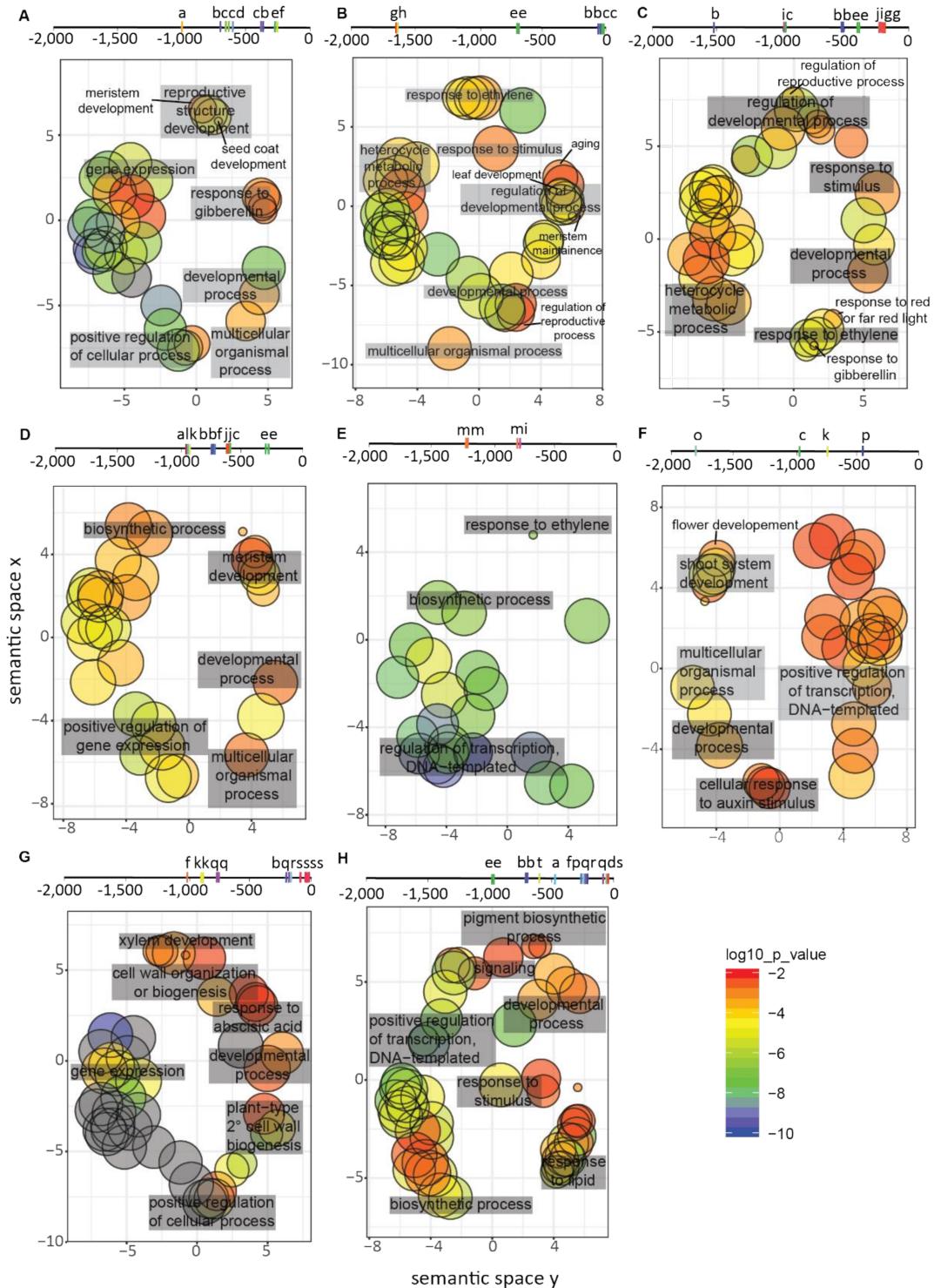*GhMFT-L2p:uidA* plants. Scale bars = 1 mm.

Figure 5.3 Cotton *CETS* genes were predicted to have conserved regulation in developmental

and signaling pathways. 2.0 kb promoter sequences of orthologous cotton *CETS* genes from

four *Gossypium* genome assemblies were analyzed for conserved regulation using the

Regulation Prediction tool at PlantRegMap. Sets of *G. raimondii* TFs conserved in regulation of

all four orthologous cotton promoters were then analyzed for GO term enrichment. Shown are

line diagrams of predicted binding sites for TFs conserved in regulation of *CETS* orthologs.

Letters indicate TF families of predicted binding factors (a complete list of predicted binding

factors is found in table 5.2): a) WOX, b) Dof, c) MIKC-MADS, d) trihelix, e) TCP, f) MYB/G2-like,

g) BBR-BPC, h) ARR-B, i) GRAS, j) AP2, k) NAC, l) Nin-like, m) ERF, n) TALE, o) Arf, p) BES1, q)

bZIP, r) bHLH, s) C2H2 and t) HD-ZIP. Below, scatterplots show cluster representatives in a 2-D

space based upon GO term semantic similarity: (A) *SFT*, (B) *SP*, (C) *TFL1-L1*, (D) *TFL1-L2*, (E) *BFT-L1*, (F) *BFT-L2*, (G) *MFT-L1* and (H) *MFT-L2*. Circle size represents GO term generality with larger

circles correlating to more general terms. Circle color is based upon GO enrichment q-values

using a Fisher's exact test. Several cotton *CETS* are predicted to have conserved regulation by

binding factors associated with plant development and response to endogenous or exogenous

signals associated with plant flowering pathways. Conserved predicted binding factors of *MFT-L1* and *MFT-L2* genes were not enriched in terms associated with flowering pathways.

Table 5.2 Prediction of conserved regulation of *Gossypium* orthologous *CETS* promoter sequences. *Gossypium* orthologous *CETS* promoters from *G. raimondii, G. arboreum,* and *G. hirsutum* genomes were used as input sequences for regulation prediction to predict conserved regulation of cotton *CETS* genes using a *G. raimondii* TF database. TF descriptions were found at Phytozome.net and closest Arabidopsis homologs are based upon BLASTp searches (homologs in red indicate orthologous genes based upon reciprocal BLASTp searches), For binding site locations, 0 = 2,000 upstream of ATG start codons and ending +/- indicates strand binding.

Gossypium Promoter Sequence	Conserved TF Identifier	Conserved TF Description (At homolog)	TF Family	Bound Promoter	Binding Site
SFT	(a) Gorai.001G199200	WUSCHEL-related homeobox 10-related (AtWOX13)	WOX	GaSFTp	1053 - 1062 +
				GhSFT-Ap	1054 - 1063 +
				GhSFT-Dp	1039 - 1048 +
				GrSFTp	1036 - 1045 +
	(a) Gorai.006G188900 (b) Gorai.009G162500 (c) Gorai.009G319300 (d) Gorai.011G040200 (e) Gorai.011G067500	(AtDof5.6) (AtCOGWHEEL1) Dof2.4 (AtDof2.4) (AtCOGWHEEL1) (AtDof2)	Dof	GaSFTp	1290 - 1310 +
				GhSFT-Ap	1291 - 1311 +
				GhSFT-Dp	1289 - 1309 +
				GrSFTp	1288 - 1308 +
	(a) Gorai.005G087600	PISTILLATA (AtPI)	MIKC-MADS	GaSFTp	1352 - 1365 +
				GhSFT-Ap	1353 - 1366 +
				GhSFT-Dp	1351 - 1364 +
				GrSFTp	1350 - 1363 +
	(a) Gorai.005G087600	PISTILLATA (AtPI)	MIKC-MADS	GaSFTp	1377 - 1390 +
				GhSFT-Ap	1378 - 1391 +
				GhSFT-Dp	1376 - 1389 +

				GrSFTp	1375 - 1388 +
	(a) Gorai.009G333800	GT-1 related (AtGT-1)	trihelix	GaSFTp	1392 - 1399 +/-
				GhSFT-Ap	1393 - 1400 +/-
				GhSFT-Dp	1391 - 1398 +/-
				GrSFTp	1390 - 1397 +/-
	(a) Gorai.008G115200	(AtSOC1)	MIKC-MADS	GaSFTp	1616 - 1636 -
				GhSFT-Ap	1617 - 1637 -
				GhSFT-Dp	1617 - 1637 -
				GrSFTp	1619 - 1639 -
	(a) Gorai.011G067500 (b) Gorai.011G040200 (c) Gorai.009G319300 (d) Gorai.009G162500 (e) Gorai.006G188900 (f) Gorai.001G067000 (g) Gorai.003G021000 (h) Gorai.006G173700 (i) Gorai.008G193700 (j) Gorai.009G100100	(AtDof2) (AtCOGWHEEL1) Dof2.4 (AtDof2.4) (AtCOGWHEEL1) (AtDof5.6) (AtOBF BINDING PROTEIN 4) Dof 3.4 (AtOBF BINDING PROTEIN 1) (AtDof2.4) CELLGROWTH DEFECT FACTOR2 related (AtCDF3) (AtOBF BINDING PROTEIN 3)	Dof	GaSFTp	1615(25) - 1635 (46) +
				GhSFT-Ap	1615(25) - 1635 (46) +
				GaSFTp	1615(25) - 1635 (46) +
				GhSFT-Ap	1615(25) - 1635 (46) +
	(a) Gorai.001G072200	(AtTCP14)	tcp	GaSFTp	1713 - 1732 -
				GhSFT-Ap	1714 - 1733 -
				GhSFT-Dp	1714 - 1733 -
				GrSFTp	1716 - 1735 -
	(a) Gorai.008G086400	REGULATOR OF AXILLARY MERISTEMS2 (AtRAX3)	MYB	GaSFTp	1720 - 1732 -
				GhSFT-Ap	1721 - 1733 -
				GhSFT-Dp	1721 - 1733 -
				GrSFTp	1723 - 1735 -
SP	(a) Gorai.002G124900	BPC6-related (AtBPC6)	BBR-BPC	GaSPp	438 - 458 -
				GhSP-Ap	373 - 393 -

			GhSP-Dp	334 - 354 -
			GrSPp	258 - 278 -
(a) Gorai.002G171200	BPC1-related (AtBPC2)	BBR-BPC	GaSPp	436 -459 +
			GhSP-Ap	371 - 394 +
			GhSP-Dp	332 - 355 +
			GrSPp	258 - 281 +
(a) Gorai.006G259700	ARABIDOPSIS RESPONSE REGULATOR1-related (AtARR1)	ARR-B	GaSPp	453 - 462 -
			GhSP-Ap	388 - 397 -
			GhSP-Dp	349 - 358 -
			GrSPp	338 -347 -
(a) Gorai.011G086900	TCP20-related (AtTCP20)	TCP	GaSPp	1294 - 1314 +
			GhSP-Ap	1294 - 1314 +
			GhSP-Dp	1290 - 1310 +
			GrSPp	1277 - 1297 +
(a) Gorai.006G197000	TCP19 (AtTCP19)	TCP	GaSPp	1304 - 1311 -
(b) Gorai.009G153900	TCP2-related (AtTCP2)		GhSP-Ap	1304 - 1311 -
(c) Gorai.005G211900	TCP21-related (AtTCP7)		GhSP-Dp	1300 - 1307 -
(d) Gorai.006G043800	TCP20-related (AtTCP20)		GrSPp	1287 - 1294 -
(e) Gorai.013G068600	TCP21-related (AtTCP7)			
(a) Gorai.001G200400	TCP1 (AtTCP1)	TCP	GaSPp	1303 - 1332 +
			GhSP-Ap	1303 - 1332 +
			GhSP-Dp	1299 - 1328 +
			GrSPp	1286 - 1315 +
(a) Gorai.003G021000	Dof 3.4 (AtOBF BINDING PROTEIN 1)	Dof	GaSPp	1941 - 1961 -
(b) Gorai.009G100100	(AtOBF BINDING PROTEIN 3)		GhSP-Ap	1941 - 1961 -
			GhSP-Dp	1940 - 1960 -
			GrSPp	1941 - 1961 -
(a) Gorai.011G067500	(AtDof2)	Dof	GaSPp	1950 - 1970 +
			GhSP-Ap	1950 - 1970 +

				GhSP-Dp	1949 - 1969 +
				GrSPp	1950 - 1970 +
	(a) Gorai.008G115200 (b) Gorai.N017200	(AtSOC1) AGAMOUS-like AGL11 (AtAG)	MIKC_MADS	GaSPp	1941(44) - 1961(62) -
				GhSP-Ap	1941(44) - 1961(62) -
				GhSP-Dp	1940(43) - 1960(61) -
				GrSPp	1941(47) - 1961(62) -
	(a) Gorai.005G087600	PISTILLATA (AtPI)	MIKC_MADS	GaSPp	1947 - 1960 +
				GhSP-Ap	1947 - 1960 +
				GhSP-Dp	1946 - 1959 +
				GrSPp	1947 - 1960 +
TFL1-L1	(a) Gorai.009G319300	Dof2.4 (AtDof2.4)	Dof	GaTFL1-L1p	471 - 491 +
				GhTFL-L1-Ap	476 - 496 +
				GhTFL1-L1-Dp	480 - 500 +
				GrTFL1-L1p	475 - 495 +
	(a) Gorai.002G177000	DELLA protein (AtREPPRESSOR OF GA)	GRAS	GaTFL1-L1p	1011 -1030 +
				GhTFL-L1-Ap	1016 - 1035 +
				GhTFL1-L1-Dp	1012 - 1031 +
				GrTFL1-L1p	1007 - 1026 +
	(a) Gorai.008G115200	(AtSOC1)	MIKC-MADS	GaTFL1-L1p	1012 - 1032 -
				GhTFL-L1-Ap	1017 - 1037 -
				GhTFL1-L1-Dp	1013 - 1033 -
				GrTFL1-L1p	1008 - 1028 -
	(a) Gorai.009G319300	Dof2.4 (AtDof2.4)	Dof	GaTFL1-L1p	1480 - 1500 -
				GhTFL-L1-Ap	1480 - 1500 -
				GhTFL1-L1-Dp	1485 - 1505 -
				GrTFL1-L1p	1480 - 1500 -
	(a) Gorai.011G067500	(AtDof2)	Dof	GaTFL1-L1p	1480 - 1500 +
				GhTFL-L1-Ap	1480 -1500 +
				GhTFL1-L1-Dp	1484 - 1504 +

			GrTFL1-L1p	1479 - 1499 +
(a) Gorai.005G211900 (b) Gorai.008G181600 (c) Gorai.013G068600	TCP21-related (AtTCP7) TCP15 (AtTCP15) TCP21-related (AtTCP7)	TCP	GaTFL1-L1p GhTFL-L1-Ap GhTFL1-L1-Dp GrTFL1-L1p	1596 - 1606 - 1596 - 1606 - 1595 - 1605 - 1590 - 1600 -
(a) Gorai.007G094200 (b) Gorai.008G157300 (c) Gorai.012G084600	TCP9 (AtTCP9) TCP20-related (AtTCP20) TCP20-related (AtTCP20)	TCP	GaTFL1-L1p GhTFL-L1-Ap GhTFL1-L1-Dp GrTFL1-L1p	1597 - 1606 + 1597 - 1606 + 1596 - 1605 + 1591 - 1600 +
(a) Gorai.004G263900	BABYBOOM (AtBBM)	AP2	GaTFL1-L1p GhTFL-L1-Ap GhTFL1-L1-Dp GrTFL1-L1p	1781 - 1800 + 1781 - 1800 + 1780 - 1799 + 1755 - 1974 +
(a) Gorai.002G177000	DELLA protein (AtRGA)	GRAS	GaTFL1-L1p GhTFL-L1-Ap GhTFL-L1-Dp GhTFL1-L1-Dp	1789 - 1808 + 1789 - 1808 + 1791 - 1810 + 1786- 1805 +
(a) Gorai.002G124900	BPC6 (AtBPC6)	BBR-BPC	GaTFL1-L1p GhTFL-L1-Ap GhTFL1-L1-Dp GrTFL1-L1p	1789 - 1809 - 1791 - 1811 - 1786 - 1806 - 1781 - 1801 -
(a) Gorai.002G124900	BPC6 (AtBPC6)	BBR-BPC	GaTFL1-L1p GhTFL-L1-Ap GhTFL1-L1-Dp GrTFL1-L1p	1791 - 1811 - 1789 - 1809 - 1788 - 1808 - 1783 - 1803 -
(a) Gorai.002G124900	BPC6 (AtBPC6)	BBR-BPC	GaTFL1-L1p GhTFL-L1-Ap GhTFL1-L1-Dp GrTFL1-L1p	1793 - 1813 - 1793 - 1813 - 1790 - 1810 - 1785 - 1805 -
(a) Gorai.002G171200	BPC1-related (AtBPC2)	BBR-BPC	GaTFL1-L1p	1787 - 1810 +

			GhTFL-L1-Ap	1787 - 1810 +
			GhTFL1-L1-Dp	1786 - 1809 +
			GrTFL1-L1p	1781 - 1804 +
	(a) Gorai.002G171200	BPC1-related (AtBPC2)	GaTFL1-L1p	1789 - 1812 +
			GhTFL-L1-Ap	1789 - 1812 +
		BBR-BPC	GhTFL1-L1-Dp	1788 - 1811 +
			GrTFL1-L1p	1785 - 1808 +
	(a) Gorai.002G171200	BPC1-related (AtBPC2)	GaTFL1-L1p	1791 - 1814 +
			GhTFL-L1-Ap	1791 - 18114 +
		BBR-BPC	GhTFL1-L1-Dp	1790 - 1813 +
			GrTFL1-L1p	1783 - 1806 +
TFL1-L2	(a) Gorai.011G098600	WUSCHEL (AtWUS)	GaTFL-L2p	1094 - 1104 -
			GhTFL1-L2-Ap	1151 - 1125 -
		WOX	GhTFL1-L2-Dp	1079 - 1089 -
			GrTFL1-L2p	1112 - 1122 -
	(a) Gorai.002G115800	Nin-Like PROTEIN4-related (AtNLP4)	GaTFL-L2p	1083 - 1097 +
			GhTFL1-L2-Ap	1104 - 1118 +
		Nin-like	GhTFL1-L2-Dp	1068 - 1082 +
			GrTFL1-L2p	1101 - 1115 +
	(a) Gorai.009G309000	NAC protein 53-related (AtNAC2)	GaTFL-L2p	1069 - 1080 -
			GhTFL1-L2-Ap	1090 - 1101 -
		NAC	GhTFL1-L2-Dp	1054 - 1065 -
			GrTFL1-L2p	1087 - 1098 -
	(a) Gorai.002G234700 (b) Gorai.006G188900	Dof1.1-related (AtDof5.6)	GaTFL-L2p	1265(67) - 1277(85) - 1286(88) - 1298(1306) -
			GhTFL1-L2-Ap	-
		Dof	GhTFL1-L2-Dp	1250(52) - 1262(70) -
			GrTFL1-L2p	1285 - 1295(1305) -
	(a) Gorai.009G319300	Dof2.4 (AtDof2.4)	GaTFL-L2p	1267 - 1287 +
		Dof	GhTFL1-L2-Ap	1288 - 1308 +

				GhTFL1-L2-Dp	1252 - 1272 +
				GrTFL1-L2p	1285 - 1305 +
	(a) Gorai.007G113900	CIRCADIAN CLOCK ASSOCIATED1-related (AtLATE ELOGATED HYPOCOTYL1)	MYB-related	GaTFL-L2p	1293 - 1301 -
				GhTFL1-L2-Ap	1314 - 1322 -
				GhTFL1-L2-Dp	1278 - 1286 -
				GrTFL1-L2p	1311 - 1319 -
	(a) Gorai.004G263900	BBM (AtBBM)	AP2	GaTFL-L2p	1411 - 1430 +
				GhTFL1-L2-Ap	1436 - 1455 +
				GhTFL1-L2-Dp	1397 - 1416 +
				GrTFL1-L2p	1430 - 1449 +
	(a) Gorai.004G263900	BBM (AtBBM)	AP2	GaTFL-L2p	1414 - 1433 +
				GhTFL1-L2-Ap	1438 - 1457 +
				GhTFL1-L2-Dp	1399 - 1418 +
				GrTFL1-L2p	1432 - 1451 +
	(a) Gorai.008G115200	(AtSOC1)	MIKC-MADS	GaTFL-L2p	1414 - 1434 -
				GhTFL1-L2-Ap	1434 - 1454 -
				GhTFL1-L2-Dp	1398 - 1418 -
				GrTFL1-L2p	1431 - 1451 -
	(a) Gorai.001G200400	TCP1 (AtTCP1)	TCP	GaTFL-L2p	1676 - 1705 -
				GhTFL1-L2-Ap	1692 - 1721 -
				GhTFL1-L2-Dp	1700 - 1729 -
				GrTFL1-L2p	1698 - 1727 -
	(a) Gorai.009G289000	TCP16-related (AtTCP23)	TCP	GaTFL-L2p	1696 - 1705 +
				GhTFL1-L2-Ap	1712 - 1721 +
				GhTFL1-L2-Dp	1720 - 1729 +
				GrTFL1-L2p	1718 - 1727 +
BFT-L1	(a) Gorai.008G155600	Ethylene Response Factor(ERF)087	ERF	GaBFT-L1p	774 - 788 +
				GhBFT-L1-Ap	322 - 336 +
				GhBFT-L1-Dp	938 - 952 +

				GrBFT-L1p	872 - 886 +
	(a) Gorai.011G029700	(AtERF12)	ERF/AP2	GaBFT-L1p	774 - 784 -
				GhBFT-L1-Ap	325 - 332 -
				GhBFT-L1-Dp	945 - 952 -
				GrBFT-L1p	879 - 886 -
	(a) Gorai.013G244500	SHINE (AtRelated to AP2.11)	ERF	GaBFT-L1p	774 - 788 -
				GhBFT-L1-Ap	322 - 336 -
				GhBFT-L1-Dp	938 - 952 -
				GrBFT-L1p	872 - 886 -
	(a) Gorai.002G177000	DELLA protein (AtRGA)	GRAS	GaBFT-L1p	1089 - 1108 -
				GhBFT-L1-Ap	637 - 656 -
				GhBFT-L1-Dp	1577 - 1596 -
				GrBFT-L1p	1510 - 1529 -
	(a) Gorai.001G036500	ERF003 (AtETHYLENE AND SALT INDUCEIBLE3)	ERF	GaBFT-L1p	1089 - 1108 -
				GhBFT-L1-Ap	637 - 656 -
				GhBFT-L1-Dp	1577 - 1596 -
				GrBFT-L1p	1510 - 1529 -
BFT-L2	(a) Gorai.011G238900	AUXIN RESPONSE FACTOR10-related (AtARF16)	Arf	GaBFT-L2p	71 - 91 +
				GhBFT-L2-Ap	138 - 158 +
				GhBFT-L2-Dp	341 - 361 +
				GrBFT-L2p	341 - 361 +
	(a) Gorai.009G271100 (b) Gorai.013G096100	APETELLA3 (AtAP3) AG-like AGL8-related (AtAP1)	MIKC-MADS	GaBFT-L2p	976 - 990 -
				GhBFT-L2-Ap	986 - 1000 -
				GhBFT-L2-Dp	1115 - 1126 -
				GrBFT-L2p	1115 - 1126 -
	(a) Gorai.004G186700 (b) Gorai.007G112500	(AtNAC071) NAC-43-related (AtNac SECONDARY WALL THICKENING1)	NAC	GaBFT-L2p	1249 - 1263 -
				GhBFT-L2-Ap	1257 - 1271 -
				GhBFT-L2-Dp	1361 - 1375 -
				GrBFT-L2p	1361 - 1375 -
	(a) Gorai.008G218300	Beta Amylase 8 (AtBeta Amylase 8)	BES1	GaBFT-L2p	1545 - 1559 +

				GhBFT-L2-Ap	1553 - 1567 +
				GhBFT-L2-Dp	1548 - 1562 +
				GrBFT-L2p	1548 - 1562 +
MFT-L1	(a) Gorai.007G274700 (b) Gorai.008G185500	MYB-related (AtPhytochrome-Dependent Late-Flowering) MYB-related (AtHHO2)	G2-like	GaMFT-L1p	993(5) - 1005(9) +
				GhMFT-L1-Ap	1000(2) - 1012(16) +
				GhMFT-L1-Dp	1005(7) - 1017(21) +
				GrMFT-L1p	1027(29) - 1039(43) +
	(a) Gorai.004G267300 (b) Gorai.005G195300 (c) Gorai.006G060900 (d) Gorai.007G188800 (e) Gorai.008G155200 (f) Gorai.008G236400 (g) Gorai.009G186000 (h) Gorai.011G090000	BEARSKIN1 (AtBEARSKIN2) No Apical Meristem (NAM) (AtVND INTERACTING2) NAM(AtNAC058) (AtNAC028) NAC protein 10 (AtNacSECONDARY WALL THICKENING1) NAC26-related (AtVND4) CUP SHAPE COTYLEDON 3 (AtCUC3) NAM (AtLOV1)	NAC	GaMFT-L1p	1126(29) - 1144(49) -
				GhMFT-L1-Ap	1132 (36) - 1150(55) -
				GhMFT-L1-Dp	1134(37) - 1152(57) -
				GrMFT-L1p	1156(59) - 1172(79) -
	(a) Gorai.005G195300 (b) Gorai.009G166300 (c) Gorai.010G124100 (d) Gorai.013G146300	NAM/NAC10-related (AtVNI2) NAM/NAC10 (AtSND3) NAC38-related (AtNAC038) NAC20-related (AtNAC20)	NAC	GaMFT-L1p	1129(30) - 1144(9) +
				GhMFT-L1-Ap	1135(6) - 1150(5) +
				GhMFT-L1-Dp	1137(8) - 1152(7) +
				GrMFT-L1p	1159(60) - 1174(9) +
	(a) Gorai.005G015900 (b) Gorai.011G209200	(AtbZIP69) bZIP-1 (AtbZIP18)	bZIP	GaMFT-L1p	1239 - 1249 -
				GhMFT-L1-Ap	1245 - 1255 -
				GhMFT-L1-Dp	1247 - 1257 -
				GrMFT-L1p	1269 - 1279 -
	(a) Gorai.005G015900 (b) Gorai.009G285000	(AtbZIP69) VIRE2 INTERACTING PROTEIN1 (AtVIP1)	bZIP	GaMFT-L1p	1238(40) - 1249(50) +
				GhMFT-L1-Ap	1244(6) - 1255(6) +
				GhMFT-L1-Dp	1246(8) - 1257(8) +
				GrMFT-L1p	1268(70) - 1279(80) +

138

(a) Gorai.003G021000 (b) Gorai.009G100100 (c) Gorai.009G319300	Dof 3.4 (At OBF BINDING PROTEIN 1) (At OBF BINDING PROTEIN 3) Dof2.4 (AtDof2.4)	Dof	GaMFT-L1p GhMFT-L1-Ap GhMFT-L1-Dp	1792(3) - 1812(3) - 1793(4) - 1813(4) - 1791(3) - 1811(3) -
(a) Gorai.005G015900	(AtbZIP69)	bZIP	GaMFT-L1p GhMFT-L1-Ap GhMFT-L1-Dp GrMFT-L1p	1825 - 1835 + 1826 - 1836 + 1825 - 1835 + 1825 - 1835 +
(a) Gorai.005G188700	sterol regulatory protein (AtBIG PETAL)	bHLH	GaMFT-L1p GhMFT-L1-Ap GhMFT-L1-Dp GrMFT-L1p	1836 - 1856 + 1837 - 1857 + 1836 - 1856 + 1836 - 1856 +
(a) Gorai.009G128000	GENERAL TF IIIA (AtTFIIIA)	C2H2	GaMFT-L1p GhMFT-L1-Ap GhMFT-L1-Dp GrMFT-L1p	1913 - 1931 + 1914 - 1932 + 1913 - 1931 + 1913 - 1931 +
(a) Gorai.009G128000	GENERAL TF IIIA (AtTFIIIA)	C2H2	GaMFT-L1p GhMFT-L1-Ap GhMFT-L1-Dp GrMFT-L1p	1974 - 1992 + 1975 - 1993 + 1974 - 1992 + 1974 - 1992 +
(a) Gorai.009G128000	GENERAL TF IIIA (AtTFIIIA)	C2H2	GaMFT-L1p GhMFT-L1-Ap GhMFT-L1-Dp GrMFT-L1p	1977 - 1995 + 1978 - 1996 + 1975 - 1995 + 1977 - 1995 +
(a) Gorai.009G128000 (b) Gorai.010G025200	GTFIIIA (AtTFIIIA) ZINC FINGER PROTEIN3-related (AtSALT TOLERANCE ZINC FINGER)	C2H2	GaMFT-L1p GhMFT-L1-Ap GhMFT-L1-Dp GrMFT-L1p	1979 - 1989(98) + 1980(1) - 1990(9) + 1979(80) - 1989(98) + 1979(80) - 1989(98) +
(a) Gorai.011G234400	C2H2-related	C2H2	GaMFT-L1p GhMFT-L1-Ap	1979 - 1989 - 1980 - 1990 -

139

				GhMFT-L1-Dp	1979 - 1989 -
				GrMFT-L1p	1979 - 1989 -
MFT-L2	(a) Gorai.005G211900 (b) Gorai.013G068600	TCP21-related (AtTCP7) TCP21-related (AtTCP7)	TCP	GaMFT-L2p GhMFT-L2-Ap GhMFT-L2-Dp GrMFT-L2p	1010 - 1020 - 1016 - 1026 - 1007 - 1017 - 1008 - 1018 -
	(a) Gorai.006G043800 (b) Gorai.006G197000	TCP20-related (AtTCP20) TCP19 (AtTCP19)	TCP	GaMFT-L2p GhMFT-L2-Ap GhMFT-L2-Dp GrMFT-L2p	1013 - 1020 + 1019 - 1026 + 1010 - 1017 + 1011 - 1018 +
	(a) Gorai.009G319300	Dof2.4 (AtDof2.4)	Dof	GaMFT-L2p GhMFT-L2-Ap GhMFT-L2-Dp GrMFT-L2p	1299 - 1319 - 739 - 759 - 1296 - 1316 - 1297 - 1317 -
	(a) Gorai.006G188900	(AtDof5.6)	Dof	GaMFT-L2p GhMFT-L2-Ap GhMFT-L2-Dp GrMFT-L2p	1301 - 1321 + 1301 - 1321 + 209 - 229 - 253 - 273 -
	(a) Gorai.007G051100 (b) Gorai.007G206000	ATHHB-21-related (AtHB40) ATHHB-21-related (AtHB40)	HD-ZIP	GaMFT-L2p GhMFT-L2-Ap GhMFT-L2-Dp GrMFT-L2p	1399 - 1419 + 1399 - 1419 + 1399 - 1419 + 1399 - 1419 +
	(a) Gorai.001G199200	WUSCHEL-related homeobox 10-related (AtWOX13)	WOX	GaMFT-L2p GhMFT-L2-Ap GhMFT-L2-Dp GrMFT-L2p	1516 - 1525 + 1516 - 1525 + 1516 - 1525 + 1516 - 1525 +
	(a) Gorai.001G087400	MYB113-related (AtMYB113)	MYB	GaMFT-L2p GhMFT-L2-Ap GhMFT-L2-Dp	1760 - 1770 - 1760 - 1770 - 1760 - 1770 -

			GrMFT-L2p	1760 - 1770 -
(a) Gorai.009G301100	(AtMYB33)	MYB	GaMFT-L2p	1760 - 1770 -
			GhMFT-L2-Ap	1760 - 1770 -
			GhMFT-L2-Dp	1760 - 1770 -
			GrMFT-L2p	1760 - 1770 -
(a) Gorai.009G301400	abscisic acid 5-like related 2 (AtAREB3)	bZIP	GaMFT-L2p	1781 - 1795 +
			GhMFT-L2-Ap	1781 - 1795 +
			GhMFT-L2-Dp	1781 - 1795 +
			GrMFT-L2p	1781 - 1795 +
(a) Gorai.001G183400	PIF4-related (AtPIF1)	bHLH	GaMFT-L2p	1786 - 1799 -
			GhMFT-L2-Ap	1786 - 1799 -
			GhMFT-L2-Dp	1786 - 1799 -
			GrMFT-L2p	1786 - 1799 -
(a) Gorai.008G024700	abscisic acid 5-like related (AtABF1)	bZIP	GaMFT-L2p	1786 - 1793 +
			GhMFT-L2-Ap	1786 - 1793 +
			GhMFT-L2-Dp	1786 - 1793 +
			GrMFT-L2p	1786 - 1793 +
(a) Gorai.008G024700 (b) Gorai.009G212600	abscisic acid 5-like related (AtABF1) camp-response element binding protein related (AtABF2)	bZIP	GaMFT-L2p	1783(6) - 1793(1800) -
			GhMFT-L2-Ap	1783(6) - 1793(1800) -
			GhMFT-L2-Dp	1783(6) - 1793(1800) -
			GrMFT-L2p	1783(6) - 1793(1800) -
(a) Gorai.008G218300	Beta Amylase 8 (AtBAM8)	BES1	GaMFT-L2p	1784 - 1798 +
			GhMFT-L2-Ap	1784 - 1798 +
			GhMFT-L2-Dp	1784 - 1798 +
			GrMFT-L2p	1784 - 1798 +
(a) Gorai.007G205700	G-box binding factor 1 (AtGBF1)	bZIP	GaMFT-L2p	1853 - 1862 +
			GhMFT-L2-Ap	1853 - 1862 +
			GhMFT-L2-Dp	1853 - 1862 +
			GrMFT-L2p	1853 - 1862 +

(a) Gorai.010G025200	ZINC FINGER PROTEIN3-related (AtSALT TOLERANCE ZINC FINGER	C2H2	GaMFT-L2p	1878 - 1888 +
			GhMFT-L2-Ap	1878 - 1888 +
			GhMFT-L2-Dp	1878 - 1888 +
			GrMFT-L2p	1878 - 1888 +
(a) Gorai.007G371500	aspartate kinase	Trihelix	GaMFT-L2p	1883 - 1896 -
			GhMFT-L2-Ap	1883 - 1896 -
			GhMFT-L2-Dp	1883 - 1896 -
			GrMFT-L2p	1883 - 1896 -

CHAPTER 6

SUMMARY

Plant architecture is determined by the activities of indeterminate and determinate

meristems. The outcome of these actions significantly effects productivity and crop

management. *CETS* genes, sharing homology with *PEBP* genes, have evolved to regulate plant

growth and development. *FT-* and *TFL1*-like *CETS* are major contributors in regulating the timing

and location of the transition from indeterminate to determinate growth using antagonistic

function to balance the activities of one another. *FT*-like *CETS* in Arabidopsis and other species

are key regulators in the promotion of the transition to determinate growth. *TFL1*-like

homologs compete with this action by maintaining an indeterminate state in meristems. An

established model postulates timing and placement of determinate and indeterminate growth

occurs due to the balance of FT-like and TFL1-like at each meristem. Evidence shows

domestication of desired growth habits in crops resulted from selection of modified FT/TFL1

balance.

Cotton, grown as an annual row crop in the U.S. and other developed nations, retains

perennial characteristics that complicate boll harvest and crop maintenance. Insight into the

mechanisms that regulate plant architecture is important for optimizing plant architecture to

benefit productivity. Additionally, because cotton comprises a complex branching architecture,

this insight also increases our understanding of plant growth and development. Currently, plant

architecture traits influence variety selection based upon environmental conditions as well as

crop management methods. For instance, in the windy high plains of the southwest US, where

greater than 25 percent of US cotton is produced, growing season rainfall and irrigation

capacity is restrictive and storms are recurrent and random. As a result, plants are compact and yield per acre is relatively low. In these regions, finger-stripper harvesters are a preferred harvesting strategy. Comparatively, plants grown in the eastern regions of the US, where rainfall is greater, are more robust and yields per acre are higher. For this different plant architecture, spindle pickers are a more favored harvesting strategy. Because of the high costs of harvesting equipment, a 'one-size-fits-all' plant architecture is unreasonable and plant architecture optimization strategies would benefit in the consideration of regional climatic environments. In addition, the use of robots for 'smart farming' techniques is quickly advancing (King, 2017, www.cropscience.bayer.com/en/stories/2017/high-tech-helpers-for-tomorrows-agriculture-precision-farming-is-the-future). Currently, autonomous robots spray and cultivate crops. In the coming decade, swarming harvest robots and prototype edge-of-farm seed and fiber handling technologies are expected to emerge to aid cotton harvesting. Swarming robots will be able to separate fiber from seed, delivering each to its appropriate handling module. As these technologies develop, the idea of a more-annualized plant architecture altered for compatibility with these advances is envisioned (Kater Hake, Cotton, Inc., personal communication). Alterations of *CETS* gene expression in cotton could produce a finely-tuned architecture with shifts in the timing and placement of indeterminate and determine growth to meets these needs.

This study explored cotton *CETS* biology in an effort to gain understanding on how members of this gene family function in cotton to contribute to overall cotton architecture. In Chapter 2, *CETS* homologs in sequenced *G. raimondii* (D-genome), *G. arboreum* (A-genome) and *G. hirsutum* (A_t- and D_t-subgenomes) assemblies were identified. *CETS* genomic and protein

sequence were analyzed and compared with characterized *CETS* in other species to make

hypotheses regarding function. Cotton genomes comprise eight putative *CETS* genes whose

exon/intron structures were similar to *CETS* in Arabidopsis, maize, tomato and other species.

One putative *CETS* in cotton genomes shared similar genetic structure and amino acid

conservation of key residues with *FT*-like genes that promote the transition to determinate

growth. Five *CETS* comprised genetic structures and conservation of key residues with *TFL1*-like

CETS. Phylogenetic analysis assigned cotton CETS into the three generally accepted subfamilies

of *CETS* in angiosperms: FT-like, TFL1-like, and MFT-like. Cotton FT-like subfamily with the sole

member, SFT, is reduced in comparison to other closely related species while its TFL1-like

subfamily is expanded since its divergence from other *Malvaceae.* Presumably, this expansion

resulted from documented genome duplication before the divergence of the A- and D-genomes

within the *Gossypium* lineage. Similarly, cotton includes two MFT-like homologs which

represents an increase in this subfamily in comparison to some close relatives; again, this

increase is likely due to *Gossypium*-specific genome duplication.

Due to the recalcitrance of cotton to stable transformation, this study used

heterologous expression in Arabidopsis to explore the function of cotton *CETS.* In Chapter 3,

cotton *CETS* were expressed by a constitutive *2xCaMV35S$_{pro}$* to analyze the potential of cotton

CETS gene product for regulation of the transition from indeterminate to determinate growth.

While prior evidence has shown that changing even one amino acid in a CETS protein can alter

protein function (Hanzawa et al., 2005), *Gossypium hirsutum CETS* acted according to

phylogenetic classification. Namely, heterologous expression of *GhSFT* accelerated determinate

growth both before and after the transition to reproduction. These plants were early flowering

and some lines possessed terminal flowers lacking the three outer whorls and having instead

multiple unfused carpels surrounding a fused carpel. These phenotypes represent a quickening

towards determinacy throughout development. In opposition, each of the five heterologously

expressed cotton *TFL1*-like genes (*GhSP, GhTFL1-L1, GhTFL1-L2, GhBFT-L1,* and *GhBFT-L2*)

retarded determinate growth. Transformed plants were very late in flowering and generated a

previously described I1* stage in which the primary inflorescence produced axillary branches

unsubtended by cauline leaves. An interpretation of the I1* structure might be seen as

competition between indeterminate and determinate factors in which constitutively expressed

indeterminate factors resist the action of endogenous determinate factors resulting in

phenotypes that are intermediate to indeterminate and determinate growth. Finally, cotton

MFT-like genes, similar to other characterized genes, had trivial effect on the transitions from

determinate to indeterminate growth and the architecture of observed plants appeared very

similar to WT controls (Table 6.1).

To more completely understand the activities of cotton *CETS* in controlling plant

architecture, genomic clones, harboring two to three kb of promoter, full exon/intron gene, and

one kb terminating sequences were transformed into Arabidopsis flowering-time mutants to

assess levels of mutant rescue. Transformed into the late flowering *ft-10* mutant, a *GhSFT*

genomic clone showed partially rescue of the mutant phenotype, primarily during vegetative

development demonstrating functionality of the gene's regulatory sequence in the Arabidopsis

flowering pathway. This result correlates with other studies showing a conservation of

flowering regulatory mechanisms. Other *CETS* genomic clones expressed in the *ft-10* mutant

background failed to show any level of rescue. Cotton *CETS* genomic clones were likewise

introduced into an early flowering *tfl1-14* mutant. While, each *TFL1*-like *CETS* delayed determinate growth under constitutive expression, only *GhSP, GhTFL1-L2,* and *GhBFT-L2* partially rescued *tfl1-14* early flowering phenotype. Paralogs, *GhTFL1-L1* and *GhBFT-L1*, failed to rescue *tfl1-14*, suggesting that their control of development and plant architecture was lost after gene duplication specifically through changes in their regulatory sequences. A *GhSFT* genomic clone in the *tfl1-14* background further accelerated plant determination, providing further evidence of its activity in accelerating the transition from determinate to indeterminate growth. Cotton *MFT*-like genomic clones showed no impact on plant architecture in the mutant background (Table 6.1).

Lastly, to investigate cotton *CETS* promoter activities, *CETS$_{pro}$:uidA* constructs were introduced into Arabidopsis and studied in the T_1 generation at three developmentally distinct phases: early rosette, late rosette and flowering. *Gossypium CETS* promoters were also computationally analyzed for conserved regulation. A 1.8 kb *GhSFT* promoter drove GUS activity predominately in the root apical meristem, but also weakly in distal minor veins of leaves in 40% of plants in which staining was observed. This staining pattern, taken together with *GhSFT* function in promoting determinate growth in heterologous Arabidopsis backgrounds, suggests that *GhSFT* shares functional conservation to plant florigens that are expressed in leaf tissues and transported via phloem to acts distally to promote determinate growth in shoot meristems. *GhSP$_{pro}$* promoted GUS activity was restricted to plant meristems while promoters of *GhTFL1-L1, GhTFL1-L2* and *GhBFT-L2* drove activity in plant vasculature tissues. Analysis of these genes suggest that they probably act to maintain indeterminate growth in cotton either locally or distally from the site of synthesis. In heterologous

147

Arabidopsis, *GhBFT-L1$_{pro}$* failed to promote GUS activity. Taken together with functional analysis, our evidence suggests that *GhBFT-L1,* a paralogous gene in the *Gossypium* lineage, does not act in regulation of plant architecture and that this loss of function is likely due to differences in regulatory sequences rather than changes in the gene products. Cotton *MFT*-like promoters weakly promoted GUS activity in Arabidopsis and their promoters were not predicted to be regulated by TFs involved in the control of plant architecture. Complete characterization of *MFT*-like genes suggests that they do act to regulate plant architecture (Table 6.1).

In future efforts, virus-based systems or CRISPR-Cas9 mutagenesis should be employed to assay gene function of *GhTFL1-L1, GhTFL1-L2, GhBFT-L1,* and *GhBFT-L2* directly in cotton systems. Additionally, since the aim of cotton plant architecture studies is to produced architectures that are finely-tuned based upon needs generated by environmental conditions and emerging technologies, CRISPR-Cas9 technologies could also be employed for the manipulation of *CETS* promoter or gene sequences to assess how manipulations to individual *CETS* gene expression might impact specific meristems activities within the complex cotton architecture system.

Table 6.1 Summary of cotton *CETS* gene characterizations and predictions of gene effect on cotton plant architecture.

Cotton *CETS* Gene	Phylogenetic Classification	Function in Arabidopsis	Expression Pattern in Arabidopsis	Prediction of Gene Function in Cotton
SFT	*FT*-like	Ectopic expression accelerated time of flowering and floral development; rescue of late flowering *ft-10*	Root apical meristems and distal minor veins of expanded leaves in early vegetative development; absent in late vegetative development and flowering plants	Regulation of flowering and sympodial growth (validated, McGarry et al., 2016)
SP	*TFL1*-like	Ectopic expression delayed time of flowering and interrupted normal floral formation; partial rescue of early flowering *tfl1-14*	All plant meristem in early and late vegetative development; immature floral buds of flowering plants	Maintenance of main stem monopodial growth and regulation of sympodial growth (validated, McGarry et al., 2016)
TFL1-L1	*TFL1*-like	Ectopic expression delayed time of flowering and interrupted normal floral formation; no rescue effects in time-of-flowering mutants	All plant vasculature tissues during early vegetative development (source-to-sink patterning in leaves); absent in late vegetative development and flowering plants	No regulation of cotton plant architecture
TFL1-L2	*TFL1*-like	Ectopic expression delayed time of flowering and interrupted normal floral formation; rescue of early flowering *tfl1-14*	All plant vasculature tissues during early and late vegetative development (source-to-sink patterning in leaves); absent in flowering plants	Maintenance of main stem monopodial growth and regulation of sympodial growth; possible systemic anti-florigenic in response to endogenous or external signals sensed in shoot or root systems

BFT-L1	*TFL1*-like	Ectopic expression delayed time of flowering and interrupted normal floral formation; no rescue effects in time-of-flowering mutants	Absent in early and late vegetative development and flowering plants	No regulation of cotton plant architecture
BFT-L2	*TFL1*-like	Ectopic expression delayed time of flowering and interrupted normal floral formation; partial rescue of early flowering *tfl1-14*	Leaf and hypocotyl vasculature tissues during early and late vegetative development (source-to-sink patterning in leaves); leaf, sepal and petal vasculature in flowering plants	Maintenance of main stem monopodial growth and regulation of sympodial growth; possible systemic anti-florigenic in response to endogenous or external signals sensed in the shoot system
MFT-L1	*MFT*-like	Ectopic expression slightly accelerated time of flowering; no rescue effects in time-of-flowering mutants	Petioles of fully expanded leaves in early vegetative development; absent in late vegetative development and flowering plants	No regulation of cotton plant architecture
MFT-L2	*MFT*-like		Petioles of fully expanded leaves in early and late vegetative development; filaments of flowers in flowering plants	No regulation of cotton plant architecture

REFERENCES

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science **309**: 1052–6

Adrian J, Farrona S, Reimer JJ, Albani MC, Coupland G, Turck F (2010) cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis. Plant Cell **22**: 1425–40

Ahn JH, Miller D, Winter VJ, Banfield MJ, Lee JH, Yoo SY, Henz SR, Brady RL, Weigel D (2006) A divergent external loop confers antagonistic activity on floral regulators FT and TFL1. EMBO J **25**: 605–14

Alvarez J, Guli CL, Yu X-H, Smyth DR (1992) terminal flower: a gene affecting inflorescence development in Arabidopsis thaliana. Plant J **2**: 103–116

Amaya I, Ratcliffe OJ, Bradley DJ (1999) Expression of CENTRORADIALIS (CEN) and CEN-like genes in tobacco reveals a conserved mechanism controlling phase change in diverse species. Plant Cell **11**: 1405–18

An H, Roussot C, Suárez-López P, Corbesier L, Vincent C, Piñeiro M, Hepworth S, Mouradov A, Justin S, Turnbull C, et al (2004) CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development **131**: 3615–3626

Anderson PR, Haj-Ahmad Y (2003) Counter-selection facilitated plasmid construction by

homologous recombination in Saccharomyces cerevisiae. Biotechniques **35**: 692–4, 696, 698

Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, et al (2011) The genome of Theobroma cacao. Nat Genet **43**: 101–108

Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature **451**: 86–89

Banfield M., Brady R. (2000) The structure of Antirrhinum centroradialis protein (CEN) suggests a role as a kinase regulator11Edited by I. A. Wilson. J Mol Biol **297**: 1159–1170

Banfield MJ, Barker JJ, Perry AC, Brady RL (1998) Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction. Structure **6**: 1245–54

Baumann K, De Paolis A, Costantino P, Gualberti G (1999) The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the rolB oncogene in plants. Plant Cell **11**: 323–34

Baumann K, Venail J, Berbel A, Domenech MJ, Money T, Conti L, Hanzawa Y, Madueno F, Bradley D (2015) Changing the spatial pattern of TFL1 expression reveals its key role in the shoot meristem in controlling Arabidopsis flowering architecture. J Exp Bot **66**: 4769–80

Becker D, Kemper E, Schell J, Masterson R (1992) New plant binary vectors with selectable markers located proximal to the left T-DNA border. Plant Mol Biol **20**: 1195–7

Benlloch R, Berbel A, Serrano-Mislata A, Madueño F (2007) Floral initiation and inflorescence architecture: a comparative view. Ann Bot **100**: 659–76

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res **7**: 1513–23

Blancke S, Van Breusegem F, De Jaeger G, Braeckman J, Van Montagu M (2015) Fatal attraction: the intuitive appeal of GMO opposition. Trends Plant Sci **20**: 414–418

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O (2006) CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science **312**: 1040–3

Bradley D, Carpenter R, Copsey L, Vincent C, Rothstein S, Coen E (1996) Control of inflorescence architecture in Antirrhinum. Nature **379**: 791–7

Carmel-Goren L, Liu YS, Lifschitz E, Zamir D (2003) The SELF-PRUNING gene family in tomato. Plant Mol Biol **52**: 1215–22

Carmona MJ, Calonje M, Martínez-Zapater JM (2007) The FT/TFL1 gene family in grapevine. Plant Mol Biol **63**: 637–50

Chailakhyan MK, Krikorian AD (1975) Forty years of research on the hormonal basis of plant development — Some personal reflections. Bot Rev **41**: 1–29

Chardon F, Damerval C (2005) Phylogenomic Analysis of the PEBP Gene Family in Cereals. J Mol Evol **61**: 579–590

Clough SJ, Bent AF (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16**: 735–743

Coelho CP, Minow MAA, Chalfun-Júnior A, Colasanti J (2014) Putative sugarcane FT/TFL1 genes delay flowering time and alter reproductive architecture in Arabidopsis. Front Plant Sci **5**: 221

Collins JJ, Gardner TS, Cantor CR (2000) Construction of a genetic toggle switch in Escherichia coli. Nature **403**: 339–342

Conti L, Bradley D (2007) TERMINAL FLOWER1 is a mobile signal controlling Arabidopsis architecture. Plant Cell **19**: 767–78

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, et al (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science **316**: 1030–3

Corbit KC, Trakul N, Eves EM, Diaz B, Marshall M, Rosner MR (2003) Activation of Raf-1 signaling by protein kinase C through a mechanism involving Raf kinase inhibitory protein. J Biol Chem **278**: 13061–8

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol **133**: 462–9

Danilevskaya ON, Meng X, Hou Z, Ananiev E V, Simmons CR (2008) A genomic and expression compendium of the expanded PEBP gene family from maize. Plant Physiol **146**: 250–64

Esumi T, Tao R, Yonemori K (2005) Isolation of LEAFY and TERMINAL FLOWER 1 homologues from six fruit tree species in the subfamily Maloideae of the Rosaceae. Sex Plant Reprod **17**: 277–287

Fornara F, de Montaigu A, Coupland G (2010) SnapShot: Control of flowering in Arabidopsis. Cell **141**: 550, 550–2

Foucher F, Morin J, Courtiade J, Cadioux S, Ellis N, Banfield MJ, Rameau C (2003) DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea. Plant Cell **15**: 2742–54

Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res **37**: 6984–90

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods **6**: 343–345

Gietz RD, Woods RA (2006) Yeast transformation by the LiAc/SS Carrier DNA/PEG method. Methods Mol Biol **313**: 107–20

Gómez-Mena C, Piñeiro M, Franco-Zorrilla JM, Salinas J, Coupland G, Martínez-Zapater JM (2001) early bolting in short days: an Arabidopsis mutation that causes early flowering and partially suppresses the floral phenotype of leafy. Plant Cell **13**: 1011–24

Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. Genome

Res **8**: 195–202

Gore UR (1935) Morphogenetic Studies on the Inflorescence of Cotton on JSTOR. Bot Gaz **97**: 118–138

Grover CE, Gallagher JP, Wendel JF (2015) Candidate Gene Identification of Flowering Time Genes in Cotton. doi: 10.3835/plantgenome2014.12.0098

Guo X, Zhao Z, Chen J, Hu X, Luo D (2006) A putative CENTRORADIALIS/TERMINAL FLOWER 1-like gene, Ljcen1, plays a role in phase transition in Lotus japonicus. J Plant Physiol **163**: 436–444

Haberman A, Ackerman M, Crane O, Kelner J-J, Costes E, Samach A (2016) Different flowering response to various fruit loads in apple cultivars correlates with degree of transcript reaccumulation of a TFL1-encoding gene. Plant J. doi: 10.1111/tpj.13190

Hanano S, Goto K (2011) Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. Plant Cell **23**: 3172–84

Hanzawa Y, Money T, Bradley D (2005) A single amino acid converts a repressor to an activator of flowering. Proc Natl Acad Sci U S A **102**: 7748–53

Haritatos E, Ayre BG, Turgeon R (2000) Identification of phloem involved in assimilate loading in leaves by the activity of the galactinol synthase promoter. Plant Physiol **123**: 929–37

Harju S, Fedosyuk H, Peterson KR (2004) Rapid isolation of yeast genomic DNA: Bust n' Grab.

BMC Biotechnol **4**: 8

Hecht V, Laurie RE, Vander Schoor JK, Ridge S, Knowles CL, Liew LC, Sussmilch FC, Murfet IC, Macknight RC, Weller JL (2011) The pea GIGAS gene is a FLOWERING LOCUS T homolog necessary for graft-transmissible specification of flowering but not for responsiveness to photoperiod. Plant Cell **23**: 147–61

Hedman H, Källman T, Lagercrantz U (2009) Early evolution of the MFT-like gene family in plants. Plant Mol Biol **70**: 359–69

Higuchi Y, Narumi T, Oda A, Nakano Y, Sumitomo K, Fukai S, Hisamatsu T (2013) The gated induction system of a systemic floral inhibitor, antiflorigen, determines obligate short-day flowering in chrysanthemums. Proc Natl Acad Sci **110**: 17137–17142

Ho WWH, Weigel D (2014) Structural features determining flower-promoting activity of Arabidopsis FLOWERING LOCUS T. Plant Cell **26**: 552–64

Huang N-C, Jane W-N, Chen J, Yu T-S (2012) Arabidopsis thaliana CENTRORADIALIS homologue (ATC) acts systemically to inhibit floral initiation in Arabidopsis. Plant J **72**: 175–84

Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and Arabidopsis. Curr Opin Plant Biol **6**: 113–120

Jaeger KE, Wigge PA (2007) FT protein acts as a long-range signal in Arabidopsis. Curr Biol **17**: 1050–4

Jang S, Marchal V, Panigrahi KCS, Wenkel S, Soppe W, Deng X-W, Valverde F, Coupland G (2008) Arabidopsis COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. EMBO J **27**: 1277–88

Jensen CS, Salchert K, Nielsen KK (2001) A TERMINAL FLOWER1-like gene from perennial ryegrass involved in floral transition and axillary meristem identity. Plant Physiol **125**: 1517–28

Jin J, He K, Tang X, Li Z, Lv L, Zhao Y, Luo J, Gao G (2015a) An *Arabidopsis* Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. Mol Biol Evol **32**: 1767–1773

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res **45**: D1040–D1045

Jin S, Jung HS, Chung KS, Lee JH, Ahn JH (2015b) FLOWERING LOCUS T has higher protein mobility than TWIN SISTER OF FT. J Exp Bot **66**: 6109–6117

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8**: 275–82

Jost P, Whitaker J, Brown SM, Bednarz C (2006) Use of Plant Growth Regulators as a Management Tool in Cotton.

Kardailsky I (1999) Activation Tagging of the Floral Inducer FT. Science (80-) **286**: 1962–1965

Karlgren A, Gyllenstrand N, Källman T, Sundström JF, Moore D, Lascoux M, Lagercrantz U (2011) Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. Plant Physiol **156**: 1967–77

Kim Y-M, Kim S, Koo N, Shin A-Y, Yeom S-I, Seo E, Park S-J, Kang W-H, Kim M-S, Park J, et al (2016) Genome analysis of *Hibiscus syriacus* provides insights of polyploidization and indeterminate flowering in woody plants. DNA Res **2**: dsw049

King A (2017) Technology: The Future of Agriculture. Nature **544**: S21–S23

Knott J (1934) Effect of a localized photoperiod on spinach. ProcAm Soc Hort Sci **31**: 152–154

Kobayashi Y (1999) A Pair of Related Genes with Antagonistic Roles in Mediating Flowering Signals. Science (80-) **286**: 1960–1962

Kojima S (2002) Hd3a, a Rice Ortholog of the Arabidopsis FT Gene, Promotes Transition to Flowering Downstream of Hd1 under Short-Day Conditions. Plant Cell Physiol **43**: 1096–1105

Kooiker M, Airoldi CA, Losa A, Manzotti PS, Finzi L, Kater MM, Colombo L (2005) BASIC PENTACYSTEINE1, a GA Binding Protein That Induces Conformational Changes in the Regulatory Region of the Homeotic Arabidopsis Gene SEEDSTICK. PLANT CELL ONLINE **17**: 722–729

Koornneef M, Alonso-Blanco C, Blankestijn-de Vries H, Hanhart CJ, Peeters AJ (1998) Genetic interactions among late-flowering mutants of Arabidopsis. Genetics **148**: 885–92

Kotoda N, Hayashi H, Suzuki M, Igarashi M, Hatsuyama Y, Kidou S-I, Igasaki T, Nishiguchi M, Yano K, Shimizu T, et al (2010) Molecular characterization of FLOWERING LOCUS T-like genes of apple (Malus x domestica Borkh.). Plant Cell Physiol **51**: 561–75

Kotoda N, Wada M (2005) MdTFL1, a TFL1-like gene of apple, retards the transition from the vegetative to reproductive phase in transgenic Arabidopsis. Plant Sci.

Kouprina N, Larionov V (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nat Protoc **3**: 371–7

Krizek BA, Meyerowitz EM (1996) The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. Development **122**: 11–22

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol **33**: 1870–1874

Längle-Rouault F, Jacobs E (1995) A method for performing precise alterations in the yeast genome using a recycable selectable marker. Nucleic Acids Res **23**: 3079–81

Larionov V, Kouprina N, Solomon G, Barrett JC, Resnick MA (1997) Direct isolation of human BRCA2 gene by transformation-associated recombination in yeast. Proc Natl Acad Sci U S A **94**: 7384–7

Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I (2000) The AGAMOUS-lIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. Genes Dev

14: 2366–2376

Lee R, Baldwin S, Kenel F, McCallum J, Macknight R (2013) FLOWERING LOCUS T genes control onion bulb formation and flowering. Nat Commun **4**: 2884

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J, et al (2015) Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nat Biotechnol **33**: 524–530

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, et al (2014) Genome sequence of the cultivated cotton Gossypium arboreum. Nat Genet **46**: 567–72

Lifschitz E, Ayre BG, Eshed Y (2014) Florigen and anti-florigen - a systemic mechanism for coordinating growth and termination in flowering plants. Front Plant Sci **5**: 465

Lifschitz E, Eviatar T, Rozman A, Shalit A, Goldshmidt A, Amsellem Z, Alvarez JP, Eshed Y (2006) The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. Proc Natl Acad Sci U S A **103**: 6398–403

Lin L, Pierce GJ, Bowers JE, Estill JC, Compton RO, Rainville LK, Kim C, Lemke C, Rong J, Tang H, et al (2010) A draft physical map of a D-genome cotton species (Gossypium raimondii). BMC Genomics **11**: 395

Lin M-K, Belanger H, Lee Y-J, Varkonyi-Gasic E, Taoka K-I, Miura E, Xoconostle-Cazares B, Gendler K, Jorgensen RA, Phinney B, et al (2007) FLOWERING LOCUS T Protein May Act as the Long-Distance Florigenic Signal in the Cucurbits. PLANT CELL ONLINE **19**: 1488–1506

Liu L-J, Zhang Y-C, Li Q-H, Sang Y, Mao J, Lian H-L, Wang L, Yang H-Q (2008) COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in Arabidopsis. Plant Cell **20**: 292–306

Ma J, Liu F, Wang Q, Wang K, Jones DC, Zhang B (2016) Comprehensive analysis of TCP transcription factors and their expression during cotton (Gossypium arboreum) fiber early development. Sci Rep **6**: 21535

Ma J, Wang Q, Sun R, Xie F, Jones DC, Zhang B (2014) Genome-wide identification and expression analysis of TCP transcription factors in Gossypium raimondii. Sci Rep **4**: 1–10

Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. Nat Biotechnol **21**: 796–802

McGarry RC, Ayre BG (2012a) Geminivirus-mediated delivery of florigen promotes determinate growth in aerial organs and uncouples flowering from photoperiod in cotton. PLoS One **7**: e36746

McGarry RC, Ayre BG (2012b) Manipulating plant architecture with members of the CETS gene family. Plant Sci **188**–**189**: 71–81

McGarry RC, Prewitt SF, Culpepper S, Eshed Y, Lifschitz E, Ayre BG (2016) Monopodial and sympodial branching architecture in cotton is differentially regulated by the Gossypium hirsutum SINGLE FLOWER TRUSS and SELF-PRUNING orthologs. New Phytol **212**: 244–258

Meeks-Wagner D, Wood JS, Garvik B, Hartwell LH (1986) Isolation of two genes that affect

mitotic chromosome transmission in S. cerevisiae. Cell **44**: 53–63

Meister RJ, Williams LA, Monfared MM, Gallagher TL, Kraft EA, Nelson CG, Gasser CS (2004) Definition and interactions of a positive regulatory element of the Arabidopsis INNER NO OUTER promoter. Plant J **37**: 426–38

Meng X, Muszynski MG, Danilevskaya ON (2011) The FT-Like ZCN8 Gene Functions as a Floral Activator and Is Involved in Photoperiod Sensitivity in Maize. Plant Cell Online 23:

Meyer L (2017) Global Growth of Cotton Mill Use Varies by Country.

Meyer L, MacDonald S (2015) Cotton and Wool Outlook.

Mimida N, Goto K, Kobayashi Y, Araki T, Ahn JH, Weigel D, Murata M, Motoyoshi F, Sakamoto W (2001) Functional divergence of the TFL1-like gene family in Arabidopsis revealed by characterization of a novel homologue. Genes Cells **6**: 327–36

Mohamed R, Wang C-T, Ma C, Shevchenko O, Dye SJ, Puzey JR, Etherington E, Sheng X, Meilan R, Strauss SH, et al (2010) Populus CEN/TFL1 regulates first onset of flowering, axillary meristem identity and dormancy release in Populus. Plant J **62**: 674–688

Monfared MM, Simon MK, Meister RJ, Roig-Villanova I, Kooiker M, Colombo L, Fletcher JC, Gasser CS (2011) Overlapping and antagonistic activities of BASIC PENTACYSTEINE genes affect a range of developmental processes in Arabidopsis. Plant J **66**: 1020–1021

Moon J, Suh SS, Lee H, Choi KR, Hong CB, Paek NC, Kim SG, Lee I (2003) The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. Plant J

35: 613–623

Oldenburg KR, Vo KT, Michaelis S, Paddon C (1997) Recombination-mediated PCR-directed

plasmid construction in vivo in yeast. Nucleic Acids Res **25**: 451–2

Oosterhuis DM (1990) Growth and development of the cotton plant. *In* WN Miley, DM

Oosterhuis, eds, Nitrogen Nutr. Cott. Pract. Issues. American Society of Agronomy,

Madison, WI, USA, pp 1–24

Ovcharenko I, Loots GG, Giardine BM, Hou M, Ma J, Hardison RC, Stubbs L, Miller W (2005)

Mulan: multiple-sequence local alignment and visualization for studying function and

evolution. Genome Res **15**: 184–94

Papi M, Sabatini S, Altamura MM, Hennig L, Schafer E, Costantino P, Vittorioso P (2002)

Inactivation of the Phloem-Specific Dof Zinc Finger Gene DAG1 Affects Response to Light

and Integrity of the Testa of Arabidopsis Seeds. PLANT Physiol **128**: 411–417

**Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu

S, Udall J, et al** (2012) Repeated polyploidization of Gossypium genomes and the evolution

of spinnable cotton fibres. Nature **492**: 423–7

Peet MM (2005) Greenhouse Tomato Production. *In* E Heuvelink, ed, Tomatoes. CABI

Publishing, Wallingford, U.K., pp 257–304

**Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ,

Worland AJ, Pelica F, et al** (1999) "Green revolution" genes encode mutant gibberellin

response modulators. Nature **400**: 256–61

Pillitteri LJ, Lovatt CJ, Walling LL (2004) Isolation and characterization of a TERMINAL FLOWER homolog and its correlation with juvenility in citrus. Plant Physiol **135**: 1540–51

Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JJL, Nilsson O (2010a) An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. Science **330**: 1397–400

Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JJL, Nilsson O (2010b) An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. Science **330**: 1397–400

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal M, Zamir D, Lifschitz E (1998) The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. Development **125**: 1979–89

Ratcliffe O, Amaya I, Vincent C, Rothstein S, Carpenter R, Coen E, Bradley D (1998) A common mechanism controls the life cycle and architecture of plants. Development **125**: 1609–1615

Raymond CK, Pownder TA, Sexson SL (1999) General method for plasmid construction using homologous recombination. Biotechniques **26**: 134–8, 140–1

Raymond CK, Sims EH, Olson M V (2002) Linker-mediated recombinational subcloning of large DNA fragments using yeast. Genome Res **12**: 190–7

Riechmann JL, Wang M, Meyerowitz EM (1996) DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res **24**: 3134–3141

Saltveit M (2005) Post harvest biology and handling. *In* E Heuvelink, ed, Tomatoes. CABI Publishing, Wallingford, U.K., pp 305–325

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G (2000) Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science **288**: 1613–6

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning: a laboratory manual.

Sangwan I, O'Brian MR (2002) Identification of a Soybean Protein That Interacts with GAGA Element Dinucleotide Repeat DNA. PLANT Physiol **129**: 1788–1794

Santi L, Wang Y, Stile MR, Berendzen K, Wanke D, Roig C, Pozzi C, Müller K, Müller J, Rohde W, et al (2003) The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene Bkn3. Plant J **34**: 813–26

Sawa M, Nusinow DA, Kay SA, Imaizumi T (2007) FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. Science **318**: 261–5

Shalit A, Rozman A, Goldshmidt A, Alvarez JP, Bowman JL, Eshed Y, Lifschitz E (2009) The flowering hormone florigen functions as a general systemic regulator of growth and termination. Proc Natl Acad Sci U S A **106**: 8392–7

Shao Z, Zhao H, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res **37**: e16

Shih PM, Vuu K, Mansoori N, Ayad L, Louie KB, Bowen BP, Northen TR, Loqué D (2016) A robust gene-stacking method utilizing yeast assembly for plant synthetic biology. Nat Commun **7**: 13215

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7**: 539

Sikorski RS, Hieter P (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics **122**: 19–27

Small RL, Wendel JF (2000) Phylogeny, Duplication, and Intraspecific Variation of Adh Sequences in New World Diploid Cottons (Gossypium L., Malvaceae). Mol Phylogenet Evol **16**: 73–84

Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature **410**: 1116–20

Supek F, Bošnjak M, Škunca N, Šmuc T, O'Donovan C (2011) REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS One **6**: e21800

Sussex IM, Kerk NM (2001) The evolution of plant architecture. Curr Opin Plant Biol **4**: 33–37

Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K (2007) Hd3a protein is a mobile flowering signal in rice. Science **316**: 1033–6

Tsuji H, Taoka K, Shimamoto K (2011) Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. Curr Opin Plant Biol **14**: 45–52

Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annu Rev Plant Biol **59**: 573–94

Vall��e BS, Coadou G, Labb�� H, Sy D, Vovelle F, Schoentgen F (2008) Peptides corresponding to the N- and C-terminal parts of PEBP are well-structured in solution: new insights into their possible interaction with partners in vivo. J Pept Res **61**: 47–57

Vent W (2008) Hallé, F., Oldeman, R. A. A. & Tomlinson, P. B., Tropical Trees and Forests - An Architectural Analysis. XVII + 441 pages, 111 figs., 10 tables. Springer-Verlag. Berlin-Heidelberg-New York 1978. ISBN 3-540-08494-0. Price: Cloth DM 125.-; US $ 62.50. Feddes Repert **93**: 146–146

Vigani M, Olper A, Hadji Fall E, Wang Y, Dallal GE, Grusak MA, Russell RM, Beale MH, Schmalensee R, Smith VK, et al (2014) GM-free private standards, public regulation of GM products and mass media. Environ Dev Econ **19**: 743–768

Wan CY, Wilkins TA (1994) A Modified Hot Borate Method Significantly Enhances the Yield of High-Quality RNA from Cotton (Gossypium hirsutum L.). Anal Biochem **223**: 7–12

Wang R, Albani MC, Vincent C, Bergonzi S, Luan M, Bai Y, Kiefer C, Castillo R, Coupland G

(2011) Aa TFL1 Confers an Age-Dependent Response to Vernalization in Perennial Arabis alpina. Plant Cell Online 23:

Wang Z, Zhou Z, Liu Y, Liu T, Li Q, Ji Y, Li C, Fang C, Wang M, Wu M, et al (2015) Functional Evolution of Phosphatidylethanolamine Binding Proteins in Soybean and Arabidopsis. Plant Cell **27**: 323–36

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLoS One **6**: e16765

Weigel D, Glazebrook J (2002) Arabidopsis: A Laboratory Manual. CSHL Press

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integration of spatial and temporal information during floral induction in Arabidopsis. Science **309**: 1056–9

Wilson RN, Heckman JW, Somerville CR (1992) Gibberellin Is Required for Flowering in Arabidopsis thaliana under Short Days. Plant Physiol **100**: 403–8

Wolabu TW, Zhang F, Niu L, Kalve S, Bhatnagar-Mathur P, Muszynski MG, Tadege M (2016) Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. New Phytol. doi: 10.1111/nph.13834

Xi W, Yu H (2010) MOTHER OF FT AND TFL1 regulates seed germination and fertility relevant to the brassinosteroid signaling pathway. Plant Signal Behav **5**: 1315–7

Yamaguchi A, Kobayashi Y, Goto K, Abe M, Araki T (2005) TWIN SISTER OF FT (TSF) acts as a

floral pathway integrator redundantly with FT. Plant Cell Physiol **46**: 1175–89

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, et al (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell **12**: 2473–2484

Yeager A (1927) Determinate growth in the tomato. J Hered **18**: 263–265

Yeung KC, Rose DW, Dhillon AS, Yaros D, Gustafsson M, Chatterjee D, McFerran B, Wyche J, Kolch W, Sedivy JM (2001) Raf kinase inhibitor protein interacts with NF-kappaB-inducing kinase and TAK1 and inhibits NF-kappaB activation. Mol Cell Biol **21**: 7207–17

Yoo SJ, Chung KS, Jung SH, Yoo SY, Lee JS, Ahn JH (2010) BROTHER OF FT AND TFL1 (BFT) has TFL1-like activity and functions redundantly with TFL1 in inflorescence meristem development in Arabidopsis. Plant J **63**: 241–53

Yoo SK, Chung KS, Kim J, Lee JH, Hong SM, Yoo SJ, Yoo SY, Lee JS, Ahn JH (2005) CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis. Plant Physiol **139**: 770–8

Yoo SY, Kardailsky I, Lee JS, Weigel D, Ahn JH (2004) Acceleration of flowering by overexpression of MFT (MOTHER OF FT AND TFL1). Mol Cells **17**: 95–101

Yu J, Jung S, Cheng C-H, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D (2014) CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids

Res **42**: D1229-36

Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L, et al (2015) The

genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the

allopolyploidization and development of superior spinnable fibres. Sci Rep **5**: 17662

Zeevaart JAD (1976) Physiology of Flower Formation. Annu Rev Plant Physiol **27**: 321–348

Zeevaart JAD (2006) Florigen coming of age after 70 years. Plant Cell **18**: 1783–9

Zeevaart JAD (2008) Leaf-produced floral signals. Curr Opin Plant Biol **11**: 541–7

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, et al

(2015) Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a

resource for fiber improvement. Nat Biotechnol **33**: 531–537

Zhang X, Wang C, Pang C, Wei H, Wang H, Song M, Fan S, Yu S (2016) Characterization and

Functional Analysis of PEBP Family Genes in Upland Cotton (Gossypium hirsutum L.). PLoS

One **11**: e0161080